Technical analysis with machine learning classification algorithms: Can it still ‘beat’ the buy-and-hold strategy?

Ba Chu*

December 28, 2023

Abstract

This paper undertakes an extensive study to search for empirical evidence of directional predictability and profitability on an aggregate stock market index by applying supervised machine learning (ML) algorithms to a large set of financial variables, technical indicators, and price patterns to generate predictions [of the moving direction of future stock price] that lead to the most profitable trading strategy. We use both symmetric and asymmetric loss function to train (and both statistical and economic scoring functions to cross-validate) a ML algorithm. We also extend the bootstrap Reality Check (RC) procedure to formally compare the performance of trading methods.

The trading strategy using one-period ahead forecasts can generate higher annualized returns than the buy-and-hold strategy when transaction cost is low. Most positive annualized excess returns (i.e., the annualized returns of our strategy in excess of those from the buy-and-hold strategy) are realized during trading sessions with high volatility. However, the trading strategy using multiple-days ahead forecasts can become less profitable. There is a strong evidence that some scoring functions used to cross validate a ML algorithm can generate more economically significant predictions than the others. Several candlestick chart patterns have a strong predictive power that can be effectively leveraged by Random Forest to increase annualized excess returns compared to using only financial variables and technical indicators as predictors.

Keywords: Machine learning, Technical indicators, Price patterns, Directional predictability, Trading strategy

JEL classification: C53, C58, G11, G17

*Department of Economics, Carleton University, 1125 Colonel By Dr., Ottawa, Ontario, Canada. Email: ba.chu@carleton.ca. Tel: +1 613-520-2600 (ext. 1546).

1 Introduction

Technical analysis (a.k.a. the technique to predict the moving direction of future asset prices based on the behaviour of past price data) has now become a standard toolkit for traders/investors and a legitimate discipline for academic research. Although the usefulness of technical analysis has been convincingly proven in practice, as indicated in the literature review below – there is still an ongoing debate: Can a trading strategy based on technical analysis outperform the simple buy-and-hold strategy over time? If the answer is 'yes', then what are the main technical predictors that drive this outperformance? The goal of this paper is to employ machine learning (ML) and a large set of all common technical indicators and price patterns to shed new light on this long-standing debate. The innovative aspect of our methodology is that we train a ML algorithm with both symmetric and asymmetric loss function and cross-validate this algorithm with both statistical and economic scoring functions to generate the most economically profitable forecasts. We also extend the scope of White's (2000) bootstrap Reality Check (RC) algorithm to compare the performance of trading methods generated by a ML model.

We start by reviewing the idea behind technical analysis and its related literature. The well-known dividend discount model states that the present fair price (or the intrinsic value) of a financial asset is the conditional expectation of the discounted value of all future cash flows returned to shareholders given the information currently available. When investors/traders as a group have rational expectations, they can use all the information available to form their expectation about the discounted value of all future cash flows so that the actual asset price (which is equal to the fair price in this case) reflects all the available information almost instantly – the price must therefore fluctuate in an unpredictable manner [see, e.g., Fama (1965); Lucas (1978); Mandelbrot (1966); Samuelson (1973)]. In reality, investors/traders' rationality is always bounded in the sense that they can use only a small subset of the information available to form their expectation about the discounted value of all future cash flows or they are limited in their computational resources to make optimal choices [see, e.g., Simon (1955)]. As a result, the asset price can deviate from the fair price for a certain period of time before they converge together while the investors/traders learn and adjust their expectation. This temporary mispricing can be caused by behavioural biases, such as underreaction or delayed overreaction – asset prices initially underreact to corporate news (such as news about earnings), the prices then continue to increase or decrease in the future, depending on the nature of the news, which creates trend and momentum followed by an eventual reversal.[1] The asset prices thus change in a somewhat predictable way so that trading on short-term trends, momentums, or reversals can generate positive excess profit (i.e., the profit of a trading strategy in excess of the profit from the buy-and-hold strategy) [see, e.g., DeBondt and Thaler (1985); Han, Zhou, and Zhu (2016); Jegadeesh and Titman (1993); Moskowitz, Ooi, and Pedersen (2012)].[2]

Behaviour finance provides theoretical underpinnings for technical analysis that is often employed by traders to predict the moving direction of future asset price (i.e., the directions of price changes can be predicted by technical indicators and price patterns at least in the short term as, in the long term, any profitable trading rule should stop making money due to investors/traders competing against each other for superior returns or due to limits to arbitrage, and the prices are eventually driven by fundamentals, such as earning expectations or other macroeconomic variables). Financial valuation ratios and macroeconomic variables (e.g., dividend yield, earning yield, book-to-market ratio, nominal interest rates, term spreads, or default spreads) can be used to predict future stock returns at a monthly or lower frequency. There are both negative and positive views [see, e.g., Ang and Bekaert (2007); Campbell and Shiller (1988); Campbell and Thompson (2008); Lander, Orphanides, and Douvogiannis (1997); Rapach, Strauss, and Zhou (2010) for empirical evidences on the out-of-sample (OoS) predictability of excess stock returns; Farmer, Schmidt, and Timmermann (2022); Pesaran and Timmermann (1995) for an evidence of time-varying predictive power of economic variables; and Welch and Goyal (2008) for an empirical evidence of rather weak OoS predictability of stock returns using financial variables]. Recent studies using ML

[1] Underreaction is caused by the slow diffusion of news [e.g., Hong and Stein (1999)], anchoring biases [e.g., Barberis, Shleifer, and Vishny (1998)], the disposition effect of holding losers too long and selling winners too early [e.g., Frazzini (2006)]. Overreaction is caused by positive feedback trading [e.g., de Long, Shleifer, Summers, and Waldmann (1990)], overconfidence and self-attribution confirmation biases [e.g., Daniel, Hirshleifer, and Subrahmanyam (1998)], the representativeness heuristics [e.g., Tversky and Kahneman (1974)], herding [e.g., Bikhchandani, Hirshleifer, and Welch (1992)], or investor sentiment [e.g., Baker and Wurgler (2007)].

The adaptive market hypothesis proposed by Lo (2004, 2005) suggests that behavioural biases can be driven by changing economic environments and the agents' adaptive responses to the changing environment. Therefore, both rationality and irrationality can coexist for a period of time as the agents learn to adapt to a new environment based on heuristics.

[2] The presence of return predictability may also be attributed to time-varying expected returns due to shocks to tastes for current versus future consumption or technological shocks (i.e., rational investors may accept lower expected returns because of lower perceived risk in good times while they may demand higher expected returns due to higher perceived risk in bad times). This is consistent with rational pricing [Fama (1991)]. Hence, Pesaran and Timmermann (1995) suggests that stock market predictability is economically significant if it can be exploited for trading/investing profitability.

methods to predict monthly excess returns with a large number of accounting variables (that provide details about firm business) confirm the economic and statistical value of stock return predictability [see, e.g., Avramov, Cheng, and Metzker (2022); Gu, Kelly, and Xiu (2020); Rossi (2018)].

There are some studies documenting that technical indicators provide a little added benefit to the OoS predictability of stock returns mostly at short predictive horizons [e.g, Sullivan, Timmermann, and White (1999) finds no evidence that trading rules based on common technical indicators (such as moving average (MA), support and resistance, channel breakouts, and on-balance volume averages) outperform a benchmark strategy after accounting for data-snooping; Bajgrowicz and Scaillet (2012) uses a new false discovery rate approach to data-snooping to establish that none of the trading rules considered in earlier studies are significant and the in-sample performance of those rules tends to be offset by low transaction costs; Li and Tsiakas (2017) shows that technical indicators do not add to the OoS R^2 in a predictive regression model when they combine with other financial variables].[3] However, many other studies confirm the usefulness of technical analysis in predicting the price moving direction of a financial asset. The gist of their argument is that there are patterns in the historical data that embody the prime movers of all market actions or crowd psychology [see, e.g., Murphy (1999)]. In exchange rate forecasting, Panopoulou and Souropanis (2019) shows that technical indicators can predict future currency movements and this prediction can be significantly improved by using economic variables; Chang and Osler (1999) identifies a common candlestick chart pattern (the head-and-shoulders pattern) as a good predictor of the moving direction of future short-term exchange rates and this pattern is the main source of profitability in a technical trading rule; LeBaron (1999) finds that technical analysis has some predictive value when it is used to predict exchange rates during periods with central bank intervention, and this predictability can dramatically decrease during periods without central bank intervention; there are also many other empirical evidences of predictability and profitability of technical trading rules in the foreign exchange market [see, e.g., Hsu, Taylor, and Wang (2016); Neely, Weller, and Dittmar (1997)]. In stock return forecasting, empirical evidences in favour of technical analysis include Allen and Karjalainen (1999); Brock, Lakonishok, and LeBaron (1992); Brock et al. (1992); Brown and Jennings (1989); Brown, Goetzmann, and Kumar

[3] *Data-snooping* bias occurs when a trading rule is selected only with the benefit of hindsight. This rule can thus work very well on an observed sample, but this performance may not repeat again in the future.

(1998); Gençay (1996); Jiang, Kelly, and Xiu (2022); Julian Andrada-Félix (2008); Lo, Mamaysky, and Wang (2000); Neely, Rapach, Tu, and Zhou (2014), among many others. [For a comprehensive review of the chronological history of technical analysis, we refer to the textbook by Lo and Hasanhodzic (2010).]

This paper extends and enriches the existing research on technical analysis with ML. To be specific, we employ a gradient boosting algorithm [initially developed by Freund and Schapire (1996) as the AdaBoost algorithm and generalized to the gradient boosting algorithm later on by Breiman (1998, 1999), and Friedman (2001a); a fast and efficient implementation boosting framework is LightGBM (LGBM) developed by Microsoft] and Random Forest (RF) [Ho (1995) and Breiman (2001)] – as described in Section 2 – to predict the moving direction of an aggregate market index [tracked by an exchange-traded fund (ETF) called SPDR S&P 500 ETF (SPY)] by eight financial variables together with 45 common technical indicators and 58 candlestick chart patterns (as tabulated in Table S.V.1 of the Supplemental Material). The predictions of the moving directions of future prices are then used to construct a simple trading strategy (i.e., our trader remains 'in' the market if the price is predicted to go up while this trader is 'out' of the market if the price is predicted to go down; and the trader pays a fixed/variable transaction cost for each trade). The reason why we focus on stock price's directional predictability is that the goal of technical trading is to use forecasts to make profit, and thus one only needs to have a good forecast of the price moving direction (instead of numerically accurate forecasts which are often more difficult to obtain given the low signal-to-noise ratio of stock returns).

Our main empirical findings are as follows. First, our trading methods using one-day ahead forecasts of the moving directions of prices can generate higher annualized (total) returns than the buy-and-hold strategy in slightly over 50% of all the trading sessions at best when transaction cost is low (note that we formed 1718 trading sessions of length 100 days and 1618 trading sessions of length 200 days). Most of this superior performance was realized during periods of high volatility (gauged by the VIX index). There is also a high variation in the daily returns generated by the trading methods during these periods. A high level of the VIX index means high uncertainty, implying a higher cost of capital and less affordable insurance against shocks. All of these will lower equity returns. An important reason for the superior performance of our trading strategy during a high-volatility period is that volatility is unusually high during market downturns and turning points are completely missed by the buy-and-hold strategy in these episodes. (Indeed, during a market rally, the buy-and-hold strategy almost mechanically leads to a better performance.) The outperformance of our strategy during a high-volatility period becomes less likely when transaction cost is high. Therefore, an important key takeaway here is that economic gains depend much on trader's ability to take risk and to select appropriate time to trade.

Second, in most trading sessions, our trading strategies can generate positive risk-adjusted returns in excess of the risk-free rate (such as the Sharpe ratio, the Calmar ratio, the Sortino ratio, or Morningstar's risk-adjusted rating as defined in Section 4). The correlation between equity curve and perfect profit (CECPP) values of our trading strategy using an optimal loss function and scoring function can exceed 70% in over 50% of trading sessions when transaction cost is low. In addition, the percentages of winning trades can surpass 60% in over 50% of trading sessions with a low level of transaction cost. This evidence suggests that the profit drawn from the predictions of the moving directions of future prices is not something random (which is also confirmed by a placebo-type experiment presented in Section S.IV).

Third, asymmetric loss functions may generate good in-sample predictions. However, our out-of-sample analysis does not provide strong evidence supporting the use of asymmetric loss functions. We find an empirical evidence that some scoring function can generate more economically profitable predictions than the others. In particular, the *CECPP* scoring function seems to be an optimal choice as it is an evaluation metric that takes into account the distribution of trades and profit relative to the actual profit potential provided by the perfect profit strategy (that is, the strategy assuming that we can accurately predict the moving directions of prices). The *Sharpe ratio* scoring function is also a good choice in several cases as it is a popular criterion to evaluate trading/investment strategies.

Fourth, several candlestick chart patterns have a strong predictive power that can be effectively leveraged by Random Forest to increase annualized excess returns compared to using only financial variables and technical indicators as predictors. These candlestick chart patterns are: 1) the three advancing white soldiers pattern (*3WHITESOLDIERS*), 2) the closing Marubozu pattern (*CLOSING-MARUBOZU*), 3) the Doji Star pattern (*DOJISTAR*), 4) the Up/Down-gap Side-by-Side White Lines pattern (*GAPSIDESIDE-WHITE*), 5) The Harami Cross pattern (*HARAMICROSS*), 6) the Identical Three Crows pattern (*IDENTICAL3CROWS*), 7) the Ladder Bottom pattern (*LADDERBOTTOM*), 8) the Mat Hold pattern (*MATHOLD*), 9) the Rickshaw Man pattern (*RICKSHAWMAN*), 10) the Stalled pattern (*STALLEDPATTERN*), and 11) the Tristar pattern (*TRISTAR*). Fifth, the annualized excess return is more likely to be positive when one trades based on shorter-horizon forecasts (say, $h \in \{1, \ldots, 4\}$). When trading based on long-horizon forecasts (say, $h > 4$), there is a high probability that our trading strategy underperforms the buy-and-hold strategy. One reason for this underperformance may be that the predictors used in this study do not have a strong predictive power to predict the price moving direction multiple days ahead. This finding confirms that technical analysis is a useful toolkit used by short-term traders/investors.

As mentioned earlier, there is a large literature debating the pros and cons of technical analysis. Therefore, we shall not attempt to discuss all the papers here. We can only discuss a few works closely related to our study. Allen and Karjalainen (1999) proposes to use genetic algorithms to select optimal trading rules from a pool of many rules based on local extrema of past prices and MAs. The rules are derived based on data in the training period and they are cross-validated using data in the selection period, then evaluated with data in the test period so that the performance in a particular test sample is purely out-of-sample (and thus, it is free from any lookahead bias). These rules can identify periods to long the S&P 500 index when daily returns are positive and less volatile and to de-invest in the index when the reverse is true. However, the rules do not outperform the buy-and-hold strategy in the OoS test period.

Julian Andrada-Félix (2008) employs a boosting algorithm (i.e., Freund and Schapire's (1999) AdaBoost) to predict the price moving direction of an aggregate stock market index by a large set of MA technical indicators. In each year, a training period of the first 100 days is used to train a boosting model which is then used to generate predictions for the remaining period. The directional forecasts are then used to decide if one will be in/out of the market in each trading day. This trading strategy can outperform the buy-and-hold strategy during periods of market downturn. In these papers, only a small subset of all the available technical indicators and patterns (mostly, MA-base rules or local extrema of past prices) is used for analysis. More recently, Jiang et al. (2022) trains a convolutional neural network (CNN) model using monochrome images of candlestick charts of stock prices over the past 5, 20, or 60 days to predict the moving direction of future prices over the subsequent 5, 20, or 60 days. It is also demonstrated that a portfolio formed by longing stocks whose prices are predicted to go up and shorting stocks whose prices are predicted to go down can be profitable.

In the aforementioned works, the role of economic scoring function in selecting models used to generate OoS forecasts, as emphasized by Satchell and Timmermann (1995), is not taken into account; and a model (that is trained in the training period and cross-validated in the selection period) used to forecast the entire test period may be slow to respond to rapid changes in the actual market environment.[4] The only reason why this trained model is held fixed throughout the entire test period is that it requires much less computational resources than a recursively re-trained model. Moreover, asymmetric loss functions have not been used by the existing methods to predict the moving directions of stock prices.

We contribute to the literature by using all common technical indicators and candlestick chart patterns available to predict the future price trends of an asset with machine learning. We focus on one-day ahead prediction (like most other works on technical analysis) instead of prediction at lower frequencies (e.g., weekly, monthly, or quarterly) that has been done in the aforementioned studies. In sum, our method differs from the methods employed in the previous studies in three main aspects: First, we use a rolling-window strategy (to be detailed in Section 3) to account for nonstationarity in the data – to make the one-period ahead forecast of the direction of a future price change, we train and cross-validate a ML model by using the last 1000 observations of the predictors (that is, to make over 1800 forecasts, the model needs to be trained and cross-validated over 1800 times). Although this procedure is computationally intensive, the obtained forecast model can be more responsive to any short-term change in the market environment because this model is selected based on a long window of past observations available up to the time when a forecast is made. Therefore, this method can enhance predictions over very short periods, typically a day, as asset prices are notably difficult to predict in the short term [see, e.g., Fama (1970)]. Second, we train a ML algorithm by minimizing either a symmetric or an asymmetric loss function [that penalizes false positives and false negatives differently]. Asymmetric loss functions may be justified in this setting where false positives (*or* incorrect predictions that stock price will go up) can cause significant losses of wealth. Therefore, the choice of loss function can affect parameter estimates, forecasts, and decisions [see, e.g., Elliott, Ghanem, and Krüger (2016)]. Third, we cross-validate the model by using either a statistical or economically-motivated evaluation metric (*or* scoring function). Economic scoring functions are more suitable than statistical ones to evaluate the economic value of predictions of nonlinear processes [see, e.g.,

[4] Cross validation is a procedure of selecting the optimal values for the hyperparameters of a ML algorithm by maximizing a scoring function evaluated on a subset of data non-overlapping with the training data.

Kelly, Malamud, and Zhou (2023); Satchell and Timmermann (1995)]. In our setting, economic scoring metrics can be the Sharpe ratio, the Calmar ratio, or the CECPP of a simple trading strategy (to be defined in Section 4).

The rest of the paper is organized as follows. Section 2 explains Gradient Boosting Machine and Random Forest [trained with a symmetric/asymmetric loss function and cross-validated with a statistical/economic scoring function] to predict the moving direction of future prices. Section 3 describes our recursive forecasting strategy [used to generate a sequence of h-period ahead forecasts over an out-of-sample period], the cross-validation procedure, and the scoring functions. In Section 4, we describe a simple trading strategy based on h-period ahead forecasts of the moving directions of future prices. Section 5 explains the dataset and presents the main empirical findings of the paper. Section 6 provides some concluding remarks. All tables and figures needed to support our main results are collected in appendices at the end of the paper. The Supplemental Material (SM) provides further details of our methodology, and extra tables and figures to corroborate the empirical findings reported in the main text.

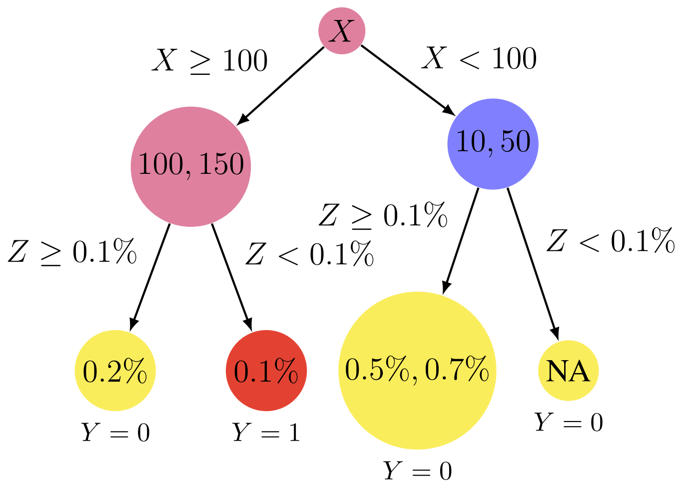
2 Predicting the Moving Direction of Stock Price with Machine Learning

Boosting, first proposed by Friedman (2001b) and Schapire (1990), is one of the most powerful machine learning methods that can be used for both classification and regression. Boosting converts weak learners (*or* simple white-box models) that predict slightly better than random guessing to a strong learner that can predict with high accuracy. In classifying stock returns (R_t, $t = 1, \ldots, T$) to positive or negative instances for a given set of predictors ($\boldsymbol{X}_t$, $t = 1, \ldots, T$), one needs to estimate the probabilities: $Prob\left(R_t \geq 0 \mid \boldsymbol{X}_t\right) \coloneqq S\left(F(\boldsymbol{X}_t)\right)$ and $Prob\left(R_t < 0 \mid \boldsymbol{X}_t\right) \coloneqq 1 - S\left(F(\boldsymbol{X}_t)\right)$, where $S(x) \coloneqq \frac{1}{1+\exp(-x)}$ is the sigmoid link function. Let $Y_t \coloneqq \begin{cases} 1 \text{ if } R_t \geq 0 \\ 0 \text{ if } R_t < 0 \end{cases}$. The goal of Boosting is to find an optimal function $F^*(\boldsymbol{x})$ [mapping from the support of $\boldsymbol{X}_t$ to $\mathbb{R}$] such that the conditional expectation of some loss function $L\left(Y, S\left(F(\boldsymbol{x})\right)\right)$ is minimized:

$$F^*(\boldsymbol{x}) \coloneqq \arg\min_{F(\boldsymbol{x})} E\left[L\left(Y_t, S\left(F(\boldsymbol{X_t})\right)\right) \mid \boldsymbol{X}_t = \boldsymbol{x}\right].$$

This loss function achieves its minimum values at $Y = 1$ and $p := Prob\,(Y = 1 \mid \boldsymbol{X} = \boldsymbol{x}) = S\,(F(\boldsymbol{x})) = 1$ *or* at $Y = 0$ and $q := Prob\,(Y = 0 \mid \boldsymbol{X} = \boldsymbol{x}) = 1 - S\,(F(\boldsymbol{x})) = 1$ (i.e., when the moving direction of price is perfectly predicted). In reality, these perfect predictions are usually unachievable. We then prefer a value of p or q which is as close to one as possible when the actual value of Y is one *or* zero respectively. There are many loss functions that can be used for this purpose. Common loss functions include Good's (1952) cross entropy loss $CE\,(Y, p) := -\left(Y \log(p) + (1 - Y) \log(1 - p)\right)$ or Brier's (1950) loss $BrierLoss\,(Y, p) := (Y - p)^2$, or Buja, Stuetzle, and Shen's (2005) Boosting loss $BoostLoss\,(Y, p) := Y\sqrt{\frac{1-p}{p}} + (1 - Y)\sqrt{\frac{p}{1-p}}$. These loss functions are symmetric. From an investor's point of view, false positives (i.e., stock return is predicted to be positive when it turns out to be negative) are much more costly than false negatives (i.e., stock return is predicted to be negative when it turns out to be positive) because a maximum drawdown is the most important risk to equity investment (an investment in a stock whose price eventually goes down by a very large amount can wipe out the whole portfolio). A symmetric loss function, which penalizes false positives and false negatives equally, may not be appropriate. Gneiting and Raftery's (2007) As2 loss function weights false positives more heavily than false negatives: $As2Loss\,(Y, p) := -Y(p - 1) - (1 - Y)\,(p + \log(1 - p))$ (in this case, investors are more adverse to losses from false positives than losses from false negatives). Elliott et al.'s (2016) As1 loss function, defined as $As1Loss\,(Y, p) := -Y\,(\log(p) - p + 1) + (1 - Y)p$, does exactly opposite to the As2 loss function in the sense that losses from false negatives are weighted more heavily than losses from false positives.

Boosting with gradient boosted trees employed in this paper uses decision trees to approximate $F^*(\boldsymbol{x})$. Approximation by decision trees is the method to generate random partitions of the predictor space by a recursive splitting procedure. To give an example of decision trees, consider a dataset of two predictors (the trading volume $X = 100, 50, 150, 10$ and the interest rate $Z = 0.1\%, 0.5\%, 0.2\%, 0.7\%$). We are also provided with the conditional probabilities: $Prob\,(Y = 1 \mid X, Z) = S\,(F^*(X, Z)) \geq 0.5$ for $X \geq 100$ and $Z \leq 0.1\%$, and $Prob\,(Y = 0 \mid X, Z) < 0.5$ for $X < 100$ or $Z > 0.1\%$. Suppose that Y is predicted to be one if $Prob\,(Y = 1 \mid X, Z) \geq 0.5$ and zero otherwise. (Note that the threshold of 0.5 is a common choice for binary classification tasks and it can also be selected by cross validation, depending on the scoring function being used. However, we chose not to do that primarily due to constraint in computational capability. We believe that the obtained results can be significantly improved if the threshold is selected by cross validation.) A decision tree can then be grown as in the following diagram:



In the above diagram, data on the predictors are recursively divided into partitions (i.e., the observations of X are split into two subsets, and given each of these subsets, the observations of Z are then split accordingly). At the end of this process, the leaf nodes are labelled according to the predicted values of Y. In this example, the root node X and the splitting rule [defined by the thresholds (100 and 0.1%) of X and Z respectively] are essentially the parameters of $F^*(X, Z)$. The maximum depth of a tree (equal to two in this case) is the hyperparameter of $F^*(X, Z)$. Thus, one can predict any value of the outcome Y for given values of X and Z by simply using the optimal mapping $F^*(X, Z)$. This mapping can effectively be learnt from data available by growing many different decision trees corresponding to various mappings, then select the mapping that minimizes the loss function $L(Y, S(F(X, Z)))$ defined above.

In light of the above discussion, the gradient boosting algorithm combines many weak learners (which are decision trees in this case) such that each new tree corrects the error made by the previous one (which is the gradient of the loss function evaluated at the value of the previous tree).[5]

Random Forest is an ensemble of many decision trees described above. However, in RF, a node of each decision tree is selected to split if it achieves the maximum information gain (the information gain of a node after it is split up is defined as the difference between the entropy of the parent node and the expectation of the entropies of its child nodes – therefore, the information gain increases as the parent node becomes as impure as possible and/or the child nodes become as pure as possible). RF is among the most

[5]The gradient is zero when a tree achieves the minimum loss.

popular and successful off-the-shelf supervised ML methods. Fernández-Delgado, Cernadas, Barro, and Amorim (2014) documented in a large-scale comparison study that RF is the best performer among 100 classifiers across over 100 different datasets. An illustrative description of the gradient boosting machine (GBM) and RF used to predict the moving direction of price h-period ahead is provided in Section S.I of the SM.

Both GBM and RF have many hyperparameters that need to be determined [before training a model] via a procedure called *cross validation* that we will elaborate in Section 3 below. The list of hyperparameters used to cross-validate a GBM or RF model is tabulated in Section S.II.

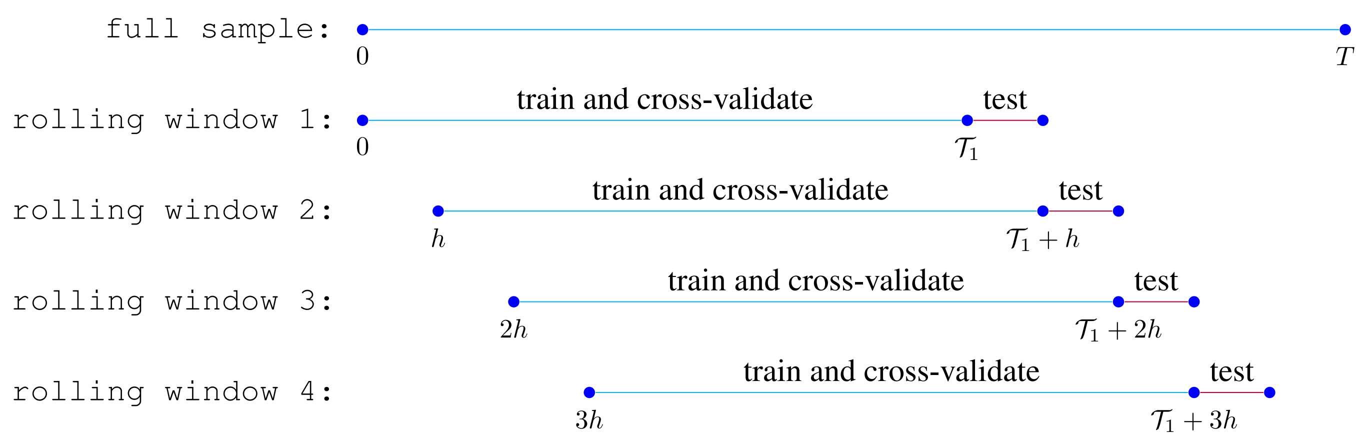
To explain predictions from a black-box ML model, we employ the SHAP framework proposed by Lundberg and Lee (2017) based on Shapley's (1953) values from cooperative game theory. The Shapley value measures the marginal contribution of each predictor towards the overall prediction performance of a ML model. Let $\nu(S_t)$ represent the probability forecast of Y_{t+1} using a subset of predictors at time t ($S_t \subset \mathcal{S}$, where $\mathcal{S}$ is the entire set of all predictors). The Shapley value of a predictor (say, i) at time t is then defined as

$$\varphi_{i,t} := \sum_{S_t \subseteq \mathcal{S} \setminus i} \frac{|\mathcal{S}|!\,(|\mathcal{S}| - |S_t| - 1)!}{|\mathcal{S}|!} \left(\nu \left(S_t \bigcup i \right) - \nu(S_t) \right),$$

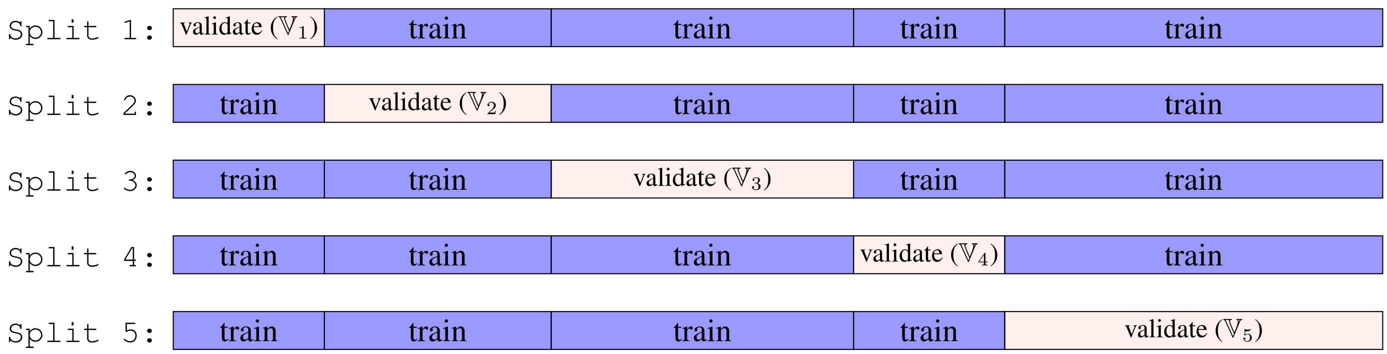
where $\mathcal{S} \setminus i$ contains all the elements of $\mathcal{S}$ except i, $|S|$ is the cardinality of a set, and $S_t \bigcup i$ is the union set of S_t and i. In the above equation, $\nu(S_t \bigcup i) - \nu(S_t)$ represents the marginal effect of the predictor i (relative to other predictors in S_t) towards the probability forecast. The weight of this marginal effect is given by $\frac{|\mathcal{S}|!(|\mathcal{S}|-|S_t|-1)!}{|\mathcal{S}|!}$ which is inversely proportional to the number of $|S_t|$-combinations in $\mathcal{S}$. This weighting scheme indicates that the marginal effect of a predictor relative to a small subset of the other predictors will receive less weight than the marginal effect of the same predictor relative to a larger subset (i.e., the marginal effect of a predictor in a large sub-model should be weighted more than in a small sub-model). The global Shapley value of the predictor i is then the time average of $\varphi_{i,t}$.

3 Recursive Forecasting Strategy

To evaluate the OoS performance of a predictive model over time, we construct M rolling windows from a given sample. We can then produce a h-period ahead forecast of the future price moving direction in the test set of each rolling window. In the end, we obtain M h-period ahead forecasts which can be used to compute the rolling OoS performance metrics (to be defined below). As illustrated in the following diagram, `rolling window 2` is formed by moving `rolling window 1` forward h observations, `rolling window 3` is formed by moving `rolling window 2` forward h observations, and so on. The number of rolling windows that can be created by this method is then $M = \left\lfloor \frac{T-\mathcal{T}_1}{h} \right\rfloor$.



Observations in each rolling window is then split into two parts (the first part used to train and cross-validate a model and the second part is used to test this trained model). Cross validation is an important device employed to reduce the risk of overfitting for a ML algorithm. We use the stratified k-fold cross validation, which is a variation of the k-fold cross validation that generates non-overlapping subsets (so-called folds) with roughly the same class ratio as the original set by a stratification procedure (see, e.g., Arlot and Celisse (2010)). The ML model is then trained on $k-1$ folds and validated on the fold left out. For example, the 5-fold cross validation splits the first part [consisting of observations used to train and validate a model] of a rolling window into five folds (four of which are used to train the model and the one left out is used to cross-validate the model) as in the following diagram:



We take the following steps to produce a h-period ahead forecast in each rolling window:

Step 1: For each split, `Split j`, $j = 1, \ldots, 5$, we estimate the model with the observations in the four folds used for training, and at the same time, use the fold left out for cross validation to choose an optimal combination of hyperparameter values so as to maximize an evaluation metric (*or* a scoring function) of the forecasts [of validation data]. We will define our scoring functions below. [6]

Step 2: Use the optimal hyperparameter values obtained in *Step 1* to re-estimate the model with the sample formed by concatenating the training and validation samples together, and this refitted model is then employed to produce a h-period ahead forecast.

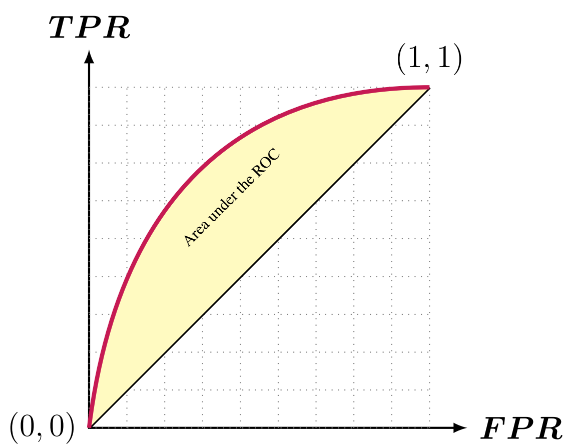
Let $\widehat{p}_{j,t+h}$ for all t in the 'validate' fold $\mathbb{V}_j$, $j = 1, \ldots, 5$, represent the [h-period ahead] probability predictions associated with the actual price moving directions $Y_{j,t+h}$, $j = 1, \ldots, 5$, for all indices t in the $k-1$ 'train' folds. The performance of these predictions in $\mathbb{V}_j$, $j = 1, \ldots, 5$, can be evaluated based on a statistical evaluation metric or an economic evaluation metric. Statistical evaluation metrics include, apart from the cross entropy score, the Brier score, the Boosting score, the As1 score, and the As2 score (which are associated with the loss functions defined earlier),

[6]Optimal hyperparameters (for instance, in LGBM, they include the boosting type, the max depth of each tree, the number of leaves in a leaf node, the learning rate, or the number of boosting iterations, as described in Table S.II.1) are often selected using a popular data-driven method called cross validation (CV). In CV, the joint domain of hyperparameters is divided into grid points (and each grid point is associated with a combination of hyperparameter values). The model is estimated using only the training data for each combination of hyperparameter values (corresponding to each grid point). This estimated model is then used to make predictions on the validating sample. The accuracy/profitability of these predictions is evaluated by a scoring function. The optimal hyperparameter values are thus the values that maximize this scoring function. Therefore, the main idea behind CV is to check the actual forecast performance of a method using a dataset that is different from the one used to train the model.

1. Accuracy rate: $Accuracy_j := \frac{TP_j + TN_j}{|\mathbb{V}_j|}$, where TP_j is the number of true positives (i.e., the number of positive directions correctly predicted) for observations in $\mathbb{V}_j$, TN_j is the number of true negatives (i.e., the number of negative directions correctly predicted) for observations in $\mathbb{V}_j$, and $|\cdot|$ denotes the cardinality of a set.

2. Precision score: $Precision_j := \frac{TP_j}{TP_j + FP_j}$, where FP_j is the number of false positives (i.e., the number of negative directions incorrectly predicted as positive directions) for observations in $\mathbb{V}_j$.

3. The area under the Receiver Operation Characteristic (ROC) curve: The ROC curve is the plot of the rate of true positives $TPR_j := \frac{TP_j}{TP_j + FN_j}$, where FN_j is the number of false negatives (i.e., the number of positive directions incorrectly predicted as negative directions) for observations in $\mathbb{V}_j$, against the rate of false positives $FPR_j := \frac{FP_j}{FP_j + TN_j}$ for $j = 1, \ldots, 5$. In the diagram below, if all the points $j = 1, \ldots, 5$ lie on the diagonal line, then observations are classified into two completely random classes. The top left corner of the diagram is the "ideal" point – a false positive rate of zero, and a true positive rate of one. Therefore, a larger area under the ROC curve means better directional predictions.



The predictions of price moving directions in each 'validate' fold, $\mathbb{V}_j$, $j = 1, \ldots, 5$, will be used to create a profitable trading strategy (which we will discuss in Section 4). We can then score the predictions of price moving directions based on profit/loss made by this trading strategy. We will employ the following economic scoring functions:

1. The gain to pain ratio (proposed by Schwager (2012, p. 513)):

$$GPR_j := \frac{\text{net return}_j}{\text{the absolute value of the sum of all losses}_j}, \ j = 1, \dots, 5,$$

where *net return* is the sum of all positive returns minus negative returns.

2. The Sharpe ratio: $SR_j := \frac{\text{average excess return}_j}{\text{standard deviation of returns}_j}$, $j = 1, \dots, 5$, where the average excess return is the mean of returns in excess of a risk-free rate. The Sharpe ratio is a measure of risk-adjusted return.

3. The Calmar Ratio (proposed by Young (1991)): $CR_j := \frac{\text{average excess return}_j}{DD_j}$, $j = 1, \dots, 5$, where the maximum drawdown $DD_j := \max_{t_1 < t_2 \in \mathbb{V}_j} \left(cumret_{j,t_1} - cumret_{j,t_2} \right)$, where $cumret_{j,t}$ is the cumulative return up to time t in $\mathbb{V}_j$.

4. The Sortino ratio (proposed by Sortino and van der Meer (1991)):

$$STR_j := \frac{\text{average excess return}_j}{\text{standard deviation of negative returns}_j},$$

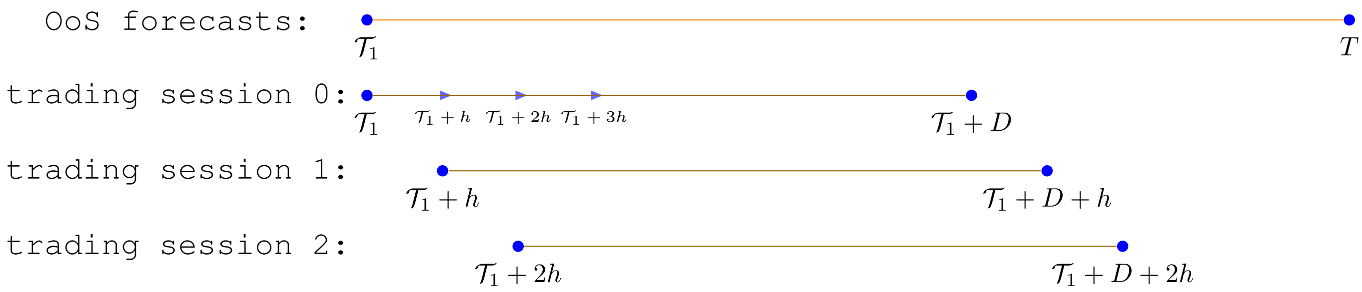
$j = 1, \dots, 5$,

5. The correlation between equity curve and perfect profit (CECPP): an equity curve is the plot of profit/losses over time being realized by a trading strategy [see, e.g., Kisela, Virdzek, and Vajda (2015)], and perfect profit is the potential profit that could be realized by an idealized trading strategy – buying/holding when the price is about to go up and selling when the price is about to go down. The CECPP ranges between -1 and 1 (a value of 1 suggests that the model can make very good predictions.)

4 Trading Strategies

We shall use the h-period ahead forecasts of price moving directions to construct our trading strategy. This strategy is intentionally simple, and we believe that there are more sophisticated strategies that can better use the probability predictions of price movements.

The rolling-window forecasting strategy [described in Section 3] creates $\left\lfloor \frac{T-\mathcal{T}_1}{h} \right\rfloor$ forecasts. To evaluate the performance of a trading strategy in changing economic environments, we define many trading sessions of length D (days) starting on different dates in the interval $[\mathcal{T}_1, T]$ as shown in the following diagram:



There is a maximum of $\lceil \frac{D}{h} \rceil$ trades in each trading session, $j \in \left\{0, \left\lfloor \frac{T-\mathcal{T}_1-D}{h} \right\rfloor \right\}$. For example, in trading session 0, we have $\lceil \frac{D}{h} \rceil$ h-period ahead forecasts at maximum with the first trade taking place at time $\mathcal{T}_1$, the second trade taking place at time $\mathcal{T}_1 + h$, the third trade taking place at time $\mathcal{T}_1 + 2h$, and so on. Afterwards, the whole position is liquidated at time $\mathcal{T}_1 + D$.

In each trading session $j \in \left\{0, \left\lfloor \frac{T-\mathcal{T}_1-D}{h} \right\rfloor \right\}$, the trader initiates the strategy by investing an initial endowment, say $W_{\mathcal{T}_1+jh}$, in a stock or an ETF (e.g., *SPY*) on day $d = \mathcal{T}_1 + jh$. The number of shares (assumed to be a decimal) that this trader can buy is then defined as

$$N_{\mathcal{T}_1+jh} := \begin{cases} \frac{W_{\mathcal{T}_1+jh}-c}{P_{\mathcal{T}_1+jh}} & if \quad c \text{ is the amount of fixed transaction cost,} \\ \frac{W_{\mathcal{T}_1+jh}}{P_{\mathcal{T}_1+jh}(1+c)} & if \quad c \text{ is a variable transaction cost as the percentage of the stock price} \end{cases}$$

Next, *on any other day* $d = \mathcal{T}_1 + jh + \ell h$, $\ell \in \left\{1, \left\lceil \frac{D}{h} \right\rceil \right\}$, if the predicted probability [that the stock price on day $d + h$ will move up] is greater than 0.5, there are two possibilities:

Possibility 1: If the position on the last day $d - h$ was in stock, the trader will keep on holding the stock. The number of shares held by the trader on day d is $N_d = N_{d-h}$. The total wealth assuming that the stock position is liquidated at this point in time is

$$W_d := \begin{cases} N_d P_d - c & if \quad c \text{ is the amount of fixed transaction cost,} \\ N_d P_d (1-c) & if \quad c \text{ is a variable transaction cost} \end{cases}$$

Possibility 2: If the position on the last day $d-h$ was in cash, the total wealth accrued by holding cash till day d is $W_d := W_{d-h} \prod_{m=1}^{h}(1+r^f_{d-m})$, where r^f_d is the daily risk-free interest rate available on day d. [Note that we use the daily risk-free interest rates obtained from Kenneth R. French - Data Library in our empirical study.] The trader then invests all this wealth in the stock. The number of shares that the trader can buy on day d can then be defined as

$$N_d := \begin{cases} \frac{W_d - c}{P_d} & if \quad c \text{ is the amount of fixed transaction cost,} \\ \frac{W_d}{P_d(1+c)} & if \quad c \text{ is a variable transaction cost} \end{cases}$$

On the day $d = \mathcal{T}_1 + jh + \ell h,\ \ell \in \left\{1, \left\lceil \frac{D}{h} \right\rceil \right\}$, if the predicted probability [that the stock price on day $d+h$ will move up] is less than 0.5, there are two possibilities:

Possibility 1: If the position on the last day $d-h$ was in stock, the trader will liquidate the stock position and put the proceeds in a savings account. The number of shares held on day d is $N_d = 0$. The total wealth of the trader on day d is

$$W_d := \begin{cases} N_{d-h} P_d - c & if \quad c \text{ is the amount of fixed transaction cost,} \\ N_{d-h} P_d (1-c) & if \quad c \text{ is a variable transaction cost} \end{cases}$$

Possibility 2: If the position on the last day $d-h$ was in cash, the trader will keep on holding cash. The number of shares held on day d is $N_d = N_{d-h} = 0$. The total wealth accrued by holding cash till day d is then $W_d := W_{d-h} \prod_{m=1}^{h}(1+r^f_{d-m})$.

On the last day of the trading session, the trader always liquidate the stock position if the current position is still in stock.

REMARK 4.1. *The buy-and-hold strategy is akin to a trading strategy where* $N_d = N_{d-h}$ *for all* $h > 0$. *Put it differently, this strategy is like using a classifier that always attaches probability one to a price increase.*

The following metrics are used to evaluate a trading strategy or to compare trading strategies in each trading session $j \in \left\{0, \left\lfloor \frac{T - \mathcal{T}_1 - D}{h} \right\rfloor \right\}$ (of D days) with $\left\lceil \frac{D}{h} \right\rceil$ trades:

1. The average number of trades (*ave_num_trades*) is the ratio of the number of trades (defined as $\#\left\{N_d > 0,\ d \in \left\{\mathcal{T}_1 + jh + \tau h :\ \tau = 0, \ldots, \left\lceil \frac{D}{h} \right\rceil - 1\right\}\right\}$) over the number of h-period ahead forecasts ($\left\lceil \frac{D}{h} \right\rceil$) in a trading session (*or* a holding period);

2. The percentage of winning trades (*percent_wins*) is the ratio of the number of winning trades (defined as $\#\left\{ret_d > 0,\ d \in \left\{\mathcal{T}_1 + jh + \tau h :\ \tau = 0, \ldots, \left\lceil \frac{D}{h} \right\rceil\right\}\right\}$, where $ret_d := \frac{W_d - W_{d-h}}{W_{d-h}}$) over the number of trades in a trading session;

3. The largest return (*largest_ret* $:= \max\left\{ret_d,\ d \in \left\{\mathcal{T}_1 + jh + \tau h :\ \tau = 0, \ldots, \left\lceil \frac{D}{h} \right\rceil\right\}\right\}$) and the smallest return (*smallest_ret* $:= \min\left\{ret_d,\ d \in \left\{\mathcal{T}_1 + jh + \tau h :\ \tau = 0, \ldots, \left\lceil \frac{D}{h} \right\rceil\right\}\right\}$);

4. The ratio of the average wealth won from winning trades over the average wealth lost from losing trades (*ratio_win_loss* $:= \frac{average\left\{ret_d > 0,\ d \in \left\{\mathcal{T}_1 + jh + \tau h:\ \tau = 0, \ldots, \left\lceil \frac{D}{h} \right\rceil\right\}\right\}}{\left|average\left\{ret_d < 0,\ d \in \left\{\mathcal{T}_1 + jh + \tau h:\ \tau = 0, \ldots, \left\lceil \frac{D}{h} \right\rceil\right\}\right\}\right|}$);

5. The maximum number of consecutive winners (*max_num_consecutive_winners*) is the maximum number of positive returns realized in a row, and the maximum number of consecutive losers (*max_num_consecutive_losers*) is the number of negative returns realized in a row;

6. The annualized return (*annualized_return* $:= \left(\prod_{\tau=0}^{\left\lceil \frac{D}{h} \right\rceil} \left(1 + ret_{\mathcal{T}_1 + jh + \tau h}\right)\right)^{\frac{252}{D}} - 1$, where 252 is the average number of trading days per year in the NYSE and NASDAQ);

7. The annualized excess return (*annualized_excess_return*) is the difference between the annualized return of a trading strategy and the annualized return of the buy-and-hold strategy (which is defined as $(1 + \text{the simple return of the buy-and-hold strategy})^{\frac{252}{D}} - 1$);

8. The annualized standard deviation (*annualized_std* $:= \sqrt{\frac{252}{h} * \widehat{Var}\left(ret_d\right)}$, where $\widehat{Var}(\cdot)$ represents the sample variance);

9. The maximum drawdown (*max_dd*), Schwager's (2012) gain to pain ratio, the Sharpe ratio, the Calmar ratio, the Sortino ratio, and the correlation between equity curve and perfect profit (CECPP) (as defined in Section 3);

10. Morningstar's risk-adjusted rating (*mrar*) is essentially a utility function with constant relative risk aversion whose degree of risk aversion is set to two. Ingersoll, Spiegel, Goetzmann, and

Welch (2007) shows that this measure is robust to dynamic manipulation (e.g., using derivatives to create a nonlinear payoff can lead to an inaccurate performance measure when evaluated by other metrics, but not by *mrar*). This performance measure is defined as

$$mrar := \left(\frac{1}{\lceil \frac{D}{h} \rceil} \sum_{\tau=0}^{\lceil \frac{D}{h} \rceil} \frac{1}{\left(1 + ret_{\mathcal{T}_1+jh+\tau h} - r^f_{\mathcal{T}_1+jh+\tau h}\right)^2} \right)^{-\frac{1}{2}} - 1.$$

5 Empirical Results

We use the daily data on *SPDR S&P 500 ETF Trust* (*SPY*) which is one of the most popular ETFs that tracks the market capitalization-weighted S&P 500 stock market index. We also use daily (fundamental) financial data [e.g., dividend-price ratio (*dp*), the 3-month T-Bill rate (*T-bill*), ICE BofA US corporate bond total return index (*corp_bond_ind*), term spread (*ts*), S&P 500 earnings per share (*eps*), default spread (*df*), and the 10-Year breakeven inflation rate (*inflation*)]. [7] A detailed description of all the data variables is provided in Table S.V.1 (in the SM). The entire time period of our dataset ranges from 3/30/2011 to 7/15/2022. First, we use rolling windows of size $\mathcal{T}_1 = 1000$ to generate 1818 [one-day ahead] forecasts of the moving directions of next day prices (the date of the first forecast is 4/9/2015 and the date of the last forecast is 7/15/2022). We then create multiple trading sessions based on those forecasts: There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days. We consider five levels of fixed transaction cost (\$0.05, \$0.1, \$0.5, \$1, and \$5) and an initial endowment of \$1000 for each holding period with length $D = 100$ or 200 days. We do not report the results of trading sessions with variable transaction cost in this section as they are similar to those obtained for fixed transaction cost.

Two sets of predictors are used to predict price moving directions:

Dataset I: Financial variables together with technical indicators constructed from the last {14, 24, 34, 54, 104}-day price/volume data comprise 181 predictors in total;

Dataset II: Dataset I together with candlestick chart (price) patterns comprise 239 predictors in total.

[7] All data were obtained from Yahoo! Finance, FRED, and Nasdaq Data Link.

Note that the choice of these five number of days in a technical indicator is quite ad hoc. The most commonly used numbers of days in existing studies range from 2 to 200 days for stocks [see, e.g., Brock et al. (1992); Sullivan et al. (1999)] and from 1 to 60 days for exchange rates [see, e.g., Chang and Osler (1999)]. These numbers of days may vary from one paper to another, and there is no general guideline for their optimal selection. In fact, they can serve as hyperparameters in a ML algorithm that can be optimally selected by cross validation. However, this cross validation process will be extremely time-consuming as we have over 1800 one-day ahead forecasts to make (thus, we need to repeat the same number of cross-validations that we have performed in this paper over 1800 times).

We implement the trading strategy based on one-day ahead forecasts by the LGBM models trained with one of the loss functions $\{CE,\ Brier,\ Boost,\ As1,\ As2\}$ and cross-validated with one of the scoring functions {*Accuracy*, *AUC*, *Schwager's gain/pain ratio*, *Calmar ratio*, *Sharpe ratio*, *Sortino ratio*, *CECPP*}. We also implement the trading strategy based on forecasts by the RF models cross-validated with one of these seven scoring functions. We will also consider trading based on multi-day ahead forecasts in Section 5.5.

We shall present our findings by using both box plots (which consist of the median marked by the line within a box, the first and third quartiles which are the edges of the box, and the minimum and maximum values depicted by the two whiskers below and above the box respectively) and tables (tabulating the median and interquartile range (IQR)) of the values [of each performance measure] of trading sessions based on one-day ahead forecasts generated by a ML algorithm.

5.1 Which scoring function generates the best strategy for a given loss function?

In this subsection, we report the performance of the trading strategy with fixed transaction cost by four metrics (*namely*, the annualized excess return, the Sharpe ratio, Morningstar's risk-adjusted rating, and the percentage of winning trades) as the other metrics will give a quite similar information. We also employ a bootstrap RC method (explained in Section S.III of the SM) to compare the performance of trading methods. The key takeaways here are: (i) the bootstrap RC p-values for the test of the null hypothesis that a (benchmark) trading method performs at least as well as all the other ones in the collection of 70 LGBM-based trading methods show that the *CECPP* or *Sharpe ratio* scoring function tends to yield higher annualized excess returns than the other scoring functions across sets of predictors and loss functions (i.e., Figure A.V.1 shows that the *CECPP* gives five p-values greater than 10%, the *Sharpe ratio* gives four p-values greater than 10%, the *Calmar ratio* gives about three p-values greater than 10%, and all the other scoring functions give about two p-values greater than 10%). Also, for RF-based trading methods, the bootstrap RC p-values in Figure A.V.2 show a clear evidence that economic scoring functions can generate higher annualized excess returns than statistical scoring functions across sets of predictors; (ii) the trading strategy using the *CECPP* or *Sharpe ratio* scoring function yields better performance than the other scoring functions in terms of median and IQR across the two sets of predictors considered; and (iii) the predictive power of candlestick chart patterns can be leveraged in a trading strategy to yield higher annualized excess returns than the case where only financial variables and technical indicators are used (i.e., Figure A.V.2 shows that there are much higher p-values for the benchmark RF-based trading methods employing all the predictors.)

Now we start by comparing the annualized excess returns of trading sessions across the seven scoring functions for each loss function. Figures A.I.1 and A.I.2 present the box plots of the annualized excess returns of trading sessions using the fixed transaction cost strategy investing in *SPY* for 100 days based on one-day ahead forecasts made by the LGBM models trained with a loss function and cross-validated with seven different scoring functions. Figures A.I.1(a) and A.I.2(a) clearly show that, using the *CE* loss, the *Sharpe ratio* scoring function yields a higher median excess return than all the other scoring functions. Figures A.I.1(b) and A.I.2(b) demonstrate that, using the *Brier* loss, the CECPP yields a much higher median excess return than the other scoring functions for different amounts of transaction cost and holding periods across the two sets of predictors (which is corroborated in Table A.I.1 that reports the median (and IQR) of the annualized excess returns of trading sessions with the fixed transaction cost strategy investing in SPY based on one-day ahead forecasts by the LGBM models trained with the *Brier* loss function and cross-validated with seven different scoring functions for different amounts of transaction cost across various holding periods). Figures A.I.1(c) and A.I.2(c) show that, using the *Boost* loss, all the scoring functions yield negative annualized excess returns and there does not seem to be the best scoring function. Figures A.I.1(d) and A.I.2(d) demonstrate that, using the *As1* loss, the *Sharpe ratio* scoring function yields the lowest IQR of excess returns. Figures A.I.1(e) and A.I.2(e) show that, using the *As2* loss, the *Sharpe* *ratio* scoring function yields the greatest median excess return overall across the two sets of predictors.

We observe roughly the same patterns explained above when we compare the Sharpe ratios of trading sessions across the seven scoring functions for each loss function as shown by the box plots in Figures S.VI.1 and S.VI.2 (in the SM). For the *Brier* or *Boost* loss, the *Calmar ratio* scoring function can yield a slightly higher median Sharpe ratio value than both the *CECPP* and *Sharpe ratio* scoring functions (as confirmed in Table S.VI.1). However, the former can have more negative outliers than the latter.

Next, we compare Morningstar's risk-adjusted ratings of trading sessions across the seven scoring functions for each loss function. The box plots in Figures S.VI.3 and S.VI.4 show that: (a) for the *CE* loss, the *Sharpe ratio* scoring function clearly generates a higher median *mrar* value than all the other scoring functions while the IQR values are quite similar; (b) for the *Brier* loss, the *CECPP* and *Calmar ratio* scoring functions yield quite similar median *mrar* values as confirmed in Table S.VI.2; (c) for the *Boost*, *As1*, or *As2* loss function, there is no clear evidence that the *Sharpe ratio* or *CECPP* scoring function generates a higher median *mrar* than the other scoring functions.

Finally, we examine the percentages of winning trade of trading sessions across the seven scoring function for each loss function. The box plots in Figures S.VI.5 and S.VI.6 show that the *CECPP* or *Sharpe ratio* scoring function can give significantly lower IQRs of *percent_wins* values than all the other scoring functions when the *Brier* or *As1* loss function is used for training (cf. Table S.VI.3). However, this is not the case for the other loss functions.

The outperformance of the *CECPP* and *Sharpe ratio* scoring functions, which is established by GBM, can be corroborated by RF. Note that, as explained above, RF employs an impurity measure to split a parent node to leaf nodes in a decision tree [so that the information gain is maximized]. Figure A.I.3 and Table A.I.2 show box plots of the annualized excess returns of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day forecasts by RF models cross-validated with seven different scoring functions. Two main conclusions can be drawn from these figures: (a) the *CECPP* and *Sharpe ratio* scoring functions clearly provide superior annualized excess returns in terms of both median and IQR, and (b) with these scoring functions, the trading strategy using Dataset II are more likely to outperform the buy-and-hold strategy in terms of annualized returns than the one using Dataset I.

The box plots in Figure S.VI.7 and Table S.VI.4 show that the *CECPP* scoring function can generate slightly better Sharpe ratio values in terms of either median or IQR than the other scoring functions, especially at higher levels of transaction cost. The box plots in Figure S.VI.8 and Table S.VI.5 show the same patterns observed for Morningstar's risk-adjusted rating. The box plots in Figure S.VI.9 and Table S.VI.6 suggest that both the *CECPP* and *Sharpe ratio* scoring functions can deliver slightly better *percent_wins* values in terms of median and IQR than the other scoring functions, but this outperformance does not seem significant. It is also important to note at this point that the median *percent_wins* values and the median *annualized_excess_return* values obtained from LGBM (as reported in Tables S.VI.3 and A.I.1 respectively) are significantly higher than those obtained from RF.

5.2 Which loss function generates the best strategy for a given scoring function?

In this subsection, we report the performance of the fixed transaction cost strategy across five loss functions (*namely*, *CE*, *Brier*, *Boost*, *As1*, and *As2*) for a given scoring function. The key takeaway is that the *Brier* or *As1* loss function can yield a better performance [measured by the annualized excess returns] in terms of median and IQR than the other loss functions across the two sets of predictors that we consider.

We start by comparing the annualized excess returns of trading sessions across the above five loss functions for each scoring function. Figures A.II.1 and A.II.2 show box plots of the annualized excess returns of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the LGBM models trained with five different loss functions for each scoring function. When using *AUC*, *Calmar ratio*, or *Sharpe ratio* scoring function for cross validation, the *As1* loss function clearly provides better *annualized_excess_return* values in terms of median and IQR than all the other loss functions. However, As suggested by Table A.II.1 (which reports the median (and IQR) of the annualized excess returns of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the LGBM models trained with five different loss functions for each scoring function), the *Brier* loss function yield a much better median *annualized_excess_return* than the other loss functions when the *CECPP* scoring function is used for cross validation. The *As1* loss function seems to provide the smallest IQR for every scoring function.

Next, we compare the Sharpe ratios across the trading methods. The box plots in Figures S.VII.1 and S.VII.2 show that, when using *CECPP* scoring function for cross validation, the *Brier* and *As1* loss functions still generate a higher median Sharpe ratio value than the other loss functions (as confirmed in Table S.VII.1). However, this may not be a clear case for the other scoring functions. The same conclusion can also be reached for Morningstar's risk-adjusted ratings of the trading sessions as shown by the box plots in Figures S.VII.3 and S.VII.4 and Table S.VII.2,.

Finally, we turn to the percentages of winning trades. The box plots in Figures S.VII.5 and S.VII.6 show that, for every scoring function, the *Brier* and *As1* loss functions yield the lowest IQRs at the cost of higher medians over different amounts of transaction cost and holding periods across the two sets of predictors. Tables S.VII.3 and A.II.1 suggest that a combination of a loss function and a scoring function that generate a low variation of *percent_wins* can lead to an outperforming *annualized_excess_return* in terms of median or IQR.

5.3 Performance of the trading strategy as transaction cost increases

Figures A.III.1 and A.III.2 show the box plots of the annualized excess returns of 1718 trading sessions with the fixed transaction cost strategy investing in SPY for 100 days based on one-day ahead forecasts by a LGBM model over two levels of transaction cost ($0.05 and $0.5). We can see that the buy-and-hold strategy will be more likely to generate higher returns than our trading strategy as transaction cost increases, and the IQR of the annualized excess returns of the trading strategy is not very susceptible to changes in transaction costs. This conclusion holds for every loss and scoring function across the two sets of predictors and holding periods (as corroborated in Tables A.III.1, A.III.2, A.III.3, and A.III.4 which report the median (and IQR) of the annualized excess returns of trading sessions based on one-day ahead forecasts by the LGBM models trained with the *Brier* or *As1* loss function and cross-validated with seven scoring functions for different amounts of transaction cost across various holding periods).

The box plots in Figures S.VIII.1 and S.VIII.2 show that the Sharpe ratio of the trading strategy will be more likely to decrease as transaction cost increases for every set of predictors. This conclusion holds for every loss and scoring function across the two sets of predictors and holding periods (as corroborated in Tables S.VIII.1, S.VIII.2, S.VIII.3, and S.VIII.4). Moreover, the box plots in Figures S.VIII.3 and S.VIII.4 suggest that the percentage of winning trades does not much depend on transaction cost (which is true since only the forecasts of the moving directions of prices determine whether or not a trade happens)

These empirical findings can also be verified by trading sessions based on forecasts made by RF. The box plots in Figure A.III.3 shows the superior performance of the buy-and-hold strategy in terms of its likelihood to produce high returns as transaction costs increase. Tables A.III.5 and A.III.6 confirm this tendency for every scoring function and trading session across the two sets of predictors. Moreover, the Sharpe ratio of our trading strategy can decrease when either the transaction cost or holding period increases. It is also important to notice that the RF models cross-validated with the *CECPP* scoring function can generate positive median Sharpe ratio values for a very high level of transaction cost and a long holding period across the two sets of predictors; and this is not the case for the RF models cross-validated with all the other scoring functions (see Figure S.VIII.5 and Tables S.VIII.5 and S.VIII.6). The percentage of winning trades also does not significantly depend on transaction cost in this case (see Figure S.VIII.6 and Tables S.VIII.7 and S.VIII.8).

5.4 Performance of the trading strategy over time

Figures S.IX.1 and S.IX.2 show the box plots of the performance statistics of 1718 trading sessions with the fixed transaction cost strategy investing in SPY for 100 days based on one-day ahead forecasts by a LGBM model trained using a set of predictors *Dataset I* or *Dataset II*. We find that, by using the LGBM to forecast price moving directions, (i) our trading strategy is more likely to generate a higher profit than holding cash in a savings account (to earn interests) as indicated by the positive first quartiles/medians of the Sharpe ratio, Sortino ratio, and Morningstar's risk-adjusted rating; (ii) the performance of our strategy is not attributable to randomness since the percentages of winning trades, the win/loss ratios, and the CECPP values all have medians above 50% (and Schwager's gain/pain ratio has a highly positive median) when using the *Brier* or *As1* loss function for training and the *CECPP* or *Sharpe ratio* scoring function for cross validation. To corroborate this second point, we also conduct a placebo-type experiment (detailed in Section S.IV) by generating logarithmic closing stock prices from a random walk (RW) and drawing random volume data from an integer uniform distribution for the same time period (from 3/30/2011 to 7/15/2022) of the *SPY* dataset. This experiment shows that the performance of the trading strategy based on predictions of the directions of entirely random price changes tends to be much more random than that based on predictions of the directions of price changes with some predictive content; and (iii) the trading strategy can yield higher annualized returns than the buy-and-hold strategy in several periods of time as suggested by the box plots of annualized excess returns and standard deviations. To elaborate more on the last point (iii), we also draw line plots of all the performance statistics obtained by using the *Brier* or *As1* loss function for training and the *CECPP* or *Sharpe ratio* scoring function for cross validation in Figures S.IX.3, S.IX.4, S.IX.5, and S.IX.6. We now describe each of the performance statistics plotted in those figures:

- The average numbers of trades look quite stable across all the trading sessions (i.e., all are above 50% except in the sessions ending from mid-2016 till early 2017);
- The percentages of winning trades tend to decline for the trading sessions ending from early 2020 onwards (these sessions are associated with significantly high volatility and bear markets, proxied by the VIX index). However, Tables S.IX.1 and S.IX.2 suggest that the median *percent_wins* values are well above 50% across all the levels of transaction cost and holding periods;
- The trading strategy can yield very high maximum returns as well as very low minimum returns during periods of very high volatility;
- The win/loss ratios tend to decline very steeply when entering a period of high volatility (such as in the trading sessions ending around early 2020). The median win/loss ratio values can exceed 80% for several levels of transaction cost across the two holding periods, as indicated in Tables S.IX.1 and S.IX.2;
- The maximum numbers of consecutive winners tend to decline in the trading sessions ending from early 2020 onwards while the maximum numbers of consecutive losers tend to peak during periods of high volatility (such as in the sessions ending around early 2019 or early 2020);
- The annualized excess returns vary around zero in the trading sessions ending before early 2020. The sessions ending from February 2020 to late June 2020 seem to enjoy high positive excess returns (simply because the stock prices are very low during those sessions). The annualized excess returns decline steeply in the trading sessions ending around early July 2020, then gradually recover afterwards;

- The annualized standard deviations and maximum drawdowns of annualized excess returns spike quite high in the trading sessions ending from early 2020 to mid 2020 (when the VIX index is high);
- Schwager's gain/pain, Morningstar's risk-adjusted rating, the Calmar, Sharpe, and Sortino ratios are positive for most of time, and they all decline steeply in the trading sessions ending around early 2020 when the VIX index is extremely high;
- The CECPP is mostly above 50% except in the trading sessions where the VIX index is very high, implying that our trading strategy is highly correlated with the perfect profit trading strategy in every single trading session.

The above results can also be confirmed by employing the same trading strategy based on forecasts generated by RF. Figures A.IV.1 and A.IV.2 present the box plots of the performance statistics of trading sessions based on one-day ahead forecasts by RF models cross-validated with two scoring functions (i.e., the *Sharpe ratio* and *CECPP* scoring functions) for two different sets of predictors (i.e., *Dataset I* and *Dataset II*). There is a stronger evidence that our strategy does not depend on randomness as the first quartiles of Schwager's gain/pain ratio, Sharpe ratio, and Sortino ratio are all positive and the win/loss ratio can approach one for the two sets of predictors when the *CECPP* or *Sharpe ratio* scoring function is used for cross validation. Interestingly, as mentioned earlier, RF can leverage information in the candlestick chart patterns to make forecasts that can be used to generate better annualized excess returns than when using only financial variables and technical indicators.

The line plots of the performance statistics of trading sessions in Figures A.IV.3, A.IV.4, S.IX.7, and S.IX.8, for the most part, corroborate the above nine empirical findings drawn for all the performance statistics obtained from the trading strategy based on forecasts by RF models cross-validated with the *CECPP* or *Sharpe ratio* scoring function. The only difference is that the trading sessions with positive annualized excess returns extend till around early 2021, the annualized excess returns (in the trading sessions ending after that) then decline slightly before starting to climb up in the later trading sessions. Table A.IV.1 suggests that, when cross-validating with the *CECPP* scoring function, the median percentages of winning trades are always above 50% and they decrease with the amount of transaction cost across the two holding periods and the two sets of predictors considered. The median win/loss ratio values are above

80%, which are even greater than those obtained with the LGBM model. The median annualized excess returns are consistently higher for RF models using both technical indicators and candlestick chart patterns than for RF models using only technical indicators (in this case, the probability that the annualized excess return is greater than zero is slightly above 50%). The median CECPP values can be close to 80% at a low level of transaction cost.

Table S.IX.3 show that, when cross-validating with the *Sharpe ratio* scoring function, the median percentages of winning trades, win/loss ratio values, and CECPP values are quite comparable with those obtained by using the *CECPP* scoring function in Table A.IV.1. However, the median annualized excess returns are not significantly different across the two sets of predictors when using the *Sharpe ratio* scoring function for cross validation.

5.5 The performance of the trading strategy across forecasting horizons

In the previous sections, we have presented a strong empirical evidence that our simple trading strategy based on the *one-period* ahead predictions of future price moving directions generated by ML algorithms trained with an asymmetric loss function and cross-validated with an economically motivated scoring function can outperform the buy-and-hold strategy over 50 percent of the time. The final question that we want to address is: Can the performance of this trading strategy improve if we trade less often in each trading session of the same length as before?

Figure A.VI.1 shows the risk-adjusted annualized excess returns generated by our trading strategy based on the forecasts [of price moving directions over multiple horizons ($h \in \{1, \ldots, 12\}$)] made by RF models cross-validated with four scoring functions (*CECPP*, *Sharpe ratio*, *Schwager's gain/pain ratio*, and *AUC*). The bottom line is that the annualized excess return is more likely to be positive when trades are based on shorter-horizon forecasts (say, $h \in \{1, \ldots, 4\}$). When trading based on long-horizon forecasts (say, $h > 4$), there is a high probability that our trading strategy underperforms the buy-and-hold strategy across the two sets of predictors. One reason for this underperformance can be that the predictors used in this study do not have a strong predictive power to predict many days ahead. This finding confirms that technical analysis is a useful toolkit only for short-term traders/investors. Figure S.XII.1 also shows that the same finding also holds for the risk-adjusted annualized excess returns generated by the trading strategy based on multi-horizon forecasts generated by LGBM models trained with the *Brier* or *As2* loss function and cross-validated with the *CECPP* or *Sharpe ratio* scoring function (which is shown in Section 5.1 to be the best scoring function).

6 Conclusion

There is a large literature on the role of technical analysis in informing investors/traders about the moving direction of future prices of a financial asset. As reviewed in the Introduction, empirical evidences on the predictive power of technical indicators and price patterns are quite mixed. This paper contributes to the literature by employing one of the most successful supervised machine learning algorithms (i.e., [tree-based] Gradient Boosting Machine and Random Forest) to explore a large set of all popular technical indicators and candlestick chart patterns for an evidence of stock return predictability. We have established that: (1) our trading strategy based on forecasts of the moving directions of future prices can generate annualized returns above the annualized returns of the buy-and-hold strategy in slightly over 50% of all the trading sessions at best when transaction cost is low. Most of positive annualized excess returns are realized in the trading sessions created during periods of high volatility. This finding suggests that one may use ML to trade an aggregate stock market index in periods of high volatility while using the simple buy-and-hold strategy in periods of low volatility; (2) trading profit drawn from the technical analysis predictions of the moving directions of future prices is not something entirely random; (3) while we found no strong evidence in favour of using asymmetric loss functions in the out-of-sample analysis, there is an evidence that some economic scoring function can generate more economically significant predictions than the others. In particular, the CECPP scoring function seems to be an optimal choice as it is evaluated with reference to the perfect profit strategy (that is, the strategy assuming that one can accurately predict the moving direction of future prices); (4) several candlestick chart patterns may have a strong predictive power that can be effectively leveraged by RF to yield higher annualized excess returns than when we use only the financial variables and technical indicators as predictors; and (5) the trading strategy based on long-horizon forecasts may be less profitable than the simple buy-and-hold strategy as it is more difficult to predict the price moving direction with technical analysis many days ahead.

References

Allen, F. and R. Karjalainen (1999). Using genetic algorithms to find technical trading. *Journal of Financial Economics 51*(2), 245–271.

Ang, A. and G. Bekaert (2007). Stock return predictability: Is it there? *Review of Financial Studies 20*(3), 651–707.

Arlot, S. and A. Celisse (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys 4*, 40–79.

Avramov, D., S. Cheng, and L. Metzker (2022). Machine learning versus economic restrictions: Evidence from stock return predictability. forthcoming in *Management Science*.

Bajgrowicz, P. and O. Scaillet (2012). Technical trading revisited: False discoveries, persistence tests, and transaction costs. *Journal of Financial Economics 106*(3), 473–491.

Baker, M. and J. Wurgler (2007). Investor sentiment in the stock market. *Journal Of Economic Perspectives 21*(2), 129–152.

Barberis, N., A. Shleifer, and R. Vishny (1998). A model of investor sentiment. *Journal of Financial Economics 49*(3), 307–343.

Bikhchandani, S., D. Hirshleifer, and I. Welch (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy 100*(5), 992–1026.

Breiman, L. (1998). Arcing classifier. *Annals of Statistics 26*(3), 801–849.

Breiman, L. (1999). Prediction games and arcing algorithms. *Neural Computation 11*(7), 1493–1517.

Breiman, L. (2001). Random forests. *Machine Learning 45*(1), 5–32.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review 78*(1), 1–3.

Brock, W., J. Lakonishok, and B. LeBaron (1992). Simple technical trading rules and the stochastic properties of stock returns. *Journal of Finance 47*(5), 1731–1764.

Brown, D. P. and R. H. Jennings (1989). On technical analysis. *Review of Financial Studies 2*(4), 527–551.

Brown, S. J., W. N. Goetzmann, and A. Kumar (1998). The Dow theory: William Peter Hamilton's track record reconsidered. *Journal of Finance 53*(4), 1311–1333.

Buja, A., W. Stuetzle, and Y. Shen (2005). Loss functions for binary class probability estimation and classification: structure and applications. mimeo.

Campbell, J. Y. and R. J. Shiller (1988). Stock prices, earnings and expected dividends. *Journal of Finance 43*(3), 661–676.

Campbell, J. Y. and S. B. Thompson (2008). Predicting excess stock returns out of sample: can anything beat the historical average? *Review of Financial Studies 21*(4), 1509–1531.

Chang, K. P. H. and C. L. Osler (1999). Methodical madness: technical analysis and the irrationality of exchange-rate forecasts. *Economic Journal 109*(458), 636–661.

Daniel, K., D. Hirshleifer, and A. Subrahmanyam (1998). Investor psychology and security market under- and overreactions. *Journal of Finance 53*(6), 1839–1885.

de Long, B. J., A. Shleifer, L. H. Summers, and R. J. Waldmann (1990). Positive feedback investment strategies and destabilizing rational speculation. *Journal of Finance 45*(2), 379–395.

DeBondt, W. F. M. and R. Thaler (1985). Does the stock market overreact? *Journal of Finance 40*(3), 793–805.

Elliott, G., D. Ghanem, and F. Krüger (2016). Forecasting conditional probabilities of binary outcomes under misspecification. *Review of Economics and Statistics 98*(4), 742–755.

Fama, E. (1965). The behavior of stock market prices. *Journal of Business 38*(1), 34–105.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance 25*(2), 383–417.

Fama, E. F. (1991). Efficient capital markets: II. *Journal of Finance 46*(5), 1575–1617.

Farmer, L. E., L. Schmidt, and A. Timmermann (2022). Pockets of predictability. forthcoming in *Journal of Finance*.

Fernández-Delgado, M., E. Cernadas, S. Barro, and D. Amorim (2014). Do we need hundreds of classifiers to solve real world classification problems? *The Journal of Machine Learning Research 15*(1), 3133–3181.

Frazzini, A. (2006). The disposition effect and underreaction to news. *Journal of Finance 61*(4), 2017–2046.

Freund, Y. and R. E. Schapire (1996). Experiments with a new boosting algorithm. In L. Saitta (Ed.), *Proceedings of the Thirteenth International Conference on Machine Learning (ICML 1996)*, San Francisco, California, pp. 148–156. Morgan Kaufmann.

Freund, Y. and R. E. Schapire (1999). A short introduction to boosting. *Journal of Japanese Society for Artificial Intelligence 14*(5), 771–780.

Friedman, J. H. (2001a). Greedy function approximation: A gradient boosting machine. *Annal of Statistics 29*(5), 1189–1232.

Friedman, J. H. (2001b). Greedy function approximation: a gradient boosting machine. *Annals of statistics 5*(29), 1189–1232.

Gençay, R. (1996). Nonlinear prediction of security returns with moving average rules. *Journal of Forecasting 15*(3), 165–174.

Gneiting, T. and A. E. Raftery (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association 102*(477), 359–378.

Good, I. J. (1952). Rational decisions. *Journal of the Royal Statistical Society, Series B 14*(1), 107–114.

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

Han, Y., G. Zhou, and Y. Zhu (2016). A trend factor: Any economic gains from using information over investment horizons? *Journal of Financial Economics 122*(2), 352–375.

Ho, T. K. (1995). Random decision forests. *Proceedings of the 3rd International Conference on Document Analysis and Recognition 1*(1), 278–282. Montreal, QC, Canada.

Hong, H. and J. Stein (1999). A unified theory of underreaction, momentum trading and overreaction in asset markets. *Journal of Finance 54*(6), 2143–2184.

Hsu, P.-H., M. P. Taylor, and Z. Wang (2016). Technical trading: Is it still beating the foreign exchange market? *Journal of International Economics 102*, 188–208.

Ingersoll, J., M. Spiegel, W. Goetzmann, and I. Welch (2007). Portfolio performance manipulation and manipulation-proof performance measures. *Review of Financial Studies 20*(5), 1503–1546.

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

Jiang, J., B. T. Kelly, and D. Xiu (2022). (Re-)Imag(in)ing price trends. forthcoming in *Journal of Finance*.

Julian Andrada-Félix, F. F.-R. (2008). Improving moving average trading rules with boosting and statistical learning methods. *Journal of Forecasting 27*(5), 433–449.

Kelly, B. T., S. Malamud, and K. Zhou (2023). The virtue of complexity in return prediction. forthcoming in *Journal of Finance*.

Kisela, P., T. Virdzek, and V. Vajda (2015). Trading the equity curves. *Procedia Economics and Finance 32*, 50–55.

Lander, J., A. Orphanides, and M. Douvogiannis (1997). Earnings forecasts and the predictability of stock returns: Evidence from trading the s&p. *Journal of Portfolio Management 23*(4), 24–35.

LeBaron, B. (1999). Technical trading rule profitability and foreign exchange intervention. *Journal of International Economics 49*(1), 125–143.

Li, J. and I. Tsiakas (2017). Equity premium prediction: The role of economic and statistical constraints. *Journal of Financial Markets 36*, 56–75.

Lo, A. W. (2004). The adaptive markets hypothesis. *Journal of Portfolio Management 30*(5), 15–29.

Lo, A. W. (2005). Reconciling efficient markets with behavioral finance: The adaptive markets hypothesis. *Journal of Investment Consulting 7*(2), 21–44.

Lo, A. W. and J. Hasanhodzic (2010). *The Evolution of Technical Analysis: Financial Prediction from Babylonian Tablets to Bloomberg Terminals*. New Jersey: John Wiley & Sons.

Lo, A. W., H. Mamaysky, and J. Wang (2000). Foundations of technical analysis: computational algorithms, statistical inference, and empirical implementation. *Journal of Finance 55*(4), 1705–1765.

Lucas, R. E. (1978). Asset prices in an exchange economy. *Econometrica 46*(6), 1429–1445.

Lundberg, S. M. and S.-I. Lee (2017). A unified approach to interpreting model predictions. In U. von Luxburg, I. Guyon, S. Bengio, H. Wallach, and R. Fergus (Eds.), *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, California, USA, pp. 4768–4777. Curran Associates Inc.

Mandelbrot, B. (1966). Forecasts of future prices, unbiased markets, and “martingale" models. *The Journal of Business 39*(1), 242–255.

Moskowitz, T. J., Y. H. Ooi, and L. H. Pedersen (2012). Time series momentum. *Journal of Financial Economics 104*(2), 228–250.

Murphy, J. J. (1999). *Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications*. New Jersey: Prentice Hall Press.

Neely, C., P. Weller, and R. Dittmar (1997). Is technical analysis in the foreign exchange market profitable? A genetic programming approach. *Journal of Financial and Quantitative Analysis 32*(4), 405–426.

Neely, C. J., D. E. Rapach, J. Tu, and G. Zhou (2014). Forecasting the equity risk premium: the role of technical indicators. *Management Science 60*(7), 1772–1791.

Panopoulou, E. and I. Souropanis (2019). The role of technical indicators in exchange rate forecasting. *Journal of Empirical Finance 53*, 197–221.

Pesaran, M. H. and A. Timmermann (1995). Predictability of stock returns: Robustness and economic significance. *Journal of Finance 50*(4), 1201–1228.

Rapach, D. E., J. K. Strauss, and G. Zhou (2010). Out-of-sample equity premium prediction: combination forecasts and links to the real economy. *Review of Financial Studies 23*(2), 821–862.

Rossi, A. G. (2018). Predicting stock market returns with machine learning. mimeo.

Samuelson, P. A. (1973). Proof that properly discounted present values of assets vibrate randomly. *The Bell Journal of Economics and Management Science 4*(2), 369–374.

Satchell, S. and A. Timmermann (1995). An assessment of the economic value of non-linear foreign exchange rate forecasts. *Journal of Forecasting 14*(6), 477–497.

Schapire, R. E. (1990). The strength of weak learnability. *Machine Learning 5*, 197–227.

Schwager, J. D. (2012). *Hedge Fund Market Wizards: How Winning Traders Win* (1 ed.). New Jersey: John Wiley & Sons.

Shapley, L. S. (1953). A value for n-person games. In H. Kuhn and A. Tucker (Eds.), *Contributions to the Theory of Games*, Volume II, Princeton, pp. 307–317. Princeton University Press.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics 69*(1), 99–118.

Sortino, F. A. and R. van der Meer (1991). Downside risk. *Journal of Portfolio Management 17*(4), 27–31.

Sullivan, R., A. Timmermann, and H. White (1999). Data-snooping, technical trading rule performance, and the bootstrap. *Journal of Finance 54*(5), 1647–1691.

Tversky, A. and D. Kahneman (1974). Judgment under uncertainty: heuristics and biases. *Science 185*(4157), 1124–1131.

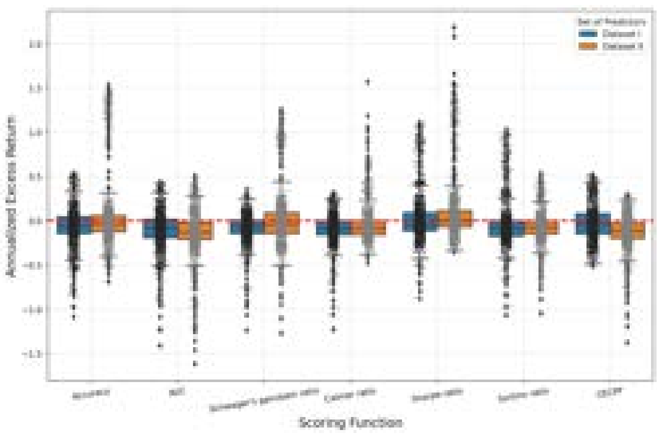
Welch, I. and A. Goyal (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies 21*(4), 1455–1508.

White, H. (2000). A reality check for data snooping. *Econometrica 68*(5), 1097–1126.

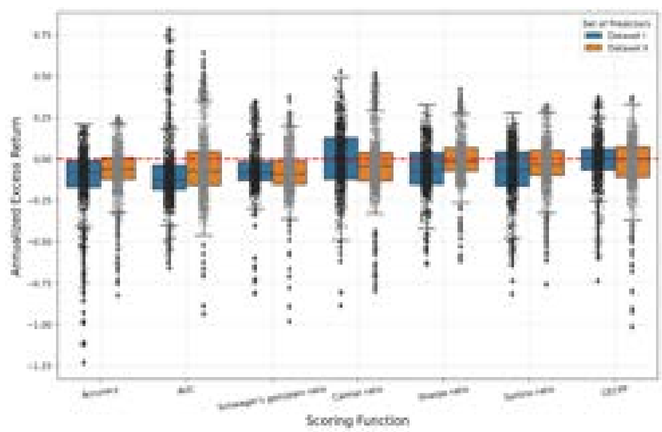
Young, T. W. (1991). Calmar ratio: A smoother tool. *Futures 20*(1), 40.

A.I Performance of the Trading Strategy across Scoring Functions

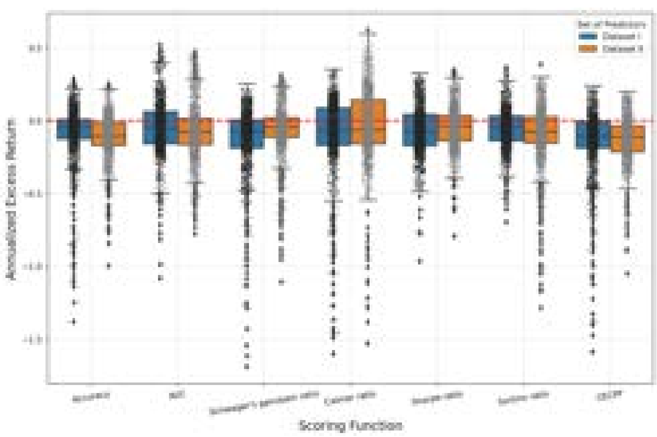
Figure A.I.1: The annualized excess returns of 1718 trading sessions with the fixed transaction cost strategy investing in *SPY* for *100 days* based on one-day ahead forecasts by the LGBM models cross-validated with seven different scoring functions



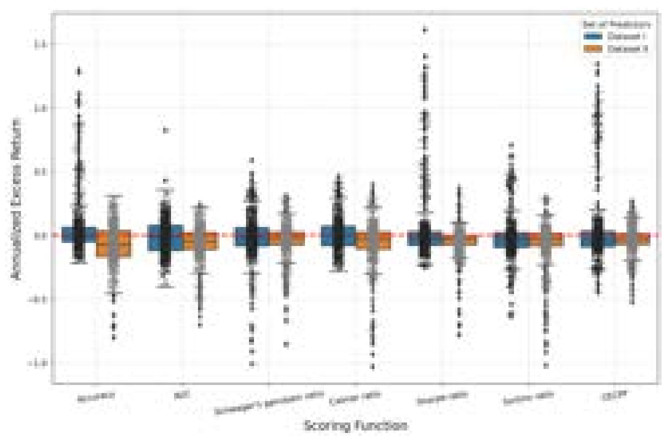
(a) *CE* loss



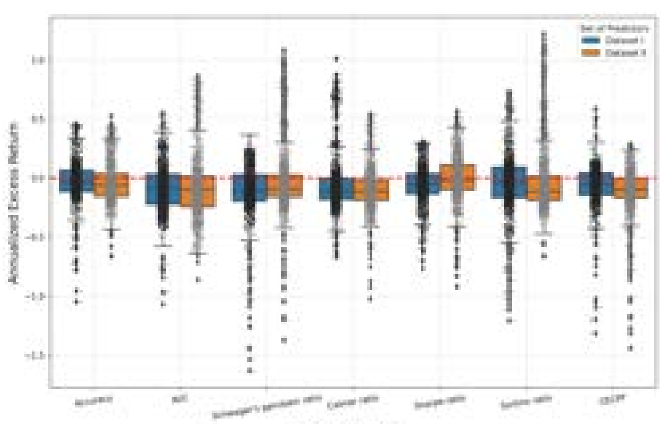
(b) *Brier* loss



(c) *Boost* loss



(d) *As1* loss

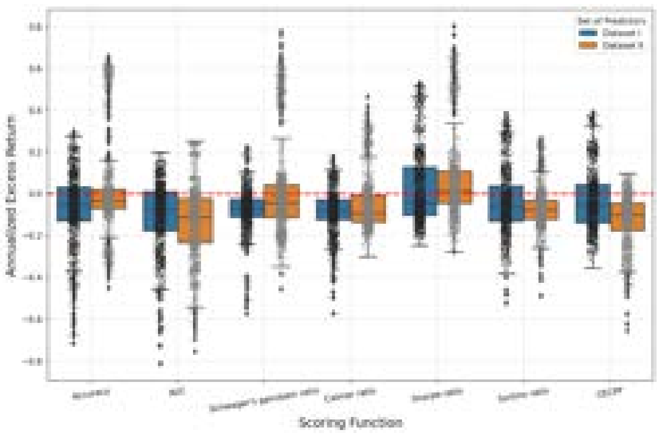


(e) *As2* loss

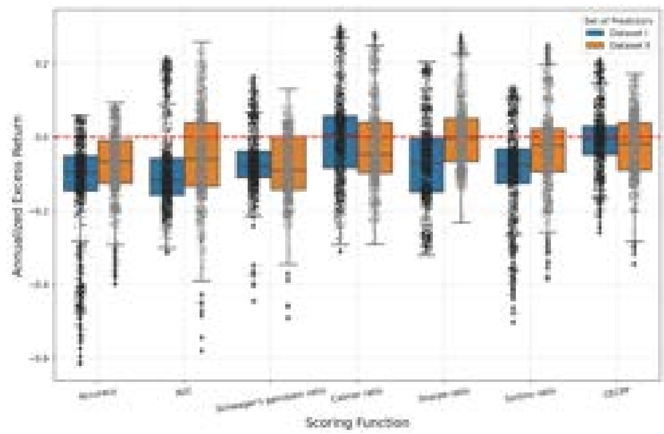
i The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

ii Two sets of predictors are used: *Dataset I* (including the financial variables and technical indicators) and *Dataset II* (including *Dataset I* and candlestick chart patterns as described in Table S.V.1).

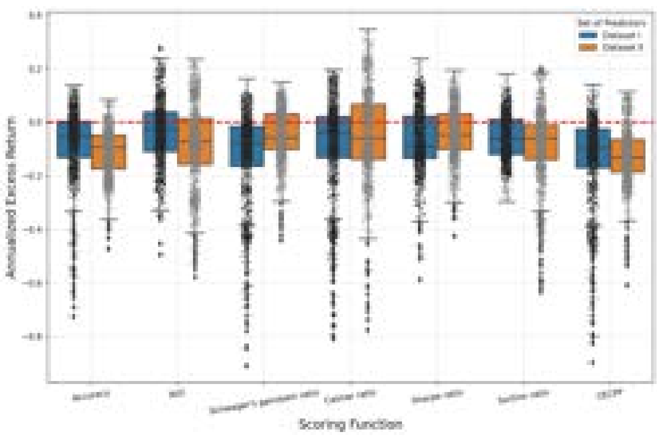
Figure A.I.2: The annualized excess returns of 1618 trading sessions with the fixed transaction cost strategy investing in *SPY* for *200 days* based on one-day ahead forecasts by the LGBM models cross-validated with seven different scoring functions



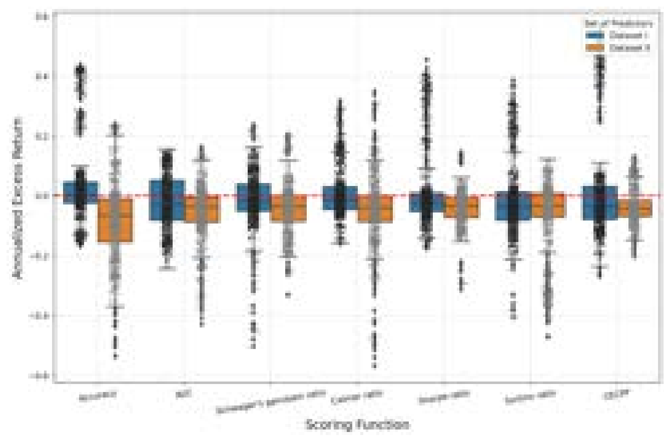
(a) *CE* loss



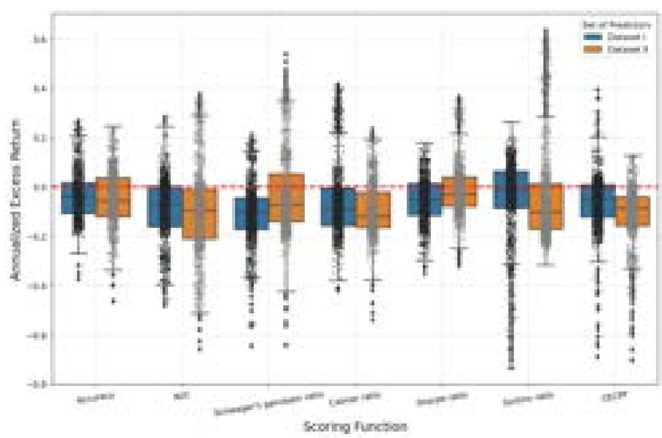
(b) *Brier* loss



(c) *Boost* loss



(d) *As1* loss



(e) *As2* loss

i The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

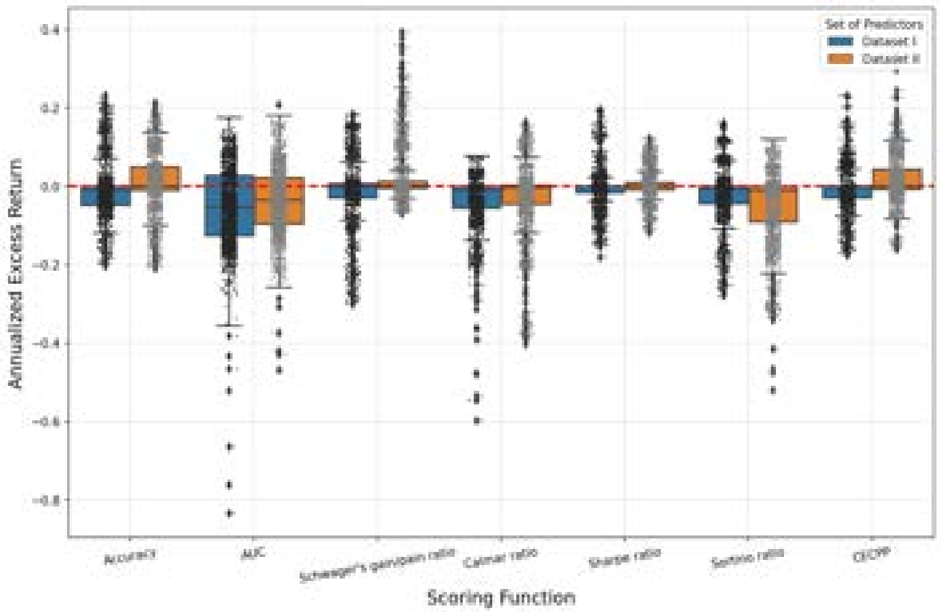
ii Two sets of predictors are used: *Dataset I* (including the financial variables and technical indicators) and *Dataset II* (including *Dataset I* and candlestick chart patterns as described in Table S.V.1).

Table A.I.1: The median (and IQR) of the annualized excess returns of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the LGBM models trained with the *Brier* loss function and cross-validated with seven different scoring functions (i.e., *Accuracy, AUC, Schwager's gain/pain ratio, Calmar ratio, Sharpe ratio, Sortino ratio,* or *CECPP*) for different amounts of transaction cost across various holding periods (in days)

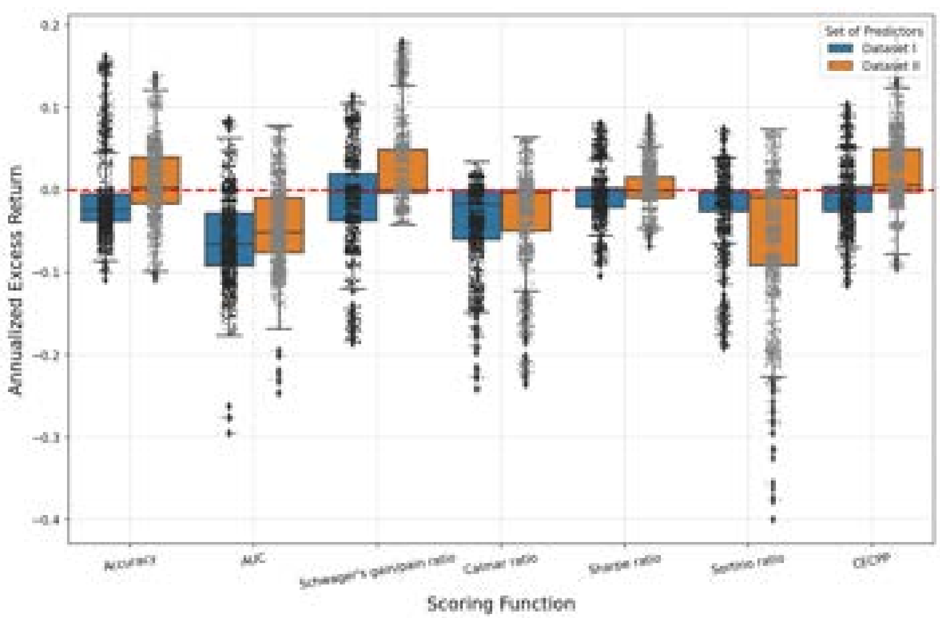
Holding period	Trans. cost	Scoring function used for cross validation													
		Accuracy		AUC		Schwager's gain/pain ratio		Calmar ratio		Sharpe ratio		Sortino ratio		CECPP	
		Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II
100	0.05	-0.08 (0.1613)	-0.06 (0.1325)	-0.1263 (0.1474)	-0.0707 (0.2043)	-0.0806 (0.115)	-0.0879 (0.1416)	-0.0048 (0.2516)	-0.0405 (0.1706)	-0.0714 (0.193)	-0.0143 (0.1378)	-0.0701 (0.2081)	-0.0335 (0.152)	0.0013 (0.1267)	-0.0014 (0.1834)
	0.1	-0.0851 (0.1634)	-0.0654 (0.1336)	-0.1321 (0.1476)	-0.0774 (0.2027)	-0.0852 (0.1146)	-0.0927 (0.1411)	-0.0113 (0.2519)	-0.0463 (0.1703)	-0.0754 (0.1929)	-0.0181 (0.1396)	-0.074 (0.2077)	-0.0367 (0.1525)	-0.0031 (0.1274)	-0.0051 (0.1839)
	0.5	-0.1251 (0.1702)	-0.1016 (0.1367)	-0.1779 (0.1526)	-0.1286 (0.201)	-0.1274 (0.1119)	-0.1301 (0.14)	-0.0626 (0.2559)	-0.0986 (0.1674)	-0.107 (0.1977)	-0.0482 (0.1464)	-0.1133 (0.2056)	-0.0753 (0.1574)	-0.0317 (0.1412)	-0.0347 (0.1744)
	1	-0.1733 (0.1797)	-0.1449 (0.1452)	-0.2348 (0.1611)	-0.1894 (0.1999)	-0.1769 (0.1217)	-0.1777 (0.1439)	-0.1248 (0.2592)	-0.1594 (0.1636)	-0.1418 (0.1983)	-0.08 (0.1546)	-0.1611 (0.2073)	-0.119 (0.1645)	-0.0689 (0.1599)	-0.0805 (0.169)
	5	-0.4687 (0.2059)	-0.4516 (0.237)	-0.608 (0.2251)	-0.5839 (0.2128)	-0.5083 (0.2044)	-0.5108 (0.202)	-0.5075 (0.2704)	-0.5588 (0.2294)	-0.4112 (0.2673)	-0.3355 (0.275)	-0.4606 (0.2361)	-0.4291 (0.2155)	-0.3377 (0.3041)	-0.3638 (0.2041)
200	0.05	-0.0955 (0.0945)	-0.0659 (0.1149)	-0.1166 (0.0999)	-0.0598 (0.1731)	-0.078 (0.0709)	-0.0896 (0.1425)	0.0062 (0.1418)	-0.0459 (0.1388)	-0.0691 (0.1488)	-0.0023 (0.1166)	-0.0749 (0.0927)	-0.0223 (0.1169)	-0.0047 (0.0773)	-0.018 (0.1304)
	0.1	-0.1005 (0.0951)	-0.0703 (0.1156)	-0.122 (0.0996)	-0.0661 (0.1735)	-0.0831 (0.0704)	-0.0946 (0.142)	0.0009 (0.1434)	-0.051 (0.1386)	-0.0732 (0.1508)	-0.0059 (0.1189)	-0.0795 (0.0924)	-0.027 (0.1176)	-0.0082 (0.0785)	-0.0232 (0.1306)
	0.5	-0.1382 (0.0969)	-0.106 (0.1238)	-0.1659 (0.102)	-0.1162 (0.1731)	-0.1221 (0.0752)	-0.135 (0.1418)	-0.0457 (0.1502)	-0.0973 (0.1456)	-0.1036 (0.163)	-0.0348 (0.1308)	-0.1181 (0.0974)	-0.0689 (0.1195)	-0.0373 (0.0892)	-0.0647 (0.1335)
	1	-0.1882 (0.0999)	-0.1515 (0.1379)	-0.2204 (0.1088)	-0.1792 (0.1759)	-0.1728 (0.0823)	-0.1874 (0.1431)	-0.1047 (0.1601)	-0.1562 (0.1555)	-0.1408 (0.1684)	-0.0694 (0.156)	-0.1681 (0.1045)	-0.1205 (0.1315)	-0.0754 (0.1073)	-0.107 (0.1405)
	5	-0.5527 (0.1976)	-0.4969 (0.1716)	-0.6554 (0.1629)	-0.6424 (0.1904)	-0.5561 (0.1828)	-0.5521 (0.1693)	-0.5495 (0.198)	-0.5999 (0.2048)	-0.4289 (0.2086)	-0.3304 (0.2914)	-0.5086 (0.1895)	-0.4667 (0.2132)	-0.3507 (0.2264)	-0.4051 (0.1806)

[a] *Dataset I* comprises of the financial variables and technical indicators while *Dataset II* contains both *Dataset I* and the candlestick chart (price) patterns as described in Table S.V.1.
[b] The initial endowment is $1000.
[c] The IQR values are shown in parentheses.
[d] There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

Figure A.I.3: The annualized excess returns of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the RF models cross-validated with seven different scoring functions



(a) Holding period (D): 100 days



(b) Holding period (D): 200 days

[i] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

[ii] Two sets of predictors are used: *Dataset I* (including the financial variables and technical indicators) and *Dataset II* (including *Dataset I* and candlestick chart patterns as described in Table S.V.1).

[iii] There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

Table A.I.2: The median (and IQR) of the annualized excess returns of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the RF models cross-validated with seven different scoring functions (i.e., *Accuracy, AUC, Schwager's gain/pain ratio, Calmar ratio, Sharpe ratio, Sortino ratio,* or *CECPP*) for different amounts of transaction cost across various holding periods (in days)

Holding period	Trans. cost	Accuracy		AUC		Schwager's gain/pain ratio		Calmar ratio		Sharpe ratio		Sortino ratio		CECPP	
		Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II
100	0.05	-0.009 (0.0479)	0.0 (0.06)	-0.0523 (0.1547)	-0.0348 (0.1171)	-0.0 (0.0376)	0.0 (0.0176)	-0.0089 (0.0553)	-0.0033 (0.0488)	-0.0 (0.0156)	-0.0 (0.0186)	-0.004 (0.0438)	-0.0103 (0.0905)	-0.0011 (0.0301)	0.0 (0.0506)
	0.1	-0.0101 (0.0486)	0.0 (0.0577)	-0.0552 (0.1535)	-0.036 (0.1184)	-0.0 (0.0373)	0.0 (0.0151)	-0.0101 (0.0567)	-0.0037 (0.0521)	-0.0 (0.0174)	-0.0 (0.0183)	-0.0043 (0.0447)	-0.0109 (0.0933)	-0.0014 (0.0312)	0.0 (0.0496)
	0.5	-0.0262 (0.0575)	-0.0 (0.0451)	-0.0787 (0.1425)	-0.051 (0.1171)	-0.0063 (0.039)	-0.0 (0.0099)	-0.0238 (0.078)	-0.007 (0.0722)	-0.0 (0.0395)	-0.0039 (0.0276)	-0.0078 (0.0555)	-0.0183 (0.1081)	-0.0037 (0.0474)	-0.0 (0.0438)
	1	-0.036 (0.0811)	-0.0036 (0.0492)	-0.1091 (0.1465)	-0.0798 (0.12)	-0.0164 (0.0588)	-0.0 (0.0259)	-0.0372 (0.1096)	-0.0153 (0.0934)	-0.006 (0.0661)	-0.0098 (0.0479)	-0.0213 (0.0674)	-0.0307 (0.1246)	-0.0162 (0.0727)	-0.0 (0.0516)
	5	-0.0969 (0.2168)	-0.0672 (0.178)	-0.3136 (0.2196)	-0.2726 (0.2395)	-0.1366 (0.2364)	-0.0357 (0.1617)	-0.1675 (0.2404)	-0.0918 (0.2646)	-0.0365 (0.2723)	-0.0709 (0.1929)	-0.0974 (0.1875)	-0.1485 (0.2467)	-0.0898 (0.2284)	-0.0418 (0.2116)
200	0.05	-0.0236 (0.0332)	0.0024 (0.0562)	-0.0657 (0.0638)	-0.0507 (0.0672)	-0.0017 (0.0568)	0.0 (0.0523)	-0.0174 (0.0596)	-0.0032 (0.05)	-0.0 (0.0234)	-0.0006 (0.0252)	-0.0018 (0.0261)	-0.0085 (0.0911)	-0.0025 (0.0308)	0.0057 (0.0517)
	0.1	-0.0243 (0.0336)	0.0012 (0.0586)	-0.0689 (0.0646)	-0.0537 (0.0677)	-0.0018 (0.0537)	0.0 (0.0516)	-0.0196 (0.0615)	-0.0034 (0.0513)	-0.0 (0.0242)	-0.0007 (0.0241)	-0.0025 (0.0275)	-0.0089 (0.092)	-0.0027 (0.03)	0.0017 (0.0525)
	0.5	-0.0273 (0.0416)	-0.0002 (0.0706)	-0.0974 (0.0677)	-0.0804 (0.0767)	-0.0095 (0.0425)	-0.0 (0.0506)	-0.0387 (0.0729)	-0.0254 (0.0665)	-0.0028 (0.0379)	-0.0041 (0.0179)	-0.0178 (0.0367)	-0.0228 (0.1021)	-0.0161 (0.0478)	-0.0 (0.0574)
	1	-0.0325 (0.0623)	-0.0067 (0.0842)	-0.1314 (0.0715)	-0.1138 (0.0901)	-0.0265 (0.0527)	-0.0 (0.0336)	-0.0573 (0.0942)	-0.0492 (0.083)	-0.0235 (0.0522)	-0.0179 (0.0344)	-0.0222 (0.0575)	-0.0431 (0.1137)	-0.0293 (0.0676)	-0.0043 (0.0615)
	5	-0.1071 (0.2201)	-0.1022 (0.2099)	-0.3291 (0.1594)	-0.3122 (0.2145)	-0.1333 (0.2092)	-0.0473 (0.2395)	-0.1837 (0.2181)	-0.1654 (0.3036)	-0.1125 (0.2866)	-0.0976 (0.2352)	-0.0986 (0.215)	-0.1707 (0.2729)	-0.1026 (0.2316)	-0.0653 (0.2202)

Scoring function used for cross validation (spanning header over all scoring-function columns).

[a] *Dataset I* comprises of the financial variables and technical indicators while *Dataset II* contains both *Dataset I* and the candlestick chart (price) patterns as described in Table S.V.1.

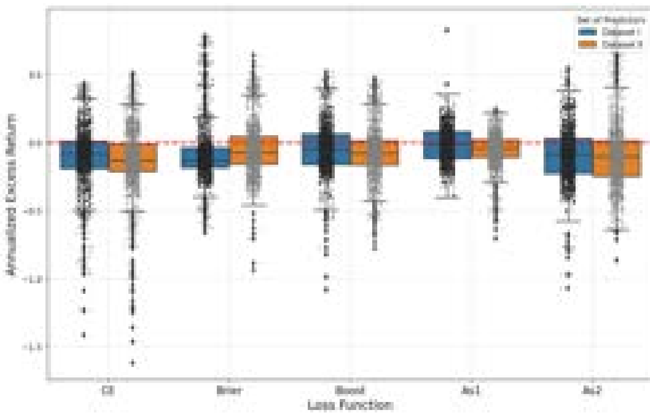
[b] The initial endowment is $1000.

[c] The IQR values are shown in parentheses.

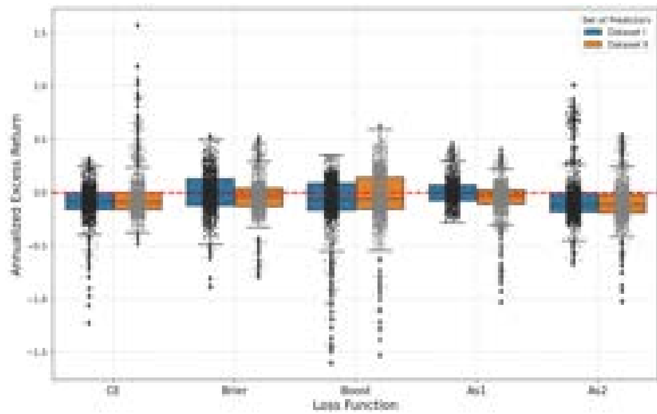
[d] There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

A.II Performance of the Trading Strategy across Loss Functions

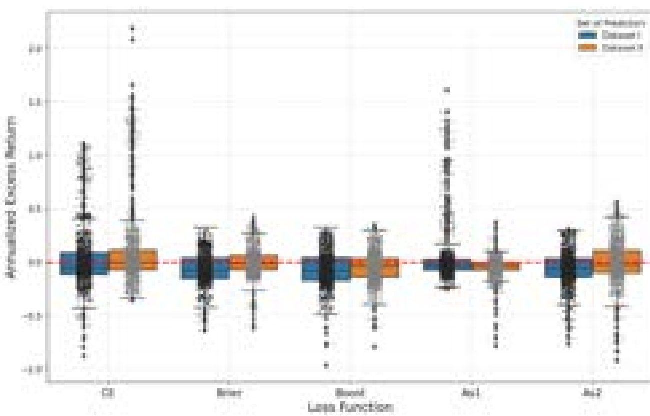
Figure A.II.1: The annualized excess returns of 1718 trading sessions with the fixed transaction cost strategy investing in *SPY* for 100 days based on one-day ahead forecasts by the LGBM models trained with five different loss functions



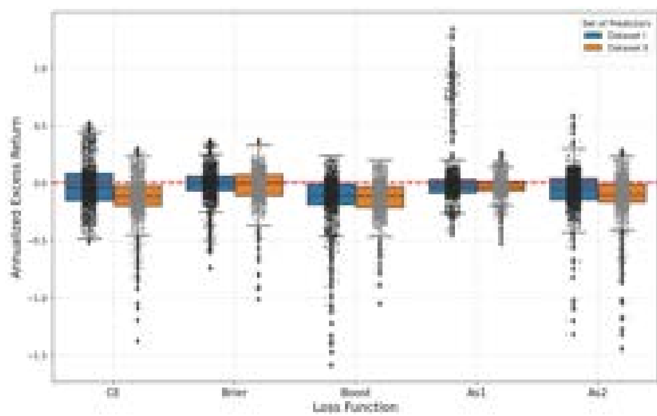
(a) *AUC* score



(b) *Calmar ratio* score



(c) *Sharpe ratio* score

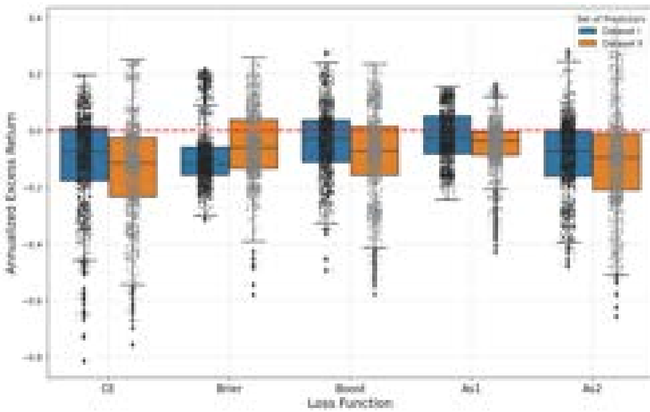


(d) *CECPP* score

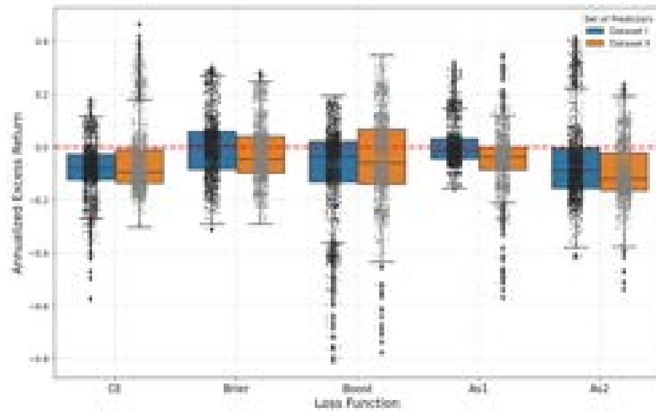
[i] The initial endowment is $1000, and the amount of fixed transaction cost $c = \$0.05$.

[ii] Two sets of predictors are used: *Dataset I* (including the financial variables and technical indicators) and *Dataset II* (including *Dataset I* and candlestick chart patterns as described in Table S.V.1).

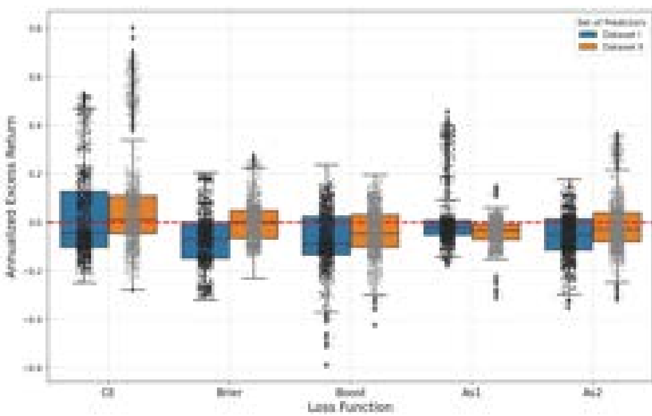
Figure A.II.2: The annualized excess returns of 1618 trading sessions with the fixed transaction cost strategy investing in *SPY* for 200 days based on one-day ahead forecasts by the LGBM models trained with five different loss functions



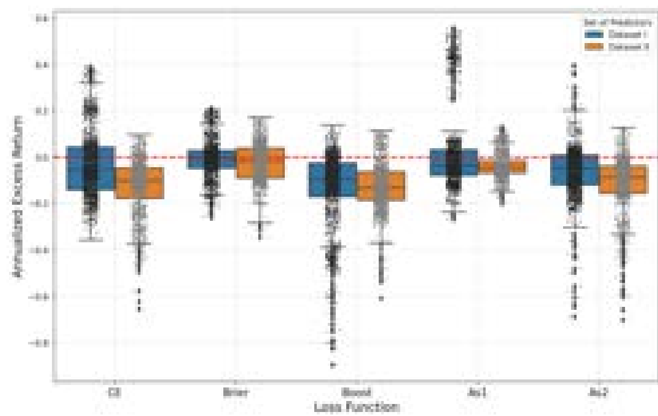
(a) *AUC* score



(b) *Calmar ratio* score



(c) *Sharpe ratio* score



(d) *CECPP* score

[i] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

[ii] Two sets of predictors are used: *Dataset I* (including the financial variables and technical indicators) and *Dataset II* (including *Dataset I* and candlestick chart patterns as described in Table S.V.1).

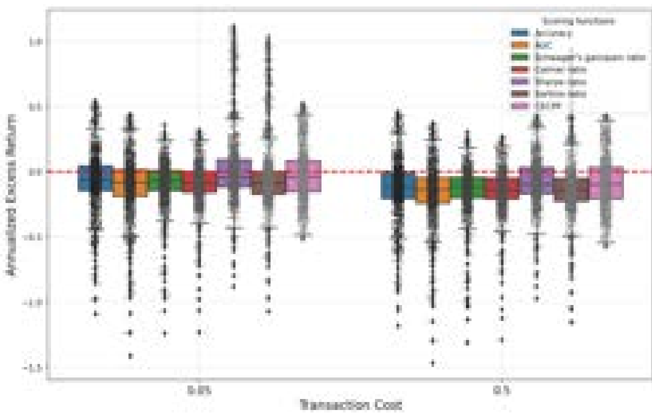
Table A.II.1: The median (and IQR) of the annualized excess returns of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the LGBM models trained with five loss functions (i.e., *CE, Brier, Boost, As1,* or *As2*), then cross-validated with the *CECPP* scoring function for different amounts of transaction cost across various holding periods (in days)

Holding period	Trans. cost	Loss function used for training									
		CE		Brier		Boost		As1		As2	
		Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II
100	0.05	-0.0387 (0.2367)	-0.1147 (0.1732)	0.0013 (0.1267)	-0.0014 (0.1834)	-0.1033 (0.1836)	-0.111 (0.1811)	-0.032 (0.1185)	-0.033 (0.0843)	-0.0525 (0.1894)	-0.1001 (0.1619)
	0.1	-0.0443 (0.2372)	-0.1204 (0.173)	-0.0031 (0.1274)	-0.0051 (0.1839)	-0.1095 (0.1825)	-0.1166 (0.1817)	-0.0336 (0.118)	-0.0354 (0.0831)	-0.0581 (0.1879)	-0.1058 (0.1626)
	0.5	-0.0939 (0.2408)	-0.165 (0.1762)	-0.0317 (0.1412)	-0.0347 (0.1744)	-0.1542 (0.1801)	-0.1622 (0.183)	-0.0507 (0.1189)	-0.0556 (0.0806)	-0.1039 (0.1863)	-0.1483 (0.1631)
	1	-0.1471 (0.2428)	-0.2167 (0.1744)	-0.0689 (0.1599)	-0.0805 (0.169)	-0.2087 (0.1787)	-0.2182 (0.1898)	-0.0723 (0.1266)	-0.0792 (0.0784)	-0.1574 (0.1932)	-0.2022 (0.1673)
	5	-0.5392 (0.2884)	-0.5635 (0.1887)	-0.3377 (0.3041)	-0.3638 (0.2041)	-0.5818 (0.2206)	-0.5847 (0.2315)	-0.2566 (0.1835)	-0.2544 (0.163)	-0.5155 (0.2177)	-0.5489 (0.2434)
200	0.05	-0.0546 (0.1856)	-0.1012 (0.1317)	-0.0047 (0.0773)	-0.018 (0.1304)	-0.0954 (0.1454)	-0.1323 (0.1228)	-0.0345 (0.1108)	-0.0428 (0.0532)	-0.0467 (0.1303)	-0.0865 (0.1171)
	0.1	-0.0609 (0.1871)	-0.107 (0.1323)	-0.0082 (0.0785)	-0.0232 (0.1306)	-0.1012 (0.1456)	-0.1383 (0.124)	-0.037 (0.1104)	-0.0461 (0.0533)	-0.0528 (0.1295)	-0.0925 (0.1169)
	0.5	-0.1066 (0.1971)	-0.1521 (0.1352)	-0.0373 (0.0892)	-0.0647 (0.1335)	-0.1477 (0.1474)	-0.1863 (0.1325)	-0.0549 (0.1186)	-0.066 (0.0603)	-0.0983 (0.1244)	-0.1391 (0.1219)
	1	-0.1627 (0.2083)	-0.2083 (0.1374)	-0.0754 (0.1073)	-0.107 (0.1405)	-0.2066 (0.1442)	-0.2456 (0.141)	-0.0748 (0.1187)	-0.0929 (0.0715)	-0.1518 (0.1215)	-0.1969 (0.1218)
	5	-0.5924 (0.2152)	-0.6421 (0.1401)	-0.3507 (0.2264)	-0.4051 (0.1806)	-0.6716 (0.2005)	-0.7013 (0.1976)	-0.2498 (0.1363)	-0.3043 (0.1481)	-0.5961 (0.1415)	-0.6365 (0.1425)

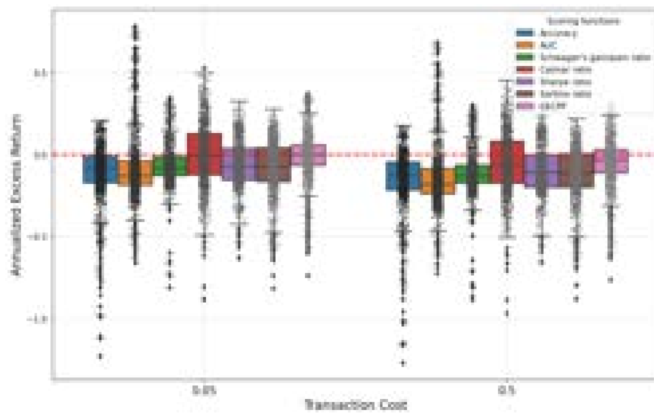
[a] *Dataset I* comprises of the financial variables and technical indicators while *Dataset II* contains both *Dataset I* and the candlestick chart (price) patterns as described in Table S.V.1.
[b] The initial endowment is $1000.
[c] The IQR values are shown in parentheses.
[d] There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

A.III Performance of the Trading Strategy for Different Amounts of Transaction Cost

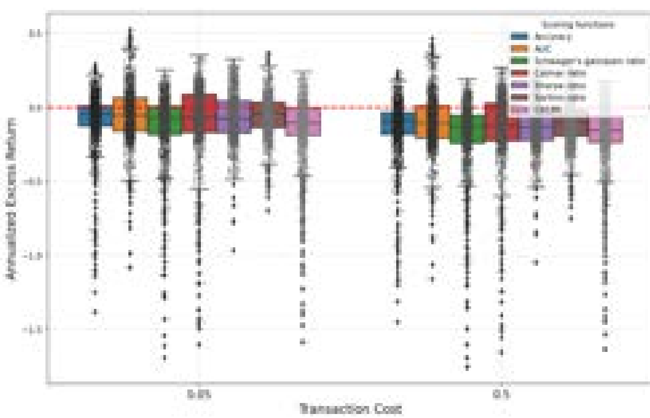
Figure A.III.1: The annualized excess returns of 1718 trading sessions with the fixed transaction cost strategy investing in *SPY* for 100 days based on one-day ahead forecasts by a LGBM model trained using the data on financial variables and technical indicators (*Dataset I*) described in Table S.V.1



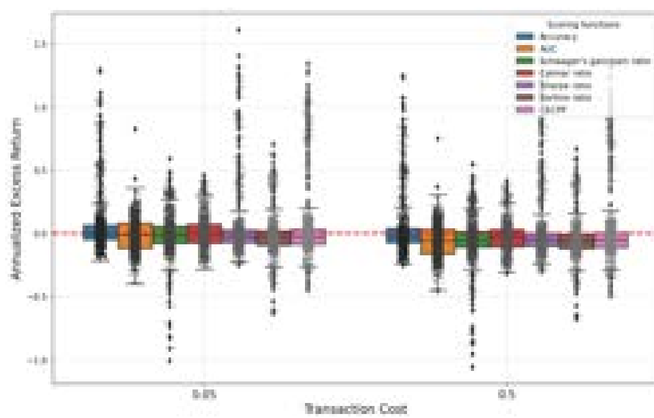
(a) *CE* loss



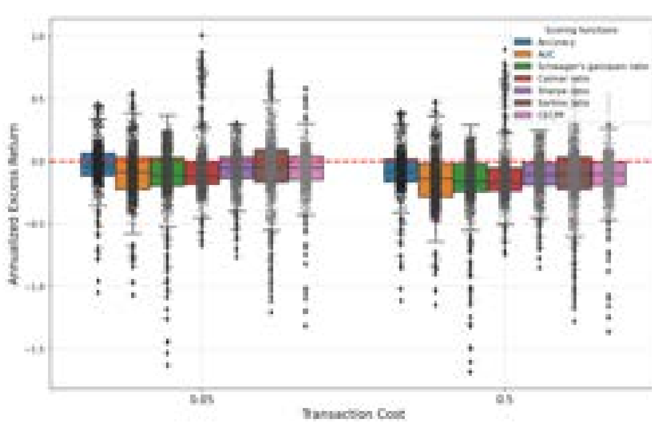
(b) *Brier* loss



(c) *Boost* loss



(d) *As1* loss



(e) *As2* loss

[i] The initial endowment is $1000.

Table A.III.1: The median (and IQR) of the annualized excess returns of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the LGBM models trained with the *Brier* loss function, then cross-validated with seven scoring functions (i.e., *Accuracy, AUC, Schwager's gain/pain ratio, Calmar ratio, Sharpe ratio, Sortino ratio,* or *CECPP*) for different amounts of transaction cost across various holding periods (in days). The predictors used: financial variables and technical indicators (*Dataset I*) described in Table S.V.1

Holding period	Trans. cost	Scoring function used for cross validation						
		Accuracy	AUC	Schwager's gain/pain ratio	Calmar ratio	Sharpe ratio	Sortino ratio	CECPP
100	0.05	-0.08 (0.1613)	-0.1263 (0.1474)	-0.0806 (0.115)	-0.0048 (0.2516)	-0.0714 (0.193)	-0.0701 (0.2081)	0.0013 (0.1267)
	0.1	-0.0851 (0.1634)	-0.1321 (0.1476)	-0.0852 (0.1146)	-0.0113 (0.2519)	-0.0754 (0.1929)	-0.074 (0.2077)	-0.0031 (0.1274)
	0.5	-0.1251 (0.1702)	-0.1779 (0.1526)	-0.1274 (0.1119)	-0.0626 (0.2559)	-0.107 (0.1977)	-0.1133 (0.2056)	-0.0317 (0.1412)
	1	-0.1733 (0.1797)	-0.2348 (0.1611)	-0.1769 (0.1217)	-0.1248 (0.2592)	-0.1418 (0.1983)	-0.1611 (0.2073)	-0.0689 (0.1599)
	5	-0.4687 (0.2059)	-0.608 (0.2251)	-0.5083 (0.2044)	-0.5075 (0.2704)	-0.4112 (0.2673)	-0.4606 (0.2361)	-0.3377 (0.3041)
200	0.05	-0.0955 (0.0945)	-0.1166 (0.0999)	-0.078 (0.0709)	0.0062 (0.1418)	-0.0691 (0.1488)	-0.0749 (0.0927)	-0.0047 (0.0773)
	0.1	-0.1005 (0.0951)	-0.122 (0.0996)	-0.0831 (0.0704)	0.0009 (0.1434)	-0.0732 (0.1508)	-0.0795 (0.0924)	-0.0082 (0.0785)
	0.5	-0.1382 (0.0969)	-0.1659 (0.102)	-0.1221 (0.0752)	-0.0457 (0.1502)	-0.1036 (0.163)	-0.1181 (0.0974)	-0.0373 (0.0892)
	1	-0.1882 (0.0999)	-0.2204 (0.1088)	-0.1728 (0.0823)	-0.1047 (0.1601)	-0.1408 (0.1684)	-0.1681 (0.1045)	-0.0754 (0.1073)
	5	-0.5527 (0.1976)	-0.6554 (0.1629)	-0.5561 (0.1828)	-0.5495 (0.198)	-0.4289 (0.2086)	-0.5086 (0.1895)	-0.3507 (0.2264)

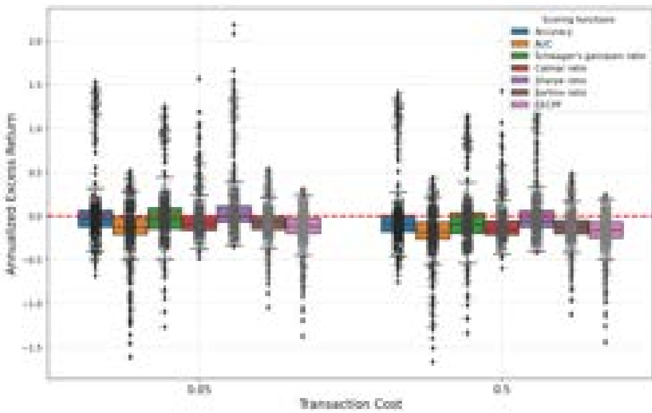
[a] The initial endowment is $1000.
[b] The IQR values are shown in parentheses.
[c] There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

Table A.III.2: The median (and IQR) of the annualized excess returns of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the LGBM models trained with the *As1* loss function, then cross-validated with seven scoring functions (i.e., *Accuracy, AUC, Schwager's gain/pain ratio, Calmar ratio, Sharpe ratio, Sortino ratio,* or *CECPP*) for different amounts of transaction cost across various holding periods (in days). The predictors used: financial variables and technical indicators (*Dataset I*) described in Table S.V.1

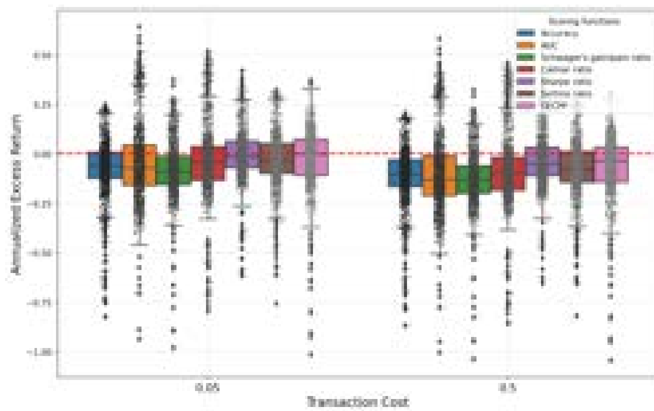
Holding period	Trans. cost	Scoring function used for cross validation						
		Accuracy	AUC	Schwager's gain/pain ratio	Calmar ratio	Sharpe ratio	Sortino ratio	CECPP
100	0.05	-0.0035 (0.1138)	-0.0107 (0.1982)	-0.0168 (0.1423)	0.0002 (0.1499)	-0.0278 (0.1028)	-0.0301 (0.1154)	-0.032 (0.1185)
	0.1	-0.0068 (0.1141)	-0.0159 (0.1989)	-0.0216 (0.1418)	-0.0035 (0.149)	-0.0298 (0.1022)	-0.0342 (0.1145)	-0.0336 (0.118)
	0.5	-0.0293 (0.1152)	-0.0575 (0.1966)	-0.0549 (0.1232)	-0.0289 (0.1384)	-0.051 (0.1002)	-0.0614 (0.1143)	-0.0507 (0.1189)
	1	-0.0557 (0.1205)	-0.1072 (0.1948)	-0.0908 (0.114)	-0.0613 (0.1245)	-0.0742 (0.1021)	-0.0929 (0.1134)	-0.0723 (0.1266)
	5	-0.2662 (0.2481)	-0.4495 (0.226)	-0.3586 (0.1472)	-0.3301 (0.1644)	-0.2629 (0.1566)	-0.3086 (0.1733)	-0.2566 (0.1835)
200	0.05	0.0049 (0.07)	-0.0273 (0.1343)	-0.0175 (0.0898)	-0.0147 (0.076)	-0.0225 (0.0614)	-0.0369 (0.0929)	-0.0345 (0.1108)
	0.1	0.0008 (0.0694)	-0.0328 (0.1338)	-0.0215 (0.0901)	-0.0185 (0.0762)	-0.0253 (0.0615)	-0.0407 (0.0927)	-0.037 (0.1104)
	0.5	-0.0218 (0.0779)	-0.0725 (0.1341)	-0.0508 (0.0855)	-0.0465 (0.0712)	-0.0459 (0.0652)	-0.0642 (0.0929)	-0.0549 (0.1186)
	1	-0.0526 (0.0896)	-0.1195 (0.1372)	-0.0882 (0.0776)	-0.0848 (0.0695)	-0.0716 (0.0695)	-0.0941 (0.0995)	-0.0748 (0.1187)
	5	-0.295 (0.2314)	-0.5003 (0.1949)	-0.3725 (0.0915)	-0.3678 (0.1422)	-0.2658 (0.1259)	-0.3069 (0.1577)	-0.2498 (0.1363)

[a] The initial endowment is $1000.
[b] The IQR values are shown in parentheses.
[c] There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

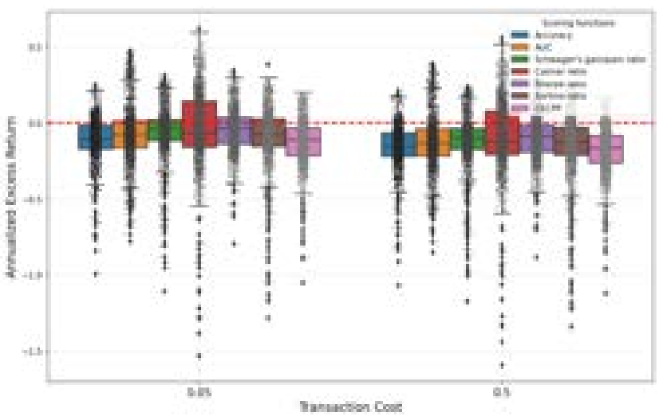
Figure A.III.2: The annualized excess returns of 1718 trading sessions with the fixed transaction cost strategy investing in *SPY* for 100 days based on one-day ahead forecasts by a LGBM model trained using the data on financial variables, technical indicators, and candlestick chart patterns (*Dataset II*) described in Table S.V.1



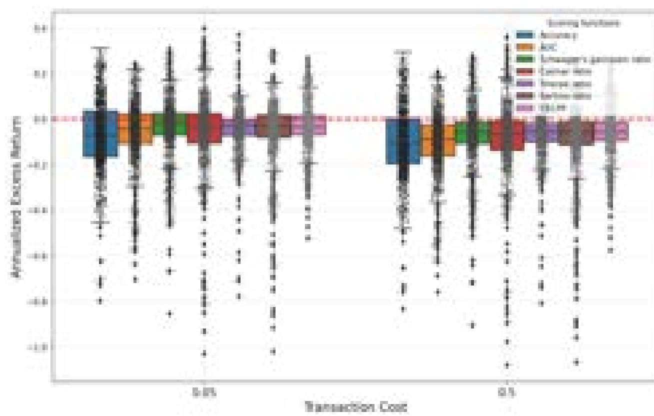
(a) *CE* loss



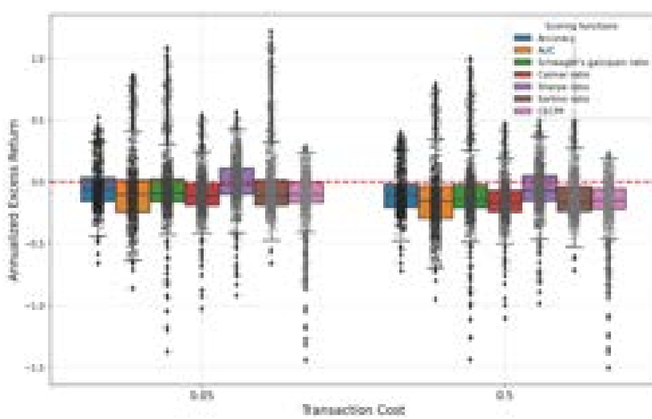
(b) *Brier* loss



(c) *Boost* loss



(d) *As1* loss



(e) *As2* loss

[i] The initial endowment is $1000.

Table A.III.3: The median (and IQR) of the annualized excess returns of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the LGBM models trained with the *Brier* loss function, then cross-validated with seven scoring functions (i.e., *Accuracy, AUC, Schwager's gain/pain ratio, Calmar ratio, Sharpe ratio, Sortino ratio,* or *CECPP*) for different amounts of transaction cost across various holding periods (in days). The predictors used: financial variables, technical indicators, and candlestick chart patterns (*Dataset II*) described in Table S.V.1

Holding period	Trans. cost	Scoring function used for cross validation						
		Accuracy	AUC	Schwager's gain/pain ratio	Calmar ratio	Sharpe ratio	Sortino ratio	CECPP
100	0.05	-0.06 (0.1325)	-0.0707 (0.2043)	-0.0879 (0.1416)	-0.0405 (0.1706)	-0.0143 (0.1378)	-0.0335 (0.152)	-0.0014 (0.1834)
	0.1	-0.0654 (0.1336)	-0.0774 (0.2027)	-0.0927 (0.1411)	-0.0463 (0.1703)	-0.0181 (0.1396)	-0.0367 (0.1525)	-0.0051 (0.1839)
	0.5	-0.1016 (0.1367)	-0.1286 (0.201)	-0.1301 (0.14)	-0.0986 (0.1674)	-0.0482 (0.1464)	-0.0753 (0.1574)	-0.0347 (0.1744)
	1	-0.1449 (0.1452)	-0.1894 (0.1999)	-0.1777 (0.1439)	-0.1594 (0.1636)	-0.08 (0.1546)	-0.119 (0.1645)	-0.0805 (0.169)
	5	-0.4516 (0.237)	-0.5839 (0.2128)	-0.5108 (0.202)	-0.5588 (0.2294)	-0.3355 (0.275)	-0.4291 (0.2155)	-0.3638 (0.2041)
200	0.05	-0.0659 (0.1149)	-0.0598 (0.1731)	-0.0896 (0.1425)	-0.0459 (0.1388)	-0.0023 (0.1166)	-0.0223 (0.1169)	-0.018 (0.1304)
	0.1	-0.0703 (0.1156)	-0.0661 (0.1735)	-0.0946 (0.142)	-0.051 (0.1386)	-0.0059 (0.1189)	-0.027 (0.1176)	-0.0232 (0.1306)
	0.5	-0.106 (0.1238)	-0.1162 (0.1731)	-0.135 (0.1418)	-0.0973 (0.1456)	-0.0348 (0.1308)	-0.0689 (0.1195)	-0.0647 (0.1335)
	1	-0.1515 (0.1379)	-0.1792 (0.1759)	-0.1874 (0.1431)	-0.1562 (0.1555)	-0.0694 (0.156)	-0.1205 (0.1315)	-0.107 (0.1405)
	5	-0.4969 (0.1716)	-0.6424 (0.1904)	-0.5521 (0.1693)	-0.5999 (0.2048)	-0.3304 (0.2914)	-0.4667 (0.2132)	-0.4051 (0.1806)

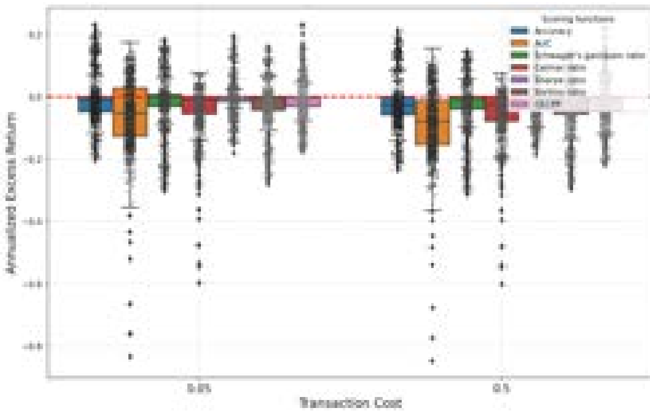
[a] The initial endowment is \$1000.
[b] The IQR values are shown in parentheses.
[c] There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

Table A.III.4: The median (and IQR) of the annualized excess returns of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the LGBM models trained with the *As1* loss function, then cross-validated with seven scoring functions (i.e., *Accuracy, AUC, Schwager's gain/pain ratio, Calmar ratio, Sharpe ratio, Sortino ratio,* or *CECPP*) for different amounts of transaction cost across various holding periods (in days). The predictors used: financial variables, technical indicators, and candlestick chart patterns (*Dataset II*) described in Table S.V.1

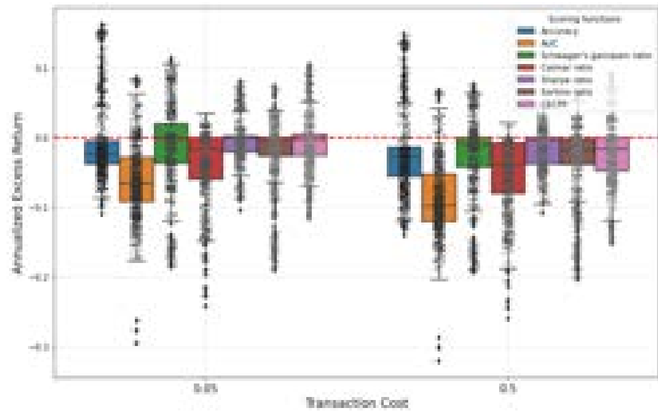
Holding period	Trans. cost	Scoring function used for cross validation						
		Accuracy	AUC	Schwager's gain/pain ratio	Calmar ratio	Sharpe ratio	Sortino ratio	CECPP
100	0.05	-0.0708 (0.201)	-0.0428 (0.13)	-0.0217 (0.097)	-0.0338 (0.1309)	-0.0268 (0.0717)	-0.029 (0.098)	-0.033 (0.0843)
	0.1	-0.0745 (0.2003)	-0.0482 (0.1303)	-0.0252 (0.0968)	-0.0374 (0.1304)	-0.0293 (0.0733)	-0.0325 (0.0994)	-0.0354 (0.0831)
	0.5	-0.1043 (0.1987)	-0.0877 (0.1383)	-0.0569 (0.0951)	-0.0653 (0.1293)	-0.055 (0.0822)	-0.0561 (0.1005)	-0.0556 (0.0806)
	1	-0.1352 (0.1951)	-0.1334 (0.151)	-0.0968 (0.0958)	-0.1001 (0.1305)	-0.079 (0.0916)	-0.0854 (0.1015)	-0.0792 (0.0784)
	5	-0.3551 (0.2353)	-0.4823 (0.2485)	-0.3718 (0.1487)	-0.3465 (0.1527)	-0.2616 (0.1707)	-0.3378 (0.1656)	-0.2544 (0.163)
200	0.05	-0.0651 (0.1436)	-0.0354 (0.0829)	-0.0342 (0.0837)	-0.0377 (0.0841)	-0.0323 (0.0646)	-0.0318 (0.0796)	-0.0428 (0.0532)
	0.1	-0.0691 (0.145)	-0.0403 (0.0841)	-0.0384 (0.0843)	-0.0414 (0.0845)	-0.0351 (0.0654)	-0.0352 (0.0802)	-0.0461 (0.0533)
	0.5	-0.1009 (0.1504)	-0.0824 (0.0938)	-0.0695 (0.0891)	-0.0705 (0.084)	-0.0563 (0.0675)	-0.0638 (0.0835)	-0.066 (0.0603)
	1	-0.1333 (0.1594)	-0.1341 (0.1043)	-0.1097 (0.0952)	-0.1039 (0.0829)	-0.0822 (0.0655)	-0.0968 (0.089)	-0.0929 (0.0715)
	5	-0.3874 (0.1843)	-0.5306 (0.1731)	-0.4082 (0.1134)	-0.3748 (0.0857)	-0.29 (0.0962)	-0.3536 (0.1408)	-0.3043 (0.1481)

[a] The initial endowment is $1000.
[b] The IQR values are shown in parentheses.
[c] There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

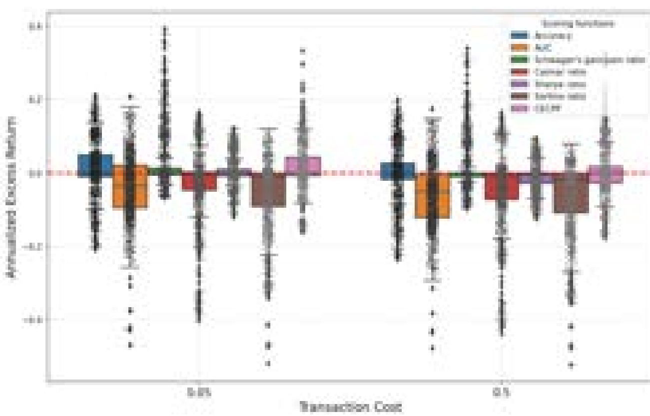
Figure A.III.3: The annualized excess returns of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by RF models cross-validated with seven different scoring functions for two different sets of predictors



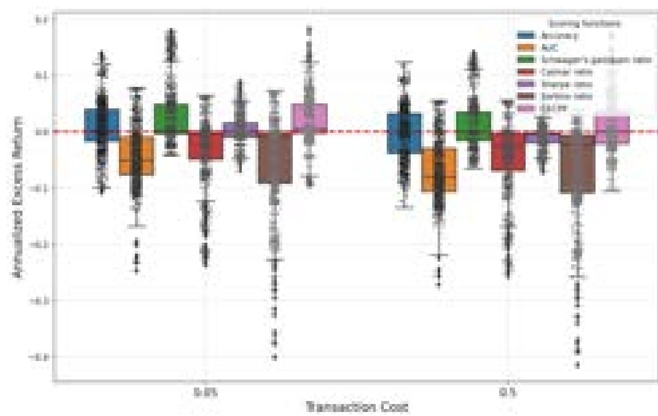
(a) *Dataset I*: Holding period (D) = 100 days



(b) *Dataset I*: Holding period (D) = 200 days



(c) *Dataset II*: Holding period (D) = 100 days



(d) *Dataset II*: Holding period (D) = 200 days

[i] *Dataset I* comprises of the financial variables and technical indicators while *Dataset II* contains both *Dataset I* and the candlestick chart (price) patterns as described in Table S.V.1.

[ii] The initial endowment is $1000.

[iii] There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

Table A.III.5: The median (and IQR) of the annualized excess returns of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the RF models cross-validated with seven scoring functions (i.e., *Accuracy, AUC, Schwager's gain/pain ratio, Calmar ratio, Sharpe ratio, Sortino ratio,* or *CECPP*) for different amounts of transaction cost across various holding periods (in days). The predictors used: financial variables and technical indicators (*Dataset I*) described in Table S.V.1

Holding period	Trans. cost	Scoring function used for cross validation						
		Accuracy	AUC	Schwager's gain/pain ratio	Calmar ratio	Sharpe ratio	Sortino ratio	CECPP
100	0.05	-0.009 (0.0479)	-0.0523 (0.1547)	-0.0 (0.0376)	-0.0089 (0.0553)	-0.0 (0.0156)	-0.004 (0.0438)	-0.0011 (0.0301)
	0.1	-0.0101 (0.0486)	-0.0552 (0.1535)	-0.0 (0.0373)	-0.0101 (0.0567)	-0.0 (0.0174)	-0.0043 (0.0447)	-0.0014 (0.0312)
	0.5	-0.0262 (0.0575)	-0.0787 (0.1425)	-0.0063 (0.039)	-0.0238 (0.078)	-0.0 (0.0395)	-0.0078 (0.0555)	-0.0037 (0.0474)
	1	-0.036 (0.0811)	-0.1091 (0.1465)	-0.0164 (0.0588)	-0.0372 (0.1096)	-0.006 (0.0661)	-0.0213 (0.0674)	-0.0162 (0.0727)
	5	-0.0969 (0.2168)	-0.3136 (0.2196)	-0.1366 (0.2364)	-0.1675 (0.2404)	-0.0365 (0.2723)	-0.0974 (0.1875)	-0.0898 (0.2284)
200	0.05	-0.0236 (0.0332)	-0.0657 (0.0638)	-0.0017 (0.0568)	-0.0174 (0.0596)	-0.0 (0.0234)	-0.0018 (0.0261)	-0.0025 (0.0308)
	0.1	-0.0243 (0.0336)	-0.0689 (0.0646)	-0.0018 (0.0537)	-0.0196 (0.0615)	-0.0 (0.0242)	-0.0025 (0.0275)	-0.0027 (0.03)
	0.5	-0.0273 (0.0416)	-0.0974 (0.0677)	-0.0095 (0.0425)	-0.0387 (0.0729)	-0.0028 (0.0379)	-0.0178 (0.0367)	-0.0161 (0.0478)
	1	-0.0325 (0.0623)	-0.1314 (0.0715)	-0.0265 (0.0527)	-0.0573 (0.0942)	-0.0235 (0.0522)	-0.0222 (0.0575)	-0.0293 (0.0676)
	5	-0.1071 (0.2201)	-0.3291 (0.1594)	-0.1333 (0.2092)	-0.1837 (0.2181)	-0.1125 (0.2866)	-0.0986 (0.215)	-0.1026 (0.2316)

[a] The initial endowment is $1000.
[b] The IQR values are shown in parentheses.
[c] There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

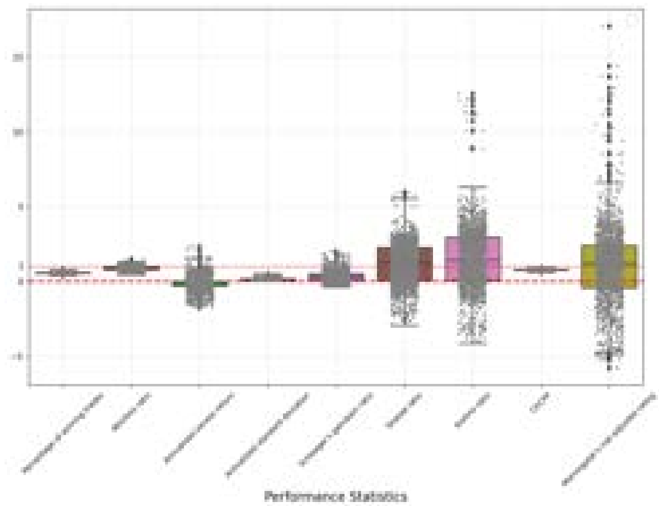
Table A.III.6: The median (and IQR) of the annualized excess returns of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the RF models cross-validated with seven scoring functions (i.e., *Accuracy, AUC, Schwager's gain/pain ratio, Calmar ratio, Sharpe ratio, Sortino ratio,* or *CECPP*) for different amounts of transaction cost across various holding periods (in days). The predictors used: financial variables, technical indicators, and candlestick chart patterns (*Dataset II*) described in Table S.V.1

Holding period	Trans. cost	Scoring function used for cross validation						
		Accuracy	AUC	Schwager's gain/pain ratio	Calmar ratio	Sharpe ratio	Sortino ratio	CECPP
100	0.05	0.0 (0.06)	-0.0348 (0.1171)	0.0 (0.0176)	-0.0033 (0.0488)	-0.0 (0.0186)	-0.0103 (0.0905)	0.0 (0.0506)
	0.1	0.0 (0.0577)	-0.036 (0.1184)	0.0 (0.0151)	-0.0037 (0.0521)	-0.0 (0.0183)	-0.0109 (0.0933)	0.0 (0.0496)
	0.5	-0.0 (0.0451)	-0.051 (0.1171)	-0.0 (0.0099)	-0.007 (0.0722)	-0.0039 (0.0276)	-0.0183 (0.1081)	-0.0 (0.0438)
	1	-0.0036 (0.0492)	-0.0798 (0.12)	-0.0 (0.0259)	-0.0153 (0.0934)	-0.0098 (0.0479)	-0.0307 (0.1246)	-0.0 (0.0516)
	5	-0.0672 (0.178)	-0.2726 (0.2395)	-0.0357 (0.1617)	-0.0918 (0.2646)	-0.0709 (0.1929)	-0.1485 (0.2467)	-0.0418 (0.2116)
200	0.05	0.0024 (0.0562)	-0.0507 (0.0672)	0.0 (0.0523)	-0.0032 (0.05)	-0.0006 (0.0252)	-0.0085 (0.0911)	0.0057 (0.0517)
	0.1	0.0012 (0.0586)	-0.0537 (0.0677)	0.0 (0.0516)	-0.0034 (0.0513)	-0.0007 (0.0241)	-0.0089 (0.092)	0.0017 (0.0525)
	0.5	-0.0002 (0.0706)	-0.0804 (0.0767)	-0.0 (0.0506)	-0.0254 (0.0665)	-0.0041 (0.0179)	-0.0228 (0.1021)	-0.0 (0.0574)
	1	-0.0067 (0.0842)	-0.1138 (0.0901)	-0.0 (0.0336)	-0.0492 (0.083)	-0.0179 (0.0344)	-0.0431 (0.1137)	-0.0043 (0.0615)
	5	-0.1022 (0.2099)	-0.3122 (0.2145)	-0.0473 (0.2395)	-0.1654 (0.3036)	-0.0976 (0.2352)	-0.1707 (0.2729)	-0.0653 (0.2202)

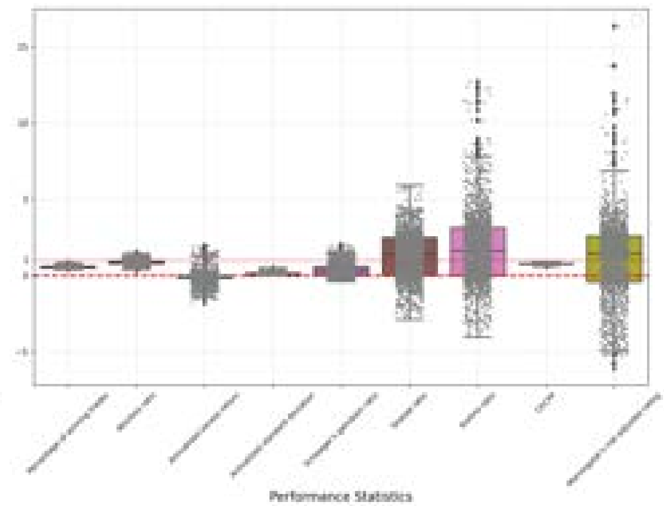
[a] The initial endowment is $1000.
[b] The IQR values are shown in parentheses.
[c] There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

A.IV Performance Statistics of the Trading Strategy over Time

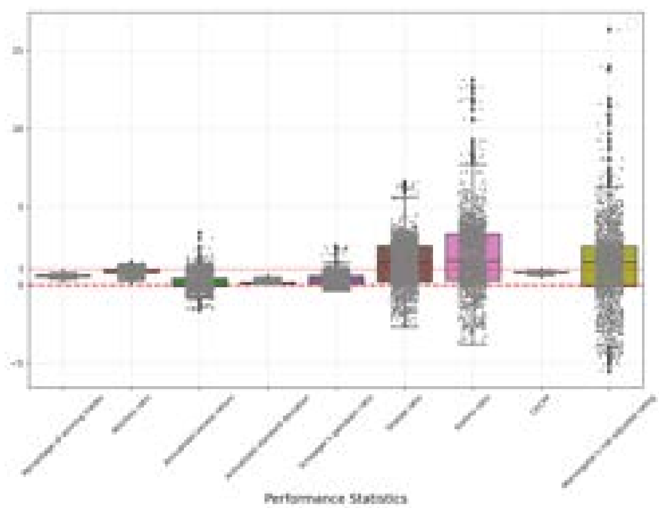
Figure A.IV.1: The performance statistics of 1718 trading sessions with the fixed transaction cost strategy investing in *SPY* for 100 days based on one-day ahead forecasts by RF models cross-validated with two scoring functions for two different sets of predictors



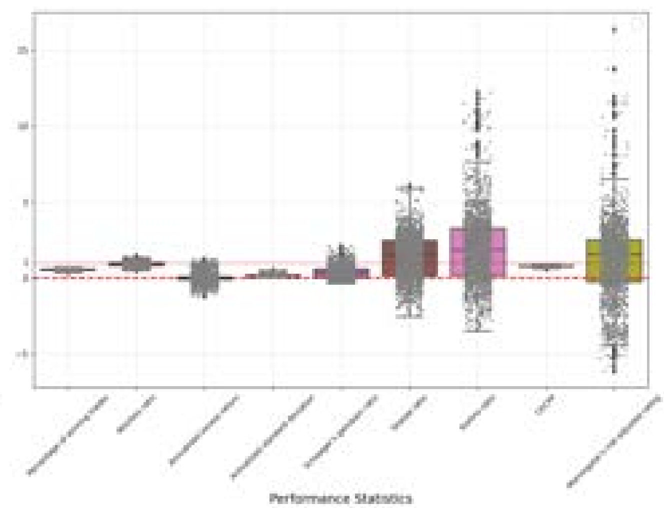
(a) *Dataset I*: *CECPP* score



(b) *Dataset I*: *Sharpe ratio* score



(c) *Dataset II*: *CECPP* score



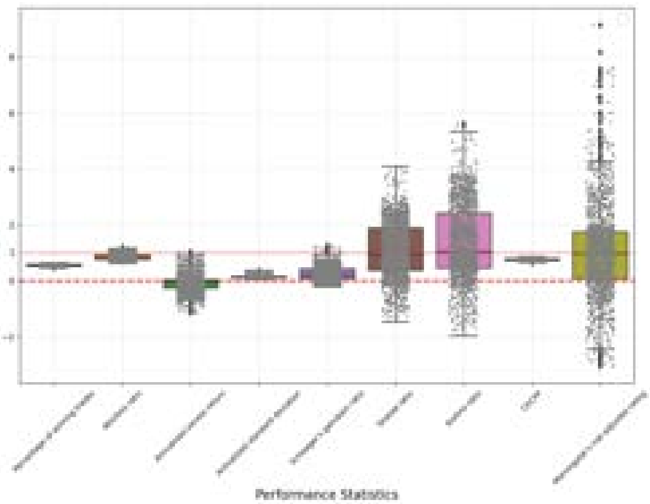
(d) *Dataset II*: *Sharpe ratio* score

[i] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

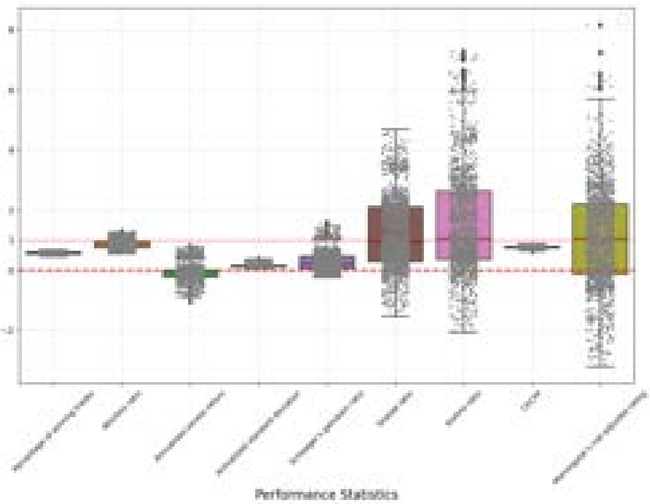
[ii] The values of the annualized excess return and Morningstar's risk-adjusted rating are scaled up ten times to make them more visible in this box plot.

[iii] *Dataset I* comprises of the financial variables and technical indicators while *Dataset II* contains both *Dataset I* and the candlestick chart (price) patterns as described in Table S.V.1.

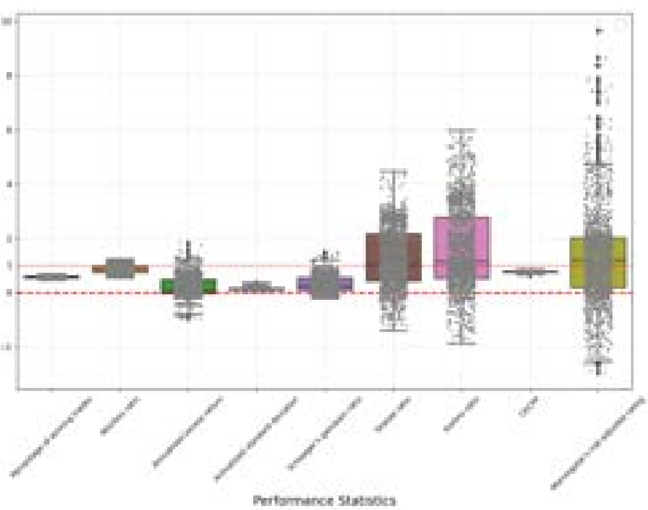
Figure A.IV.2: The performance statistics of 1618 trading sessions with the fixed transaction cost strategy investing in *SPY* for 200 days based on one-day ahead forecasts by RF models cross-validated with two scoring functions for two different sets of predictors



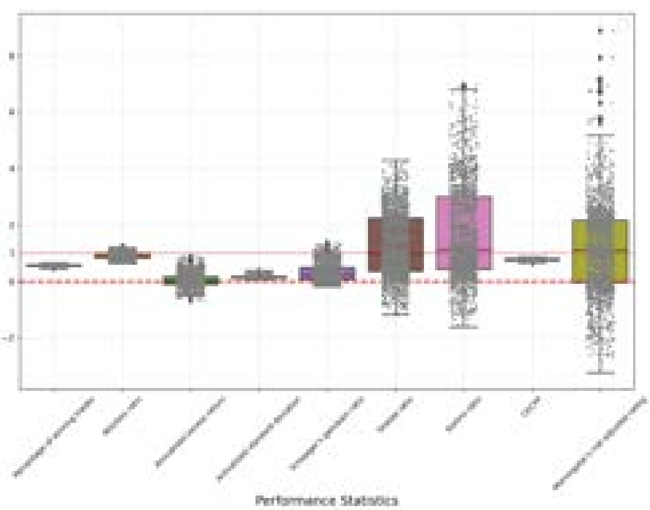
(a) *Dataset I*: *CECPP* score



(b) *Dataset I*: *Sharpe ratio* score



(c) *Dataset II*: *CECPP* score



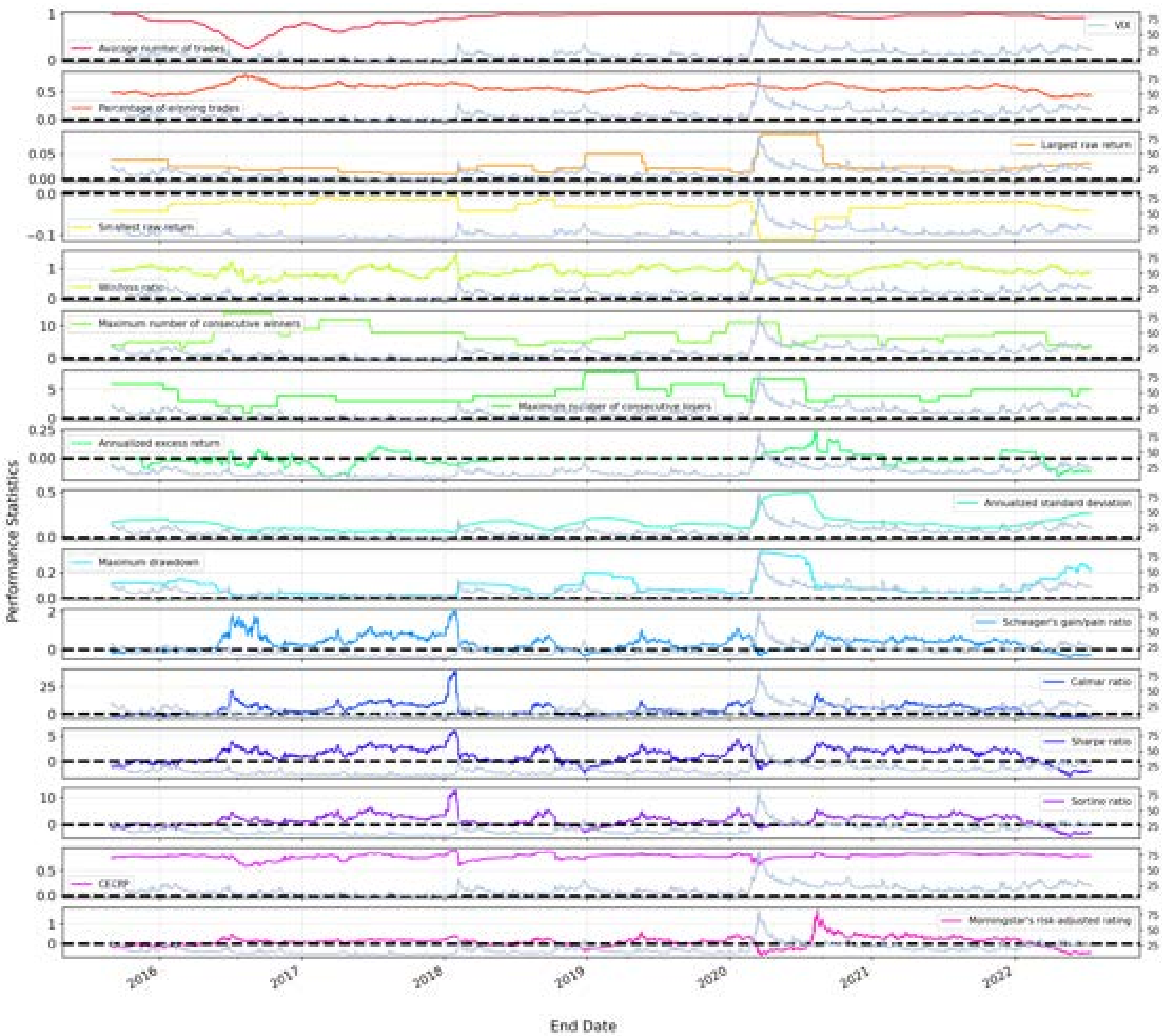
(d) *Dataset II*: *Sharpe ratio* score

[i] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

[ii] The values of the annualized excess return and Morningstar's risk-adjusted rating are scaled up ten times to make them more visible in this box plot.

[iii] *Dataset I* comprises of the financial variables and technical indicators while *Dataset II* contains both *Dataset I* and the candlestick chart (price) patterns as described in Table S.V.1.

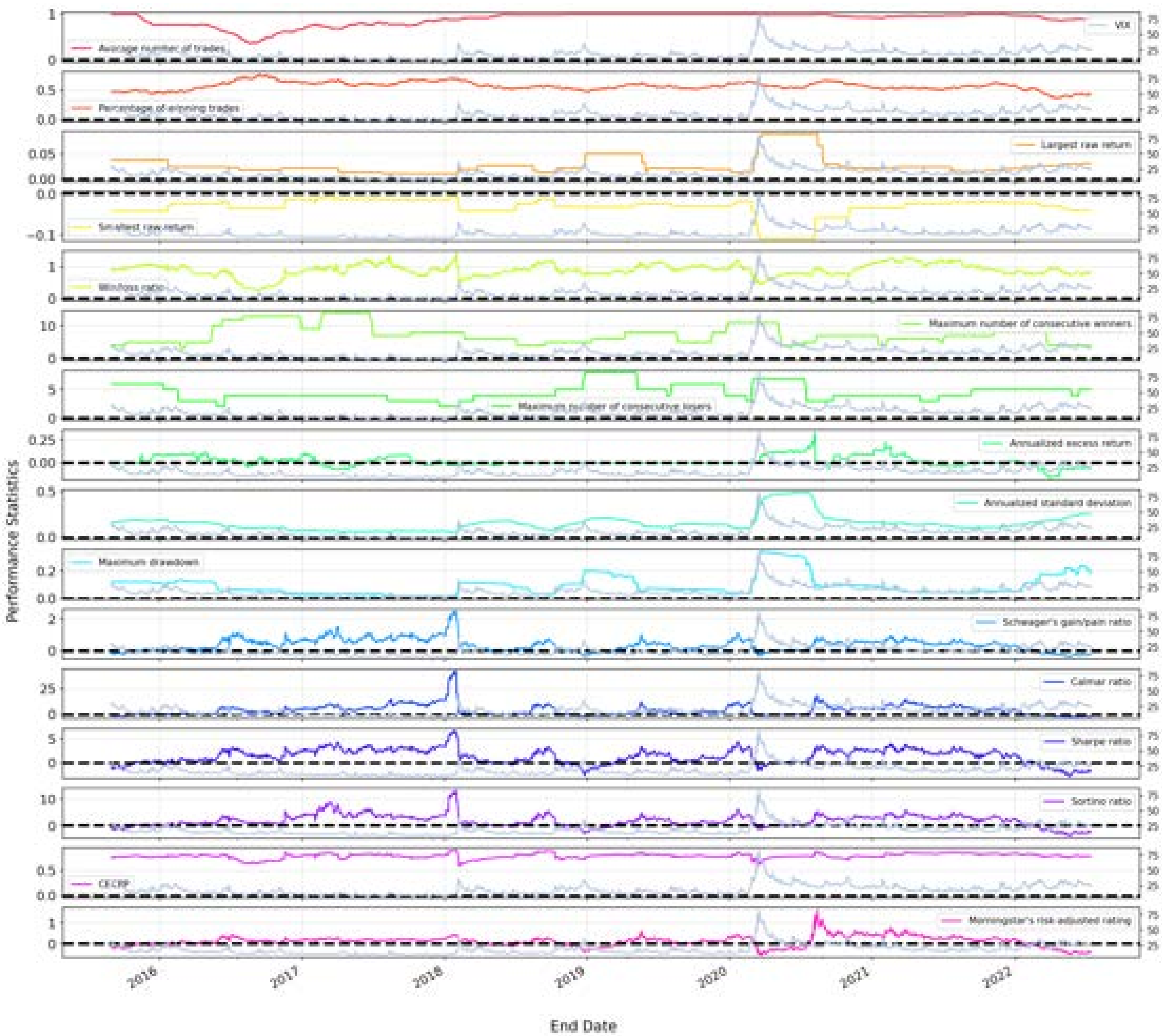
Figure A.IV.3: The performance statistics of 1718 trading sessions with the fixed transaction cost strategy investing in *SPY* for 100 days based on one-day ahead forecasts by a RF model (cross-validated with the *CECPP* scoring function) using the data on financial variables and technical indicators (*Dataset I*) described in Table S.V.1



[a] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

[b] The horizonal axis contains the end dates of the trading sessions.

Figure A.IV.4: The performance statistics of 1718 trading sessions with the fixed transaction cost strategy investing in *SPY* for 100 days based on one-day ahead forecasts by a RF model (cross-validated with the *CECPP* scoring function) using the data on financial variables, technical indicators, and candlestick chart patterns (*Dataset II*) described in Table S.V.1



[a] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

[b] The horizonal axis contains the end dates of the trading sessions.

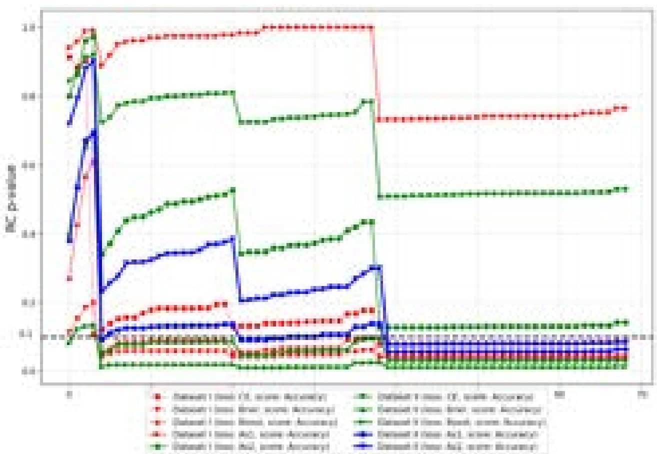
Table A.IV.1: The median (and IQR) of all the performance metrics of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the RF model cross-validated with the *CECPP* scoring function for different amounts of transaction cost across various holding periods (in days)

Holding period	Trans. cost	Percent. of win. trades		Win/loss ratio		Annualized excess ret.		Annualized std. dev.		Schwager's gain/pain ratio		Sharpe ratio		Sortino ratio		CECPP		MRAR	
		Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II
100	0.05	0.5625 (0.0744)	0.567 (0.0918)	0.8981 (0.2249)	0.9016 (0.2335)	-0.0011 (0.0301)	0.0 (0.0506)	0.1399 (0.0815)	0.1394 (0.0747)	0.2911 (0.506)	0.351 (0.5528)	1.2931 (2.1883)	1.4575 (2.2785)	1.4336 (2.844)	1.5541 (2.9895)	0.7956 (0.0501)	0.7928 (0.0531)	0.1231 (0.2838)	0.147 (0.2552)
	0.1	0.5625 (0.0744)	0.567 (0.0918)	0.8969 (0.2242)	0.8992 (0.2314)	-0.0014 (0.0312)	0.0 (0.0496)	0.1399 (0.0815)	0.1394 (0.0748)	0.2894 (0.5027)	0.3461 (0.5459)	1.278 (2.1704)	1.4355 (2.2572)	1.4116 (2.8254)	1.5518 (2.9762)	0.7946 (0.0504)	0.792 (0.0529)	0.1213 (0.2836)	0.1456 (0.2549)
	0.5	0.5612 (0.0709)	0.5652 (0.0944)	0.8987 (0.2272)	0.8923 (0.2313)	-0.0037 (0.0474)	-0.0 (0.0438)	0.1398 (0.0813)	0.1394 (0.0758)	0.2547 (0.4737)	0.3124 (0.4929)	1.1069 (2.0965)	1.2798 (2.0926)	1.2425 (2.7217)	1.461 (2.7656)	0.7849 (0.0536)	0.7854 (0.0511)	0.0986 (0.2795)	0.1242 (0.2485)
	1	0.5567 (0.0663)	0.5612 (0.0921)	0.8932 (0.2213)	0.8832 (0.2263)	-0.0162 (0.0727)	-0.0 (0.0516)	0.1399 (0.0807)	0.1397 (0.0762)	0.2009 (0.434)	0.2492 (0.4327)	0.8725 (1.9721)	1.0594 (1.9502)	1.0239 (2.5852)	1.2761 (2.56)	0.7729 (0.063)	0.7695 (0.0551)	0.0708 (0.2828)	0.0976 (0.2505)
	5	0.5417 (0.0622)	0.5464 (0.0669)	0.8442 (0.275)	0.8242 (0.285)	-0.0898 (0.2284)	-0.0418 (0.2116)	0.1422 (0.0731)	0.1437 (0.0725)	-0.0162 (0.4874)	-0.0062 (0.4619)	-0.1559 (2.7718)	-0.0875 (2.5042)	-0.1837 (3.4445)	-0.1111 (3.1747)	0.4624 (0.276)	0.4643 (0.2737)	-0.0777 (0.3736)	-0.0667 (0.3579)
200	0.05	0.5707 (0.051)	0.5736 (0.0689)	0.8649 (0.1582)	0.8635 (0.2227)	-0.0025 (0.0308)	0.0057 (0.0517)	0.1399 (0.0493)	0.1401 (0.0475)	0.2125 (0.3502)	0.2467 (0.4518)	0.926 (1.508)	1.0371 (1.7375)	1.0548 (1.9676)	1.1573 (2.2586)	0.7817 (0.051)	0.7816 (0.0457)	0.0972 (0.1682)	0.1213 (0.1828)
	0.1	0.5707 (0.0508)	0.5729 (0.0689)	0.8633 (0.1592)	0.8658 (0.2215)	-0.0027 (0.03)	0.0017 (0.0525)	0.1399 (0.0493)	0.1401 (0.0474)	0.2104 (0.345)	0.2447 (0.4451)	0.9187 (1.4935)	1.0199 (1.7066)	1.0471 (1.9538)	1.1511 (2.2525)	0.7807 (0.0509)	0.781 (0.0453)	0.0945 (0.1658)	0.12 (0.1801)
	0.5	0.5685 (0.0448)	0.5696 (0.0652)	0.8669 (0.1558)	0.8606 (0.2131)	-0.0161 (0.0478)	-0.0 (0.0574)	0.1401 (0.0497)	0.1401 (0.0473)	0.1934 (0.3091)	0.2226 (0.388)	0.8288 (1.3409)	0.9465 (1.5594)	0.9675 (1.708)	1.0715 (2.1382)	0.7744 (0.0477)	0.7727 (0.042)	0.0802 (0.1541)	0.1041 (0.1616)
	1	0.5657 (0.042)	0.5657 (0.0588)	0.8648 (0.1566)	0.8661 (0.1999)	-0.0293 (0.0676)	-0.0043 (0.0615)	0.1401 (0.0499)	0.14 (0.0474)	0.1672 (0.2724)	0.195 (0.3117)	0.6937 (1.1951)	0.8329 (1.3347)	0.7737 (1.5191)	0.939 (1.8367)	0.7604 (0.052)	0.7602 (0.0428)	0.056 (0.1418)	0.0838 (0.1451)
	5	0.5561 (0.0392)	0.5563 (0.0428)	0.8318 (0.2244)	0.8134 (0.2639)	-0.1026 (0.2316)	-0.0653 (0.2202)	0.1475 (0.0515)	0.1434 (0.0472)	0.0182 (0.336)	0.0325 (0.3209)	0.0013 (1.7615)	0.0669 (1.6571)	0.002 (2.1848)	0.0774 (1.9634)	0.4884 (0.2188)	0.5007 (0.225)	-0.0534 (0.2662)	-0.0361 (0.2185)

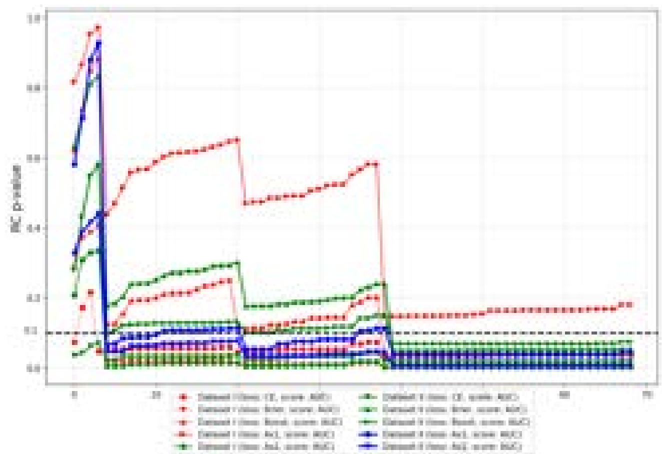
[a] *Dataset I* comprises of the financial variables and technical indicators while *Dataset II* contains both *Dataset I* and the candlestick chart (price) patterns as described in Table S.V.1.
[b] The initial endowment is $1000.
[c] The IQR values are shown in parentheses.
[d] There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

A.V Bootstrap Reality Check Evaluation of the Performance of Trading Methods

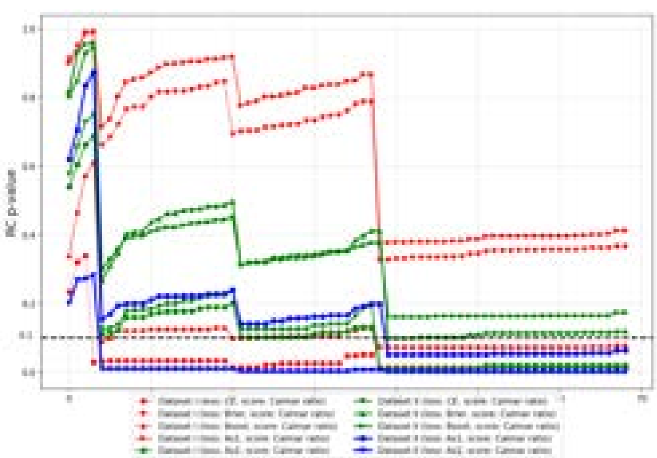
Figure A.V.1: The bootstrap Reality Check (RC) p-values for the test of the null hypothesis that a (benchmark) trading method performs at least as well as all the other trading methods: using the annualized excess return as a performance metric of the fixed transaction cost strategy [invested in *SPY*] based on one-day ahead forecasts by a LGBM model (trained with one of the following five different loss functions: *CE*, *Brier*, *Boost*, *As1*, and *As2*; and cross-validated with one of the following seven scoring functions: *Accuracy*, *AUC*, *Calmar ratio*, *Schwager's gain/pain ratio*, *Sortino ratio*, *Sharpe ratio*, and *CECPP*)



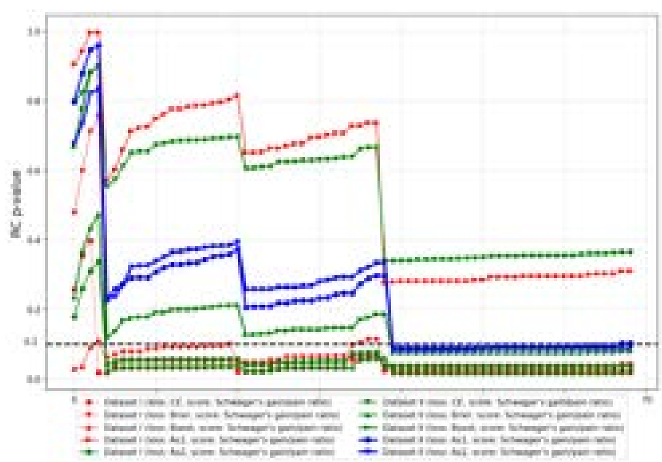
(a) *Accuracy* score



(b) *AUC* score



(c) *Calmar ratio* score



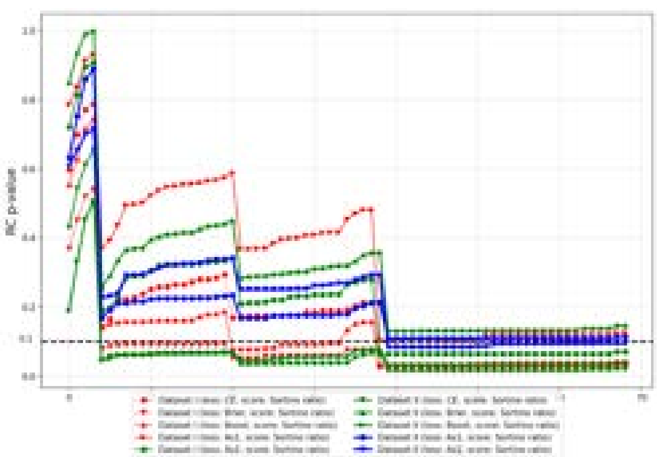
(d) *Schwager's (2012) gain to pain ratio* score

[a] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

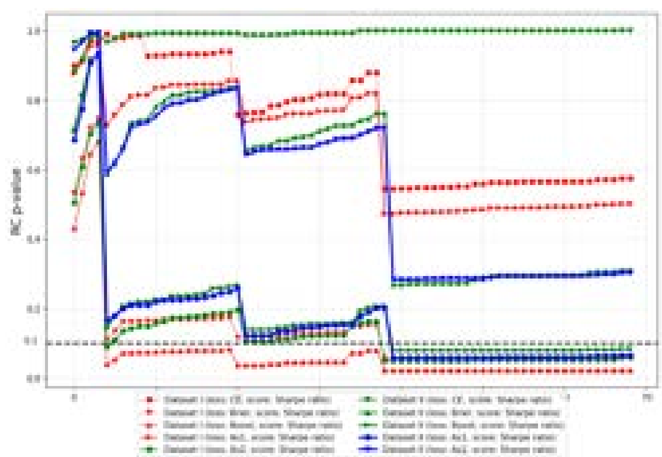
[b] *Dataset I* consists of financial variables and technical indicators while *Dataset II* consists of *Dataset I* and candlestick chart patterns (as described in Table S.V.1).

[c] There are 70 trading methods in total (as detailed in Section S.III). The benchmark trading methods are listed in the legend box of this plot.

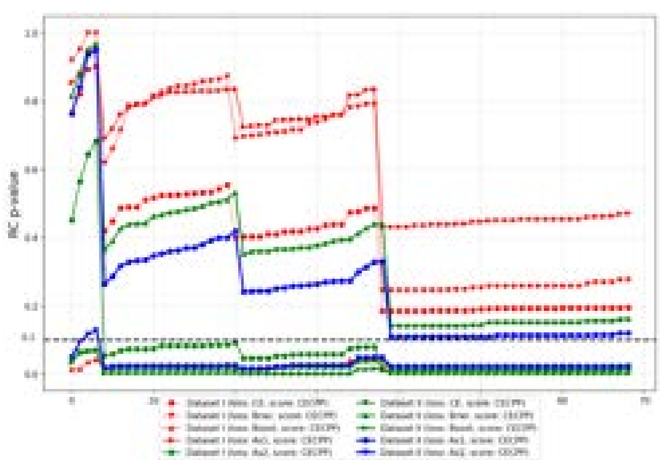
Figure A.V.1 (continued): The bootstrap Reality Check (RC) p-values for the test of the null hypothesis that a (benchmark) trading method performs at least as well as all the other trading methods: using the annualized excess return as a performance metric of the fixed transaction cost strategy [invested in *SPY*] based on one-day ahead forecasts by a LGBM model (trained with one of the following five different loss functions: *CE*, *Brier*, *Boost*, *As1*, and *As2*; and cross-validated with one of the following seven scoring functions: *Accuracy*, *AUC*, *Calmar ratio*, *Schwager's gain/pain ratio*, *Sortino ratio*, *Sharpe ratio*, and *CECPP*)



(e) *Sortino ratio* score



(f) *Sharpe ratio* score



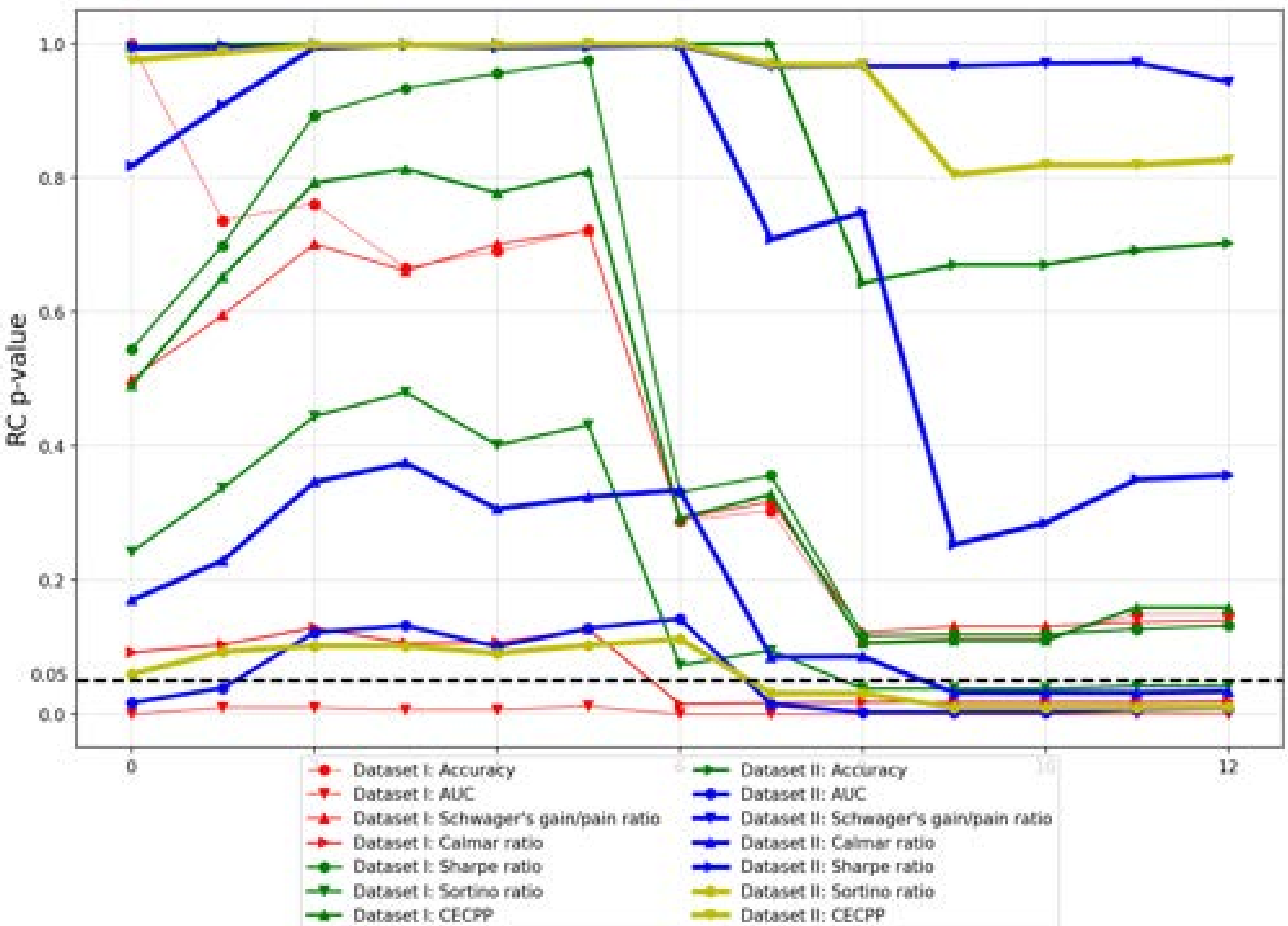
(g) *CECPP* score

[a] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

[b] *Dataset I* consists of financial variables and technical indicators while *Dataset II* consists of *Dataset I* and candlestick chart patterns (as described in Table S.V.1).

[c] There are 70 trading methods in total (as detailed in Section S.III). The benchmark trading methods are listed in the legend box of this plot.

Figure A.V.2: The bootstrap Reality Check (RC) p-values for the test of the null hypothesis that a (benchmark) trading method performs at least as well as all the other trading methods: using the annualized excess return as a performance metric of the fixed transaction cost strategy [invested in *SPY*] based on one-day ahead forecasts by a RF model (cross-validated with one of the following seven scoring functions: *Accuracy*, *AUC*, *Calmar ratio*, *Schwager's gain/pain ratio*, *Sortino ratio*, *Sharpe ratio*, and *CECPP*)



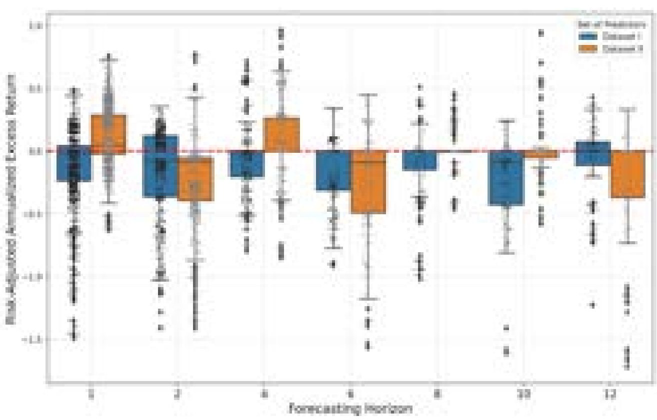
[a] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

[b] *Dataset I* consists of financial variables and technical indicators while *Dataset II* consists of *Dataset I* and candlestick chart patterns (as described in Table S.V.1).

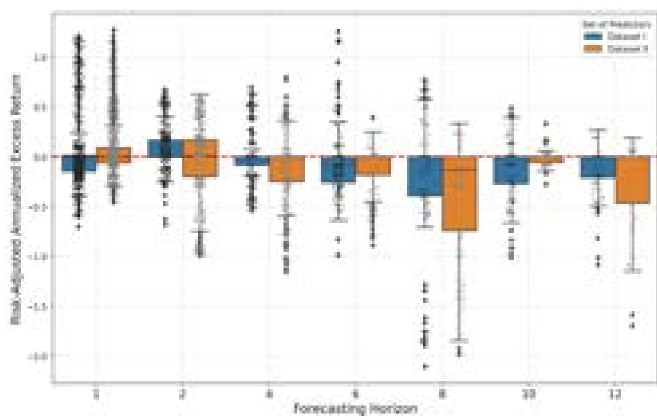
[c] There are 14 trading methods in total (as detailed in Section S.III). The benchmark trading methods are listed in the legend box of this plot.

A.VI Risk-Adjusted Annualized Excess Returns across Forecasting Horizons

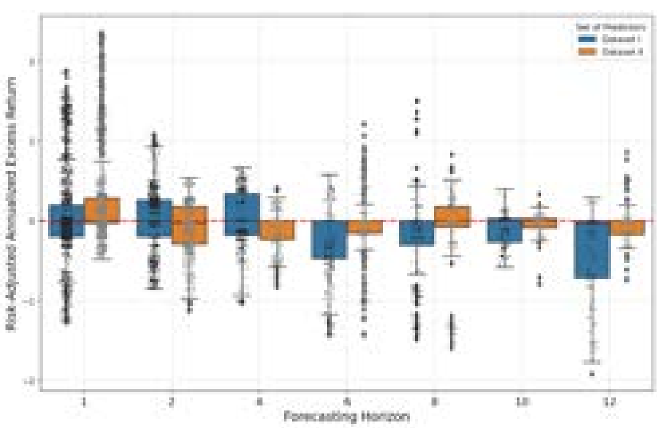
Figure A.VI.1: The risk-adjusted annualized excess returns (*or* the ratio of the annualized excess return over the annualized standard deviation) of the fixed transaction cost strategy investing in *SPY* for *200 days* based on multi-horizon forecasts [of price moving directions] by the RF models cross-validated with four different scoring functions



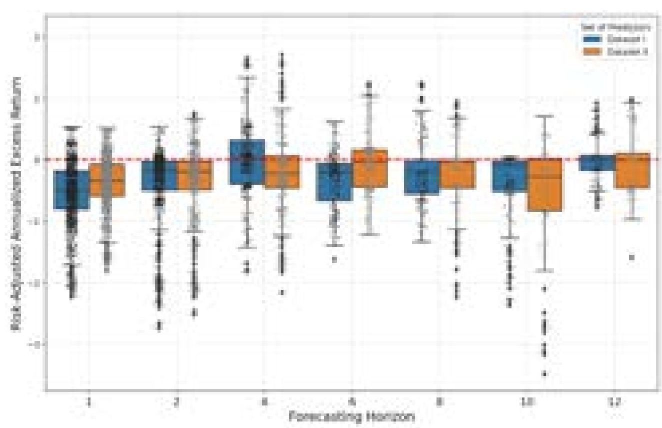
(a) *CECPP* score



(b) *Sharpe ratio* loss



(c) *Schwager's (2012) gain to pain ratio* score



(d) *AUC* score

[a] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

[b] Two sets of predictors are used: *Dataset I* (including the financial variables and technical indicators) and *Dataset II* (including *Dataset I* and candlestick chart patterns as described in Table S.V.1).

Technical analysis with machine learning classification algorithms: Can it still 'beat' the buy-and-hold strategy? – Supplemental Material –

Ba Chu*

December 28, 2023

Abstract

This Supplemental Material (SM) appendix contains (1) an illustrative description of the machine learning algorithms employed in this paper (S.I); (2) the list of hyperparameters used to cross-validate a Gradient Boosting Machine (GBM) or Random Forest (RF) model (S.II); (3) the details of the bootstrap Reality Check evaluation of the performance of trading methods (S.III); (4) a placebo-type experiment where prices are generated from a random walk to show that the performance of our trading strategy is not attributable to randomness (S.IV); (5) the definitions of all financial variables, technical indicators, and price patterns (S.V); and extra tables and line/box-plots used to show: (6) the performance metrics of the trading strategy across various scoring functions (S.VI); (7) the performance metrics of the trading strategy across various loss functions (S.VII); (8) the performance metrics of the trading strategy for different amounts of transaction cost (S.VIII); (9) the performance metrics of the trading strategy over time (S.IX); (10) the heat map of the global Shapley values of top predictors over time (S.X); (11) the values of each scoring function evaluated on the predictions of validation subsamples (S.XI); and (12) the risk-adjusted annualized excess returns generated by the trading strategy across various forecasting horizons (S.XII).

Keywords: Machine learning, Technical indicators, Price patterns, Directional predictability, Trading strategy
JEL classification: C53, C58, G11, G17

*Department of Economics, Carleton University, 1125 Colonel By Dr., Ottawa, Ontario, Canada. Email: ba.chu@carleton.ca. Tel: +1 613-520-2600 (ext. 1546).

S.I Machine Learning (ML) Algorithms

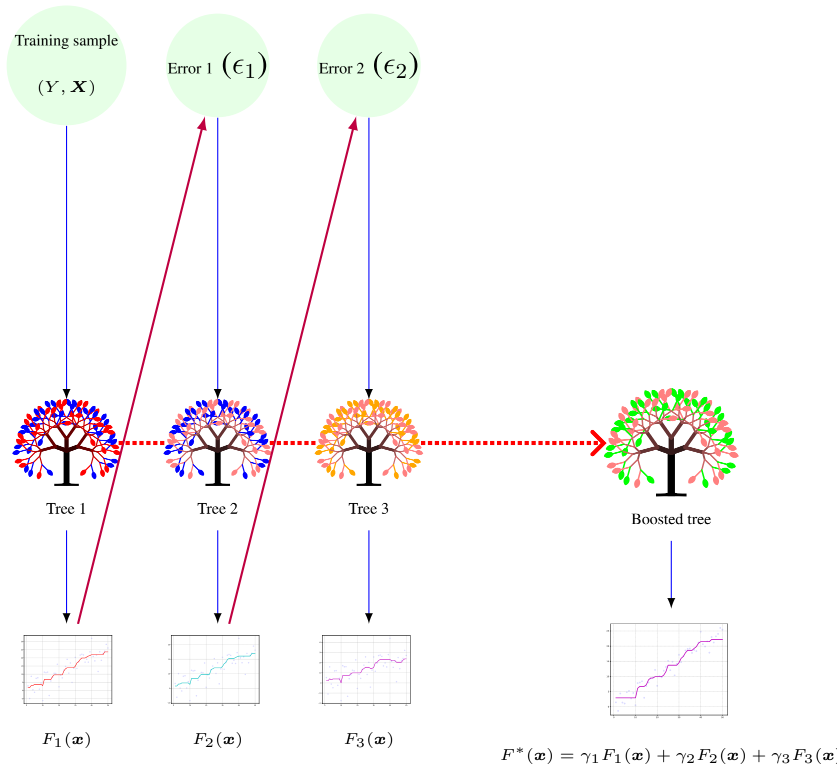
Gradient Boosting Machine (GBM)

1. Use the tree-growing procedure [described in the main text] to fit an initial decision tree [defined via the mapping $F_1(\boldsymbol{x})$] to a dataset on $(Y_{t+h}, \boldsymbol{X}_t)$, then compute the gradient (namely, the pseudo-residual) $\epsilon_{1,t+h} := -\left.\frac{\partial L(Y_{t+h}, F(\boldsymbol{X}_t))}{\partial F(\boldsymbol{X}_t)}\right|_{F(\boldsymbol{X}_t)=F_1(\boldsymbol{X}_t)}$ This residual captures any predictive content in Y_{t+h} unexplained by F_1. [This pseudo-residual becomes the conventional residual if the loss function is the mean squared error.]

2. Fit another decision tree [defined via the mapping $F_2(\boldsymbol{x})$] to the target $\epsilon_{1,t+h}$ using the same feature $\boldsymbol{X}_t$, then calculate the pseudo-residual $\epsilon_{2,t+h} := -\left.\frac{\partial L\left(\epsilon_{1,t+h}, F(\boldsymbol{X}_t)\right)}{\partial F(\boldsymbol{X}_t)}\right|_{F(\boldsymbol{X}_t)=F_2(\boldsymbol{X}_t)}$. This residual captures any predictive content in Y_{t+h} unexplained by both F_1 and F_2.

3. We continue to apply the above procedure until the iteration $n-1$, where we obtain another pseudo-residual, $\epsilon_{n-1,t+h}$. Fitting the n-th decision tree [defined via the mapping $F_n(\boldsymbol{x})$] to $\epsilon_{n-1,t+h}$ with the predictor $\boldsymbol{X}_t$ yields the n-th pseudo-residual $\epsilon_{n,t+h} - \left.\frac{\partial L\left(\epsilon_{1,t+h}, F(\boldsymbol{X}_t)\right)}{\partial F(\boldsymbol{X}_t)}\right|_{F(\boldsymbol{X}_t)=F_n(\boldsymbol{X}_t)}$.

4. Given n decision trees grown in the above steps, we score them in terms of their prediction accuracy with scoring weights, γ_i for $i = 1, \ldots, n$. The final boosting prediction of Y_{t+h} at a value of the predictor $\boldsymbol{X}_t = \boldsymbol{x}$ can then be generated through the conditional probability $Prob\left(Y_{t+h} = 1 \mid \boldsymbol{X}_t = \boldsymbol{x}\right) = S\left(\sum_{i=1}^{n} \gamma_i F_i(\boldsymbol{x})\right)$.



Random Forest

1. Given a dataset on $(Y_{t+h}, \boldsymbol{X}_t)$, $t = 1, \ldots, T$, draw a bootstrap sample of size less than or equal to the original sample size (T).
2. Grow a decision tree from each bootstrap sample. At each node,
   a. Randomly select a subset of features without replacement.
   b. Split the node using the feature that provides the best split using the threshold obtained by maximizing the information gain.
3. Repeat the above two steps k times to create k independent trees.
4. Aggregate the predictions produced by those k independent trees via majority voting. In order words, for any input, every decision tree produces a h-period ahead forecast of the moving direction of future price, and the majority wins.

S.II Hyperparameter Tuning for ML Algorithms

Table S.II.1: Hyperparameters used to cross-validate a LightGBM (LGBM) model

Parameter [a]	Searchable Values[b]
boosting_type	* a tree type: {the standard Gradient Boosting Decision tree ('gbdt'), the Dropouts meet Multiple Additive Regression tree ('dart'), or the Random Forest tree ('rf')}
max_depth	* the maximum depth of each decision tree (set it to a lower value to prevent overfitting): $Uniform\,\{1,\ldots,6\}$
num_leaves	* the maximum number of leaves in each tree: $Uniform\,\{10,20,30,40,\ldots,500\}$
learning_rate	* the learning rate used by the stochastic gradient descent algorithm for minimization of a loss function: $Uniform(0,100]$
min_child_samples	* the minimum number of observations in a child (leaf) node (set it to a bigger value to prevent overfitting): $Uniform\{10,\ldots,50\}$
feature_fraction	* the subsample ratio of columns when growing each tree: $Uniform(0,1)$
bagging_fraction	* the subsample ratio of training examples (set it to a smaller value to prevent overfitting): $Uniform(0,1)$
n_estimators	* the number of boosting iterations to be performed: $Uniform\,\{10,20,30,40,\ldots,200\}$

[a] A detailed description of all hyperparameters in LGBM can be found in the LGBM document.

[b] We use the Hyperband pruning algorithm with a random sampler (used to draw uniformly distributed hyperparameter values from their search domain) for hyperparameter optimization [Li et al. (2018)].

Table S.II.2: Hyperparameters used to cross-validate a Random Forest (RF) model

Parameter [a]	Searchable Values [b]
n_estimators	the number of decision trees used in RF: $Uniform\,\{10,20,30,\ldots,200\}$
max_features	the maximum number of randomly sampled features that are used by RF to grow a decision tree: $Uniform\,\{1,2,\ldots,\text{n\_features}\}$, where n_features is the number of features in the dataset
max_depth	the maximum depth of each decision tree (set it to a lower value to prevent overfitting): $Uniform\,\{1,\ldots,6\}$
min_samples_split	the minimum number of observations required to further split an internal node (set it to a bigger value to prevent overfitting): $Uniform\,\{2,12,22,\ldots,102\}$
min_samples_leaf	the minimum number of observations in a child (leaf) node (set it to a bigger value to prevent overfitting): $Uniform\,\{1,11,21,\ldots,101\}$

[a] A detailed description of all hyperparameters in RF can be found in the scikit-learn document.

[b] We use the Hyperband pruning algorithm with a random sampler (used to draw uniformly distributed hyperparameter values from their search domain) for hyperparameter optimization [Li et al. (2018)].

S.III Bootstrap Reality Check Evaluation of the Performance of Trading Methods

Let $perf_{\mathcal{T}_1+jh:\mathcal{T}_1+D+jh}$, $j = 0, \dots, \left\lfloor \frac{T-\mathcal{T}_1-D}{h} \right\rfloor$, represent a performance metric (such as the *annualized excess return* or the *Sharpe ratio* defined in the main text) of a trading strategy for a trading session, $j \in \left\{0, \left\lfloor \frac{T-\mathcal{T}_1-D}{h} \right\rfloor\right\}$, starting at $\mathcal{T}_1 + jh$ and ending at $\mathcal{T}_1 + D + jh$. By construction, this performance metric for the j-th trading session is a function of rolling sub-samples indexed by $\{jh + \tau h, \dots, \mathcal{T}_1 + jh + (\tau - 1)h\}$, $\tau = 0, \dots, \lceil \frac{D}{h} \rceil$, used to train and cross-validate a ML algorithm used to generate one-day ahead predictions. The OoS sub-sample indexed by $\mathcal{T}_1 + jh, \dots, \mathcal{T}_1 + D + jh$ is used to calculate the performance of the trading strategy. Suppose that we have K trading methods executed in the same trading session. Let $perf_{k,\mathcal{T}_1+jh:\mathcal{T}_1+D+jh}$, $k = 1, \dots, K$, denote the performance metrics of those trading methods. (In this case, a trading method is our trading strategy [defined in Section 4] using predictions made by a forecasting method.) We are interested in verifying the null hypothesis that a (benchmark) trading method, say $\ell \in \{1, \dots, K\}$, can perform at least as well as the best method in the complement set $\{1, \dots, K\} \setminus \ell$ on an average trading session:

$$H_0 : \max_{k \in \{1,\dots,K\} \setminus \ell} E \left[perf_{k,\mathcal{T}_1+jh:\mathcal{T}_1+D+jh} - perf_{\ell,\mathcal{T}_1+jh:\mathcal{T}_1+D+jh} \right] \leq 0. \tag{III-1}$$

This null hypothesis is different from the null hypothesis [of no predictive superiority over a benchmark model] used for White's (2000) Reality Check in one important aspect: our null hypothesis involves a performance metric calculated from a finite sample of observations while the null hypothesis in White (2000) involves a criterion function of OoS observations and the true values of model parameters. Therefore, as remarked in Giacomini and White (2006), the null hypothesis defined by (III-1) is used for comparing empirical methods or algorithms while the null hypothesis in White (2000) is suitable for comparing population parametric models.

We define the following statistic:

$$\bar{f}_{k,\ell} := \frac{1}{\left\lfloor \frac{T-\mathcal{T}_1-D}{h} \right\rfloor} \sum_{j=0}^{\left\lfloor \frac{T-\mathcal{T}_1-D}{h} \right\rfloor} f_{k,\ell,j}, \text{ where } f_{k,\ell,j} := perf_{k,\mathcal{T}_1+jh:\mathcal{T}_1+D+jh} - perf_{\ell,\mathcal{T}_1+jh:\mathcal{T}_1+D+jh}. \tag{III-2}$$

Theorem S.1 *Let the underlying data generating process* $(Y_t, \boldsymbol{X}_t)$, $t = 1, 2, \dots$, *be a strictly stationary, strongly mixing process with the mixing coefficient* $\alpha(\tau) = O\left(\tau^{-b}\right)$ *for some* $b > 3(6+\epsilon)/\epsilon$ *with some* $\epsilon > 0$. *Suppose that: (i)* $\max_{k \in \{1,\dots,K\}} E \left| perf_{k,\mathcal{T}_1+jh:\mathcal{T}_1+D+jh} \right|^{6+\epsilon} < \infty$ *and (ii)* $0 < \max(\mathcal{T}_1, D) < \infty$. *Then,*

$$\sup_{\boldsymbol{x}} \left| Prob\left\{ \sqrt{T} \boldsymbol{\Omega}_\ell^{-1/2} \left(\bar{\boldsymbol{f}}_\ell - E\left[\bar{\boldsymbol{f}}_\ell\right]\right) \leq \boldsymbol{x} \right\} - \Phi(\boldsymbol{x}) \right| \to 0,$$

where $\bar{\boldsymbol{f}}_\ell := \left\{\bar{f}_{k,\ell} :\ k \in \{1,\ldots,K\} \setminus \ell\right\}$, $\Phi(\boldsymbol{x})$ *represents the probability density function of a multivariate standard normal random variable, and* $\boldsymbol{\Omega}_\ell := \lim_{T\uparrow\infty} var\left(\sqrt{T}\bar{\boldsymbol{f}}_\ell\right)$ *is a positive definite variance-covariance matrix with finite maximum eigenvalue.*

Proof. Theorem S.1 immediately follows from Politis and Romano (1994, Theorem 2) by noticing that, if the underlying time series $(Y_t, \boldsymbol{X}_t)$, $t = 1, 2, \ldots,$ is strictly stationary and strongly mixing with the mixing coefficient $\alpha(\tau)$, then the performance metric of a trading method during a trading session, $j \in \left\{0, \ldots, \left\lfloor \frac{T-\mathcal{T}_1-D}{h} \right\rfloor\right\}$, defined by $perf_{\mathcal{T}_1+jh:\mathcal{T}_1+D+jh}$ as a measurable function of finite subsamples with indices $\{jh, \ldots, \mathcal{T}_1 + D + jh\}$ will also be strong mixing with a mixing coefficient, say $\alpha^{\diamond}(\tau)$, such that

$$\alpha^{\diamond}(\tau) = \begin{cases} 1 & \text{if } 0\leq\tau\leq\mathcal{T}_1+D \text{ or } -\mathcal{T}_1-D\leq\tau\leq 0, \\ \alpha(\tau-\mathcal{T}_1-D) & \text{if } \tau>\mathcal{T}_1+D, \\ \alpha(|\tau+\mathcal{T}_1+D|) & \text{if } \tau<-\mathcal{T}_1-D. \end{cases}$$

Since both $\mathcal{T}_1$ and D are finite, $\alpha(\tau) = O\left(\tau^{-b}\right)$ implies $\alpha^{\diamond}(\tau) = O\left(\tau^{-b}\right)$. ■

The conditions set out in Theorem S.1 need further discussion. The strong-mixing coefficient decaying at a polynomial rate is a standard condition used to guarantee weak dependence in the time series data. The moment condition (i) is required so that Politis and Romano (1994, Theorem 2) can be applied to obtain the main result. Condition (ii) is an important condition – the size of a rolling window and the duration of a trading session are finite. This condition is satisfied for estimation schemes with rolling windows (but not with recursive windows); and in practice, the length of a day-trading session is often quite short. Note that Condition (ii) can be relaxed by allowing $\mathcal{T}_1$ and D grow slowly relative to T at the cost of more elaborate mathematical arguments.

By White (2000, Proposition 2.2), Theorem S.1 implies that

$$\max_{k\in\{1,\ldots,K\}\setminus\ell}\left(\sqrt{T}\left\{\bar{f}_{k,\ell} - E\left[\bar{f}_{k,\ell}\right]\right\}\right) \xrightarrow{d} V_{K-1,\ell},$$

where $V_{K-1,\ell} := \max_{k\in\{1,\ldots,K\}\setminus\ell}\{\mathcal{N}_{k,\ell}\}$ and $\mathcal{N}_{k,\ell}$ is an element of a normal random vector distributed as $\boldsymbol{N}\{\boldsymbol{0}, \boldsymbol{\Omega}_\ell\}$. This limiting distribution allows one to apply White's (2000) bootstrap Reality Check (RC) procedure to test the null hypothesis defined by (III-1). Briefly speaking, the bootstrap RC p-value of the test statistic:

$$\bar{V}_{K-1,\ell} := \max_{k\in\{1,\ldots,K\}\setminus\ell}\sqrt{T}\bar{f}_{k,\ell}$$

can be obtained from the quantile of the resampled statistics:

$$\bar{V}^*_{K-1,\ell} := \max_{k\in\{1,\ldots,K\}\setminus\ell}\sqrt{T}\left(\bar{f}^*_{k,\ell} - \bar{f}_{k,\ell}\right),$$

where

$$\bar{f}^*_{k,\ell} := \frac{1}{\left\lfloor\frac{T-\mathcal{T}_1-D}{h}\right\rfloor}\sum_{j=0}^{\left\lfloor\frac{T-\mathcal{T}_1-D}{h}\right\rfloor} f^*_{k,\ell,j}, \tag{III-3}$$

where $f^*_{k,\ell,j} := perf_{k,\mathcal{T}_1+\theta(j)h:\mathcal{T}_1+D+\theta(j)h} - perf_{\ell,\mathcal{T}_1+\theta(j)h:\mathcal{T}_1+D+\theta(j)h}$ and $\theta(j)$, $j = 0, \ldots, \lfloor \frac{T-\mathcal{T}_1-D}{h} \rfloor$, is a sequence of random indices generated by Politis and Romano's (1994) stationary bootstrap.

This stationary bootstrap algorithm can generate a sequence of random observation indices, $\theta(j)$ for $j = 0, \ldots, \left\lfloor \frac{T-\mathcal{T}_1-D}{h} \right\rfloor$, by randomly drawing blocks of consecutive indices with random lengths of mean $\mu_T := 1/q_T$, where $q_T \in (0, 1]$ must satisfy $q_T \downarrow 0$ and $Tq_T \uparrow \infty$ as $T \uparrow \infty$. Following White (2000), the quantile of the resampled statistic $\bar{V}^*_{K-1,\ell}$ with a benchmark trading method, $\ell \in \{1, \ldots, K\}$, can be obtained sequentially as follows: Let $\mathcal{K} \equiv \{k_1, \ldots, k_{K-1}\} := \{1, \ldots, K\} \setminus \ell$, and suppose that we have B independent sequences of random observation indices, $\theta_b(j)$, $b = 1, \ldots, B$.

Step 1: select the first method k_1 from $\mathcal{K}$, compute $\bar{V}_{1,\ell} = \sqrt{T}\bar{f}_{k_1,\ell}$ and generate B bootstrap instances: $\bar{V}^*_{1,\ell;b} := \sqrt{T}\left(\bar{f}^*_{k_1,\ell;b} - \bar{f}_{k_1,\ell}\right)$, where

$$\bar{f}^*_{k_1,\ell;b} := \frac{1}{\left\lfloor \frac{T-\mathcal{T}_1-D}{h} \right\rfloor} \sum_{j=0}^{\left\lfloor \frac{T-\mathcal{T}_1-D}{h} \right\rfloor} f^*_{k_1,\ell,j;b}, \tag{III-4}$$

where $f^*_{k_1,\ell,j;b} := perf_{k_1,\mathcal{T}_1+\theta_b(j)h:\mathcal{T}_1+D+\theta_b(j)h} - perf_{\ell,\mathcal{T}_1+\theta_b(j)h:\mathcal{T}_1+D+\theta_b(j)h}$, $b = 1, \ldots, B$. The RC p-value in this iteration is then given by $p\text{-}value_1 := \frac{1}{B}\sum_{b=1}^{B} \mathbb{I}\left\{\bar{V}^*_{1,\ell;b} \geq \bar{V}_{1,\ell}\right\}$, where $\mathbb{I}(A)$ is an indicator function taking the value of one if A is true and zero otherwise.

Step 2: add the second method k_2 from $\mathcal{K}$, compute $\bar{V}_{2,\ell} = \max\left\{\sqrt{T}\bar{f}_{k_2,\ell}, \bar{V}_{1,\ell}\right\}$ and generate B bootstrap instances: $\bar{V}^*_{2,\ell;b} := \max\left\{\sqrt{T}\left(\bar{f}^*_{k_2,\ell;b} - \bar{f}_{k_2,\ell}\right), \bar{V}^*_{1,\ell;b}\right\}$, $b = 1, \ldots, B$. The RC p-value in this iteration is then given by $p\text{-}value_2 := \frac{1}{B}\sum_{b=1}^{B} \mathbb{I}\left\{\bar{V}^*_{2,\ell;b} \geq \bar{V}_{2,\ell}\right\}$.

Step k: continue adding the k-th method from $\mathcal{K}$ until Step $K-2$.

Step $K-1$: add the last method k_{K-1} from $\mathcal{K}$, compute $\bar{V}_{K-1,\ell} = \max\left\{\sqrt{T}\bar{f}_{k_{K-1},\ell}, \bar{V}_{K-2,\ell}\right\}$ and generate B bootstrap instances: $\bar{V}^*_{K-1,\ell;b} := \max\left\{\sqrt{T}\left(\bar{f}^*_{k_{K-1},\ell;b} - \bar{f}_{k_{K-1},\ell}\right), \bar{V}^*_{K-2,\ell;b}\right\}$, $b = 1, \ldots, B$. The RC p-value used to test the null hypothesis defined by (III-1) is then given by

$$p\text{-}value_{K-1} := \frac{1}{B}\sum_{b=1}^{B} \mathbb{I}\left\{\bar{V}^*_{K-1,\ell;b} \geq \bar{V}_{K-1,\ell}\right\}.$$

For LightGBM, we have $K = 70$ trading methods defined by enumerating the product of the list of sets of predictors, the list of loss functions, and the list of scoring functions:

$$\{\textit{Dataset I, Dataset II}\} \times \{CE, Brier, Boost, As1, As2\}$$
$$\times\{Accuracy, AUC, \textit{Schwager's gain/pain ratio}, \textit{Calmar ratio}, \textit{Sharpe ratio}, \textit{Sortino ratio}, \textit{CECPP}\}.$$

For Random Forest, we have $K = 14$ trading methods defined by enumerating the product of the list of sets of predictors and the list of scoring functions:

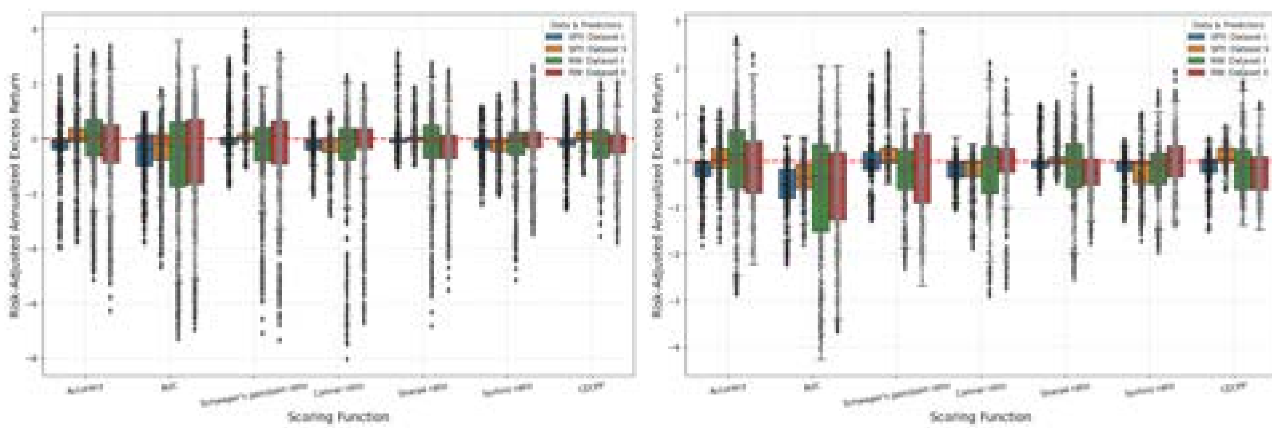
$$\{\textit{Dataset I, Dataset II}\}$$
$$\times\{Accuracy, AUC, \textit{Schwager's gain/pain ratio}, \textit{Calmar ratio}, \textit{Sharpe ratio}, \textit{Sortino ratio}, \textit{CECPP}\}\,.$$

S.IV Performance of the Trading Strategy with Random-Walk Price Data

We address the question: Can the performance of our trading strategy obtained in the previous subsections be ascribed to entire randomness? We shall conduct a placebo-type experiment by generating logarithmic closing stock prices from a random walk (RW) and drawing random volume data from an integer uniform distribution for the same time period (from 3/30/2011 to 7/15/2022) of the *SPY* dataset. Specifically, the logarithmic closing price C_t follows the RW: $C_t = C_{t-1} + \epsilon_t$, where ϵ_t are randomly sampled from a truncated $[-0.1, 0.1]$ normal distribution with mean (and standard deviation) equal to the sample mean (standard deviation) of the *SPY* returns [which are equal to 0.000519 (0.010956)]. The reason for choosing this 10% threshold is that the closing price of *SPY* did not change over 10% per day during the considered period so that the returns drawn from this RW process would be similar in magnitude to the returns of *SPY*; the logarithmic open price $O_t := C_t + u_t$, where $u_t \overset{i.i.d.}{\sim} N(0, 0.1)$ is independent of ϵ_t; the logarithmic low price $L_t = \min(O_t, C_t) - |v_t|$, where $v_t \overset{i.i.d.}{\sim} N(0, 0.2)$ is independent of both ϵ_t and u_t; and the logarithmic high price $H_t = \max(O_t, C_t) + |w_t|$, where $w_t \overset{i.i.d.}{\sim} N(0, 0.2)$ is independent of ϵ_t, u_t, and v_t. The volume data $V_t \overset{i.i.d.}{\sim} IntUniform[10, 1000]$ are also drawn independently from the price data. We use the same set of financial variables, technical indicators, and price patterns as the predictors of the directions of future changes in these randomly simulated prices. Since the stock price data are randomly generated, the predictors using historical data should have null predictive power for future price changes. A ML algorithm thus performs no better than a random guess. Therefore, the annualized excess return of the trading strategy (trading based on the predictions of the moving directions of the price of this RW stock) can have a 50% chance of being above or below zero.

We employ RF to predict the moving directions of future prices, and then use these predictions as inputs to our trading strategy. To make the returns of two different assets comparable, we use the risk-adjusted annualized excess return (i.e., the ratio of the annualized excess return over the annualized standard deviation defined in Section 4). Figure S.IV.1 shows the box plots of the risk-adjusted annualized excess returns of 1718 trading sessions with the fixed transaction cost strategy [invested in SPY vs. the RW stock] based on one-day ahead forecasts by a RF model (cross-validated with one of the following seven scoring functions: *Accuracy*, *AUC*, *Calmar ratio*, *Schwager's gain/pain ratio*, *Sortino* *ratio*, *Sharpe ratio*, and *CECPP* across two holding periods ($D = 100$ or 200 days). The risk-adjusted annualized excess returns of the RW stock clearly have a much larger IQR than those of *SPY*, and the median of the risk-adjusted annualized excess returns of *SPY* is greater than that of the RW stock for most of the considered scoring functions across the two holding periods (in this case, the better scoring functions are still *Sharpe ratio* and *CECPP*). This finding suggests that the performance of the trading strategy based on predictions of the directions of entirely random price changes tends to be much more random than that based on predictions of the directions of price changes with some predictive content. We can therefore conclude that the performance of our trading strategy is generally attributed to the possibility that the financial variables, technical indicators or price patterns do have some predictive power.

Figure S.IV.1: The risk-adjusted annualized excess returns of 1718 trading sessions with the fixed transaction cost strategy [invested in *SPY* vs. an asset with entirely random returns and trading volumes] based on one-day ahead forecasts by a RF model (cross-validated with one of the following seven scoring functions: *Accuracy*, *AUC*, *Calmar ratio*, *Schwager's gain/pain ratio*, *Sortino ratio*, *Sharpe ratio*, and *CECPP*)



(a) Holding period (D) = 100 days

(b) Holding period (D) = 200 days

[a] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

[b] As defined above, *RW* is an asset with closing prices (generated from a random walk) and trading volumes (generated from an integer uniform distribution). For each asset, we use two sets of predictors as before: *Dataset I* consists of financial variables and technical indicators while *Dataset II* consists of *Dataset I* and candlestick chart patterns (as described in Table S.V.1).

[c] The risk-adjusted annualized excess return is the ratio of the annualized excess return over the annualized standard deviation.

S.V Financial Variables, Technical Indicators, and Price Patterns

Table S.V.1: List of predictors for the total return of *SPY*

Variable and technical indicator/pattern	Description	Some references[a]
Dividend-price ratio (*dp*)	* The ratio of the dividend in a quarter released on/before a specific date over the closing stock price on that date	Campbell and Shiller (1988a,b)
Logarithmic daily trading volume (*vol*)	* The number of shares traded in a given day. Volume indicates the amount of interest the traders put in a particular security. Volume is also used to measure liquidity	Conrad et al. (1994); Avramov et al. (2006)
The 3-month T-Bill rate (*T-bill*)	* The market yield on U.S. treasury securities at 3-month constant maturity (often used as a proxy for the short-term risk-free rate). Source: https://fred.stlouisfed.org	Spiro (1990); Bessembinder (2018)
ICE BofA US corporate bond total return index (*corp_bond_ind*)	* This index tracks the performance of US dollar denominated investment grade rated corporate debt publicly issued in the U.S. market. Source: https://fred.stlouisfed.org	Baele et al. (2010); Connolly et al. (2005); Campbell et al. (1993)
Term spread (*ts*)	* The yield of the 10-year Treasury note minus the yield of the 3-month Treasury bill. In Fama and French (1989), *ts* [defined as the difference between the yield on an Aaa bond portfolio and the one-month Treasury bill rate] can track the business cycle. Source: https://fred.stlouisfed.org	Fama and French (1989)
S&P 500 earnings per share (*eps*)	* The aggregate net profit of companies listed in the index divided by the number of common shares outstanding. Source: https://data.nasdaq.com	Campbell and Shiller (1988b); Lamont (1998)
Default spread (*df*)	* The Baa corporate bond yield minus the yield of 10-Year Treasury note. In Fama and French (1989), *df* is measured by the difference between the Baa and Aaa corporate bond yields. Source: https://fred.stlouisfed.org	Fama and French (1989)
The 10-Year breakeven inflation rate (*inflation*)	* The current value of *inflation* indicates what market participants expect inflation to be in the next 10 years, on average. Equity markets tend to underperform when the expected inflation is high. Source: https://fred.stlouisfed.org	Fama (1981); Bekaert and Engstrom (2010); Wang (2010)

Continued on next page

[a] These predictors have been used by many authors. So, it is not possible to list all of the works here. We could only give a few examples.

Table S.V.1 (continued): List of predictors for the total return of *SPY*

Variable and technical indicator/pattern	Description	Some references[a]
Bollinger bands (*bbands*)	* The *bbands* at the current time are defined as: $LowerBand :=$ $SMA_T(TP) - 2SMSDEV(TP)$ and $UpperBand :=$ $SMA_T(TP) + 2SMSDEV_T(TP)$, where TP is the typical current price (i.e., the average of the current low, high, and closing prices), and SMA_T and $SMSDEV_T$ are the T-day simple MA and standard deviations respectively. A *buy* signal (+1) is generated when the current daily closing price crosses the lower band from above. A *sell* signal (-1) is generated when this price crosses the upper band from below	Bollinger (2001) and Colby (2002, p. 114)
Exponential MA (*EMA*)	* The current *EMA* is the discount sum of T observations available up to this point in time where the past observations have a diminishing contribution to the sum while the more recent ones have a greater contribution. *EMA* can smooth out the price of a stock to reveal its underlying local trend	Brock et al. (1992) and Colby (2002, p. 261)
Double EMA (*DEMA*)	* The current *DEMA* is a linear combination of a single EMA using the most recent T observations and a double EMA to produce another EMA with less lag than the original EMAs: $DEMA_T := 2EMA_T - EMA_T\,(EMA_T)$	Mulloy (1994)
Triple EMA (*TEMA*)	* The current TEMA is a trend-following indicator, which is a linear combination of a single EMA using the most recent T observations, a double EMA, and a triple EMA: $TEMA_T :=$ $3EMA_T - 3DEMA_T + EMA_T\,(DEMA_T)$	Mulloy (1994)
Kaufman adaptive MA (*KAMA*)	* Like EMA and DEMA, KAMA is another trend-following indicator, but it accounts for both the noise in and the moving direction of price. KAMA uses a fast EMA trend during a low-noise period and a slower EMA trend during a high-noise period. KAMA is calculated using the most recent T closing prices	Kaufman (2019, ch. 17)

Continued on next page

Table S.V.1 (continued): List of predictors for the total return of *SPY*

Variable and technical indicator/pattern	Description	Some references[a]
Parabolic stop and reverse (*SAR*)	* SAR is a trend-following indicator, which can be used to determine exit and entry points based on the current open/closing prices of an asset. For example, if the current price cross the SAR curve, then exit the market. SAR is also determined by the most recent T closing prices	Wilder (1978, p. 9)
Average directional movement index (*ADX*)	* ADX is a filtered momentum indicator (a high number implies the existence of a strong trend while a low number suggests the existence of a weak trend). ADX is calculated using the most recent T closing prices. ADXR is the range of ADX	Hsu et al. (2016); Wilder (1978, p. 35)
Absolute price oscillator (*APO*)	* APO is the percentage difference between a shorter and a longer MA. Thus, APO is determined by the lengths of these two MAs. A buy signal is generated when the APO crosses above zero, and a sell signal is generated when the APO crosses below zero	Colby (2002, p. 538)
The Aroon indicator (*Aroon*)	* Aroon is used to discover a new trend coming. When this indicator consists of two lines (UP and DOWN). The UP between 70 and 100 suggests an upward trend. The DOWN between 70 and 100 suggests a downward trend. A buy signal is generated when the UP crosses the DOWN from below while a sell signal is generated when the UP crosses the DOWN from above	Chande (1995); Colby (2002, p. 102)
The Aroon oscillator (*Aroonosc*)	* The difference between the Aroon UP and the Aroon DOWN. A high value of Aroonosc suggests a strong upward trend while a low value indicates a strong downward trend	Chande (1995); Colby (2002, p. 102)
The balance of power indicator (*BOP*)	* BOP measures the strength of a price trend. BOP varies in between [-1, 1], where positive values indicate strong buying pressure and negative values suggest strong selling pressure. The current BOP is calculated as $BOP_T := MA_T\left(\frac{\text{closing price - open price}}{\text{high price - low price}}\right)$, where MA_T is the moving average of the most recent T observations of daily prices	Worden (1950)

Continued on next page

Table S.V.1 (continued): List of predictors for the total return of *SPY*

Variable and technical indicator/pattern	Description	Some references[a]
The commodity channel index (*CCI*)	* CCI is used to early detect the beginning and the end of price trends. CCI values outside [-100, 100] suggest overbought/oversold conditions. the current CCI is calculated from the past T observations	Colby (2002, p. 155); Schlossberg (2006, p. 91)
The Chande momentum oscillator (*CMO*)	* The current CMO is the ratio of the net movement over the total movement of stock prices over the last T periods: $CMO_T := \frac{ups_T - downs_T}{ups_T + downs_T}$, where ups_T and $downs_T$ are the numbers of up and down movements respectively over the last T periods	Anghel (2021); Chande and Kroll (1994, p. 94)
Moving average convergence/divergence (*MACD*)	* MACD is the difference between two EMAs (say, a short EMA of length $T/2$ minus a long EMA of length T as defined in this paper), and the signal line is the EMA [of length $T/3$] of the MACD. High values of the MACD indicate overbought conditions while low values suggest oversold conditions. A buy signal is generated when the MACD line crosses the signal line from below, and a sell signal is generated when the MACD line crosses the signal line from above	Colby (2002, p. 412)
Money flow index (*MFI*)	* The current MFI approximates the ratio of money flowing into and out of a stock over the last T periods. Money flow is measured by the product of typical price and volume. MFI values fall in between 0 and 100 (MFI above 80/below 20 indicates market top/bottom respectively)	Achelis (2001); Yen and Hsu (2010)
Directional movement indices (*DM*)	* The plus DM (+*DM*) is the percentage of the true range (TR) smoothed over the last T periods that is up (TR is defined as the largest value of: high price - low price, high price - the previous period's closing price, and the previous period's closing price - low price). The minus DM (-*DM*) is the percentage of the smoothed TR that is down. A buy signal is generated when the +DM crosses the -DM from below while a sell signal is generated when the -DM crosses the +DM from below	Colby (2002, p. 212)

Continued on next page

Table S.V.1 (continued): List of predictors for the total return of *SPY*

Variable and technical indicator/pattern	Description	Some references[a]
Momentum (*MOM*)	* MOM measures the acceleration and deceleration of prices (the current MOM is the current closing price minus the closing price T periods ago)	Colby (2002, p. 400); Jegadeesh and Titman (1993)
Percentage price oscillator (*PPO*)	* PPO is the percentage difference between a long MA of length T and a short MA of length $T/2$	Colby (2002, p. 538)
Rate of change (*ROC*)	* The current ROC is the percentage difference between the current closing price and the closing price T periods ago	Colby (2002, p. 596)
Relative strength index (*RSI*)	* RSI is the ratio of the EMA of T-period gains divided by the absolute value of the EMA of T-period losses. RSI values fall in between 0 and 100. The RSI generates overbought / oversold signals when its value is over 70 / below 30 respectively	Levy (1967); Wilder (1978, p. 63)
Stochastic oscillator (*STOCH*)	* STOCH compares a stock's closing price with its price range over T periods. STOCH has two main lines: the first line is defined as $FASTK_t := \frac{P_t - P_{\min}}{P_{\max} - P_{\min}}$, where P_t is today's closing price, and $P_{\min}$ / $P_{\max}$ is the lowest / highest closing price over the last T periods respectively. The second line is just the MA of $FASTK_t$ over the last T periods. A buy / sell signal is generated when $FASTK$ rises above / below its MA	Achelis (2001); Lane (1984)
Stochastic RSI (*STOCHRSI*)	* STOCHRSI is the stochastic oscillator applied to a set of relative strength indices instead of closing price data	Chande and Kroll (1994, p. 119)
1-day ROC of a triple smooth EMA (*TRIX*)	* The current TRIX is the percent rate of change of a triple EMA of the last T periods' closing prices. TRIX oscillates around zero (a buy/sell signal can be generated when the TRIX cross above/below zero)	Colby (2002, p. 702); Hutson (1983)
Ultimate oscillator (*ULTOSC*)	* ULTOSC is a time-weighted momentum oscillator. ULTOSC values range from zero to 100 (values over 70 indicate overbought conditions while values under 30 indicate oversold conditions)	Colby (2002, p. 715); Williams (1985)
Williams' percent range (*WILLR*)	* WILLR is an inverse of *STOCH* with its values ranging from zero to 100 (values below 20 indicate overbought conditions while values above 80 indicate oversold conditions)	Achelis (2001); Williams (1985)

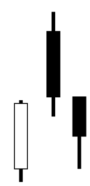
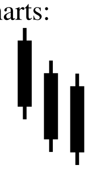
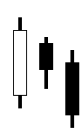
Continued on next page

Table S.V.1 (continued): List of predictors for the total return of *SPY*

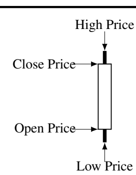
Variable and technical indicator/pattern	Description	Some references[a]
The Chaikin accumulation/distribution line (*AD*)	* AD is a volume indicator used to monitor the flow of trading volume into and out of the market. Buy/sell signals can be generated by comparing this flow to price action. The formula of AD is $AD_t := AD_{t-1} + vol_t \left(\frac{(P_t - PL_t) - (PH_t - P_t)}{PH_t - PL_t} \right)$, where P_t, PL_t, and PH_t are the closing, low, and high prices at time t	Colby (2002, p. 752); Meliones and Makrides (2019)
the Chaikin AD oscillator (*ADOSC*)	* ADOSC is the difference between a $T/2$-period long EMA and a T-period long EMA of AD. A buy signal is initiated when the ADOSC rises while in a negative territory and the current price is above its 90-day MA. A sell signal is initiated when the ADOSC declines while in a positive territory and the current price is below its 90-day MA	Colby (2002, p. 752); Meliones and Makrides (2019)
On balance volume (*OBV*)	* OBV is the cumulative total of volume relative to price change. This indicator shows if trading volume is flowing into or out of a stock	Gerritsen et al. (2020); Granville (1976)
Average true range (*ATR*)	* ATR is simply the T-period long MA of the True Range (TR). The TR is the largest value of $PH - PL$, $PH - P_{-1}$, and $P_{-1} - PL$, where PH / PL are the high / low price of the current period and P_{-1} is the closing price of the previous period. ATR measures the volatility of a stock (high / low ATR values suggest high / low volatility)	Gustafson (2001); Wilder (1978, p. 21)
The Hilbert transform - dominant cycle period (*HT_DCPERIOD*)	* HT_DCPERIOD is used to detect price trends, cycles, and even volatility. This technical indicator is constructed by employing the Hilbert transform in digital signal processing	Ehlers (2001, ch. 6); Kablan and Ng (2010)
The Hilbert transform - dominant cycle phase (*HT_DCPHA*)	* HT_DCPHA is used to detect price trends, cycles, and even volatility. This technical indicator is constructed by employing the Hilbert transform in digital signal processing	Ehlers (2001, ch. 6); Kablan and Ng (2010)
The Hilbert Transform - phasor components (*HT_PHA*)	* HT_PHA provides the phasor components of price cycles	Ehlers (2001, ch. 6); Kablan and Ng (2010)

Continued on next page

Table S.V.1 (continued): List of predictors for the total return of *SPY*

Variable and technical indicator/pattern	Description	Some references[a]
The Hilbert Transform - SineWave (*HT_SINE*)	* HT_SINE produces a sine wave plot of stock prices used to distinguish between cyclical and trending price activity	Ehlers (2001, ch. 6); Kablan and Ng (2010)
Hilbert Transform Trendline (*HT_TRENDLINE*)	* A trendline of high/low price constructed by signal processing so as to minimize noise	Ehlers (2001, ch. 6)
The Hilbert Transform - trend vs cycle mode (*HT_TREND*)	* HT_TREND indicates if a price movement is a part of a trend or a cycle	Ehlers (2001, ch. 11); Kablan and Ng (2010)
Moving standard deviation (*MSTDDEV*)	* MSTDDEV is a conventional measure of volatility	Achelis (2001)
The Two Crows pattern (*2CROWS*)	* 2CROWS predicts that price may drop after advancing during the previous day:[b] 	Bulkowski (2008, p. 853)
The Three Black Crows pattern (*3BLACKCROWS*)	* 3BLACKCROWS predicts that the up trend may reverse pretty rapidly. The performance rank of this pattern is 3 (out of 103 candles). Thus, it is one of the best candlestick charts: 	Bulkowski (2008, p. 728)
The Three Inside Down pattern (*3INSIDEDOWN*)	* 3INSIDEDOWN predicts that price may reverse after a rally on the first day. This candle appears often in a bear market. The performance rank is 53 (out of 103): 	Bulkowski (2008, p. 738)

Continued on next page

[b] A white candlestick is defined as 

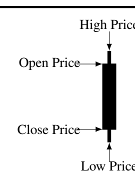
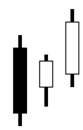
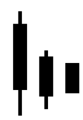
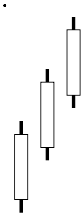
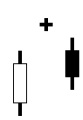
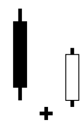
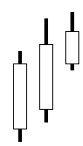
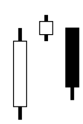
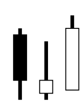
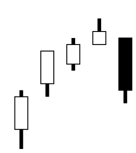
while a black candlestick is defined as 

Table S.V.1 (continued): List of predictors for the total return of *SPY*

Variable and technical indicator/pattern	Description	Some references[a]
The Three Inside UP pattern (*3INSIDEUP*)	* 3INSIDEUP predicts that price may go up after a down day. This candle also appears often in a bear market. The performance rank is 20 (out of 103): 	Bulkowski (2008, p. 747)
The Three Stars in the South pattern (*3STARSINSOUTH*)	* 3STARSINSOUTH predicts that the bear market will end in a few days and the bull market will start. This candle is quite rare in stocks. The performance rank is 103 (out of 103): 	Bulkowski (2008, p. 747)
The Three Advancing White Soldiers pattern (*3WHITESOLDIERS*)	* 3WHITESOLDIERS predicts that price will close higher each day in the next three days. This candle appears often in a bear market. The performance rank is 32 (out of 103): 	Bulkowski (2008, p. 794)
The Abandoned Baby (Bearish) pattern (*ABANDONEDBABY-BEAR*)	* ABANDONEDBABYBEAR predicts that price will go down in a bull market and continue going up in a bear market. This candle is quite rare in stocks. The overall performance rank is 64 (out of 103): 	Bulkowski (2008, p. 70)

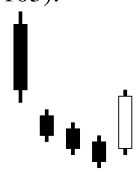
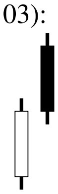
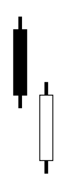
Continued on next page

Table S.V.1 (continued): List of predictors for the total return of *SPY*

Variable and technical indicator/pattern	Description	Some references[a]
The Abandoned Baby (Bullish) pattern (*ABANDONEDBABY-BULL*)	* ABANDONEDBABYBULL predicts that price will go up 70% of the time during a bull market. The performance rank is 9 (out of 103): 	Bulkowski (2008, p. 80)
The Advance Block pattern (*ADVANCEBLOCK*)	* ADVANCEBLOCK predicts that price will continue to rise possibly at a slower speed. There may be some chance of reversal in the next few days. This candle appears more often in a bear market. The performance rank is 54 (out of 103): 	Bulkowski (2008, p. 98)
The Belt-hold (Bearish) pattern (*BELTHOLDBEAR*)	* BELTHOLDBEAR predicts that an upward trend may reverse in the next few days. This candle appears very often. The performance rank is 63 (out of 103): 	Bulkowski (2008, p. 118)
The Belt-hold (Bullish) pattern (*BELTHOLDBULL*)	* BELTHOLDBULL predicts that price will break out upward from an on-going down trend. This candle appears most often in a bull market. The performance rank is 62 (out of 103): 	Bulkowski (2008, p. 128)
The Breakaway (Bearish) pattern (*BREAKAWAYBEAR*)	* BREAKAWAYBEAR predicts that price will reverse from an on-going upward trend, mostly during a bear market. This candle is quite rare in stocks. The overall performance rank is 11 (out of 103): 	Bulkowski (2008, p. 137)

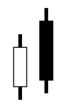
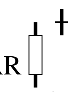
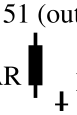
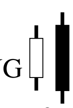
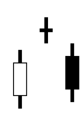
Continued on next page

Table S.V.1 (continued): List of predictors for the total return of *SPY*

Variable and technical indicator/pattern	Description	Some references[a]
The Breakaway (Bull) pattern (*BREAKAWAYBULL*)	* BREAKAWAYBULL predicts that price will break out (upward) from an on-going downward trend, mostly during a bear market. The performance rank is 45 (out of 103): 	Bulkowski (2008, p. 141)
The Closing Marubozu pattern (*CLOSING-MARUBOZU*)	* CLOSINGMARUBOZU is a continuation pattern, which suggests that a downward / upward trend will continue if the candle is black  or white 	Bulkowski (2008, pp. 504 - 551)
The Concealing Baby Swallow pattern (*CON-CEALBABYSWALL*)	* CONCEALBABYSWALL predicts that a downward trend will continue, mostly during a bull market. This candle is quite rare. Its performance in a bull market ranks third. However, the overall performance is not quite impressive (it ranks 101 out of 103): 	Bulkowski (2008, p. 178)
The Counterattack (Bearish) pattern (*COUNTERATTACK-BEAR*)	* COUNTERATTACKBEAR predicts that an upward trend will continue around 50% of the time. The performance rank is 16 (out of 103): 	Bulkowski (2008, p. 570)
The Counterattack (Bullish) pattern (*COUNTERATTACK-BULL*)	* COUNTERATTACKBULL predicts that a downward trend will continue around 50% of the time. The performance rank is 18 (out of 103): 	Bulkowski (2008, p. 579)

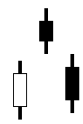
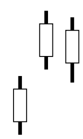
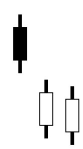
Continued on next page

Table S.V.1 (continued): List of predictors for the total return of *SPY*

Variable and technical indicator/pattern	Description	Some references[a]
The Dark Cloud Cover pattern (*DARK-CLOUDCOVER*)	* DARKCLOUDCOVER predicts the reversal of an upward trend over 60% of the time. This candle appears most often in a bull market. The performance rank is 22 (out of 103): 	Bulkowski (2008, p. 182)
The Doji Star pattern (*DOJISTAR*)	* The Bearish DOJISTAR  predicts that an upward trend will continue above 67% of the time; the overall performance rank of this pattern is 51 (out of 103). The Bullish DOJISTAR  predicts that a downward trend will continue above 60% of the time; the overall performance rank of this pattern is 49 (out of 103)	Bulkowski (2008, pp. 266 - 284)
The Dragonfly Doji pattern (*DRAGONFLYDOJI*)	* DRAGONFLYDOJI  predicts the reversal / continuation of a trend in a bull / bear market around 50% of the time. The performance rank is 98 (out of 103)	Bulkowski (2008, p. 202)
The Engulfing pattern (*ENGULFING*)	* The Bearish ENGULFING  predicts that price will go down in the next few days; this candle appears quite often in stocks; the overall performance rank is 91 (out of 103). The Bullish ENGULFING  predicts that price will go up in the next few days; this candle appears rather frequently in a bear market; the performance rank is 84 (out of 103)	Bulkowski (2008, pp. 308 - 325)
The Evening Doji Star pattern (*EVENINGDOJISTAR*)	* EVENINGDOJISTAR predicts the reversal of an upward trend. This candle appears most often in a bull market. The performance rank is 30 (out of 103): 	Bulkowski (2008, p. 326)

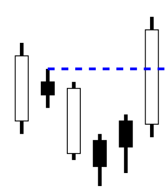
Continued on next page

Table S.V.1 (continued): List of predictors for the total return of *SPY*

Variable and technical indicator/pattern	Description	Some references[a]
The Evening Star pattern (*EVENINGSTAR*)	* EVENINGSTAR predicts the reversal of an upward trend over 70% of the time. This candle appears most often in a bear market. The performance rank is 4 (out of 103): 	Bulkowski (2008, p. 335)
The Up-gap Side-by-Side White Lines pattern (*UP-GAPSIDESIDEWHITE*)	* UP-GAPSIDESIDEWHITE predicts the continuation of an upward trend. This candle is quite rare and is more likely to appear in a bull market. The performance rank is 46 (out of 103): 	Bulkowski (2008, p. 685)
The Down-gap Side-by-Side White Lines pattern (*DOWN-GAPSIDESIDEWHITE*)	* DOWN-GAPSIDESIDEWHITE predicts the continuation of a downward trend. This candle is quite rare and is more likely to appear in a bull market. The performance rank is 29 (out of 103): 	Bulkowski (2008, p. 676)
The Gravestone Doji pattern (*GRAVESTONEDOJI*)	* GRAVESTONEDOJI  predicts that the reversal of an upward trend next day. This cadle is quite common, but its overall performance is random (which ranks 77 out of 103)	Bulkowski (2008, p. 230)
The Hammer pattern (*HAMMER*)	* HAMMER  predicts the reversal of a severe downtrend. This candle appears very often. The performance rank is 65 (out of 103)	Bulkowski (2008, p. 348)
The Hanging Man pattern (*HANGINGMAN*)	* HANGINGMAN  predicts the continuation of a strong upward trend. This candle appears very often, mostly in a bear market. The performance rank is 87 (out of 103)	Bulkowski (2008, p. 365)

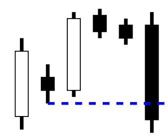
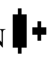
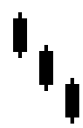
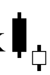
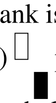
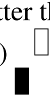
Continued on next page

Table S.V.1 (continued): List of predictors for the total return of *SPY*

Variable and technical indicator/pattern	Description	Some references[a]
The Bearish Harami pattern (*HARAMIBEAR*)	* HARAMIBEAR predicts the continuation / reversal of an upward trend in a bull / bear market. This candle appears quite often, particularly in a bear market. The performance rank is 72 (out of 103)	Bulkowski (2008, p. 374)
The Bullish Harami pattern (*HARAMIBULL*)	* HARAMIBULL predicts the reversal of a downward trend about 50% of the time. This candle appears quite often, particularly in a bear market. The overall performance rank is 38 (out of 103)	Bulkowski (2008, p. 383)
The Harami Cross pattern (*HARAMICROSS*)	* HARAMICROSS (Bearish) predicts the continuation of an upward trend over 50% of the time while HARAMICROSS (Bullish) predicts the continuation of a downward trend over 50% of the time. These candles appear often in stocks. The performance ranks of HARAMICROSS (Bearish) and HARAMICROSS (Bullish) are 80 and 50 (out of 103) respectively	Bulkowski (2008, pp. 392 - 408)
The High-Wave Candle pattern (*HIGHWAVE*)	* HIGHWAVE predicts a trend reversal. This candle appears more often in a bear market than in a bull market. The overall performance rank is 67 (out of 103)	Bulkowski (2008, p. 409)
The Bullish Hikkake pattern (*HIKKAKEBULL*)	* HIKKAKEBULL predicts the continuation of an upward trend above 50% of the time. This candle appears very often in stocks. The overall performance rank is about 80 (out of 105): 	Chesler (2004b,a)

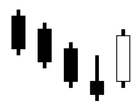
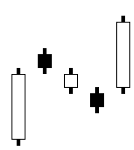
Continued on next page

Table S.V.1 (continued): List of predictors for the total return of *SPY*

Variable and technical indicator/pattern	Description	Some references[a]
The Bearish Hikkake pattern (*HIKKAKEBEAR*)	* HIKKAKEBEAR predicts the continuation of a downward trend above 50% of the time. This candle appears very often in stocks. The overall performance rank is about 80 (out of 105): 	Chesler (2004b,a)
The Homing Pigeon pattern (*HOMINGPIGEON*)	* HOMINGPIGEON  predicts the continuation of a downward trend over 50% of the time. This candle appears most often in a bull market. The performance rank is 21 (out of 103)	Bulkowski (2008, p. 418)
The Identical Three Crows pattern (*IDENTICAL3CROWS*)	* IDENTICAL3CROWS predicts the reversal of an upward trend. This candle is quite rare. The overall performance rank is 24 (out of 103): 	Bulkowski (2008, p. 427)
The In-Neck pattern (*INNECK*)	* INNECK  predicts the continuation of a downward trend. This candle is somewhat rare (it mostly appears in a bear market.) The overall performance rank is 17 (out of 103)	Bulkowski (2008, p. 436)
The Kicking pattern (*KICKING*)	* KICKING (Bearish)  predicts the reversal of an upward trend in a bull market while it predicts the continuation of a downward trend in a bear market; this candle is quite rare and it performs a little better than a random guess. KICKING (Bullish)  predicts the reversal of a downward trend next day; this candle appears a little more often than KICKING (Bearish) and it performs a little better than a random guess	Bulkowski (2008, pp. 444 - 458)

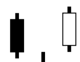
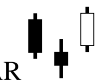
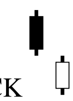
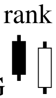
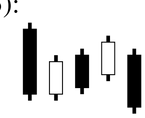
Continued on next page

Table S.V.1 (continued): List of predictors for the total return of *SPY*

Variable and technical indicator/pattern	Description	Some references[a]
The Inverted Hammer pattern (*INVERTED-HAMMER*)	* INVERTEDHAMMER predicts the continuation of a downward trend over 65% of the time. The performance rank is 6 (out of 103)	Bulkowski (2008, p. 356)
The Ladder Bottom pattern (*LADDERBOTTOM*)	* LADDERBOTTOM predicts the reversal of a downward trend in a bear market while it predicts the continuation of an upward trend in a bull market. This candle appears most often in a bull market. The overall performance rank is 41 (out of 103): 	Bulkowski (2008, p. 459)
The Long Legged Doji pattern (*LONG-LEGGEDDOJI*)	* LONGLEGGEDDOJI predicts the continuation of an upward trend in a bull market while it predicts the reversal of a downward trend in a bear market. This pattern is very common in stocks. The overall performance rank is 37 (out of 103)	Bulkowski (2008, p. 239)
The Marubozu pattern (*MARUBOZU*)	* MARUBOZU or or or or or predicts that the price trend will continue. This candle appears often in stocks. However, its overall performance rank is quite modest	Bulkowski (2008, pp. 504 - 551)
The Matching Low pattern (*MATCHINGLOW*)	* MATCHINGLOW predicts that the downward trend will continue over 60% of the time. The performance rank is 8 (out of 103)	Bulkowski (2008, pp. 561)
The Mat Hold pattern (*MATHOLD*)	* MATHOLD predicts the continuation of an upward trend over 65% of the time. This candle is quite rare. The performance rank is 86 (out of 103): 	Bulkowski (2008, p. 552)

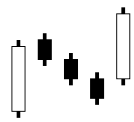
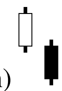
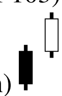
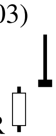
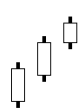
Continued on next page

Table S.V.1 (continued): List of predictors for the total return of *SPY*

Variable and technical indicator/pattern	Description	Some references[a]
The Morning Doji Star pattern (*MORNING-DOJISTAR*)	* MORNINGDOJISTAR  predicts the reversal of a downward trend over 70% of the time. This candle is quite rare in stocks. The performance rank is 25 (out of 103)	Bulkowski (2008, p. 588)
The Morning Star pattern (*MORNINGSTAR*)	* MORNINGSTAR  predicts the reversal of a downward trend over 65% of the time. This candle appears often in a bear market. The performance rank is 12 (out of 103)	Bulkowski (2008, p. 598)
The On-Neck pattern (*ONNECK*)	* ONNECK  predicts the continuation of a downward trend about 55% of the time. This candle is somewhat rare in stocks. The performance rank is 33 (out of 103)	Bulkowski (2008, p. 607)
The Piercing pattern (*PIERCING*)	* PIERCING  predicts the reversal of a downward trend over 60% of the time. This candle appears most often in a bear market. The performance rank is 13 (out of 103)	Bulkowski (2008, p. 616)
The Rickshaw Man pattern (*RICKSHAWMAN*)	* RICKSHAWMAN  predicts that the price trend will continue / reverse in a bull / bear market. This candle appears very frequently in stocks. The performance rank is 35 (out of 103)	Bulkowski (2008, p. 625)
The Falling Three Methods pattern (*FALL3METHODS*)	* FALL3METHODS predicts the continuation of a downward trend over 65% of the time. This candle is quite rare. The overall performance rank is 89 (out of 103): 	Bulkowski (2008, p. 344)

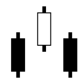
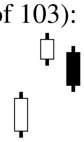
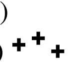
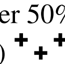
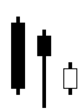
Continued on next page

Table S.V.1 (continued): List of predictors for the total return of *SPY*

Variable and technical indicator/pattern	Description	Some references[a]
The Rising Three Methods pattern (*RISE3METHODS*)	* RISE3METHODS predicts the continuation of an upward trend over 70% of the time. This candle is quite rare, but it functions well as a continuation pattern. The overall performance rank is 94 (out of 103): 	Bulkowski (2008, p. 633)
The Separating Lines pattern (*SEPARATINGLINES*)	* SEPARATINGLINES (Bearish)  predicts the continuation of a downward trend over 60% of the time; and this candle has the performance rank of 40 (out of 103). SEPARATINGLINES (Bullish)  predicts that price will continue to go up over 65% of the time; and this candle has the performance rank of 36 (out of 103)	Bulkowski (2008, pp. 642 - 659)
The Shooting Star pattern (*SHOOTINGSTAR*)	* SHOOTINGSTAR  predicts the continuation of an upward trend about 60% of the time. This candle often appears in a bull market. The performance rank is 52 (out of 103)	Bulkowski (2008, p. 668)
The Spinning Top pattern (*SPINNINGTOP*)	* SPINNINGTOP  or  predicts that the price trend will reverse about 50% of the time. This candle is very common in stocks. The performance rank is about 70 (out of 103)	Bulkowski (2008, pp. 694 - 709)
The Stalled pattern (*STALLEDPATTERN*)	* STALLEDPATTERN predicts the continuation of an upward trend over 75% of the time. This candle appears more often in a bear market. The performance rank is 93 (out of 103): 	Bulkowski (2008, p. 191)

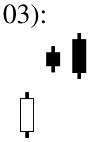
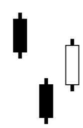
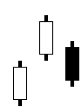
Continued on next page

Table S.V.1 (continued): List of predictors for the total return of *SPY*

Variable and technical indicator/pattern	Description	Some references[a]
The Stick Sandwich pattern (*STICKSANDWICH*)	* STICKSANDWICH predicts that price will continue to go lower over 60% of the time. This candle appears mostly in a bull market. The overall performance rank is 14 (out of 103): 	Bulkowski (2008, p. 710)
The Takuri pattern (*TAKURI*)	* TAKURI  predicts the reversal of a downward trend over 60% of the time. This candle appears quite frequently, especially in a bear market. The performance rank is 47 (out of 103)	Bulkowski (2008, p. 720)
The Tasuki Gap pattern (*TASUKIGAP*)	* TASUKIGAP predicts that price will continue to go higher over 50% of the time. The overall performance rank is 5 (out of 103): 	Bulkowski (2008, p. 888)
The Thrusting pattern (*THRUSTING*)	* THRUSTING  predicts the reversal of a downward trend over 55% of the time. This candle is quite rare in stocks. The performance rank is 15 (out of 103)	Bulkowski (2008, p. 803)
The Tristar pattern (*TRISTAR*)	* TRISTAR (Bearish)  predicts the reversal of an upward trend aver 50% of the time while TRISTAR (Bullish)  predicts the reversal of a downward trend over 55% of the time. This candle is rare in stocks. The performance rank of the bearish candle is 76 while that of the bullish candle is 28 (out of 103)	Bulkowski (2008, pp. 812 - 828)
The Unique 3 River pattern (*UNIQUE3RIVER*)	* UNIQUE3RIVER predicts that price will continue to go lower over 55% of the time. This candle appears most often in a bear market. The overall performance rank is 60 (out of 103): 	Bulkowski (2008, p. 861)

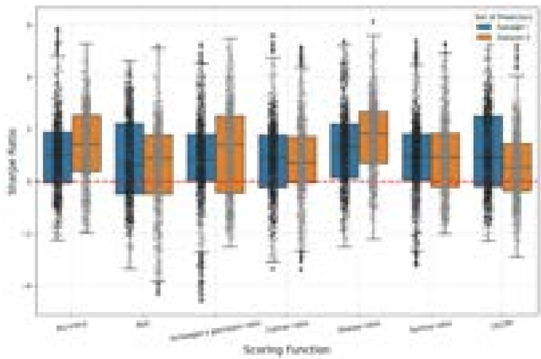
Continued on next page

Table S.V.1 (continued): List of predictors for the total return of *SPY*

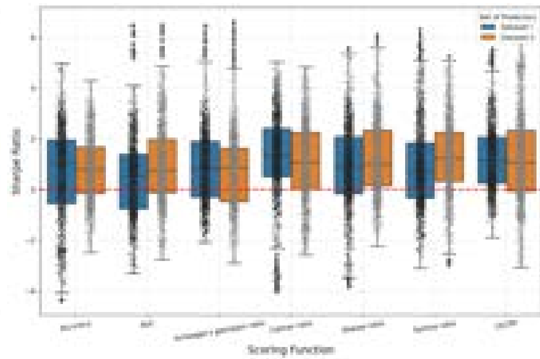
Variable and technical indicator/pattern	Description	Some references[a]
The Upside Gap Two Crows pattern (*UP-SIDEGAP2CROWS*)	* UPSIDEGAP2CROWS predicts that price will continue to go higher over 50% of the time. This candle is quite rare in stocks. The overall performance rank is 74 (out of 103): 	Bulkowski (2008, p. 879)
The Downside Gap Three Methods pattern (*DOWN-GAP3METHODS*)	* DOWNGAP3METHODS predicts the reversal of a downward trend about 60% of the time. This candle appears more often in a bear market. The performance rank is 26 (out of 103): 	Bulkowski (2008, p. 289)
The Upside Gap Three Methods pattern (*UPGAP3METHODS*)	* UPGAP3METHODS predicts the reversal of an upward trend over 59% of the time. This candle also appears more often in a bear market. The performance rank is 27 (out of 103): 	Bulkowski (2008, p. 870)

S.VI Performance of the Trading Strategy across Scoring Functions

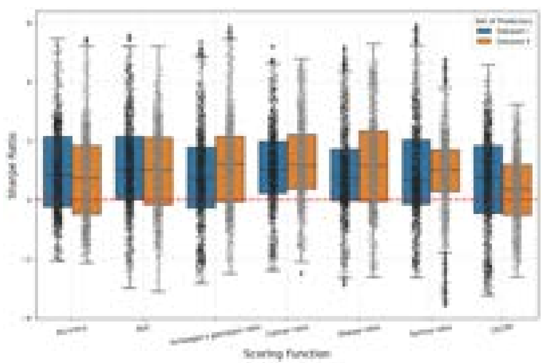
Figure S.VI.1: The Sharpe ratios of 1718 trading sessions with the fixed transaction cost strategy investing in *SPY* for 100 days based on one-day ahead forecasts by the LGBM models cross-validated with seven different scoring functions



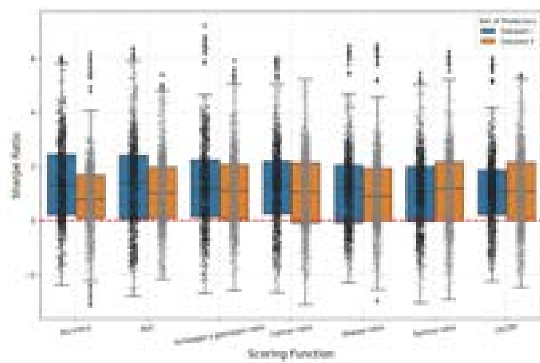
(a) *CE* loss



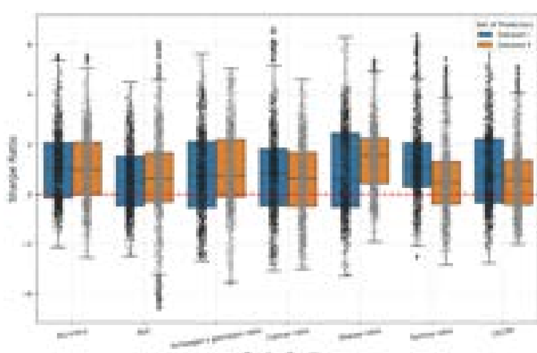
(b) *Brier* loss



(c) *Boost* loss



(d) *As1* loss

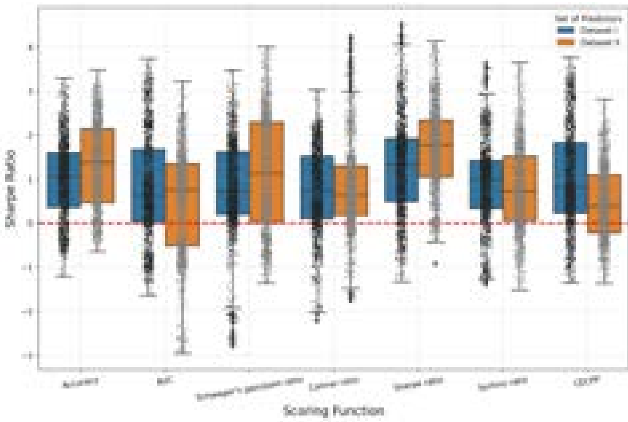


(e) *As2* loss

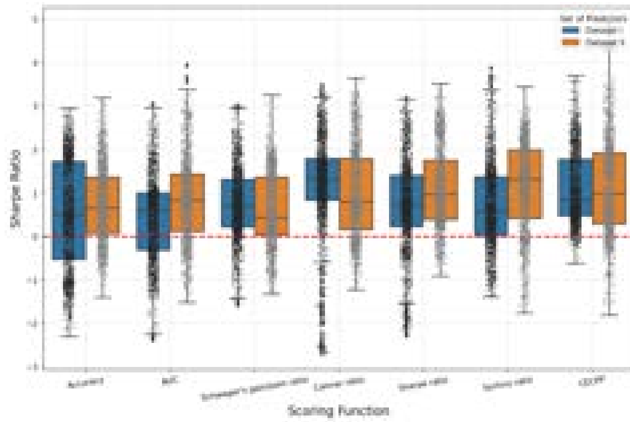
i The initial endowment is $1000, and the amount of fixed transaction cost $c = \$0.05$.

ii Two sets of predictors are used: *Dataset I* (including the financial variables and technical indicators) and *Dataset II* (including *Dataset I* and candlestick chart patterns as described in Table S.V.1 in the main text).

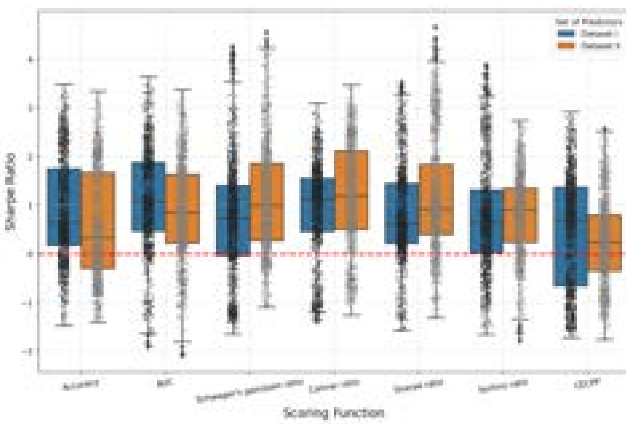
Figure S.VI.2: The Sharpe ratios of 1618 trading sessions with the fixed transaction cost strategy investing in *SPY* for 200 days based on one-day ahead forecasts by the LGBM models cross-validated with seven different scoring functions



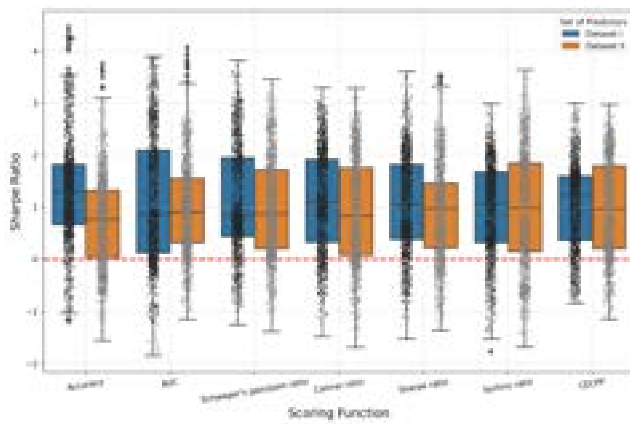
(a) *CE* loss



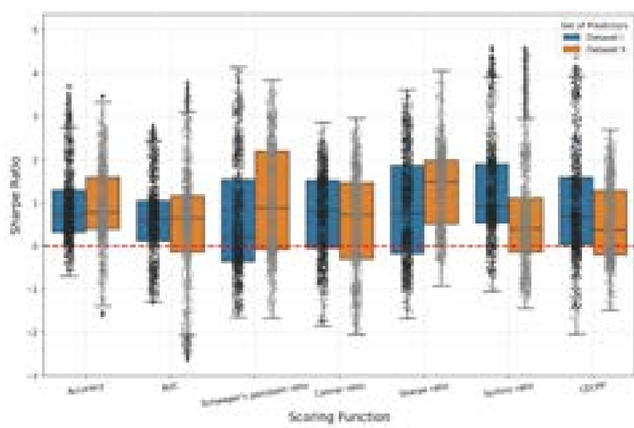
(b) *Brier* loss



(c) *Boost* loss



(d) *As1* loss



(e) *As2* loss

i The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

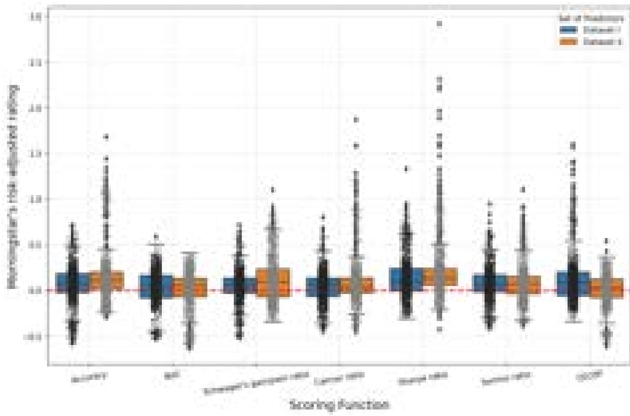
ii Two sets of predictors are used: *Dataset I* (including the financial variables and technical indicators) and *Dataset II* (including *Dataset I* and candlestick chart patterns as described in Table S.V.1).

Table S.VI.1: The median (and IQR) of the Sharpe ratio values of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the LGBM models trained with the *Brier* loss function and cross-validated with seven different scoring functions (i.e., *Accuracy, AUC, Schwager's gain/pain ratio, Calmar ratio, Sharpe ratio, Sortino ratio,* or *CECPP*) for different amounts of transaction cost across various holding periods (in days)

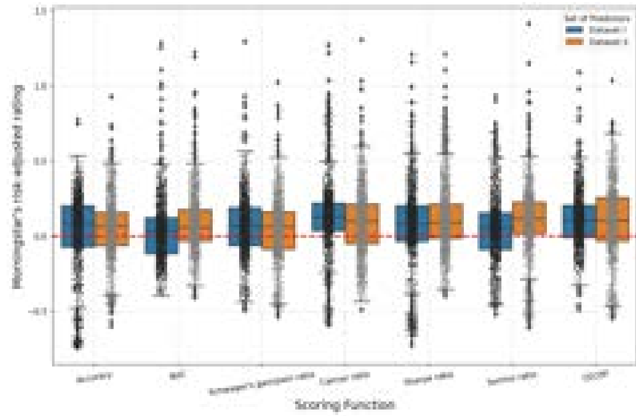
Holding period	Trans. cost	Scoring function used for cross validation													
		Accuracy		AUC		Schwager's gain/pain ratio		Calmar ratio		Sharpe ratio		Sortino ratio		CECPP	
		Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II
100	0.05	1.0328 (2.509)	0.7975 (1.8223)	0.5419 (2.1533)	0.7631 (2.0948)	0.821 (2.2033)	0.8646 (2.1063)	1.3565 (1.9179)	1.084 (2.2351)	0.8938 (2.2353)	1.0047 (2.1848)	0.7138 (2.1912)	1.2266 (1.9255)	1.1446 (1.8163)	1.07 (2.4234)
	0.1	0.9917 (2.5201)	0.7495 (1.8034)	0.4872 (2.1258)	0.6979 (2.0859)	0.7654 (2.1918)	0.7969 (2.0769)	1.3026 (1.8838)	1.0301 (2.2228)	0.8413 (2.2399)	0.9641 (2.1686)	0.6735 (2.1423)	1.1837 (1.8926)	1.1088 (1.788)	1.0271 (2.4162)
	0.5	0.5633 (2.5744)	0.345 (1.7313)	-0.1026 (2.0971)	0.2108 (2.0377)	0.3058 (1.9966)	0.3017 (2.0013)	0.8767 (1.7091)	0.5909 (2.2521)	0.5553 (2.191)	0.685 (2.0785)	0.3131 (1.8681)	0.8296 (1.8144)	0.7647 (1.6174)	0.7717 (2.4102)
	1	0.0133 (2.6088)	-0.2318 (1.7638)	-0.8631 (2.1584)	-0.3595 (1.9366)	-0.2035 (1.9066)	-0.1746 (2.0466)	0.3515 (1.5771)	0.091 (2.1229)	0.1748 (2.2917)	0.3261 (2.0432)	-0.1406 (1.6233)	0.3652 (1.7716)	0.2857 (1.6083)	0.3822 (2.3863)
	5	-3.0894 (3.7495)	-3.116 (2.8427)	-4.9142 (2.7479)	-4.7734 (2.3574)	-3.055 (2.6186)	-3.3291 (2.6863)	-3.2724 (2.4036)	-3.6847 (3.1134)	-2.0233 (2.9804)	-1.9464 (3.325)	-3.0514 (3.0838)	-2.421 (2.4307)	-1.8066 (2.7112)	-2.1631 (3.0171)
200	0.05	0.5056 (2.2499)	0.66 (1.2999)	0.5944 (1.3134)	0.8431 (1.3179)	0.692 (1.1008)	0.4394 (1.3273)	1.3298 (0.9547)	0.8019 (1.6405)	0.7257 (1.1767)	0.9745 (1.3626)	0.5807 (1.3633)	1.3174 (1.5826)	0.8486 (1.2995)	0.9904 (1.6277)
	0.1	0.4538 (2.2487)	0.6203 (1.278)	0.5162 (1.3286)	0.7835 (1.3034)	0.6498 (1.0812)	0.4051 (1.3129)	1.2803 (0.9529)	0.7548 (1.6063)	0.6901 (1.1916)	0.9397 (1.3375)	0.5519 (1.3375)	1.2875 (1.5699)	0.8077 (1.3)	0.9557 (1.6147)
	0.5	0.0951 (2.0958)	0.2421 (1.1392)	-0.1024 (1.4317)	0.2688 (1.2132)	0.3271 (0.9196)	0.104 (1.2459)	0.9323 (0.9421)	0.4894 (1.4703)	0.36 (1.3297)	0.6566 (1.2084)	0.2802 (1.0919)	0.9954 (1.2781)	0.5968 (1.2467)	0.6812 (1.5589)
	1	-0.3405 (1.7813)	-0.2303 (1.1516)	-0.7801 (1.6151)	-0.4229 (1.4025)	-0.1332 (0.8956)	-0.236 (1.2526)	0.4411 (0.8494)	-0.1307 (1.5703)	0.0344 (1.4892)	0.2666 (1.1762)	-0.0923 (0.9683)	0.4732 (1.0169)	0.3035 (0.9853)	0.2408 (1.4491)
	5	-3.2375 (3.0969)	-3.1896 (2.7293)	-5.2002 (2.1501)	-5.0484 (2.3129)	-3.2074 (2.2068)	-3.519 (2.1054)	-3.3203 (2.6938)	-3.5717 (2.5622)	-2.0269 (1.9699)	-2.0355 (3.2314)	-3.1005 (2.1393)	-2.3382 (1.6602)	-2.175 (2.7708)	-1.9933 (2.4338)

[a] *Dataset I* comprises of the financial variables and technical indicators while *Dataset II* contains both *Dataset I* and the candlestick chart (price) patterns as described in Table S.V.1.
[b] The initial endowment is $1000.
[c] The IQR values are shown in parentheses.
[d] There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

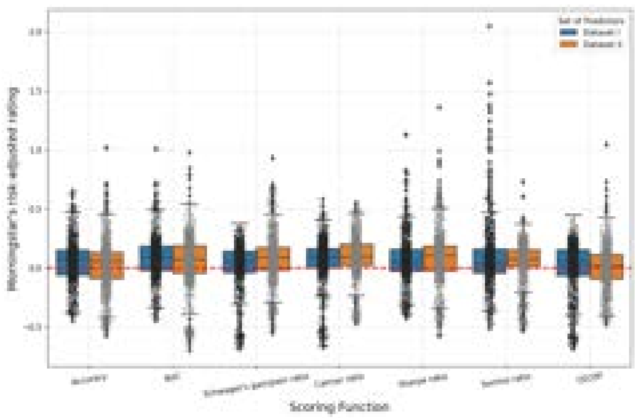
Figure S.VI.3: Morningstar's risk-adjusted ratings of 1718 trading sessions with the fixed transaction cost strategy investing in *SPY* for 100 days based on one-day ahead forecasts by the LGBM models cross-validated with seven different scoring functions



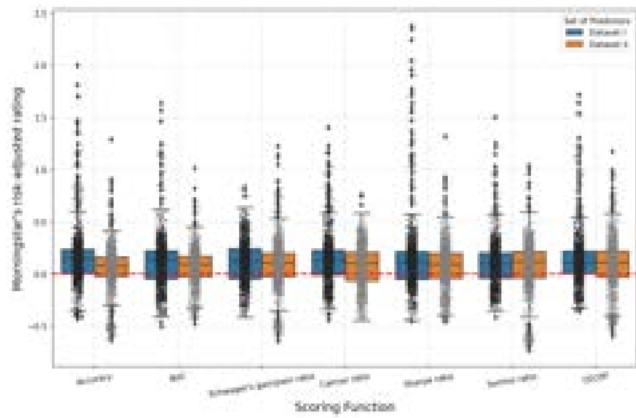
(a) *CE* loss



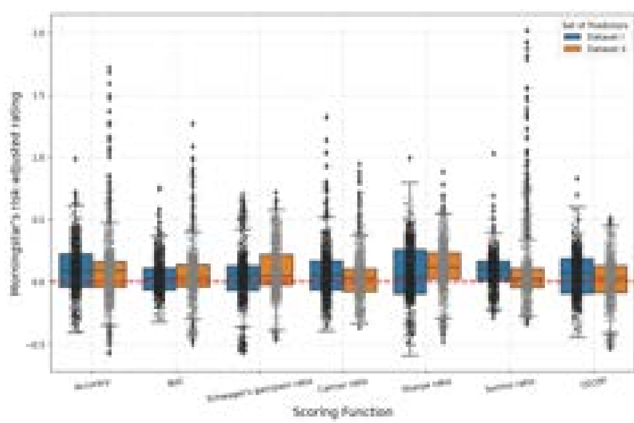
(b) *Brier* loss



(c) *Boost* loss



(d) *As1* loss

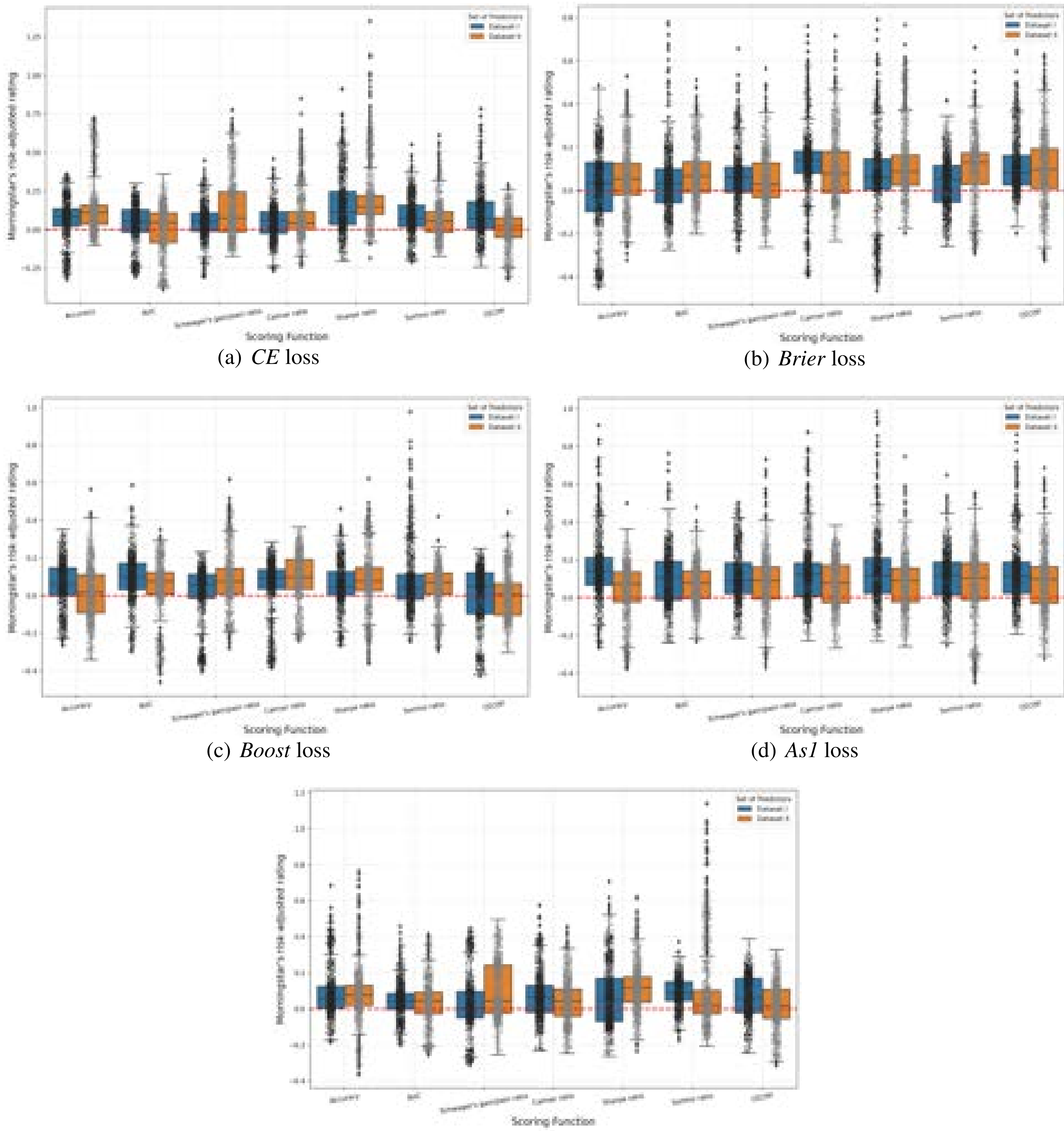


(e) *As2* loss

[i] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

[ii] Two sets of predictors are used: *Dataset I* (including the financial variables and technical indicators) and *Dataset II* (including *Dataset I* and candlestick chart patterns as described in Table S.V.1).

Figure S.VI.4: Morningstar's risk-adjusted ratings of 1618 trading sessions with the fixed transaction cost strategy investing in *SPY* for 200 days based on one-day ahead forecasts by the LGBM models cross-validated with seven different scoring functions



(a) *CE* loss

(b) *Brier* loss

(c) *Boost* loss

(d) *As1* loss

(e) *As2* loss

[i] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

[ii] Two sets of predictors are used: *Dataset I* (including the financial variables and technical indicators) and *Dataset II* (including *Dataset I* and candlestick chart patterns as described in Table S.V.1).

Table S.VI.2: The median (and IQR) of Morningstar's risk-adjusted ratings of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the LGBM models trained with the *Brier* loss function and cross-validated with seven different scoring functions (i.e., *Accuracy, AUC, Schwager's gain/pain ratio, Calmar ratio, Sharpe ratio, Sortino ratio,* or *CECPP*) for different amounts of transaction cost across various holding periods (in days)

Holding period	Trans. cost	Scoring function used for cross validation													
		Accuracy		AUC		Schwager's gain/pain ratio		Calmar ratio		Sharpe ratio		Sortino ratio		CECPP	
		Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II
100	0.05	0.0822 (0.2751)	0.0653 (0.2221)	0.035 (0.2395)	0.0587 (0.2078)	0.0649 (0.2523)	0.0624 (0.2471)	0.1278 (0.185)	0.0986 (0.261)	0.0805 (0.2354)	0.0925 (0.2265)	0.0565 (0.2446)	0.118 (0.2071)	0.1046 (0.2125)	0.1046 (0.2882)
	0.1	0.0768 (0.275)	0.0596 (0.222)	0.0286 (0.2383)	0.0526 (0.2063)	0.059 (0.2517)	0.0568 (0.2471)	0.1222 (0.1838)	0.0921 (0.2593)	0.0759 (0.2366)	0.0875 (0.2265)	0.0523 (0.2423)	0.1129 (0.2054)	0.0991 (0.212)	0.1004 (0.2858)
	0.5	0.0337 (0.2789)	0.0156 (0.2187)	-0.0231 (0.2301)	0.0032 (0.1988)	0.0134 (0.2512)	0.0132 (0.2455)	0.0757 (0.1763)	0.041 (0.251)	0.0401 (0.2415)	0.0507 (0.2292)	0.0129 (0.2243)	0.0705 (0.1927)	0.0581 (0.2174)	0.0665 (0.2753)
	1	-0.017 (0.2861)	-0.0374 (0.2186)	-0.085 (0.2197)	-0.0553 (0.1884)	-0.0399 (0.2498)	-0.0404 (0.2416)	0.0179 (0.1738)	-0.0199 (0.2402)	-0.0026 (0.247)	0.0103 (0.229)	-0.0344 (0.2078)	0.0187 (0.1859)	0.0091 (0.2174)	0.0218 (0.267)
	5	-0.3704 (0.2887)	-0.3994 (0.2182)	-0.4971 (0.1455)	-0.4597 (0.1507)	-0.3872 (0.2154)	-0.3832 (0.2129)	-0.3753 (0.1556)	-0.4268 (0.2063)	-0.2909 (0.2986)	-0.2795 (0.2957)	-0.3902 (0.2128)	-0.3314 (0.1822)	-0.2826 (0.2694)	-0.2979 (0.2752)
200	0.05	0.0345 (0.2284)	0.055 (0.1466)	0.0334 (0.1571)	0.0673 (0.1425)	0.061 (0.1199)	0.0299 (0.1598)	0.1391 (0.1026)	0.079 (0.1929)	0.0621 (0.1435)	0.088 (0.1396)	0.0494 (0.1685)	0.1311 (0.1461)	0.0832 (0.1366)	0.098 (0.1837)
	0.1	0.03 (0.2271)	0.05 (0.1465)	0.0274 (0.1565)	0.061 (0.142)	0.0563 (0.1201)	0.0248 (0.1601)	0.1332 (0.1016)	0.0726 (0.1928)	0.0565 (0.1444)	0.0833 (0.1397)	0.0444 (0.167)	0.1262 (0.1454)	0.079 (0.1369)	0.0936 (0.1834)
	0.5	-0.0088 (0.2172)	0.008 (0.146)	-0.0224 (0.1539)	0.0104 (0.1414)	0.0162 (0.1198)	-0.0138 (0.1616)	0.0844 (0.1014)	0.0242 (0.1917)	0.0204 (0.1528)	0.0471 (0.1398)	0.0071 (0.1563)	0.0826 (0.1407)	0.0463 (0.1346)	0.0544 (0.1797)
	1	-0.0585 (0.2129)	-0.0443 (0.1454)	-0.0814 (0.1495)	-0.0509 (0.1379)	-0.0367 (0.1259)	-0.0646 (0.1652)	0.0252 (0.1029)	-0.0372 (0.1895)	-0.0206 (0.1659)	0.0042 (0.1507)	-0.0411 (0.1467)	0.0286 (0.1334)	0.0021 (0.1291)	0.0056 (0.1838)
	5	-0.437 (0.2231)	-0.4242 (0.1462)	-0.539 (0.1162)	-0.5061 (0.1066)	-0.4548 (0.1876)	-0.4464 (0.1893)	-0.4202 (0.1224)	-0.474 (0.1627)	-0.3217 (0.2828)	-0.2941 (0.2555)	-0.4163 (0.1251)	-0.3407 (0.1658)	-0.3128 (0.1974)	-0.336 (0.2175)

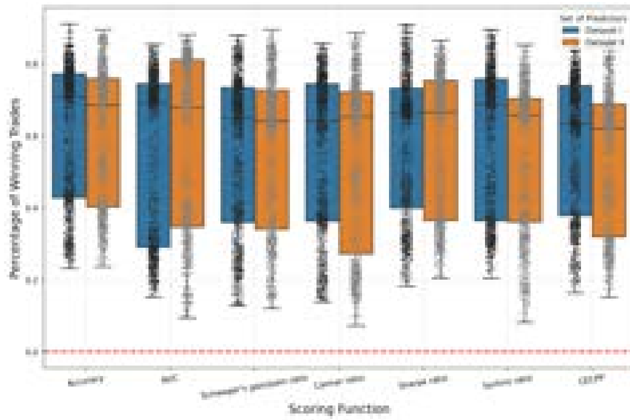
[a] *Dataset I* comprises of the financial variables and technical indicators while *Dataset II* contains both *Dataset I* and the candlestick chart (price) patterns as described in Table S.V.1.

[b] The initial endowment is $1000.

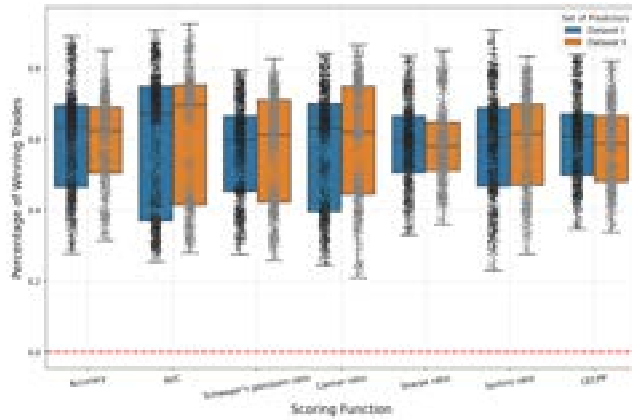
[c] The IQR values are shown in parentheses.

[d] There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

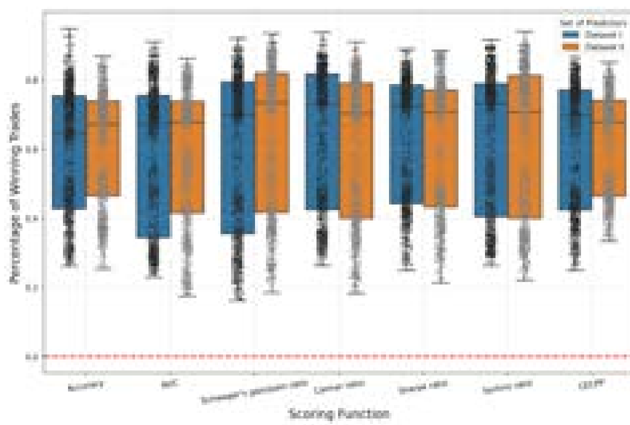
Figure S.VI.5: The percentages of winning trades of 1718 trading sessions with the fixed transaction cost strategy investing in *SPY* for 100 days based on one-day ahead forecasts by the LGBM models cross-validated with seven different scoring functions



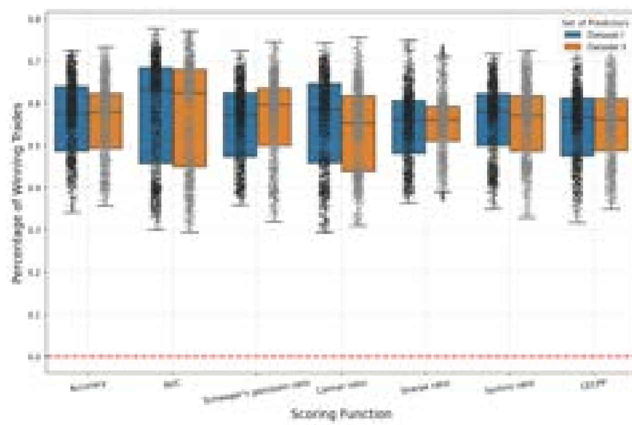
(a) *CE* loss



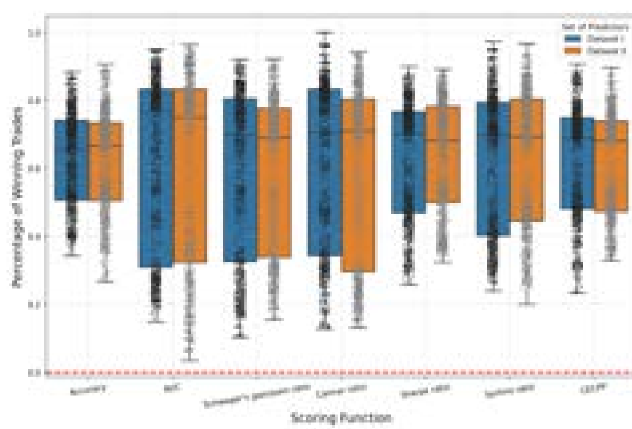
(b) *Brier* loss



(c) *Boost* loss



(d) *As1* loss

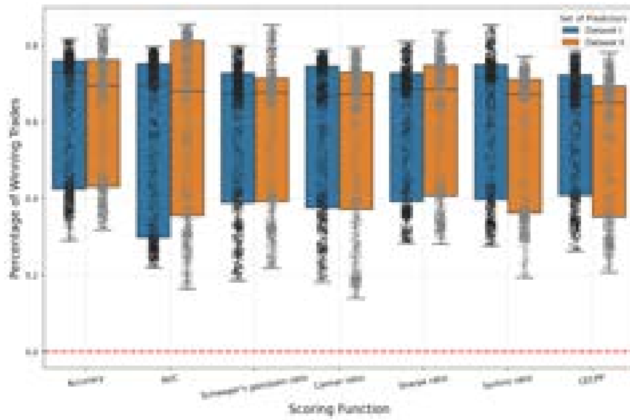


(e) *As2* loss

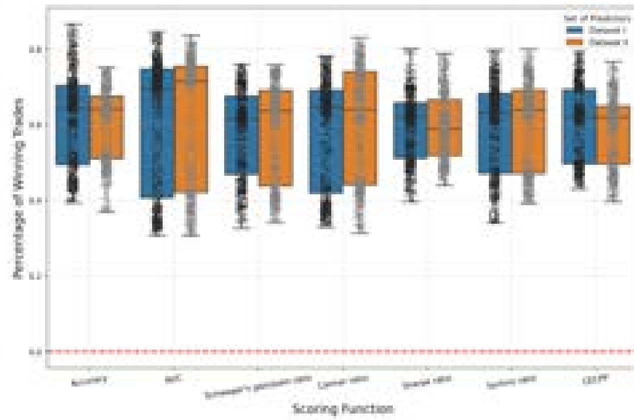
[i] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

[ii] Two sets of predictors are used: *Dataset I* (including the financial variables and technical indicators) and *Dataset II* (including *Dataset I* and candlestick chart patterns as described in Table S.V.1).

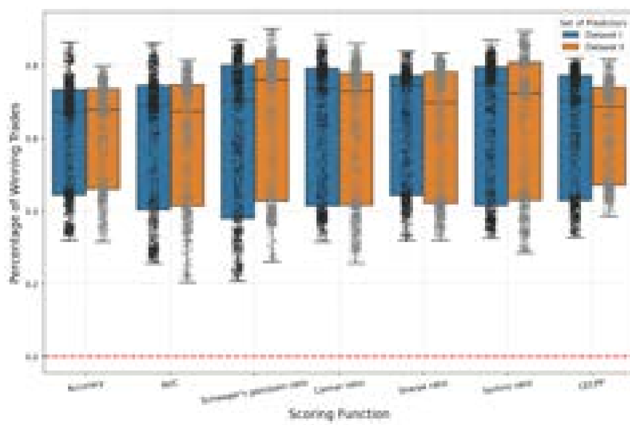
Figure S.VI.6: The percentages of winning trades of 1618 trading sessions with the fixed transaction cost strategy investing in *SPY* for 200 days based on one-day ahead forecasts by the LGBM models cross-validated with seven different scoring functions



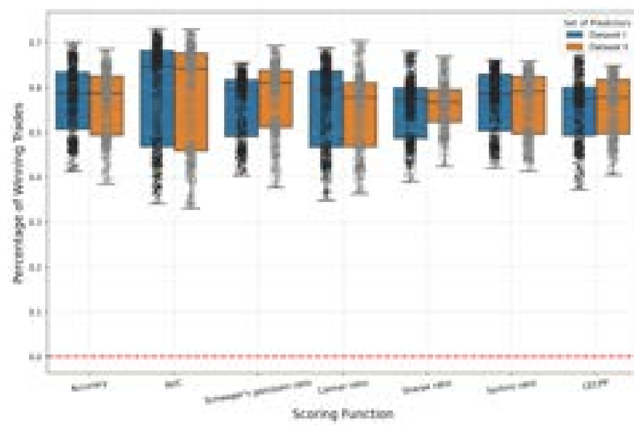
(a) *CE* loss



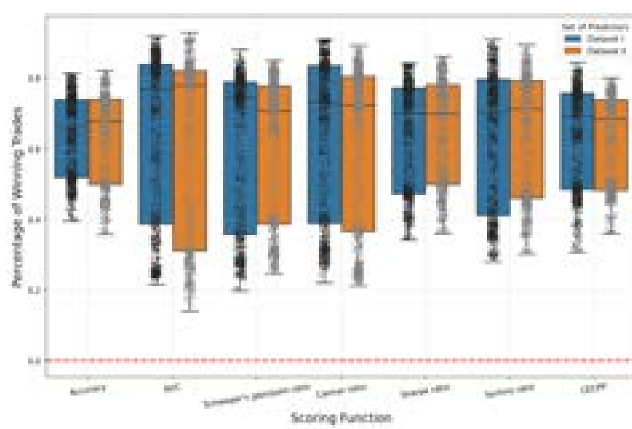
(b) *Brier* loss



(c) *Boost* loss



(d) *As1* loss



(e) *As2* loss

i The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

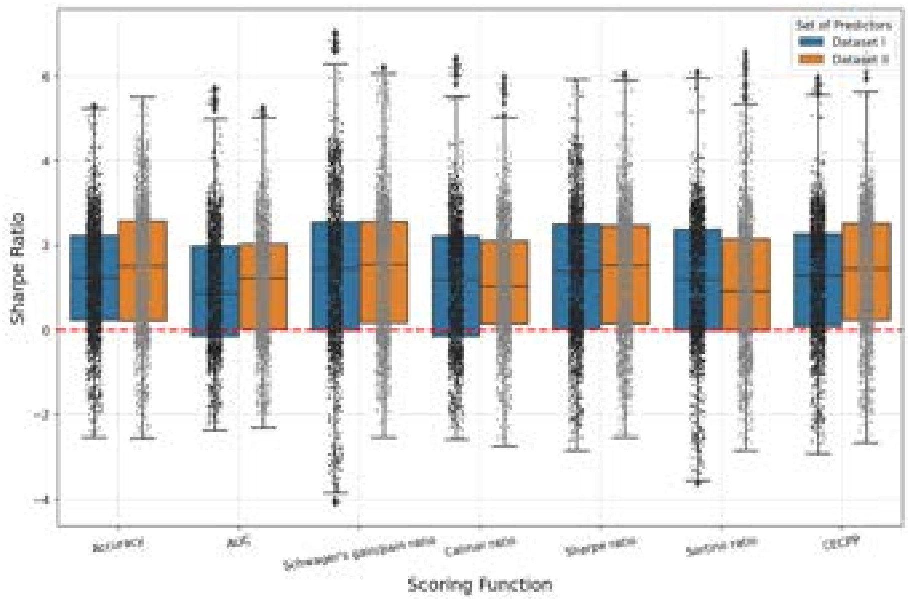
ii Two sets of predictors are used: *Dataset I* (including the financial variables and technical indicators) and *Dataset II* (including *Dataset I* and candlestick chart patterns as described in Table S.V.1).

Table S.VI.3: The median (and IQR) of the percentages of winning trades of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the LGBM models trained with the *Brier* loss function and cross-validated with seven different scoring functions (i.e., *Accuracy, AUC, Schwager's gain/pain ratio, Calmar ratio, Sharpe ratio, Sortino ratio,* or *CECPP*) for different amounts of transaction cost across various holding periods (in days)

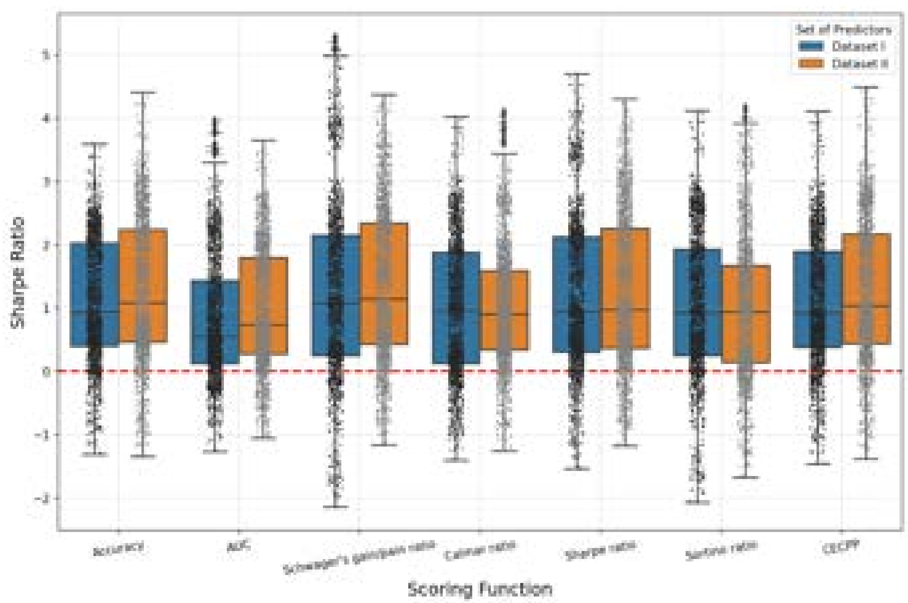
Holding period	Trans. cost	Scoring function used for cross validation													
		Accuracy		AUC		Schwager's gain/pain ratio		Calmar ratio		Sharpe ratio		Sortino ratio		CECPP	
		Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II
100	0.05	0.6351 (0.2301)	0.625 (0.1844)	0.6727 (0.379)	0.6957 (0.339)	0.5989 (0.2133)	0.6133 (0.2885)	0.631 (0.3034)	0.6197 (0.3056)	0.5954 (0.1606)	0.5814 (0.1365)	0.6048 (0.2218)	0.6171 (0.2279)	0.6092 (0.1719)	0.5889 (0.1853)
	0.1	0.6296 (0.2268)	0.6216 (0.1834)	0.6723 (0.3741)	0.6949 (0.339)	0.5977 (0.2133)	0.6133 (0.2838)	0.6305 (0.3019)	0.6184 (0.3013)	0.5952 (0.1563)	0.5814 (0.1349)	0.5946 (0.22)	0.6171 (0.2275)	0.6053 (0.1716)	0.5879 (0.1853)
	0.5	0.6232 (0.2222)	0.6027 (0.1774)	0.6552 (0.3636)	0.6552 (0.3323)	0.5811 (0.207)	0.6026 (0.2732)	0.622 (0.2824)	0.6051 (0.2737)	0.5854 (0.1447)	0.5698 (0.1238)	0.5909 (0.211)	0.6081 (0.2179)	0.5955 (0.172)	0.5802 (0.1742)
	1	0.6098 (0.2181)	0.5929 (0.1604)	0.65 (0.3651)	0.6346 (0.3324)	0.5772 (0.1935)	0.589 (0.2769)	0.6143 (0.2853)	0.5951 (0.2586)	0.5765 (0.1332)	0.5647 (0.1189)	0.5833 (0.2139)	0.6047 (0.2199)	0.5871 (0.1676)	0.5691 (0.1675)
	5	0.5493 (0.1741)	0.5294 (0.161)	0.5676 (0.3413)	0.5686 (0.3126)	0.5238 (0.1891)	0.5338 (0.2522)	0.5556 (0.3064)	0.53 (0.2611)	0.5287 (0.1253)	0.5301 (0.0983)	0.525 (0.1806)	0.5437 (0.185)	0.5385 (0.134)	0.522 (0.1422)
200	0.05	0.6429 (0.2105)	0.642 (0.1643)	0.6972 (0.3405)	0.7155 (0.3316)	0.6154 (0.2056)	0.637 (0.2528)	0.6471 (0.2729)	0.6412 (0.2984)	0.6205 (0.1467)	0.589 (0.1478)	0.6329 (0.2116)	0.6401 (0.2203)	0.6207 (0.1964)	0.6209 (0.1531)
	0.1	0.6403 (0.2081)	0.6376 (0.1643)	0.6944 (0.3365)	0.7155 (0.3316)	0.6149 (0.2056)	0.6364 (0.25)	0.6468 (0.2695)	0.6392 (0.2984)	0.619 (0.1447)	0.5882 (0.1452)	0.625 (0.2097)	0.64 (0.2203)	0.6181 (0.1957)	0.6199 (0.1524)
	0.5	0.631 (0.2045)	0.6218 (0.1581)	0.6697 (0.3174)	0.6885 (0.3182)	0.6012 (0.1908)	0.625 (0.2451)	0.6289 (0.2681)	0.6214 (0.2742)	0.6118 (0.131)	0.5774 (0.1449)	0.6194 (0.2)	0.635 (0.2166)	0.6087 (0.1784)	0.6127 (0.1538)
	1	0.6115 (0.2047)	0.6111 (0.1443)	0.6636 (0.3153)	0.6572 (0.3102)	0.5963 (0.1754)	0.6115 (0.2408)	0.625 (0.2558)	0.6091 (0.2723)	0.5979 (0.1213)	0.5767 (0.1204)	0.6148 (0.2014)	0.628 (0.2072)	0.6011 (0.1735)	0.6011 (0.1373)
	5	0.5548 (0.1582)	0.5478 (0.1446)	0.5818 (0.2876)	0.5809 (0.2709)	0.5463 (0.1734)	0.5556 (0.2251)	0.5563 (0.2891)	0.5427 (0.2606)	0.5509 (0.1113)	0.5427 (0.0939)	0.5548 (0.1757)	0.5617 (0.1955)	0.5549 (0.1241)	0.548 (0.127)

[a] *Dataset I* comprises of the financial variables and technical indicators while *Dataset II* contains both *Dataset I* and the candlestick chart (price) patterns as described in Table S.V.1.
[b] The initial endowment is $1000.
[c] The IQR values are shown in parentheses.
[d] There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

Figure S.VI.7: The Sharpe ratios of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the RF models cross-validated with seven different scoring functions



(a) Holding period (D): 100 days



(b) Holding period (D): 200 days

i The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

ii Two sets of predictors are used: *Dataset I* (including the financial variables and technical indicators) and *Dataset II* (including *Dataset I* and candlestick chart patterns as described in Table S.V.1).

iii There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

Table S.VI.4: The median (and IQR) of the Sharpe ratios of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the RF models cross-validated with seven different scoring functions (i.e., *Accuracy, AUC, Schwager's gain/pain ratio, Calmar ratio, Sharpe ratio, Sortino ratio,* or *CECPP*) for different amounts of transaction cost across various holding periods (in days)

		Scoring function used for cross validation													
Holding period	Trans. cost	Accuracy		AUC		Schwager's gain/pain ratio		Calmar ratio		Sharpe ratio		Sortino ratio		CECPP	
		Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II
100	0.05	1.2579 (1.999)	1.5178 (2.3477)	0.8696 (2.1242)	1.2256 (2.0039)	1.4526 (2.5603)	1.5648 (2.3861)	1.1892 (2.3567)	1.0421 (1.9504)	1.3977 (2.4593)	1.535 (2.3288)	1.1913 (2.3908)	0.9312 (2.1703)	1.2931 (2.1883)	1.4575 (2.2785)
	0.1	1.234 (1.9789)	1.4945 (2.3289)	0.8412 (2.1068)	1.2062 (1.987)	1.4348 (2.5655)	1.5568 (2.3688)	1.1856 (2.34)	1.0244 (1.9435)	1.3694 (2.4405)	1.51 (2.3199)	1.1874 (2.3926)	0.9207 (2.1883)	1.278 (2.1704)	1.4355 (2.2572)
	0.5	1.0156 (1.8582)	1.2584 (2.2371)	0.6438 (1.9297)	0.9281 (1.7861)	1.3038 (2.5563)	1.4031 (2.2688)	1.0818 (2.2656)	0.9141 (1.8403)	1.2448 (2.2974)	1.3724 (2.2058)	1.1096 (2.3162)	0.8268 (2.1533)	1.1069 (2.0965)	1.2798 (2.0926)
	1	0.7094 (1.8139)	0.9922 (2.1723)	0.3379 (1.7859)	0.5233 (1.6292)	1.154 (2.4231)	1.2004 (2.124)	0.9492 (2.222)	0.7591 (1.7169)	1.039 (2.1669)	1.1949 (2.0998)	0.9629 (2.1724)	0.7144 (2.0866)	0.8725 (1.9721)	1.0594 (1.9502)
	5	-0.4552 (2.4311)	-0.3422 (2.4236)	-1.5438 (1.9034)	-1.3596 (1.73)	-0.2843 (2.4137)	-0.1041 (2.5419)	-0.3732 (2.3416)	-0.3335 (2.3182)	-0.2484 (2.4719)	-0.2585 (2.5271)	-0.2144 (2.4345)	-0.5779 (2.237)	-0.1559 (2.7718)	-0.0875 (2.5042)
200	0.05	0.9431 (1.6251)	1.0814 (1.7677)	0.5522 (1.3064)	0.7332 (1.5225)	1.0714 (1.903)	1.1447 (1.9185)	0.9675 (1.752)	0.9046 (1.2561)	0.9458 (1.8189)	0.9759 (1.8913)	0.9279 (1.6777)	0.9417 (1.5213)	0.926 (1.508)	1.0371 (1.7375)
	0.1	0.931 (1.6041)	1.0733 (1.7661)	0.5329 (1.295)	0.7153 (1.5083)	1.0611 (1.8941)	1.1312 (1.902)	0.9542 (1.7371)	0.8954 (1.2506)	0.9365 (1.7926)	0.9696 (1.8832)	0.9219 (1.6625)	0.9302 (1.5147)	0.9187 (1.4935)	1.0199 (1.7066)
	0.5	0.8284 (1.3469)	0.9636 (1.707)	0.3489 (1.1739)	0.5631 (1.2887)	0.9701 (1.7591)	1.0907 (1.7462)	0.8332 (1.5899)	0.798 (1.1868)	0.8859 (1.6337)	0.8951 (1.7668)	0.8315 (1.5784)	0.8141 (1.4479)	0.8288 (1.3409)	0.9465 (1.5594)
	1	0.5868 (1.073)	0.7569 (1.4483)	0.1643 (0.8914)	0.3433 (1.0036)	0.8661 (1.5596)	0.9829 (1.5931)	0.6182 (1.5231)	0.617 (1.1105)	0.7981 (1.5078)	0.814 (1.5847)	0.62 (1.4747)	0.6094 (1.4032)	0.6937 (1.1951)	0.8329 (1.3347)
	5	-0.2844 (1.6006)	-0.3032 (1.6236)	-1.6829 (1.6398)	-1.2842 (1.398)	-0.0391 (1.5868)	0.0804 (1.6093)	-0.1337 (1.8162)	-0.1457 (1.6929)	-0.0067 (1.872)	-0.0022 (1.6402)	0.0652 (1.9633)	-0.2754 (1.6414)	0.0013 (1.7615)	0.0669 (1.6571)

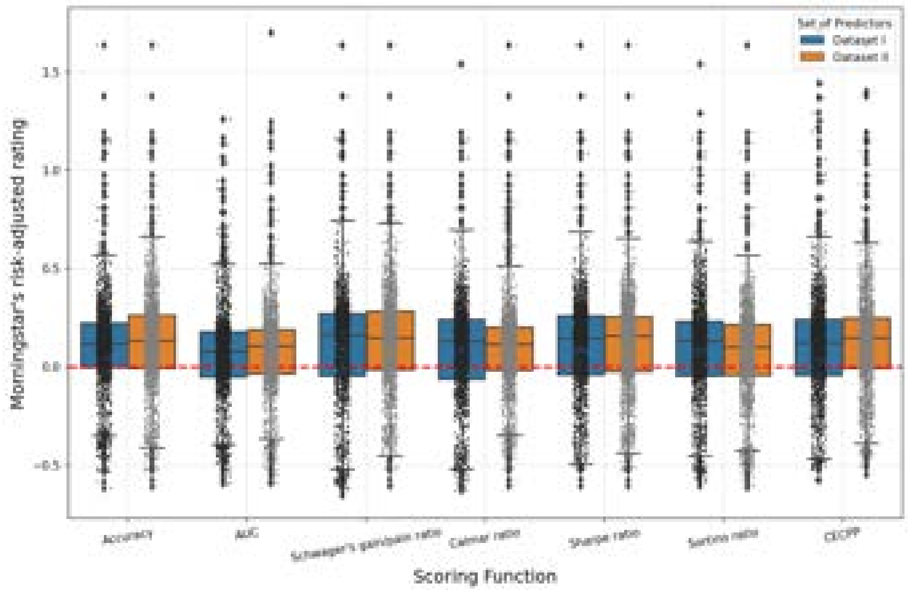
[a] *Dataset I* comprises of the financial variables and technical indicators while *Dataset II* contains both *Dataset I* and the candlestick chart (price) patterns as described in Table S.V.1.

[b] The initial endowment is $1000.

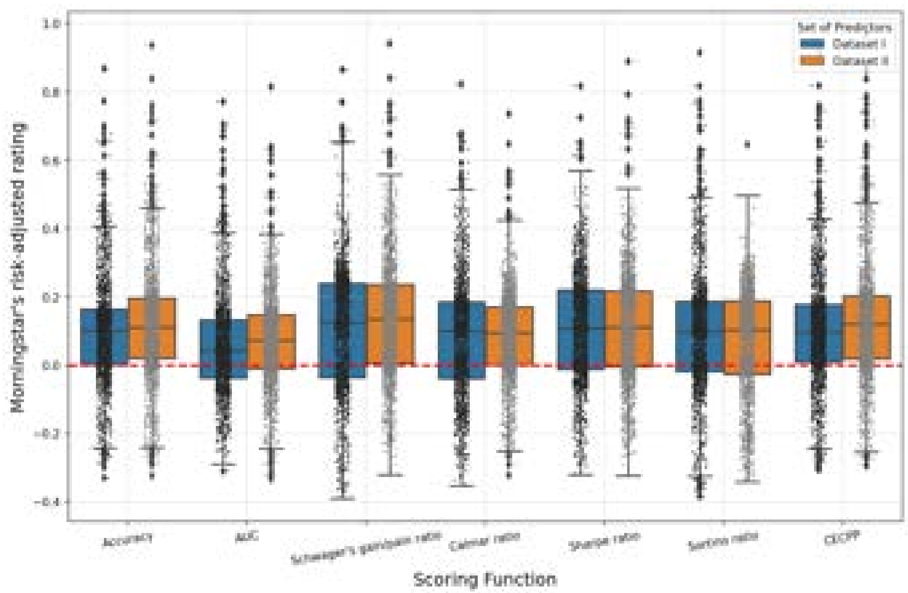
[c] The IQR values are shown in parentheses.

[d] There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

Figure S.VI.8: Morningstar's risk-adjusted ratings of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the RF models cross-validated with seven different scoring functions



(a) Holding period (D): 100 days



(b) Holding period (D): 200 days

i The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

ii Two sets of predictors are used: *Dataset I* (including the financial variables and technical indicators) and *Dataset II* (including *Dataset I* and candlestick chart patterns as described in Table S.V.1).

iii There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

Table S.VI.5: The median (and IQR) of Morningstar's risk-adjusted ratings of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the RF models cross-validated with seven different scoring functions (i.e., *Accuracy, AUC, Schwager's gain/pain ratio, Calmar ratio, Sharpe ratio, Sortino ratio,* or *CECPP*) for different amounts of transaction cost across various holding periods (in days)

Holding period	Trans. cost	Scoring function used for cross validation													
		Accuracy		AUC		Schwager's gain/pain ratio		Calmar ratio		Sharpe ratio		Sortino ratio		CECPP	
		Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II
100	0.05	0.1175 (0.2298)	0.132 (0.271)	0.0765 (0.2299)	0.106 (0.2237)	0.1619 (0.3171)	0.1486 (0.2987)	0.1285 (0.3042)	0.1138 (0.2185)	0.1462 (0.3033)	0.1565 (0.2787)	0.13 (0.276)	0.0992 (0.2569)	0.1231 (0.2838)	0.147 (0.2552)
	0.1	0.1142 (0.2298)	0.1298 (0.2696)	0.0733 (0.2281)	0.102 (0.2217)	0.1593 (0.3155)	0.1456 (0.2985)	0.126 (0.3015)	0.1119 (0.2169)	0.145 (0.3004)	0.1546 (0.2774)	0.1273 (0.2744)	0.0972 (0.2564)	0.1213 (0.2836)	0.1456 (0.2549)
	0.5	0.0894 (0.2234)	0.1092 (0.2624)	0.0503 (0.2144)	0.0727 (0.2091)	0.1443 (0.3032)	0.1278 (0.2895)	0.1099 (0.2921)	0.0939 (0.2111)	0.1229 (0.283)	0.1334 (0.2726)	0.1118 (0.2711)	0.0871 (0.2587)	0.0986 (0.2795)	0.1242 (0.2485)
	1	0.0533 (0.2219)	0.0855 (0.2524)	0.0154 (0.2036)	0.0339 (0.2032)	0.1243 (0.2874)	0.107 (0.2748)	0.0883 (0.2886)	0.0679 (0.205)	0.0984 (0.27)	0.1054 (0.2675)	0.0881 (0.272)	0.069 (0.2621)	0.0708 (0.2828)	0.0976 (0.2505)
	5	-0.105 (0.3271)	-0.0997 (0.2988)	-0.2132 (0.2249)	-0.2081 (0.1985)	-0.0684 (0.3336)	-0.0607 (0.3582)	-0.083 (0.3096)	-0.1004 (0.2721)	-0.0906 (0.3188)	-0.0838 (0.3443)	-0.0722 (0.3485)	-0.1161 (0.3)	-0.0777 (0.3736)	-0.0667 (0.3579)
200	0.05	0.1009 (0.1628)	0.1106 (0.1757)	0.0435 (0.1701)	0.073 (0.1579)	0.1245 (0.2761)	0.1345 (0.2305)	0.1004 (0.2235)	0.0955 (0.1691)	0.1079 (0.2332)	0.1112 (0.2199)	0.0983 (0.2051)	0.103 (0.2142)	0.0972 (0.1682)	0.1213 (0.1828)
	0.1	0.0981 (0.1611)	0.1086 (0.1749)	0.0405 (0.1688)	0.0705 (0.1557)	0.1228 (0.2735)	0.1334 (0.2274)	0.0977 (0.224)	0.0933 (0.1681)	0.1065 (0.2304)	0.1101 (0.216)	0.0968 (0.2049)	0.1013 (0.2127)	0.0945 (0.1658)	0.12 (0.1801)
	0.5	0.0753 (0.1514)	0.091 (0.1625)	0.0176 (0.1543)	0.0471 (0.1399)	0.1085 (0.2546)	0.1192 (0.2103)	0.0784 (0.2197)	0.0776 (0.1538)	0.0961 (0.2069)	0.0988 (0.1956)	0.0783 (0.1972)	0.0796 (0.2058)	0.0802 (0.1541)	0.1041 (0.1616)
	1	0.0407 (0.1432)	0.0661 (0.1556)	-0.0116 (0.1345)	0.0164 (0.1261)	0.0898 (0.234)	0.0976 (0.1881)	0.0543 (0.2129)	0.0518 (0.1352)	0.0788 (0.188)	0.085 (0.1803)	0.0534 (0.1892)	0.0554 (0.2009)	0.056 (0.1418)	0.0838 (0.1451)
	5	-0.0972 (0.2004)	-0.0872 (0.2078)	-0.2494 (0.1556)	-0.2273 (0.1633)	-0.0792 (0.2316)	-0.0456 (0.2228)	-0.0761 (0.2549)	-0.0833 (0.1912)	-0.0676 (0.2386)	-0.063 (0.2176)	-0.048 (0.2513)	-0.1006 (0.2373)	-0.0534 (0.2662)	-0.0361 (0.2185)

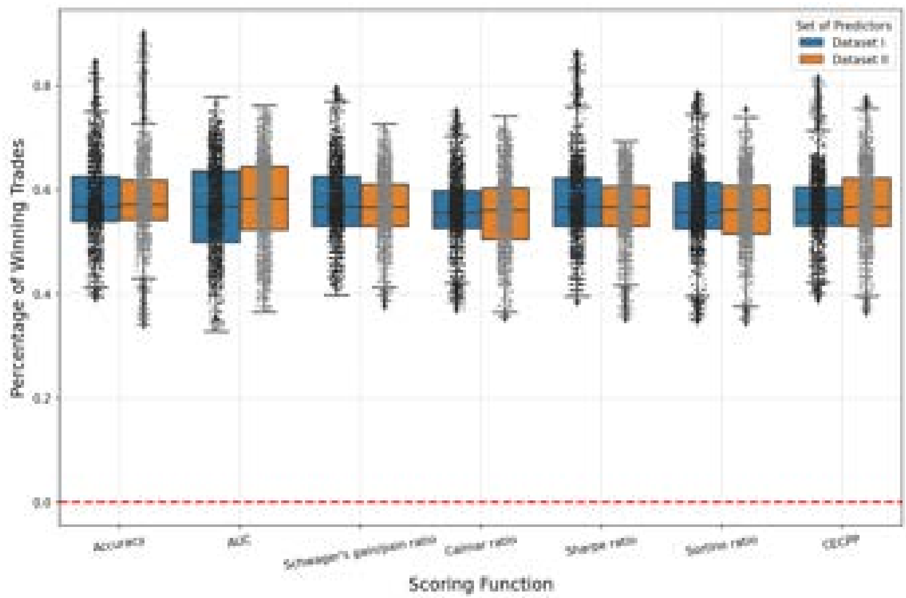
[a] *Dataset I* comprises of the financial variables and technical indicators while *Dataset II* contains both *Dataset I* and the candlestick chart (price) patterns as described in Table S.V.1.

[b] The initial endowment is $1000.

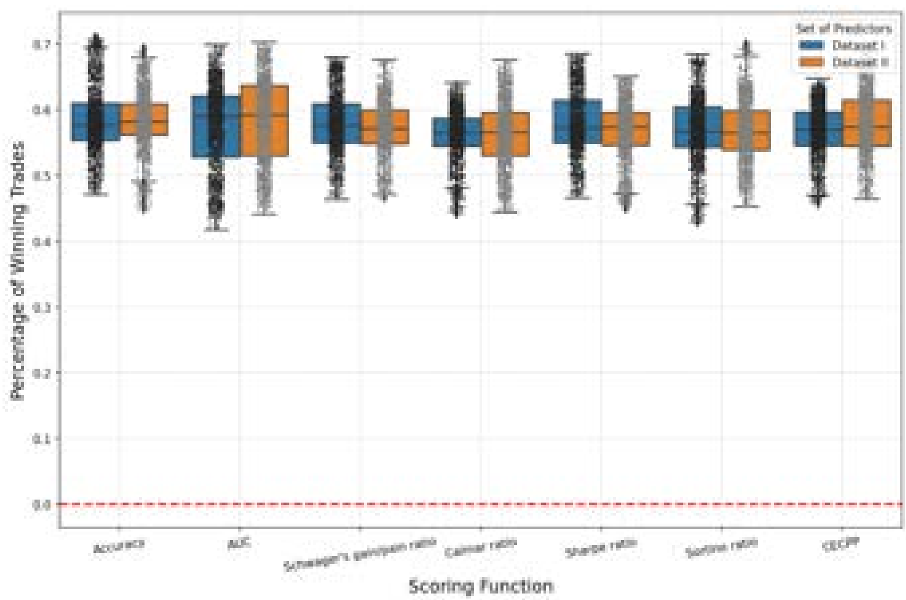
[c] The IQR values are shown in parentheses.

[d] There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

Figure S.VI.9: The percentages of winning trades of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the RF models cross-validated with seven different scoring functions



(a) Holding period (D): 100 days



(b) Holding period (D): 200 days

[i] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

[ii] Two sets of predictors are used: *Dataset I* (including the financial variables and technical indicators) and *Dataset II* (including *Dataset I* and candlestick chart patterns as described in Table S.V.1).

[iii] There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

Table S.VI.6: The median (and IQR) of the percentages of winning trades of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the RF models cross-validated with seven different scoring functions (i.e., *Accuracy, AUC, Schwager's gain/pain ratio, Calmar ratio, Sharpe ratio, Sortino ratio,* or *CECPP*) for different amounts of transaction cost across various holding periods (in days)

Holding period	Trans. cost	Scoring function used for cross validation													
		Accuracy		AUC		Schwager's gain/pain ratio		Calmar ratio		Sharpe ratio		Sortino ratio		CECPP	
		Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II
100	0.05	0.5682 (0.0873)	0.5714 (0.0758)	0.5682 (0.1364)	0.5828 (0.1224)	0.567 (0.096)	0.567 (0.0805)	0.5567 (0.0711)	0.5612 (0.0979)	0.567 (0.0918)	0.567 (0.0758)	0.5579 (0.0887)	0.5618 (0.0928)	0.5625 (0.0744)	0.567 (0.0918)
	0.1	0.5682 (0.0879)	0.5714 (0.0758)	0.5667 (0.1341)	0.5814 (0.1217)	0.567 (0.096)	0.5668 (0.0805)	0.5567 (0.0711)	0.5612 (0.0977)	0.567 (0.0918)	0.567 (0.0758)	0.5575 (0.0887)	0.5618 (0.0928)	0.5625 (0.0744)	0.567 (0.0918)
	0.5	0.5667 (0.0864)	0.5686 (0.0777)	0.5606 (0.1342)	0.5763 (0.1128)	0.5667 (0.0833)	0.5625 (0.0776)	0.5532 (0.0793)	0.5575 (0.0948)	0.5612 (0.087)	0.5625 (0.0763)	0.5567 (0.0866)	0.5567 (0.0866)	0.5612 (0.0709)	0.5652 (0.0944)
	1	0.5635 (0.0785)	0.567 (0.0666)	0.5556 (0.1263)	0.573 (0.1049)	0.5612 (0.0763)	0.5612 (0.0803)	0.551 (0.0807)	0.5567 (0.0928)	0.5612 (0.0805)	0.5591 (0.0716)	0.5474 (0.0872)	0.5521 (0.0845)	0.5567 (0.0663)	0.5612 (0.0921)
	5	0.5567 (0.0878)	0.551 (0.0635)	0.5385 (0.1268)	0.5455 (0.1049)	0.5464 (0.0792)	0.551 (0.0672)	0.5408 (0.084)	0.5423 (0.0997)	0.551 (0.0713)	0.5417 (0.064)	0.5376 (0.0722)	0.5408 (0.0774)	0.5417 (0.0622)	0.5464 (0.0669)
200	0.05	0.5775 (0.0566)	0.5824 (0.0474)	0.5904 (0.0933)	0.5922 (0.1065)	0.5765 (0.0593)	0.5722 (0.0518)	0.5657 (0.0422)	0.5663 (0.0656)	0.5729 (0.064)	0.5736 (0.0494)	0.5668 (0.0611)	0.5657 (0.0593)	0.5707 (0.051)	0.5736 (0.0689)
	0.1	0.5775 (0.0562)	0.5819 (0.0474)	0.5899 (0.0931)	0.5909 (0.103)	0.5765 (0.0593)	0.5718 (0.0518)	0.5657 (0.0422)	0.5657 (0.0646)	0.5707 (0.0646)	0.5736 (0.0494)	0.5668 (0.0611)	0.5657 (0.0593)	0.5707 (0.0508)	0.5729 (0.0689)
	0.5	0.5753 (0.0573)	0.5789 (0.0484)	0.5812 (0.0901)	0.5856 (0.0964)	0.5736 (0.0561)	0.5707 (0.0486)	0.5625 (0.0412)	0.5641 (0.066)	0.5707 (0.0635)	0.5685 (0.0463)	0.5635 (0.0516)	0.5635 (0.0563)	0.5685 (0.0448)	0.5696 (0.0652)
	1	0.5726 (0.0543)	0.5736 (0.044)	0.5741 (0.0855)	0.578 (0.0963)	0.5707 (0.0538)	0.5684 (0.0465)	0.5606 (0.0409)	0.5635 (0.063)	0.5707 (0.0579)	0.5685 (0.0444)	0.5635 (0.052)	0.5615 (0.0528)	0.5657 (0.042)	0.5657 (0.0588)
	5	0.5645 (0.0606)	0.5604 (0.045)	0.5529 (0.0948)	0.5581 (0.085)	0.5606 (0.0427)	0.5625 (0.037)	0.5464 (0.0512)	0.5526 (0.0674)	0.5645 (0.0442)	0.5533 (0.0381)	0.5533 (0.0478)	0.5515 (0.0425)	0.5561 (0.0392)	0.5563 (0.0428)

[a] *Dataset I* comprises of the financial variables and technical indicators while *Dataset II* contains both *Dataset I* and the candlestick chart (price) patterns as described in Table S.V.1.

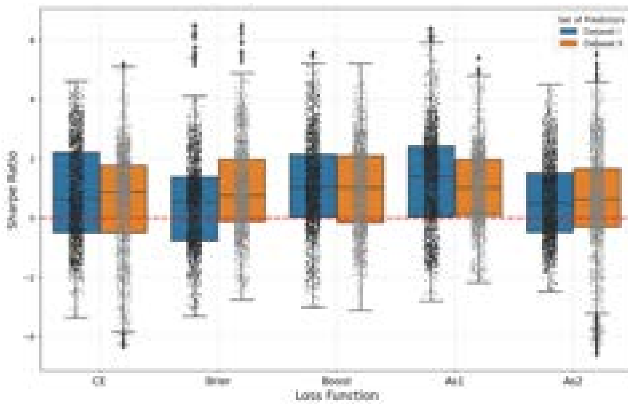
[b] The initial endowment is $1000.

[c] The IQR values are shown in parentheses.

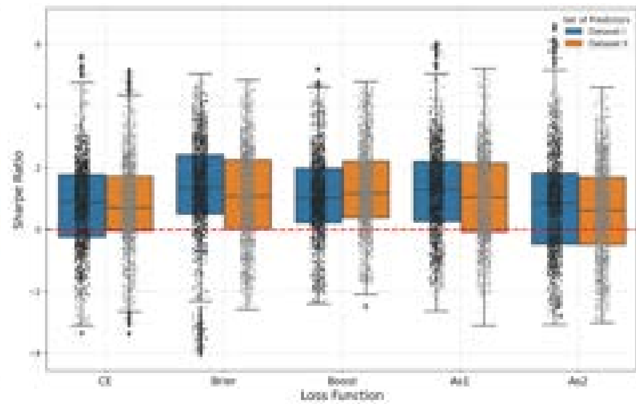
[d] There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

S.VII Performance of the Trading Strategy across Loss Functions

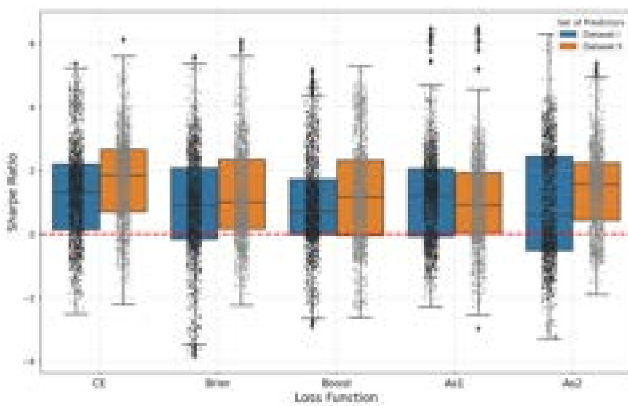
Figure S.VII.1: The Sharpe ratios of 1718 trading sessions with the fixed transaction cost strategy investing in *SPY* for 100 days based on one-day ahead forecasts by the LGBM models trained with five different loss functions



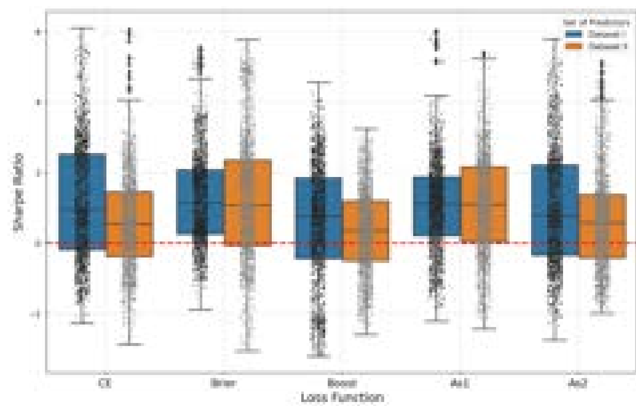
(a) *AUC* score



(b) *Calmar ratio* score



(c) *Sharpe ratio* score

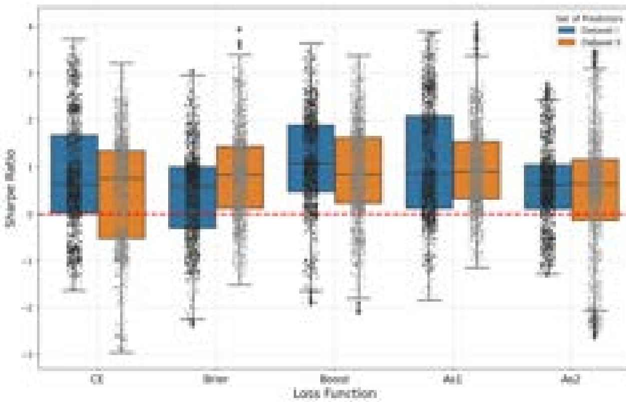


(d) *CECPP* score

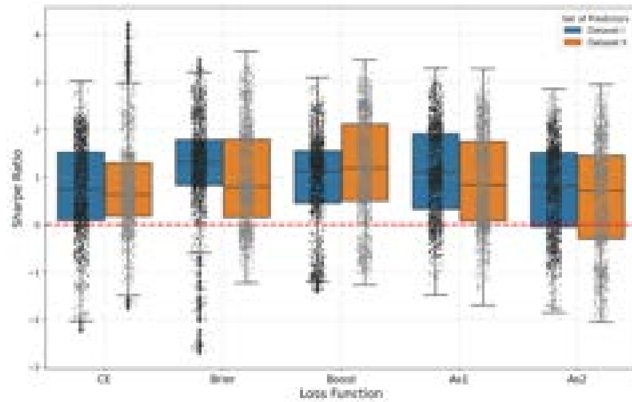
[i] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

[ii] Two sets of predictors are used: *Dataset I* (including the financial variables and technical indicators) and *Dataset II* (including *Dataset I* and candlestick chart patterns as described in Table S.V.1).

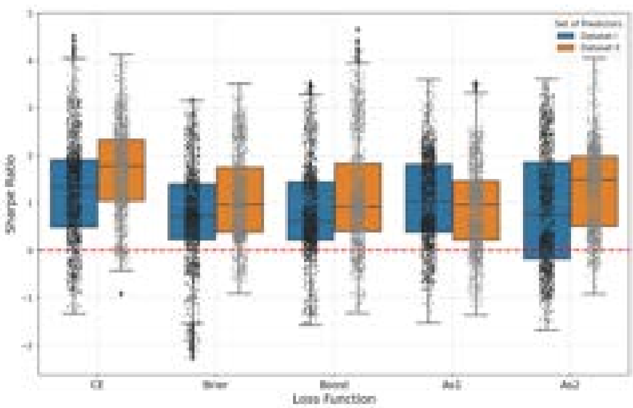
Figure S.VII.2: The Sharpe ratios of 1618 trading sessions with the fixed transaction cost strategy investing in *SPY* for 200 days based on one-day ahead forecasts by the LGBM models trained with five different loss functions



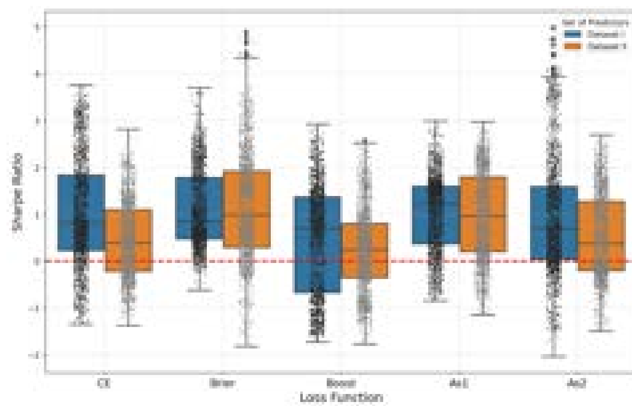
(a) *AUC* score



(b) *Calmar ratio* score



(c) *Sharpe ratio* score



(d) *CECPP* score

[i] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

[ii] Two sets of predictors are used: *Dataset I* (including the financial variables and technical indicators) and *Dataset II* (including *Dataset I* and candlestick chart patterns as described in Table S.V.1).

Table S.VII.1: The median (and IQR) of the Sharpe ratios of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the LGBM models trained with five loss functions (i.e., *CE, Brier, Boost, As1,* or *As2*), then cross-validated with the *CECPP* scoring function for different amounts of transaction cost across various holding periods (in days)

Holding period	Trans. cost	CE		Brier		Boost		As1		As2	
		Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II
100	0.05	0.9404 (2.7137)	0.5434 (1.8351)	1.1446 (1.8163)	1.07 (2.4234)	0.7677 (2.2863)	0.3576 (1.6983)	1.1239 (1.656)	1.0912 (2.1139)	0.7694 (2.5702)	0.5754 (1.7988)
	0.1	0.8981 (2.6823)	0.4755 (1.8215)	1.1088 (1.788)	1.0271 (2.4162)	0.7022 (2.2844)	0.3007 (1.6871)	1.1038 (1.6474)	1.0682 (2.1029)	0.7238 (2.5616)	0.4984 (1.7936)
	0.5	0.5129 (2.4932)	-0.0599 (1.7319)	0.7647 (1.6174)	0.7717 (2.4102)	0.1326 (2.2344)	-0.2125 (1.7685)	0.9118 (1.5925)	0.8927 (2.0276)	0.3655 (2.3795)	-0.0284 (1.798)
	1	-0.0569 (2.3101)	-0.7053 (1.62)	0.2857 (1.6083)	0.3822 (2.3863)	-0.5338 (2.2556)	-0.858 (2.152)	0.653 (1.5487)	0.6851 (1.9737)	-0.0993 (2.0603)	-0.573 (1.7278)
	5	-4.1206 (2.2963)	-4.8169 (1.8189)	-1.8066 (2.7112)	-2.1631 (3.0171)	-4.6101 (3.0771)	-4.9231 (4.0795)	-1.2306 (1.4114)	-1.0733 (1.3929)	-3.9449 (2.7145)	-3.9785 (2.7689)
200	0.05	0.8203 (1.605)	0.3919 (1.3153)	0.8486 (1.2995)	0.9904 (1.6277)	0.695 (2.0341)	0.2472 (1.1515)	1.233 (1.2335)	0.9671 (1.5858)	0.6844 (1.5599)	0.3788 (1.4811)
	0.1	0.7776 (1.5965)	0.3412 (1.2906)	0.8077 (1.3)	0.9557 (1.6147)	0.6288 (2.0229)	0.1888 (1.1506)	1.2109 (1.2343)	0.9508 (1.5843)	0.6468 (1.5604)	0.3241 (1.443)
	0.5	0.3958 (1.6221)	-0.1206 (1.0618)	0.5968 (1.2467)	0.6812 (1.5589)	0.1171 (1.9183)	-0.2449 (1.3473)	1.0291 (1.1863)	0.787 (1.535)	0.3301 (1.6597)	-0.1768 (1.3455)
	1	-0.0683 (1.57)	-0.7042 (0.7263)	0.3035 (0.9853)	0.2408 (1.4491)	-0.5701 (1.8074)	-0.7802 (1.802)	0.8031 (1.1033)	0.5836 (1.4792)	-0.0608 (1.6775)	-0.6242 (1.2621)
	5	-4.4192 (1.7632)	-5.0176 (1.7727)	-2.175 (2.7708)	-1.9933 (2.4338)	-4.6796 (2.4886)	-4.896 (4.0955)	-1.0395 (0.9313)	-1.0058 (1.1938)	-4.7635 (3.1601)	-4.4073 (3.0987)

Loss function used for training (spanning header over CE, Brier, Boost, As1, As2).

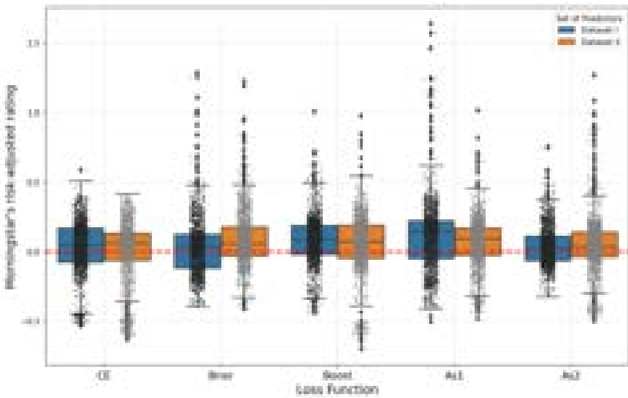
[a] *Dataset I* comprises of the financial variables and technical indicators while *Dataset II* contains both *Dataset I* and the candlestick chart (price) patterns as described in Table S.V.1.

[b] The initial endowment is $1000.

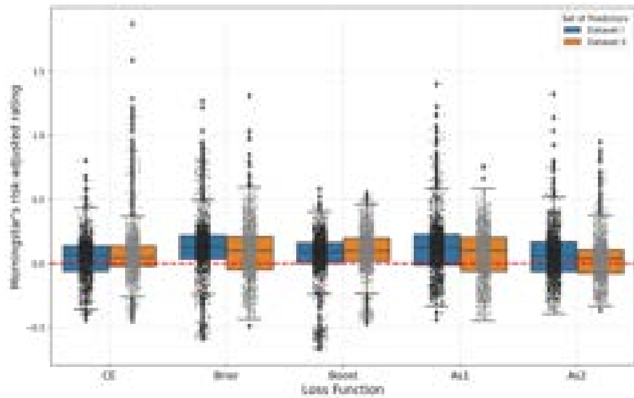
[c] The IQR values are shown in parentheses.

[d] There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

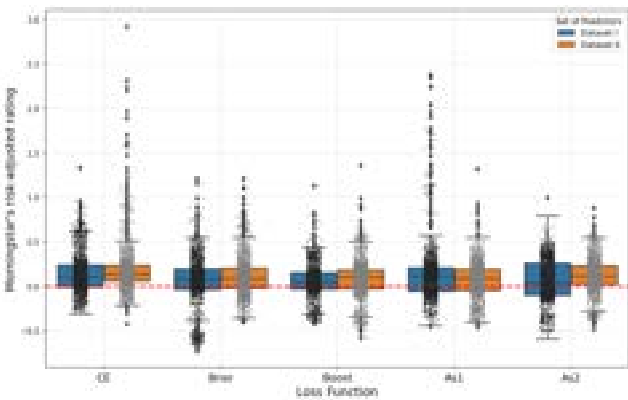
Figure S.VII.3: Morningstar's risk-adjusted ratings of 1718 trading sessions with the fixed transaction cost strategy investing in *SPY* for 100 days based on one-day ahead forecasts by the LGBM models trained with five different loss functions



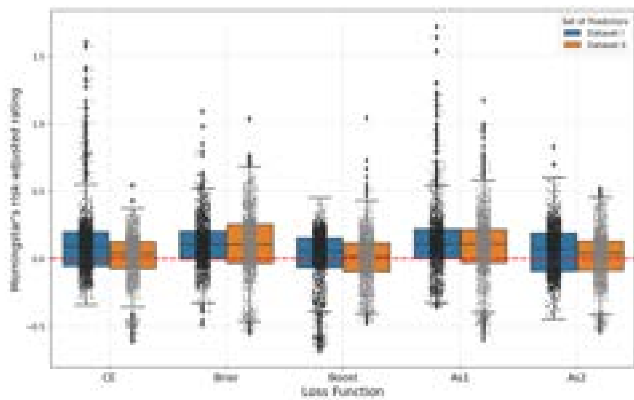
(a) *AUC* score



(b) *Calmar ratio* score



(c) *Sharpe ratio* score

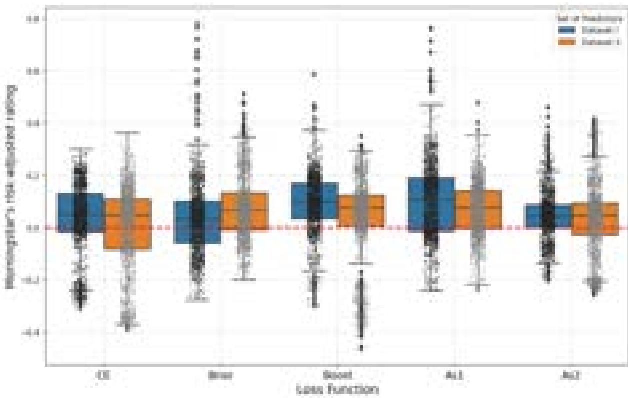


(d) *CECPP* score

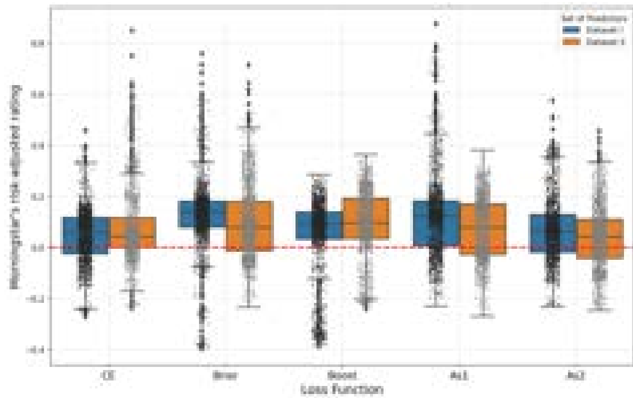
i The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

ii Two sets of predictors are used: *Dataset I* (including the financial variables and technical indicators) and *Dataset II* (including *Dataset I* and candlestick chart patterns as described in Table S.V.1).

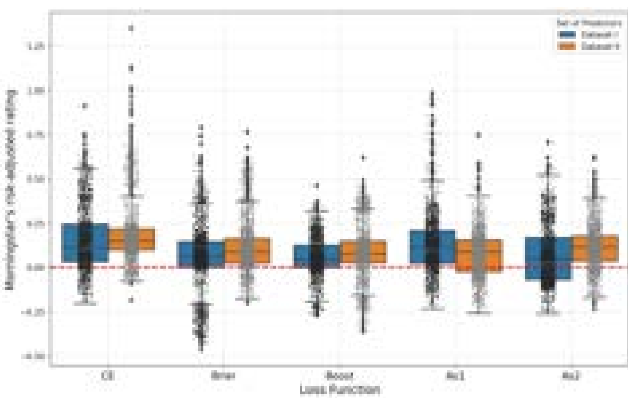
Figure S.VII.4: Morningstar's risk-adjusted ratings of 1618 trading sessions with the fixed transaction cost strategy investing in *SPY* for 200 days based on one-day ahead forecasts by the LGBM models trained with five different loss functions



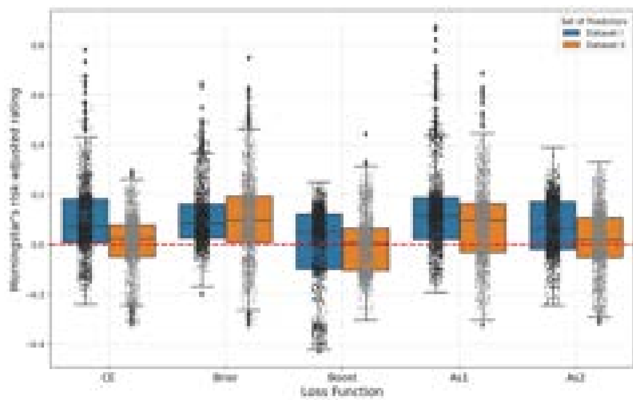
(a) *AUC* score



(b) *Calmar ratio* score



(c) *Sharpe ratio* score



(d) *CECPP* score

i The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

ii Two sets of predictors are used: *Dataset I* (including the financial variables and technical indicators) and *Dataset II* (including *Dataset I* and candlestick chart patterns as described in Table S.V.1).

Table S.VII.2: The median (and IQR) of Morningstar's risk-adjusted ratings of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the LGBM models trained with five loss functions (i.e., *CE, Brier, Boost, As1,* or *As2*), then cross-validated with the *CECPP* scoring function for different amounts of transaction cost across various holding periods (in days)

Holding period	Trans. cost	Loss function used for training									
		CE		Brier		Boost		As1		As2	
		Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II
100	0.05	0.092 (0.2511)	0.036 (0.1949)	0.1046 (0.2125)	0.1046 (0.2882)	0.0558 (0.2195)	0.0165 (0.21)	0.1109 (0.2189)	0.1101 (0.2468)	0.0753 (0.2812)	0.0382 (0.2171)
	0.1	0.0852 (0.2508)	0.03 (0.1946)	0.0991 (0.212)	0.1004 (0.2858)	0.0491 (0.2194)	0.0109 (0.2095)	0.1081 (0.2182)	0.1075 (0.2472)	0.0691 (0.2786)	0.0314 (0.2163)
	0.5	0.0318 (0.2459)	-0.0192 (0.1926)	0.0581 (0.2174)	0.0665 (0.2753)	-0.0004 (0.2154)	-0.035 (0.211)	0.0847 (0.2076)	0.0861 (0.2398)	0.021 (0.2653)	-0.018 (0.2081)
	1	-0.0297 (0.2355)	-0.0778 (0.1906)	0.0091 (0.2174)	0.0218 (0.267)	-0.0615 (0.2084)	-0.0924 (0.2134)	0.0558 (0.1969)	0.0604 (0.2323)	-0.0341 (0.2499)	-0.0766 (0.1973)
	5	-0.4406 (0.1808)	-0.4653 (0.1598)	-0.2826 (0.2694)	-0.2979 (0.2752)	-0.4729 (0.1707)	-0.4879 (0.2503)	-0.1723 (0.1732)	-0.1613 (0.2442)	-0.4346 (0.2)	-0.4693 (0.1756)
200	0.05	0.0729 (0.1751)	0.0246 (0.1293)	0.0832 (0.1366)	0.098 (0.1837)	0.0502 (0.2191)	0.0066 (0.1691)	0.1139 (0.1668)	0.0984 (0.1948)	0.0621 (0.1918)	0.0192 (0.1603)
	0.1	0.0679 (0.1749)	0.019 (0.1287)	0.079 (0.1369)	0.0936 (0.1834)	0.0442 (0.2181)	0.0008 (0.1706)	0.1113 (0.165)	0.0958 (0.1952)	0.0564 (0.1917)	0.0134 (0.1594)
	0.5	0.025 (0.1728)	-0.028 (0.1237)	0.0463 (0.1346)	0.0544 (0.1797)	-0.0045 (0.2144)	-0.0457 (0.1815)	0.0923 (0.1545)	0.0763 (0.2014)	0.0078 (0.1918)	-0.035 (0.1553)
	1	-0.0261 (0.1662)	-0.086 (0.1209)	0.0021 (0.1291)	0.0056 (0.1838)	-0.0637 (0.2076)	-0.1053 (0.1947)	0.0664 (0.1449)	0.0504 (0.2077)	-0.0506 (0.191)	-0.0935 (0.1522)
	5	-0.4631 (0.1474)	-0.5102 (0.1284)	-0.3128 (0.1974)	-0.336 (0.2175)	-0.5257 (0.1492)	-0.5592 (0.2637)	-0.1577 (0.1259)	-0.1672 (0.2188)	-0.473 (0.1756)	-0.5262 (0.1407)

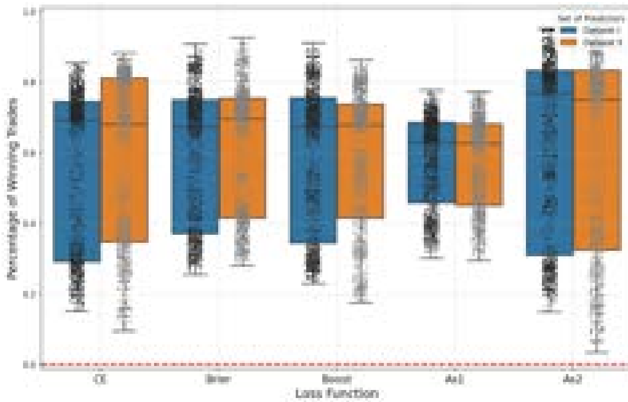
[a] *Dataset I* comprises of the financial variables and technical indicators while *Dataset II* contains both *Dataset I* and the candlestick chart (price) patterns as described in Table S.V.1.

[b] The initial endowment is $1000.

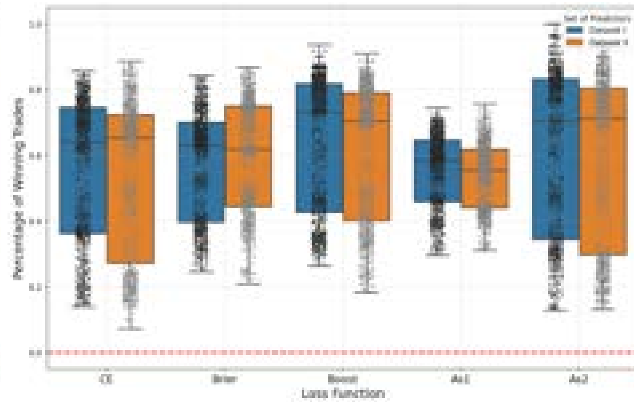
[c] The IQR values are shown in parentheses.

[d] There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

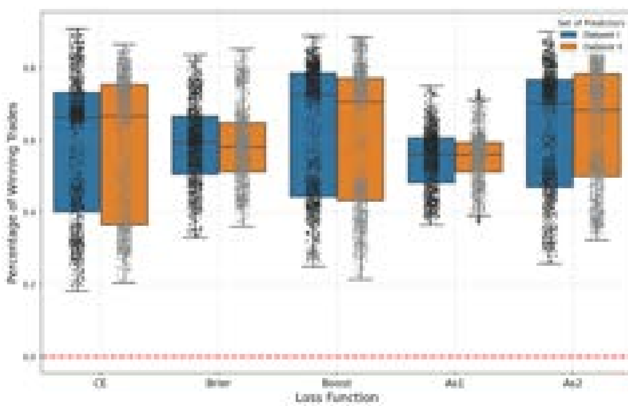
Figure S.VII.5: The percentages of winning trades of 1718 trading sessions with the fixed transaction cost strategy investing in *SPY* for 100 days based on one-day ahead forecasts by the LGBM models trained with five different loss functions



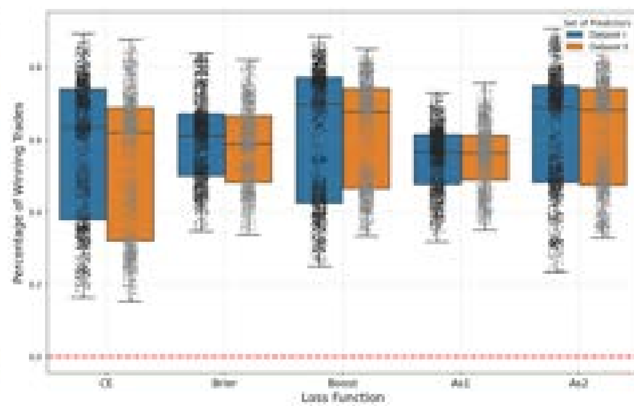
(a) *AUC* score



(b) *Calmar ratio* score



(c) *Sharpe ratio* score

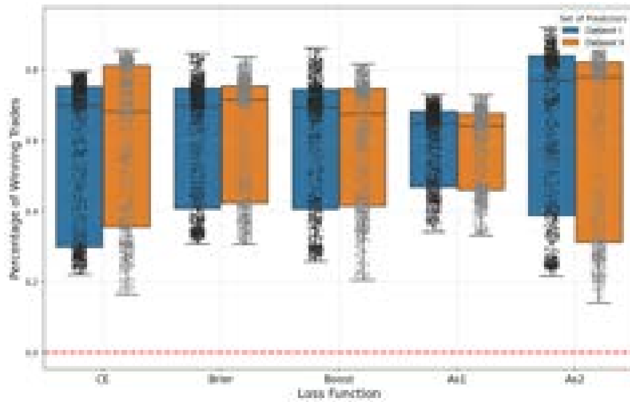


(d) *CECPP* score

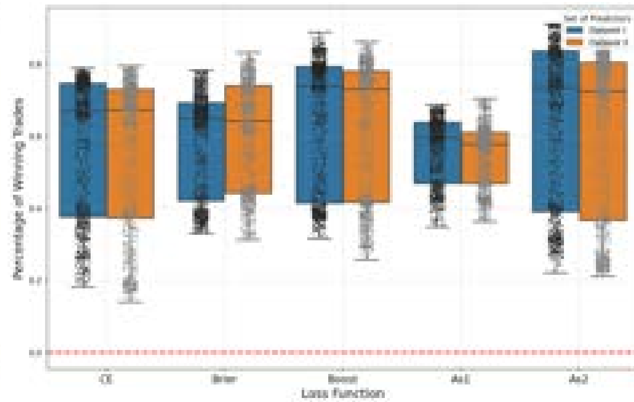
i The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

ii Two sets of predictors are used: *Dataset I* (including the financial variables and technical indicators) and *Dataset II* (including *Dataset I* and candlestick chart patterns as described in Table S.V.1).

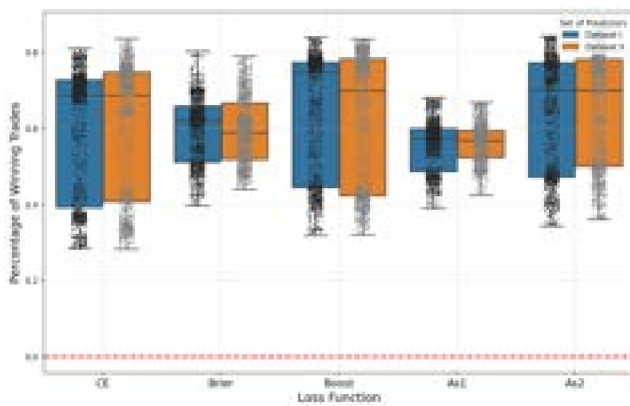
Figure S.VII.6: The percentages of winning trades of 1618 trading sessions with the fixed transaction cost strategy investing in *SPY* for 200 days based on one-day ahead forecasts by the LGBM models trained with five different loss functions



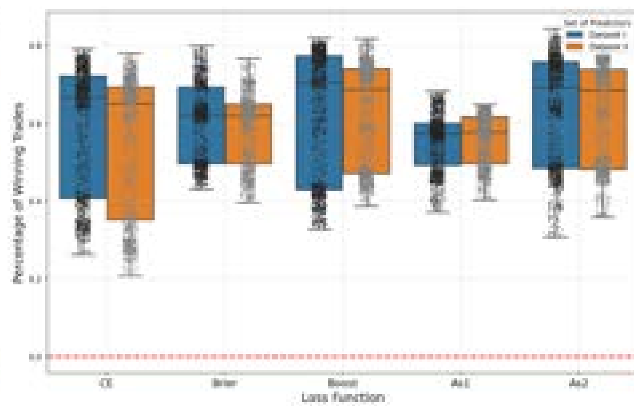
(a) *AUC* score



(b) *Calmar ratio* score



(c) *Sharpe ratio* score



(d) *CECPP* score

[i] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

[ii] Two sets of predictors are used: *Dataset I* (including the financial variables and technical indicators) and *Dataset II* (including *Dataset I* and candlestick chart patterns as described in Table S.V.1).

Table S.VII.3: The median (and IQR) of the percentages of winning trades of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the LGBM models trained with five loss functions (i.e., *CE, Brier, Boost, As1,* and *As2*), then cross-validated with the *CECPP* scoring function for different amounts of transaction cost across various holding periods (in days)

Holding period	Trans. cost	Loss function used for training									
		CE		Brier		Boost		As1		As2	
		Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II
100	0.05	0.6351 (0.3611)	0.6203 (0.3678)	0.6092 (0.1719)	0.5889 (0.1853)	0.6981 (0.347)	0.6777 (0.276)	0.5667 (0.1374)	0.5629 (0.1237)	0.6885 (0.2681)	0.6815 (0.2656)
	0.1	0.6351 (0.3611)	0.6203 (0.3678)	0.6053 (0.1716)	0.5879 (0.1853)	0.6981 (0.347)	0.6777 (0.2721)	0.5667 (0.1361)	0.5629 (0.1237)	0.6885 (0.2681)	0.6786 (0.2646)
	0.5	0.6328 (0.3604)	0.5987 (0.3511)	0.5955 (0.172)	0.5802 (0.1742)	0.6538 (0.3244)	0.6731 (0.2647)	0.5618 (0.1355)	0.5604 (0.1298)	0.6813 (0.2684)	0.6596 (0.2464)
	1	0.6275 (0.3499)	0.5797 (0.3475)	0.5871 (0.1676)	0.5691 (0.1675)	0.6415 (0.3174)	0.6604 (0.2548)	0.5568 (0.131)	0.5556 (0.122)	0.6591 (0.2656)	0.624 (0.2469)
	5	0.5532 (0.3442)	0.5085 (0.344)	0.5385 (0.134)	0.522 (0.1422)	0.5818 (0.2967)	0.5614 (0.2121)	0.5349 (0.1313)	0.5233 (0.1181)	0.5667 (0.2176)	0.5565 (0.2187)
200	0.05	0.6667 (0.3138)	0.6502 (0.3422)	0.6207 (0.1964)	0.6209 (0.1531)	0.7054 (0.3443)	0.6882 (0.2699)	0.5746 (0.1086)	0.5785 (0.1201)	0.6923 (0.2735)	0.6847 (0.2568)
	0.1	0.6667 (0.3138)	0.6502 (0.3422)	0.6181 (0.1957)	0.6199 (0.1524)	0.7054 (0.3443)	0.6882 (0.2634)	0.5746 (0.1086)	0.5785 (0.1201)	0.6923 (0.2735)	0.6847 (0.2563)
	0.5	0.6629 (0.2986)	0.6378 (0.3294)	0.6087 (0.1784)	0.6127 (0.1538)	0.6578 (0.3311)	0.6809 (0.2561)	0.5694 (0.1058)	0.5777 (0.1223)	0.6855 (0.2706)	0.6753 (0.2361)
	1	0.65 (0.2991)	0.6198 (0.3181)	0.6011 (0.1735)	0.6011 (0.1373)	0.6505 (0.3233)	0.6724 (0.2466)	0.5651 (0.102)	0.5676 (0.118)	0.6639 (0.264)	0.6486 (0.2344)
	5	0.5744 (0.2882)	0.5403 (0.3016)	0.5549 (0.1241)	0.548 (0.127)	0.6037 (0.2832)	0.5631 (0.213)	0.547 (0.1083)	0.5385 (0.1026)	0.5826 (0.2314)	0.5658 (0.2056)

[a] *Dataset I* comprises of the financial variables and technical indicators while *Dataset II* contains both *Dataset I* and the candlestick chart (price) patterns as described in Table S.V.1.

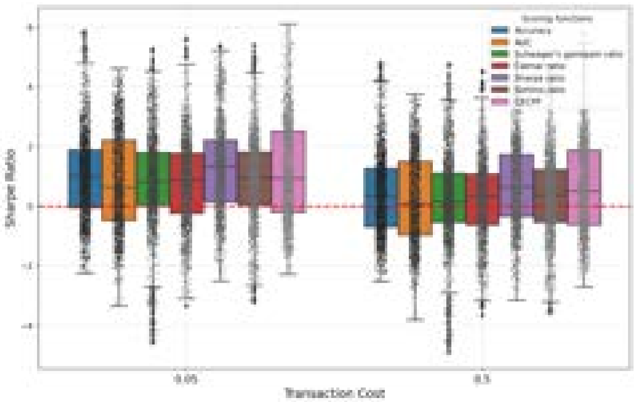
[b] The initial endowment is $1000.

[c] The IQR values are shown in parentheses.

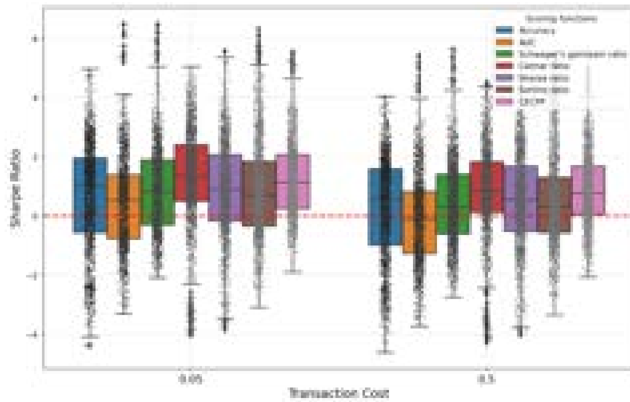
[d] There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

S.VIII Performance of the Trading Strategy for Different Amounts of Transaction Cost

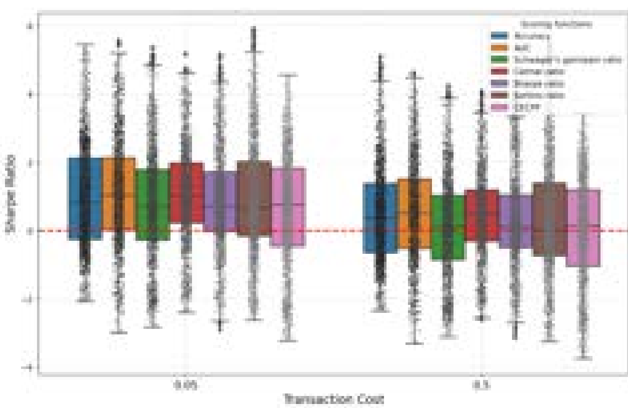
Figure S.VIII.1: The Sharpe ratios of 1718 trading sessions with the fixed transaction cost strategy investing in *SPY* for 100 days based on one-day ahead forecasts by a LGBM model trained using the data on financial variables and technical indicators (*Dataset I*) described in Table S.V.1



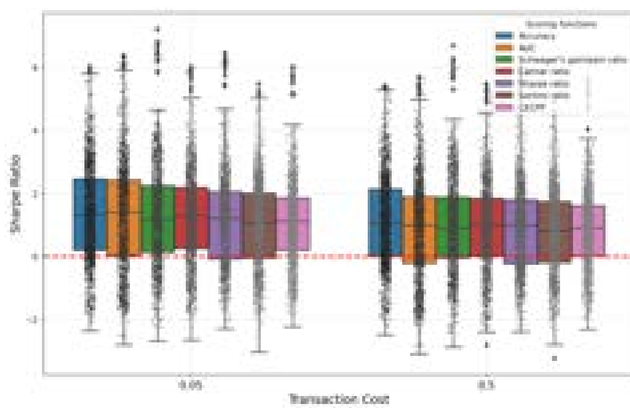
(a) *CE* loss



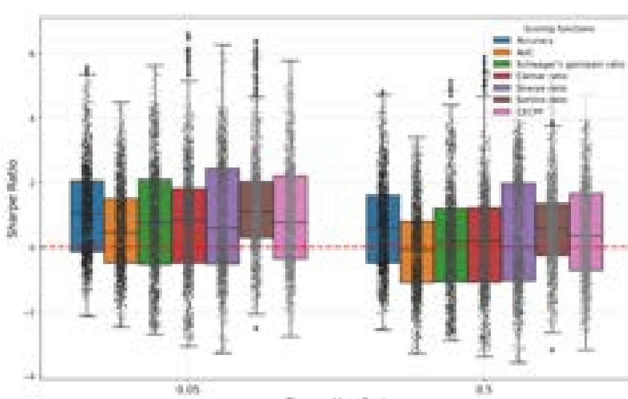
(b) *Brier* loss



(c) *Boost* loss



(d) *As1* loss



(e) *As2* loss

[i] The initial endowment is $1000.

Table S.VIII.1: The median (and IQR) of the Sharpe ratios of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the LGBM models trained with the *Brier* loss function, then cross-validated with seven scoring functions (i.e., *Accuracy, AUC, Schwager's gain/pain ratio, Calmar ratio, Sharpe ratio, Sortino ratio,* or *CECPP*) for different amounts of transaction cost across various holding periods (in days). The predictors used: financial variables and technical indicators (*Dataset I*) described in Table S.V.1

Holding period	Trans. cost	Scoring function used for cross validation						
		Accuracy	AUC	Schwager's gain/pain ratio	Calmar ratio	Sharpe ratio	Sortino ratio	CECPP
100	0.05	1.0328 (2.509)	0.5419 (2.1533)	0.821 (2.2033)	1.3565 (1.9179)	0.8938 (2.2353)	0.7138 (2.1912)	1.1446 (1.8163)
	0.1	0.9917 (2.5201)	0.4872 (2.1258)	0.7654 (2.1918)	1.3026 (1.8838)	0.8413 (2.2399)	0.6735 (2.1423)	1.1088 (1.788)
	0.5	0.5633 (2.5744)	-0.1026 (2.0971)	0.3058 (1.9966)	0.8767 (1.7091)	0.5553 (2.191)	0.3131 (1.8681)	0.7647 (1.6174)
	1	0.0133 (2.6088)	-0.8631 (2.1584)	-0.2035 (1.9066)	0.3515 (1.5771)	0.1748 (2.2917)	-0.1406 (1.6233)	0.2857 (1.6083)
	5	-3.0894 (3.7495)	-4.9142 (2.7479)	-3.055 (2.6186)	-3.2724 (2.4036)	-2.0233 (2.9804)	-3.0514 (3.0838)	-1.8066 (2.7112)
200	0.05	0.5056 (2.2499)	0.5944 (1.3134)	0.692 (1.1008)	1.3298 (0.9547)	0.7257 (1.1767)	0.5807 (1.3633)	0.8486 (1.2995)
	0.1	0.4538 (2.2487)	0.5162 (1.3286)	0.6498 (1.0812)	1.2803 (0.9529)	0.6901 (1.1916)	0.5519 (1.3375)	0.8077 (1.3)
	0.5	0.0951 (2.0958)	-0.1024 (1.4317)	0.3271 (0.9196)	0.9323 (0.9421)	0.36 (1.3297)	0.2802 (1.0919)	0.5968 (1.2467)
	1	-0.3405 (1.7813)	-0.7801 (1.6151)	-0.1332 (0.8956)	0.4411 (0.8494)	0.0344 (1.4892)	-0.0923 (0.9683)	0.3035 (0.9853)
	5	-3.2375 (3.0969)	-5.2002 (2.1501)	-3.2074 (2.2068)	-3.3203 (2.6938)	-2.0269 (1.9699)	-3.1005 (2.1393)	-2.175 (2.7708)

[a] The initial endowment is $1000.

[b] The IQR values are shown in parentheses.

[c] There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

Table S.VIII.2: The median (and IQR) of the Sharpe ratios of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the LGBM models trained with the *As1* loss function, then cross-validated with seven scoring functions (i.e., *Accuracy, AUC, Schwager's gain/pain ratio, Calmar ratio, Sharpe ratio, Sortino ratio,* or *CECPP*) for different amounts of transaction cost across various holding periods (in days). The predictors used: financial variables and technical indicators (*Dataset I*) described in Table S.V.1

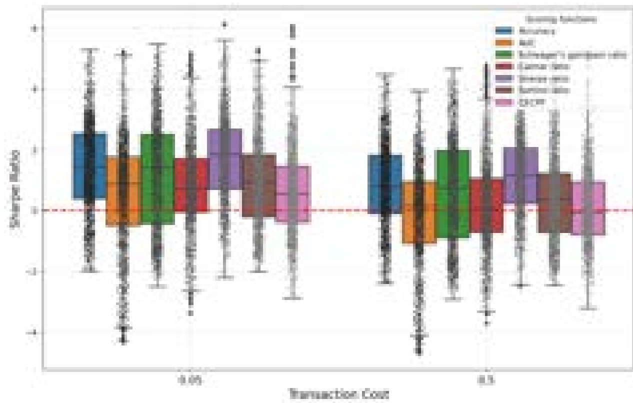
Holding period	Trans. cost	Scoring function used for cross validation						
		Accuracy	AUC	Schwager's gain/pain ratio	Calmar ratio	Sharpe ratio	Sortino ratio	CECPP
100	0.05	1.3171 (2.2395)	1.4092 (2.3657)	1.1983 (2.1132)	1.2913 (1.9424)	1.2038 (2.1317)	1.0698 (2.0481)	1.1239 (1.656)
	0.1	1.2935 (2.2276)	1.3682 (2.3394)	1.1633 (2.0827)	1.2625 (1.9198)	1.1807 (2.1182)	1.0442 (2.0386)	1.1038 (1.6474)
	0.5	1.0739 (2.1174)	0.99 (2.0972)	0.913 (1.9243)	1.0105 (1.815)	0.9925 (2.0198)	0.8019 (1.9272)	0.9118 (1.5925)
	1	0.8099 (1.8893)	0.4736 (1.8485)	0.6092 (1.8644)	0.6656 (1.7272)	0.7527 (1.8657)	0.5371 (1.7431)	0.653 (1.5487)
	5	-1.3944 (1.4635)	-3.1465 (2.0008)	-1.9089 (1.6602)	-1.7264 (1.7391)	-1.1471 (1.3507)	-1.6728 (1.5635)	-1.2306 (1.4114)
200	0.05	1.2925 (1.1457)	0.8647 (1.9713)	0.8337 (1.5137)	1.0986 (1.5904)	1.0366 (1.4298)	1.074 (1.3454)	1.233 (1.2335)
	0.1	1.2654 (1.127)	0.847 (1.9659)	0.8116 (1.5054)	1.0714 (1.575)	1.0246 (1.4231)	1.0615 (1.3332)	1.2109 (1.2343)
	0.5	1.0435 (1.0159)	0.6866 (1.893)	0.6166 (1.4182)	0.8893 (1.4582)	0.8878 (1.3561)	0.9212 (1.2546)	1.0291 (1.1863)
	1	0.7129 (0.9784)	0.393 (1.6361)	0.3797 (1.2879)	0.683 (1.3068)	0.6874 (1.2916)	0.7274 (1.1495)	0.8031 (1.1033)
	5	-1.2334 (1.3855)	-3.4511 (1.9488)	-2.0309 (0.9236)	-1.9065 (1.2903)	-1.0233 (0.9011)	-1.531 (1.3365)	-1.0395 (0.9313)

[a] The initial endowment is $1000.

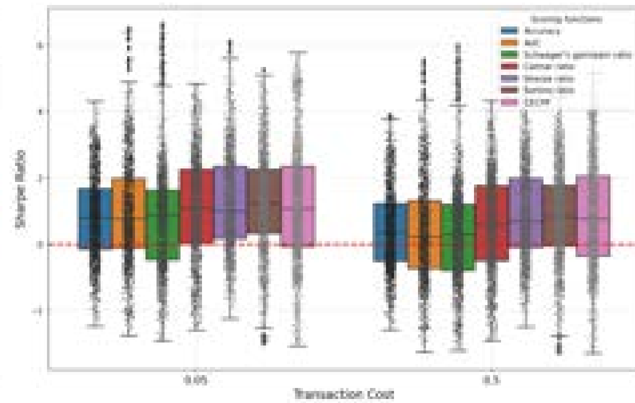
[b] The IQR values are shown in parentheses.

[c] There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

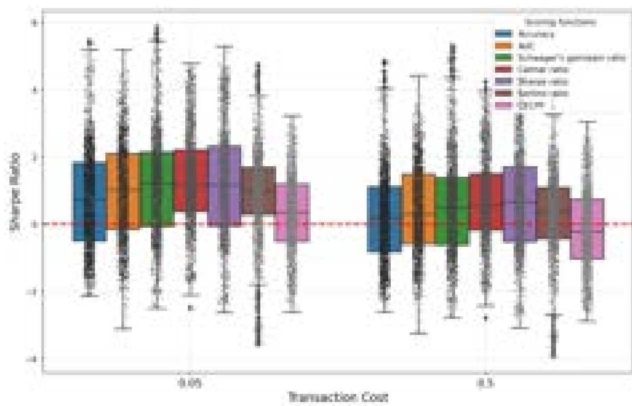
Figure S.VIII.2: The Sharpe ratios of 1718 trading sessions with the fixed transaction cost strategy investing in *SPY* for 100 days based on one-day ahead forecasts by a LGBM model trained using the data on financial variables, technical indicators, and candlestick chart patterns (*Dataset II*) described in Table S.V.1



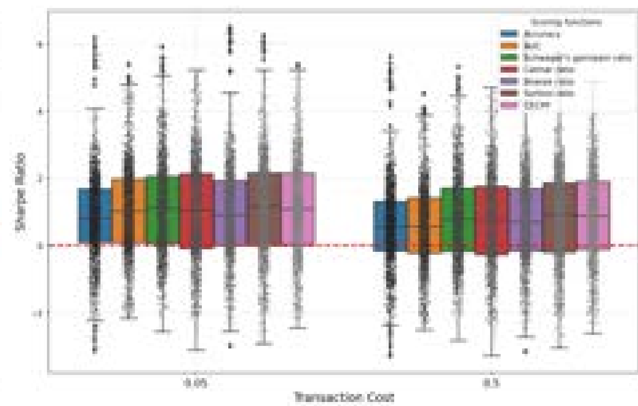
(a) *CE* loss



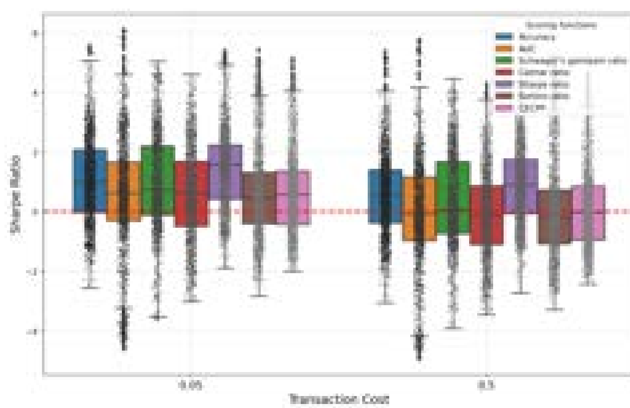
(b) *Brier* loss



(c) *Boost* loss



(d) *As1* loss



(e) *As2* loss

[i] The initial endowment is $1000.

Table S.VIII.3: The median (and IQR) of the Sharpe ratios of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the LGBM models trained with the *Brier* loss function, then cross-validated with seven scoring functions (i.e., *Accuracy, AUC, Schwager's gain/pain ratio, Calmar ratio, Sharpe ratio, Sortino ratio,* or *CECPP*) for different amounts of transaction cost across various holding periods (in days). The predictors used: financial variables, technical indicators, and candlestick chart patterns (*Dataset II*) described in Table S.V.1

Holding period	Trans. cost	Scoring function used for cross validation						
		Accuracy	AUC	Schwager's gain/pain ratio	Calmar ratio	Sharpe ratio	Sortino ratio	CECPP
100	0.05	0.7975 (1.8223)	0.7631 (2.0948)	0.8646 (2.1063)	1.084 (2.2351)	1.0047 (2.1848)	1.2266 (1.9255)	1.07 (2.4234)
	0.1	0.7495 (1.8034)	0.6979 (2.0859)	0.7969 (2.0769)	1.0301 (2.2228)	0.9641 (2.1686)	1.1837 (1.8926)	1.0271 (2.4162)
	0.5	0.345 (1.7313)	0.2108 (2.0377)	0.3017 (2.0013)	0.5909 (2.2521)	0.685 (2.0785)	0.8296 (1.8144)	0.7717 (2.4102)
	1	-0.2318 (1.7638)	-0.3595 (1.9366)	-0.1746 (2.0466)	0.091 (2.1229)	0.3261 (2.0432)	0.3652 (1.7716)	0.3822 (2.3863)
	5	-3.116 (2.8427)	-4.7734 (2.3574)	-3.3291 (2.6863)	-3.6847 (3.1134)	-1.9464 (3.325)	-2.421 (2.4307)	-2.1631 (3.0171)
200	0.05	0.66 (1.2999)	0.8431 (1.3179)	0.4394 (1.3273)	0.8019 (1.6405)	0.9745 (1.3626)	1.3174 (1.5826)	0.9904 (1.6277)
	0.1	0.6203 (1.278)	0.7835 (1.3034)	0.4051 (1.3129)	0.7548 (1.6063)	0.9397 (1.3375)	1.2875 (1.5699)	0.9557 (1.6147)
	0.5	0.2421 (1.1392)	0.2688 (1.2132)	0.104 (1.2459)	0.4894 (1.4703)	0.6566 (1.2084)	0.9954 (1.2781)	0.6812 (1.5589)
	1	-0.2303 (1.1516)	-0.4229 (1.4025)	-0.236 (1.2526)	-0.1307 (1.5703)	0.2666 (1.1762)	0.4732 (1.0169)	0.2408 (1.4491)
	5	-3.1896 (2.7293)	-5.0484 (2.3129)	-3.519 (2.1054)	-3.5717 (2.5622)	-2.0355 (3.2314)	-2.3382 (1.6602)	-1.9933 (2.4338)

[a] The initial endowment is $1000.

[b] The IQR values are shown in parentheses.

[c] There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

Table S.VIII.4: The median (and IQR) of the Sharpe ratios of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the LGBM models trained with the *As1* loss function, then cross-validated with seven scoring functions (i.e., *Accuracy, AUC, Schwager's gain/pain ratio, Calmar ratio, Sharpe ratio, Sortino ratio,* or *CECPP*) for different amounts of transaction cost across various holding periods (in days). The predictors used: financial variables, technical indicators, and candlestick chart patterns (*Dataset II*) described in Table S.V.1

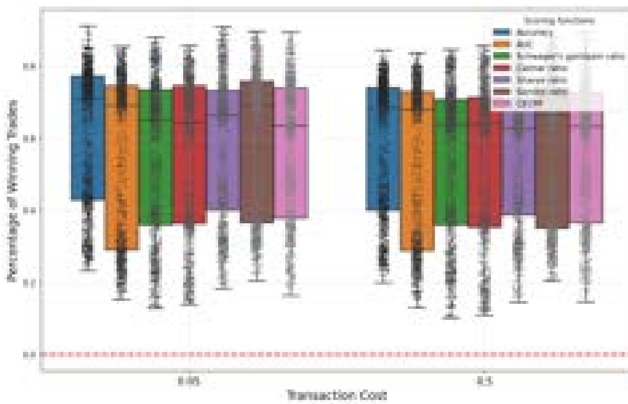
Holding period	Trans. cost	Scoring function used for cross validation						
		Accuracy	AUC	Schwager's gain/pain ratio	Calmar ratio	Sharpe ratio	Sortino ratio	CECPP
100	0.05	0.8093 (1.6012)	1.0235 (1.885)	1.116 (2.0057)	1.0598 (2.223)	0.9203 (1.9225)	1.1861 (2.1678)	1.0912 (2.1139)
	0.1	0.7809 (1.5855)	0.9687 (1.8583)	1.081 (2.0038)	1.0302 (2.2041)	0.8972 (1.9172)	1.1606 (2.1566)	1.0682 (2.1029)
	0.5	0.5649 (1.4787)	0.572 (1.6466)	0.783 (1.8849)	0.7692 (2.0359)	0.7054 (1.8518)	0.9388 (2.0654)	0.8927 (2.0276)
	1	0.2767 (1.3831)	0.0503 (1.412)	0.4036 (1.7265)	0.4123 (1.8728)	0.4875 (1.7687)	0.62 (1.9338)	0.6851 (1.9737)
	5	-1.6702 (1.692)	-3.4343 (1.8517)	-2.2264 (1.4181)	-2.0087 (1.2219)	-1.1751 (1.3952)	-1.8264 (1.3853)	-1.0733 (1.3929)
200	0.05	0.7709 (1.2818)	0.9049 (1.2278)	0.8717 (1.5071)	0.8391 (1.6778)	0.9677 (1.245)	0.9894 (1.6936)	0.9671 (1.5858)
	0.1	0.7371 (1.2765)	0.8508 (1.1952)	0.8418 (1.4917)	0.808 (1.6559)	0.9455 (1.24)	0.9657 (1.6849)	0.9508 (1.5843)
	0.5	0.5198 (1.2468)	0.4851 (1.1303)	0.6215 (1.384)	0.5909 (1.5044)	0.7307 (1.1774)	0.7614 (1.5994)	0.787 (1.535)
	1	0.2657 (1.221)	0.0866 (1.0487)	0.3212 (1.223)	0.3311 (1.3331)	0.487 (1.1318)	0.4984 (1.5086)	0.5836 (1.4792)
	5	-1.7291 (1.6619)	-3.5879 (1.4904)	-2.2425 (1.0029)	-2.0762 (0.7933)	-1.0979 (1.1878)	-1.7547 (1.1086)	-1.0058 (1.1938)

[a] The initial endowment is $1000.

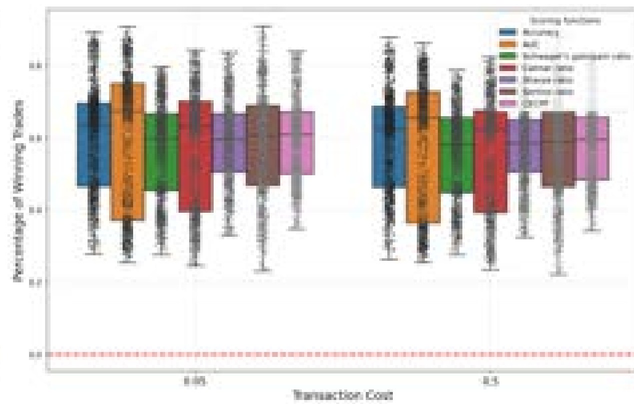
[b] The IQR values are shown in parentheses.

[c] There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

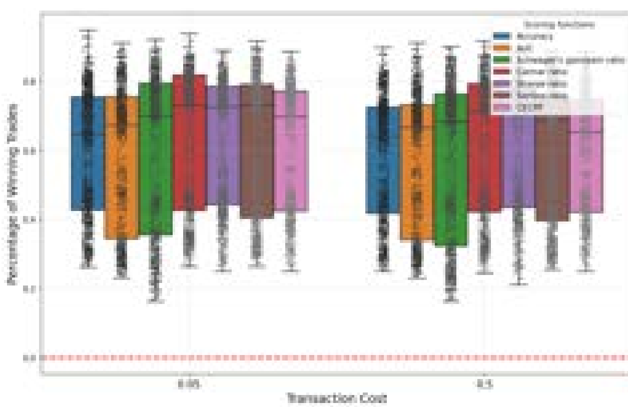
Figure S.VIII.3: The percentages of winning trades of 1718 trading sessions with the fixed transaction cost strategy investing in *SPY* for 100 days based on one-day ahead forecasts by a LGBM model trained using the data on financial variables and technical indicators (*Dataset I*) described in Table S.V.1



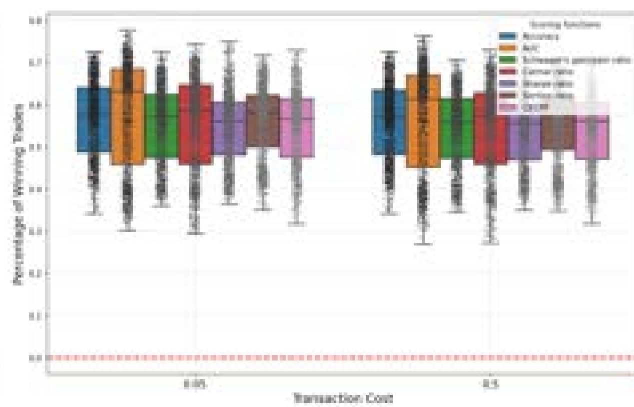
(a) *CE* loss



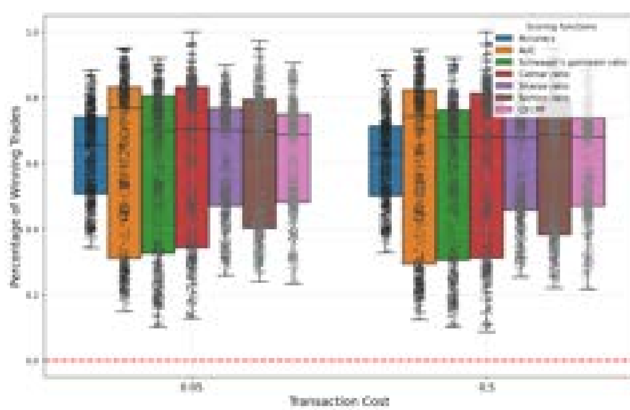
(b) *Brier* loss



(c) *Boost* loss



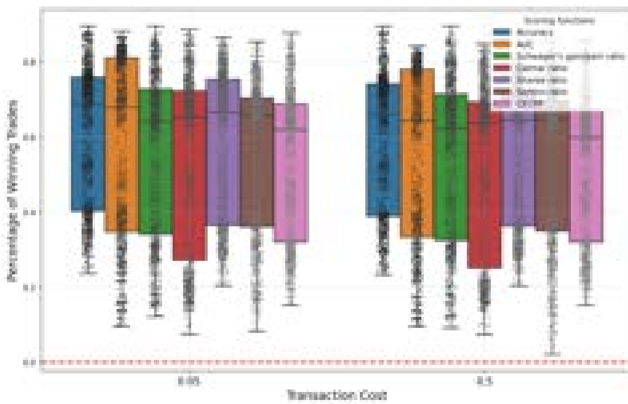
(d) *As1* loss



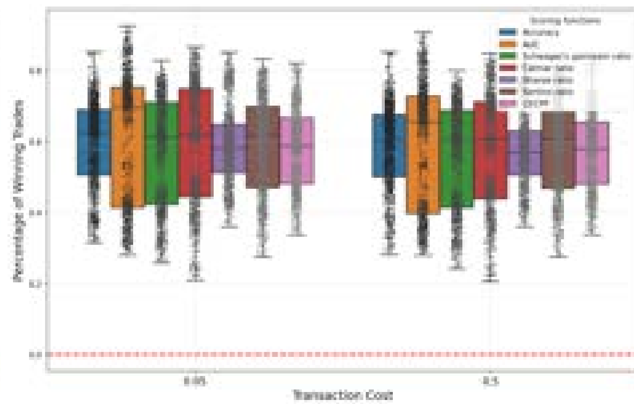
(e) *As2* loss

[i] The initial endowment is $1000.

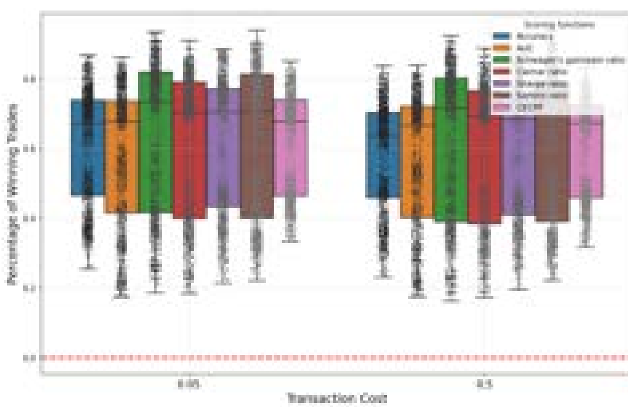
Figure S.VIII.4: The percentages of winning trades of 1718 trading sessions with the fixed transaction cost strategy investing in *SPY* for 100 days based on one-day ahead forecasts by a LGBM model trained using the data on financial variables, technical indicators, and candlestick chart patterns (*Dataset II*) described in Table S.V.1



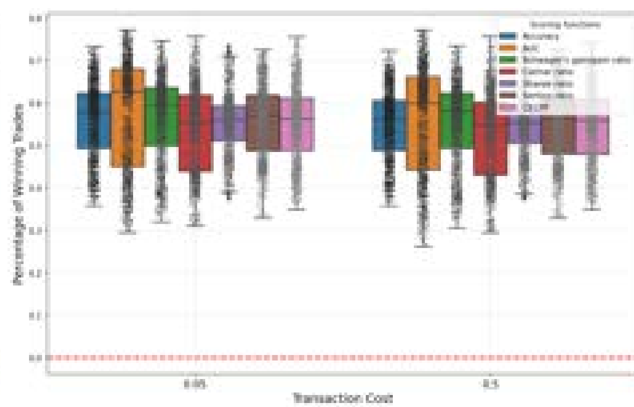
(a) *CE* loss



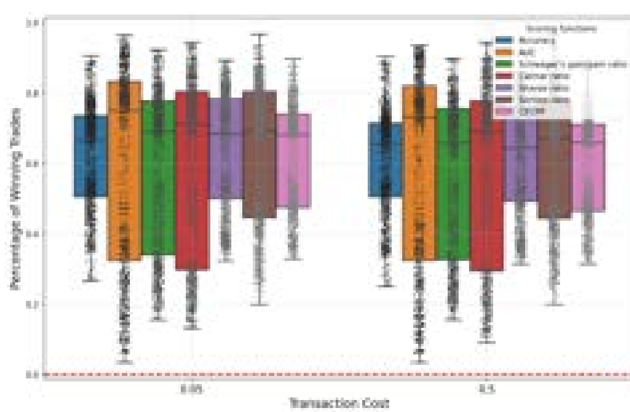
(b) *Brier* loss



(c) *Boost* loss



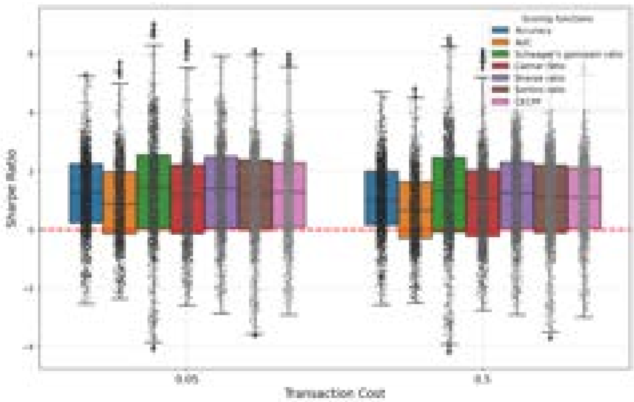
(d) *As1* loss



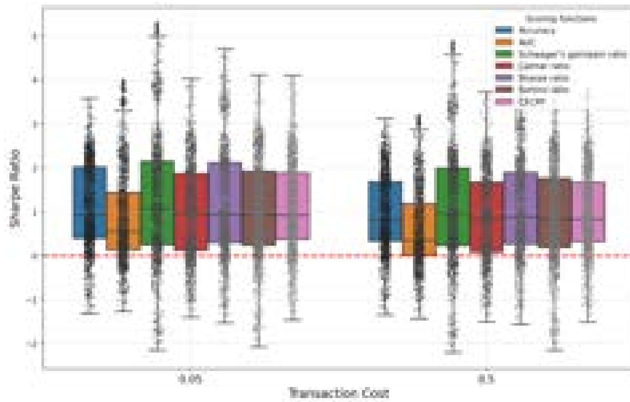
(e) *As2* loss

[i] The initial endowment is $1000.

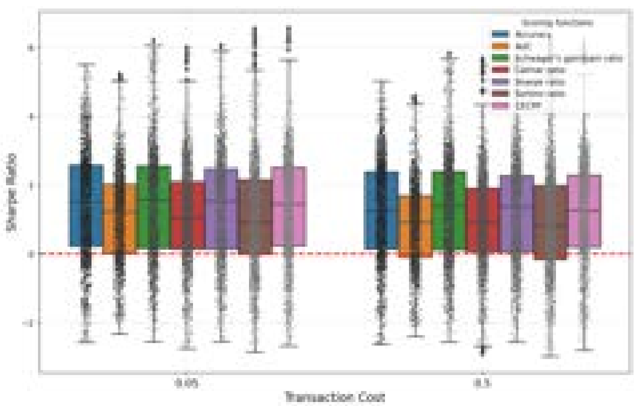
Figure S.VIII.5: The Sharpe ratios of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by RF models cross-validated with seven different scoring functions for two different sets of predictors



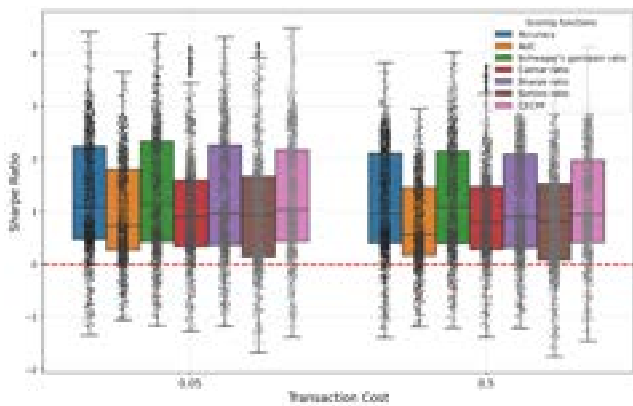
(a) *Dataset I*: Holding period (D) = 100 days



(b) *Dataset I*: Holding period (D) = 200 days



(c) *Dataset II*: Holding period (D) = 100 days



(d) *Dataset II*: Holding period (D) = 200 days

[i] *Dataset I* comprises of the financial variables and technical indicators while *Dataset II* contains both *Dataset I* and the candlestick chart (price) patterns as described in Table S.V.1.

[ii] The initial endowment is $1000.

[iii] There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

Table S.VIII.5: The median (and IQR) of the Sharpe ratios of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the RF models cross-validated with seven scoring functions (i.e., *Accuracy, AUC, Schwager's gain/pain ratio, Calmar ratio, Sharpe ratio, Sortino ratio,* or *CECPP*) for different amounts of transaction cost across various holding periods (in days). The predictors used: financial variables and technical indicators (*Dataset I*) described in Table S.V.1

Holding period	Trans. cost	Scoring function used for cross validation						
		Accuracy	AUC	Schwager's gain/pain ratio	Calmar ratio	Sharpe ratio	Sortino ratio	CECPP
100	0.05	1.2579 (1.999)	0.8696 (2.1242)	1.4526 (2.5603)	1.1892 (2.3567)	1.3977 (2.4593)	1.1913 (2.3908)	1.2931 (2.1883)
	0.1	1.234 (1.9789)	0.8412 (2.1068)	1.4348 (2.5655)	1.1856 (2.34)	1.3694 (2.4405)	1.1874 (2.3926)	1.278 (2.1704)
	0.5	1.0156 (1.8582)	0.6438 (1.9297)	1.3038 (2.5563)	1.0818 (2.2656)	1.2448 (2.2974)	1.1096 (2.3162)	1.1069 (2.0965)
	1	0.7094 (1.8139)	0.3379 (1.7859)	1.154 (2.4231)	0.9492 (2.222)	1.039 (2.1669)	0.9629 (2.1724)	0.8725 (1.9721)
	5	-0.4552 (2.4311)	-1.5438 (1.9034)	-0.2843 (2.4137)	-0.3732 (2.3416)	-0.2484 (2.4719)	-0.2144 (2.4345)	-0.1559 (2.7718)
200	0.05	0.9431 (1.6251)	0.5522 (1.3064)	1.0714 (1.903)	0.9675 (1.752)	0.9458 (1.8189)	0.9279 (1.6777)	0.926 (1.508)
	0.1	0.931 (1.6041)	0.5329 (1.295)	1.0611 (1.8941)	0.9542 (1.7371)	0.9365 (1.7926)	0.9219 (1.6625)	0.9187 (1.4935)
	0.5	0.8284 (1.3469)	0.3489 (1.1739)	0.9701 (1.7591)	0.8332 (1.5899)	0.8859 (1.6337)	0.8315 (1.5784)	0.8288 (1.3409)
	1	0.5868 (1.073)	0.1643 (0.8914)	0.8661 (1.5596)	0.6182 (1.5231)	0.7981 (1.5078)	0.62 (1.4747)	0.6937 (1.1951)
	5	-0.2844 (1.6006)	-1.6829 (1.6398)	-0.0391 (1.5868)	-0.1337 (1.8162)	-0.0067 (1.872)	0.0652 (1.9633)	0.0013 (1.7615)

[a] The initial endowment is $1000.

[b] The IQR values are shown in parentheses.

[c] There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

Table S.VIII.6: The median (and IQR) of the Sharpe ratios of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the RF models cross-validated with seven scoring functions (i.e., *Accuracy, AUC, Schwager's gain/pain ratio, Calmar ratio, Sharpe ratio, Sortino ratio,* or *CECPP*) for different amounts of transaction cost across various holding periods (in days). The predictors used: financial variables, technical indicators, and candlestick chart patterns (*Dataset II*) described in Table S.V.1

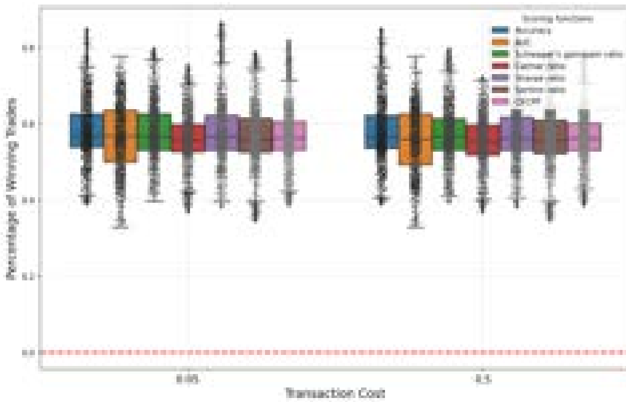
Holding period	Trans. cost	Scoring function used for cross validation						
		Accuracy	AUC	Schwager's gain/pain ratio	Calmar ratio	Sharpe ratio	Sortino ratio	CECPP
100	0.05	1.5178 (2.3477)	1.2256 (2.0039)	1.5648 (2.3861)	1.0421 (1.9504)	1.535 (2.3288)	0.9312 (2.1703)	1.4575 (2.2785)
	0.1	1.4945 (2.3289)	1.2062 (1.987)	1.5568 (2.3688)	1.0244 (1.9435)	1.51 (2.3199)	0.9207 (2.1883)	1.4355 (2.2572)
	0.5	1.2584 (2.2371)	0.9281 (1.7861)	1.4031 (2.2688)	0.9141 (1.8403)	1.3724 (2.2058)	0.8268 (2.1533)	1.2798 (2.0926)
	1	0.9922 (2.1723)	0.5233 (1.6292)	1.2004 (2.124)	0.7591 (1.7169)	1.1949 (2.0998)	0.7144 (2.0866)	1.0594 (1.9502)
	5	-0.3422 (2.4236)	-1.3596 (1.73)	-0.1041 (2.5419)	-0.3335 (2.3182)	-0.2585 (2.5271)	-0.5779 (2.237)	-0.0875 (2.5042)
200	0.05	1.0814 (1.7677)	0.7332 (1.5225)	1.1447 (1.9185)	0.9046 (1.2561)	0.9759 (1.8913)	0.9417 (1.5213)	1.0371 (1.7375)
	0.1	1.0733 (1.7661)	0.7153 (1.5083)	1.1312 (1.902)	0.8954 (1.2506)	0.9696 (1.8832)	0.9302 (1.5147)	1.0199 (1.7066)
	0.5	0.9636 (1.707)	0.5631 (1.2887)	1.0907 (1.7462)	0.798 (1.1868)	0.8951 (1.7668)	0.8141 (1.4479)	0.9465 (1.5594)
	1	0.7569 (1.4483)	0.3433 (1.0036)	0.9829 (1.5931)	0.617 (1.1105)	0.814 (1.5847)	0.6094 (1.4032)	0.8329 (1.3347)
	5	-0.3032 (1.6236)	-1.2842 (1.398)	0.0804 (1.6093)	-0.1457 (1.6929)	-0.0022 (1.6402)	-0.2754 (1.6414)	0.0669 (1.6571)

[a] The initial endowment is $1000.

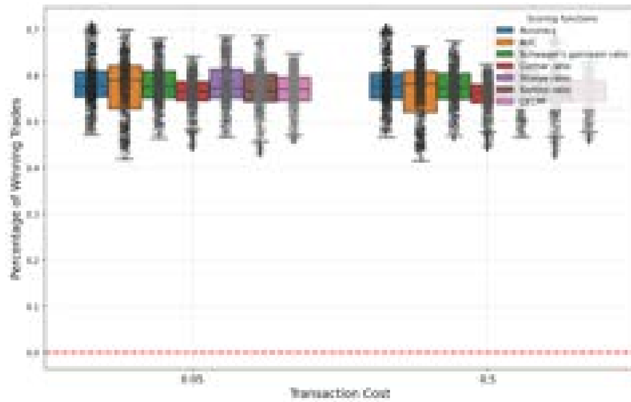
[b] The IQR values are shown in parentheses.

[c] There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

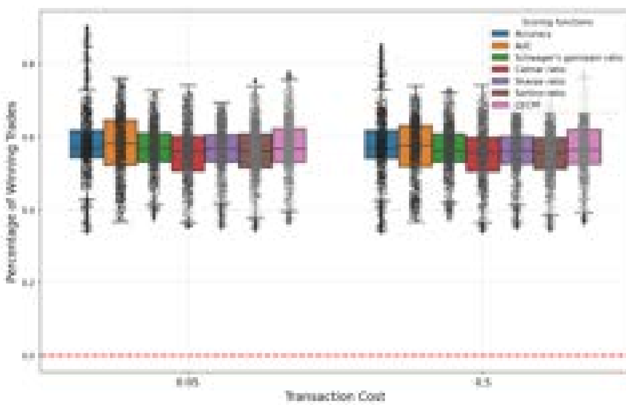
Figure S.VIII.6: The percentages of winning trades of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by RF models cross-validated with seven different scoring functions for two different sets of predictors



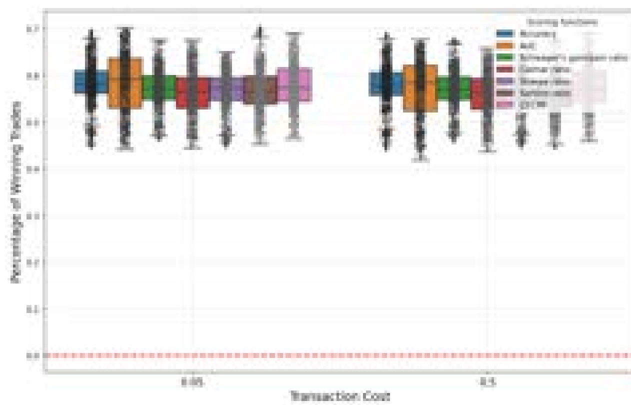
(a) *Dataset I*: Holding period (D) = 100 days



(b) *Dataset I*: Holding period (D) = 200 days



(c) *Dataset II*: Holding period (D) = 100 days



(d) *Dataset II*: Holding period (D) = 200 days

[i] *Dataset I* comprises of the financial variables and technical indicators while *Dataset II* contains both *Dataset I* and the candlestick chart (price) patterns as described in Table S.V.1.

[ii] The initial endowment is $1000.

[iii] There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

Table S.VIII.7: The median (and IQR) of the percentages of winning trades of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the RF models cross-validated with seven scoring functions (i.e., *Accuracy, AUC, Schwager's gain/pain ratio, Calmar ratio, Sharpe ratio, Sortino ratio,* or *CECPP*) for different amounts of transaction cost across various holding periods (in days). The predictors used: financial variables and technical indicators (*Dataset I*) described in Table S.V.1

Holding period	Trans. cost	Scoring function used for cross validation						
		Accuracy	AUC	Schwager's gain/pain ratio	Calmar ratio	Sharpe ratio	Sortino ratio	CECPP
100	0.05	0.5682 (0.0873)	0.5682 (0.1364)	0.567 (0.096)	0.5567 (0.0711)	0.567 (0.0918)	0.5579 (0.0887)	0.5625 (0.0744)
	0.1	0.5682 (0.0879)	0.5667 (0.1341)	0.567 (0.096)	0.5567 (0.0711)	0.567 (0.0918)	0.5575 (0.0887)	0.5625 (0.0744)
	0.5	0.5667 (0.0864)	0.5606 (0.1342)	0.5667 (0.0833)	0.5532 (0.0793)	0.5612 (0.087)	0.5567 (0.0866)	0.5612 (0.0709)
	1	0.5635 (0.0785)	0.5556 (0.1263)	0.5612 (0.0763)	0.551 (0.0807)	0.5612 (0.0805)	0.5474 (0.0872)	0.5567 (0.0663)
	5	0.5567 (0.0878)	0.5385 (0.1268)	0.5464 (0.0792)	0.5408 (0.084)	0.551 (0.0713)	0.5376 (0.0722)	0.5417 (0.0622)
200	0.05	0.5775 (0.0566)	0.5904 (0.0933)	0.5765 (0.0593)	0.5657 (0.0422)	0.5729 (0.064)	0.5668 (0.0611)	0.5707 (0.051)
	0.1	0.5775 (0.0562)	0.5899 (0.0931)	0.5765 (0.0593)	0.5657 (0.0422)	0.5707 (0.0646)	0.5668 (0.0611)	0.5707 (0.0508)
	0.5	0.5753 (0.0573)	0.5812 (0.0901)	0.5736 (0.0561)	0.5625 (0.0412)	0.5707 (0.0635)	0.5635 (0.0516)	0.5685 (0.0448)
	1	0.5726 (0.0543)	0.5741 (0.0855)	0.5707 (0.0538)	0.5606 (0.0409)	0.5707 (0.0579)	0.5635 (0.052)	0.5657 (0.042)
	5	0.5645 (0.0606)	0.5529 (0.0948)	0.5606 (0.0427)	0.5464 (0.0512)	0.5645 (0.0442)	0.5533 (0.0478)	0.5561 (0.0392)

[a] The initial endowment is $1000.

[b] The IQR values are shown in parentheses.

[c] There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

Table S.VIII.8: The median (and IQR) of the percentages of winning trades of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the RF models cross-validated with seven scoring functions (i.e., *Accuracy, AUC, Schwager's gain/pain ratio, Calmar ratio, Sharpe ratio, Sortino ratio,* or *CECPP*) for different amounts of transaction cost across various holding periods (in days). The predictors used: financial variables, technical indicators, and candlestick chart patterns (*Dataset II*) described in Table S.V.1

Holding period	Trans. cost	Scoring function used for cross validation						
		Accuracy	AUC	Schwager's gain/pain ratio	Calmar ratio	Sharpe ratio	Sortino ratio	CECPP
100	0.05	0.5714 (0.0758)	0.5828 (0.1224)	0.567 (0.0805)	0.5612 (0.0979)	0.567 (0.0758)	0.5618 (0.0928)	0.567 (0.0918)
	0.1	0.5714 (0.0758)	0.5814 (0.1217)	0.5668 (0.0805)	0.5612 (0.0977)	0.567 (0.0758)	0.5618 (0.0928)	0.567 (0.0918)
	0.5	0.5686 (0.0777)	0.5763 (0.1128)	0.5625 (0.0776)	0.5575 (0.0948)	0.5625 (0.0763)	0.5567 (0.0866)	0.5652 (0.0944)
	1	0.567 (0.0666)	0.573 (0.1049)	0.5612 (0.0803)	0.5567 (0.0928)	0.5591 (0.0716)	0.5521 (0.0845)	0.5612 (0.0921)
	5	0.551 (0.0635)	0.5455 (0.1049)	0.551 (0.0672)	0.5423 (0.0997)	0.5417 (0.064)	0.5408 (0.0774)	0.5464 (0.0669)
200	0.05	0.5824 (0.0474)	0.5922 (0.1065)	0.5722 (0.0518)	0.5663 (0.0656)	0.5736 (0.0494)	0.5657 (0.0593)	0.5736 (0.0689)
	0.1	0.5819 (0.0474)	0.5909 (0.103)	0.5718 (0.0518)	0.5657 (0.0646)	0.5736 (0.0494)	0.5657 (0.0593)	0.5729 (0.0689)
	0.5	0.5789 (0.0484)	0.5856 (0.0964)	0.5707 (0.0486)	0.5641 (0.066)	0.5685 (0.0463)	0.5635 (0.0563)	0.5696 (0.0652)
	1	0.5736 (0.044)	0.578 (0.0963)	0.5684 (0.0465)	0.5635 (0.063)	0.5685 (0.0444)	0.5615 (0.0528)	0.5657 (0.0588)
	5	0.5604 (0.045)	0.5581 (0.085)	0.5625 (0.037)	0.5526 (0.0674)	0.5533 (0.0381)	0.5515 (0.0425)	0.5563 (0.0428)

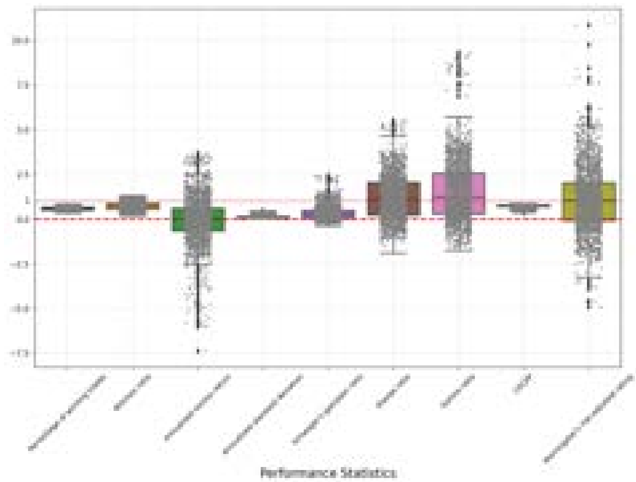
[a] The initial endowment is $1000.

[b] The IQR values are shown in parentheses.

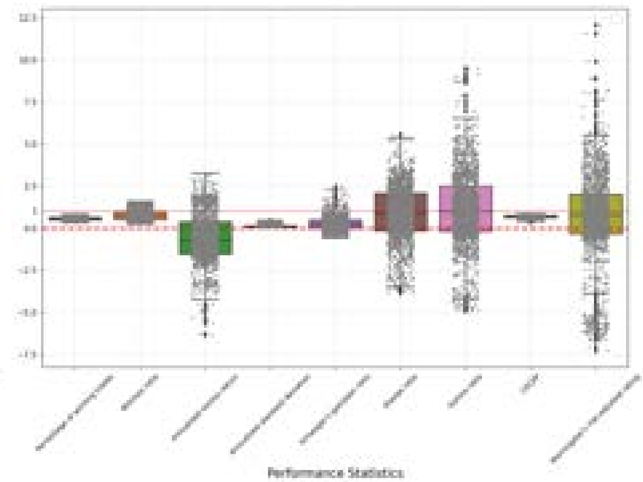
[c] There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

S.IX Performance Statistics of the Trading Strategy over Time

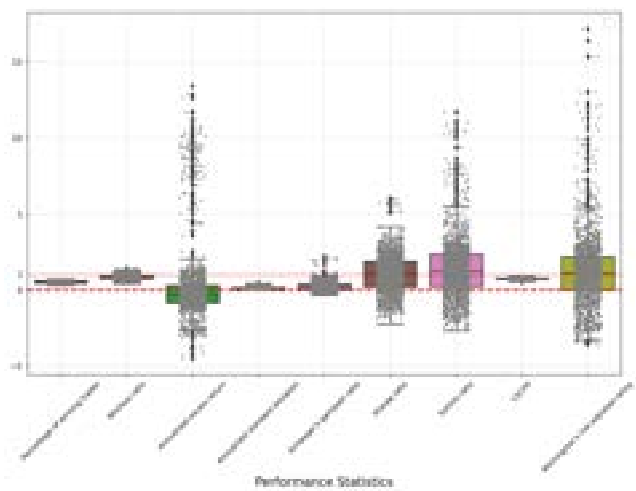
Figure S.IX.1: The performance statistics of 1718 trading sessions with the fixed transaction cost strategy investing in *SPY* for 100 days based on one-day ahead forecasts by a LGBM model trained using the data on financial variables and technical indicators (*Dataset I*) described in Table S.V.1



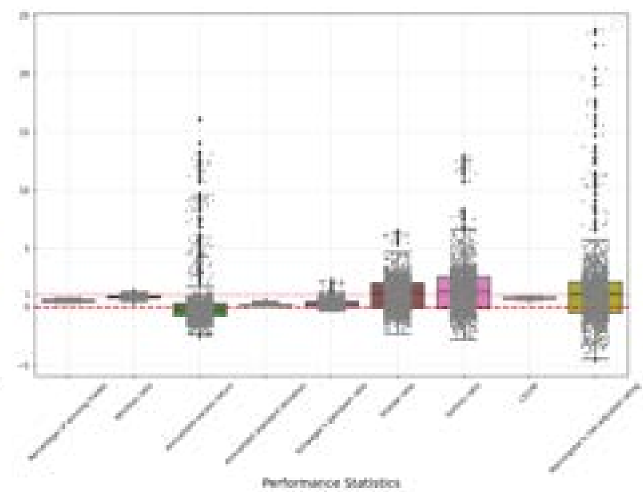
(a) *Brier* loss and *CECPP* score



(b) *Brier* loss and *Sharpe ratio* score



(c) *As1* loss and *CECPP* score

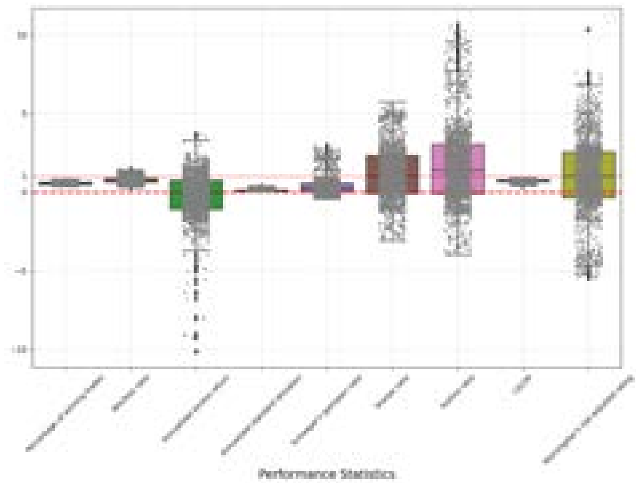


(d) *As1* loss and *Sharpe ratio* score

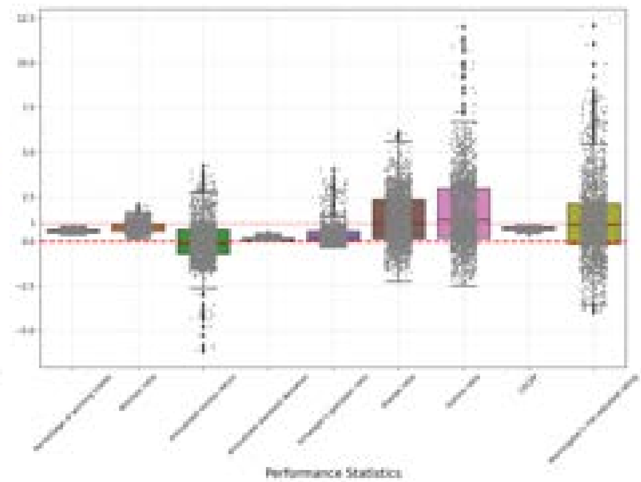
[i] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

[ii] The values of the annualized excess return and Morningstar's risk-adjusted rating are scaled up ten times to make them more visible in this box plot.

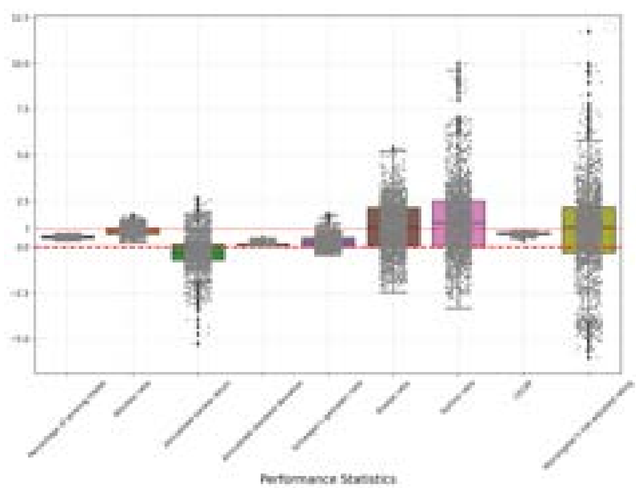
Figure S.IX.2: The performance statistics of 1718 trading sessions with the fixed transaction cost strategy investing in *SPY* for 100 days based on one-day ahead forecasts by a LGBM model trained using the data on financial variables, technical indicators, and candlestick chart patterns (*Dataset II*) described in Table S.V.1



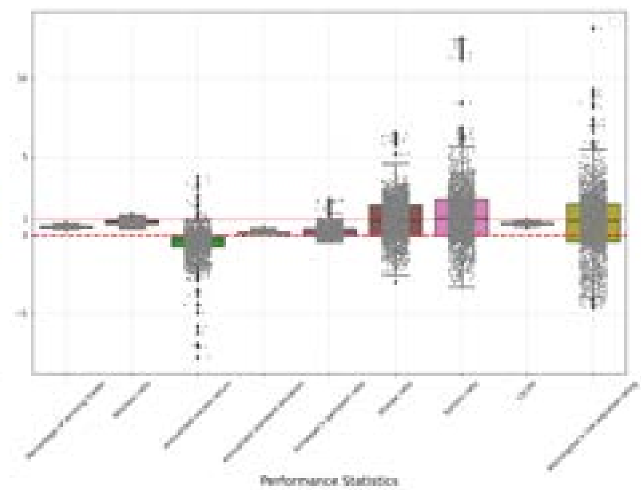
(a) *Brier* loss and *CECPP* score



(b) *Brier* loss and *Sharpe ratio* score



(c) *As1* loss and *CECPP* score

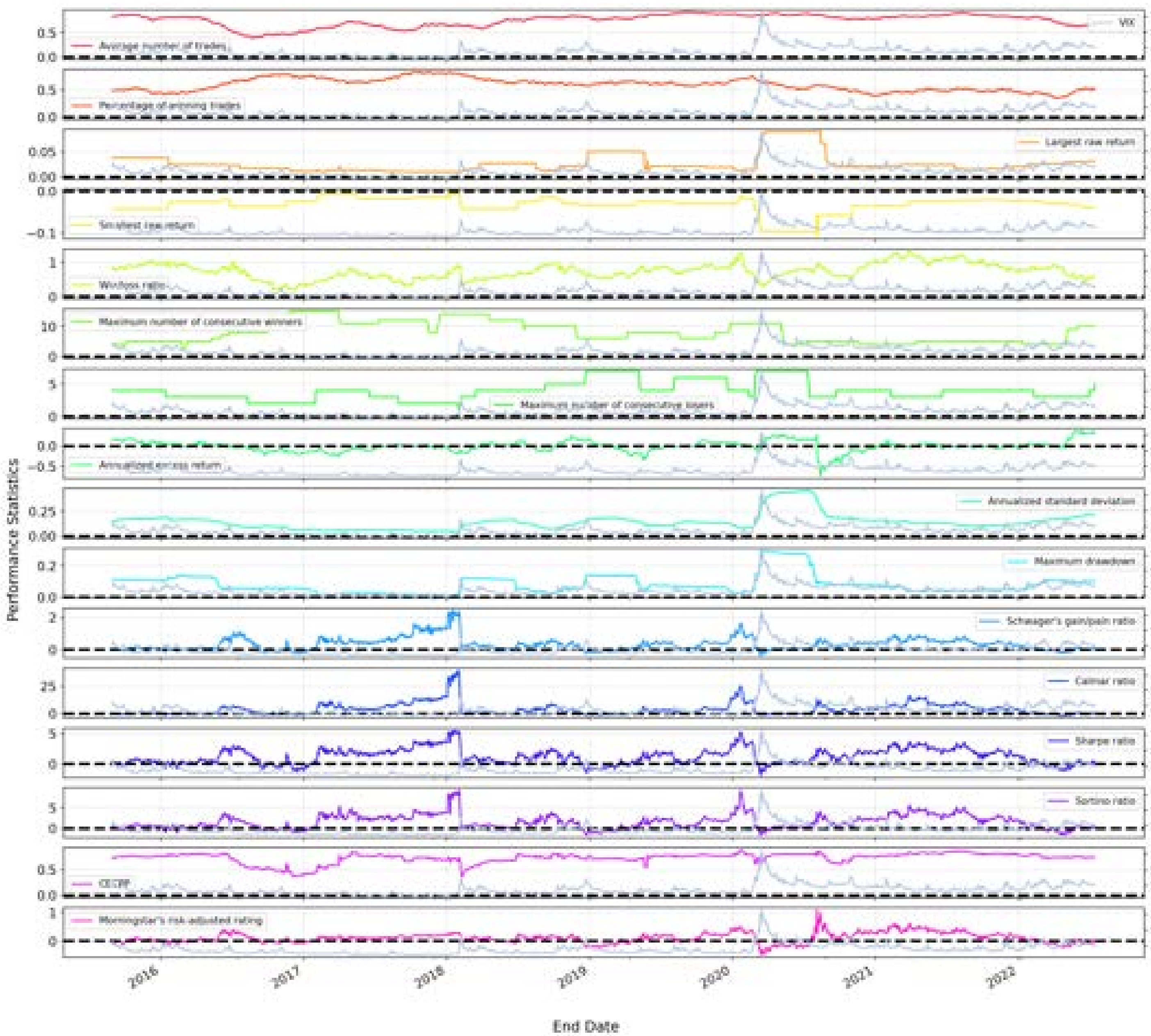


(d) *As1* loss and *Sharpe ratio* score

[i] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

[ii] The values of the annualized excess return and Morningstar's risk-adjusted rating are scaled up ten times to make them more visible in this box plot.

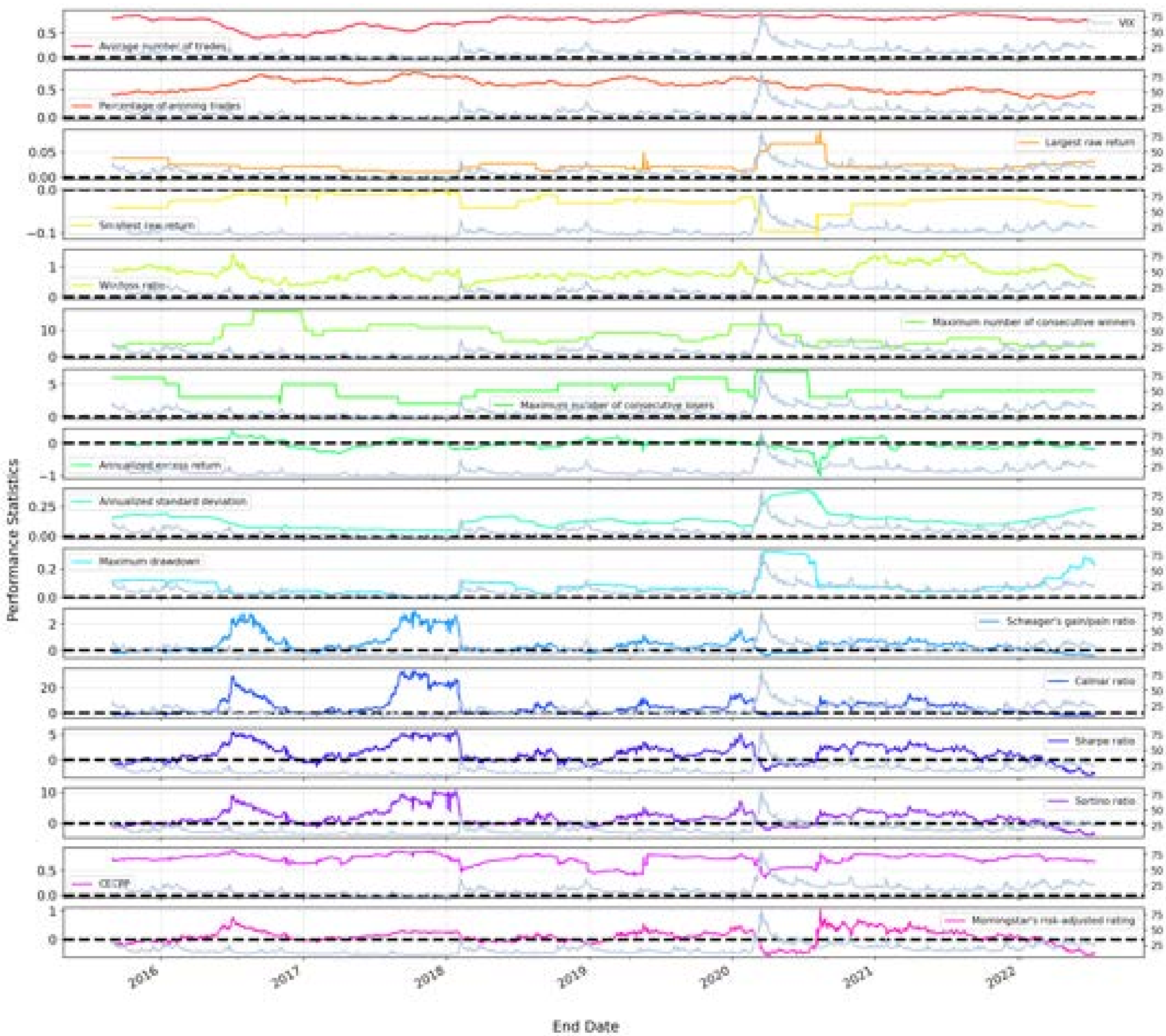
Figure S.IX.3: The performance statistics of 1718 trading sessions with the fixed transaction cost strategy investing in *SPY* for 100 days based on one-day ahead forecasts by a LGBM model (trained with the *Brier* loss function, then cross-validated with the *CECPP* scoring function) using the data on financial variables and technical indicators (*Dataset I*) described in Table S.V.1



[a] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

[b] The horizonal axis contains the end dates of the trading sessions.

Figure S.IX.4: The performance statistics of 1718 trading sessions with the fixed transaction cost strategy investing in *SPY* for 100 days based on one-day ahead forecasts by a LGBM model (trained with the *Brier* loss function, then cross-validated with the *CECPP* scoring function) using the data on financial variables, technical indicators, and candlestick chart patterns (*Dataset II*) described in Table S.V.1



[a] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

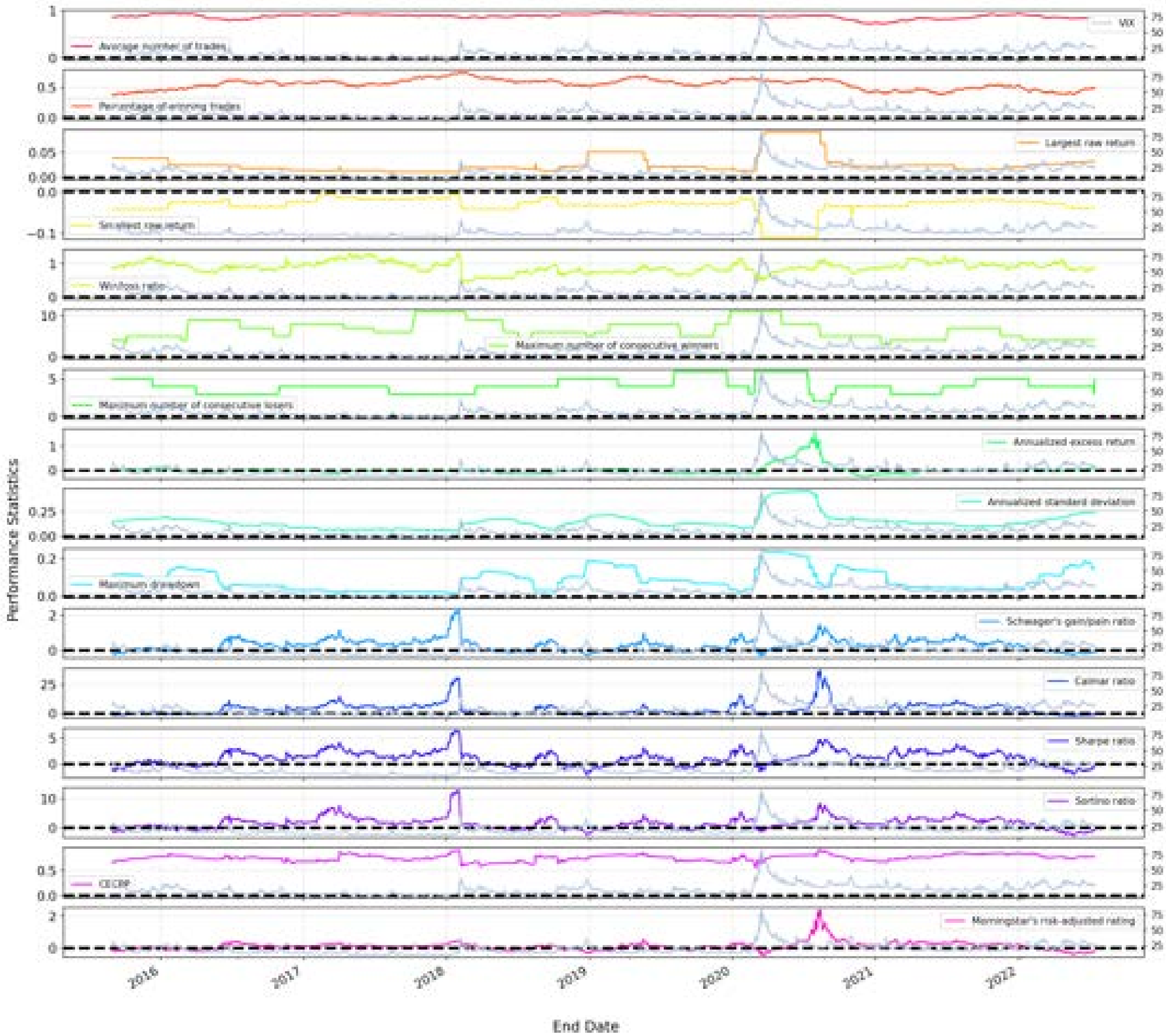
[b] The horizonal axis contains the end dates of the trading sessions.

Table S.IX.1: The median (and IQR) of all the performance metrics of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the LGBM model trained with the *Brier* loss function, then cross-validated with the *CECPP* scoring function for different amounts of transaction cost across various holding periods (in days)

Holding period	Trans. cost	Percent. of win. trades		Win/loss ratio		Annualized excess ret.		Annualized std. dev.		Schwager's gain/pain ratio		Sharpe ratio		Sortino ratio		CECPP		MRAR	
		Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset
100	0.05	0.6092 (0.1719)	0.5889 (0.1853)	0.7544 (0.3083)	0.7727 (0.238)	0.0013 (0.1267)	-0.0014 (0.1834)	0.1303 (0.072)	0.1276 (0.0738)	0.2793 (0.4448)	0.2475 (0.5906)	1.1446 (1.8163)	1.07 (2.4234)	1.2537 (2.29)	1.4112 (3.1385)	0.7627 (0.0874)	0.7458 (0.1082)	0.1046 (0.2125)	0.1046 (0.2882
	0.1	0.6053 (0.1716)	0.5879 (0.1853)	0.7524 (0.3042)	0.7679 (0.2458)	-0.0031 (0.1274)	-0.0051 (0.1839)	0.1302 (0.0719)	0.1275 (0.0738)	0.2699 (0.4363)	0.2352 (0.5848)	1.1088 (1.788)	1.0271 (2.4162)	1.2205 (2.2355)	1.341 (3.1138)	0.7622 (0.087)	0.7448 (0.1083)	0.0991 (0.212)	0.1004 (0.2858
	0.5	0.5955 (0.172)	0.5802 (0.1742)	0.7524 (0.3039)	0.7429 (0.2474)	-0.0317 (0.1412)	-0.0347 (0.1744)	0.1302 (0.0713)	0.1271 (0.0741)	0.1865 (0.3764)	0.1676 (0.5545)	0.7647 (1.6174)	0.7717 (2.4102)	0.9034 (1.9911)	0.961 (2.952)	0.752 (0.0857)	0.7372 (0.1062)	0.0581 (0.2174)	0.0665 (0.2753
	1	0.5871 (0.1676)	0.5691 (0.1675)	0.7259 (0.3343)	0.7469 (0.26)	-0.0689 (0.1599)	-0.0805 (0.169)	0.13 (0.0701)	0.127 (0.0743)	0.078 (0.3476)	0.0886 (0.5204)	0.2857 (1.6083)	0.3822 (2.3863)	0.3558 (1.9273)	0.5236 (2.9178)	0.7384 (0.1045)	0.7198 (0.1003)	0.0091 (0.2174)	0.0218 (0.267)
	5	0.5385 (0.134)	0.522 (0.1422)	0.6157 (0.4033)	0.6452 (0.3639)	-0.3377 (0.3041)	-0.3638 (0.2041)	0.1418 (0.0618)	0.1352 (0.0687)	-0.2747 (0.408)	-0.3289 (0.396)	-1.8066 (2.7112)	-2.1631 (3.0171)	-2.2357 (2.9495)	-2.6668 (3.8268)	0.4261 (0.2805)	0.3982 (0.2472)	-0.2826 (0.2694)	-0.2979 (0.2752
200	0.05	0.6207 (0.1964)	0.6209 (0.1531)	0.6855 (0.2693)	0.7333 (0.2114)	-0.0047 (0.0773)	-0.018 (0.1304)	0.1341 (0.0397)	0.1257 (0.0482)	0.2201 (0.3166)	0.2666 (0.4173)	0.8486 (1.2995)	0.9904 (1.6277)	0.9012 (1.5016)	1.1285 (2.0109)	0.7391 (0.0856)	0.7231 (0.1731)	0.0832 (0.1366)	0.098 (0.1837
	0.1	0.6181 (0.1957)	0.6199 (0.1524)	0.6905 (0.2733)	0.7266 (0.2155)	-0.0082 (0.0785)	-0.0232 (0.1306)	0.1341 (0.0397)	0.1257 (0.0482)	0.2101 (0.3138)	0.2542 (0.4125)	0.8077 (1.3)	0.9557 (1.6147)	0.8667 (1.5147)	1.093 (1.9958)	0.7379 (0.0858)	0.7227 (0.1729)	0.079 (0.1369)	0.0936 (0.1834
	0.5	0.6087 (0.1784)	0.6127 (0.1538)	0.696 (0.2839)	0.7016 (0.2361)	-0.0373 (0.0892)	-0.0647 (0.1335)	0.1344 (0.0397)	0.1255 (0.0481)	0.1513 (0.2999)	0.1846 (0.3814)	0.5968 (1.2467)	0.6812 (1.5589)	0.6355 (1.5142)	0.7421 (1.9243)	0.728 (0.0871)	0.7161 (0.1719)	0.0463 (0.1346)	0.0544 (0.1797
	1	0.6011 (0.1735)	0.6011 (0.1373)	0.6773 (0.3296)	0.7012 (0.2375)	-0.0754 (0.1073)	-0.107 (0.1405)	0.1352 (0.0399)	0.1254 (0.048)	0.0856 (0.2297)	0.0733 (0.3274)	0.3035 (0.9853)	0.2408 (1.4491)	0.333 (1.1961)	0.2785 (1.7643)	0.7098 (0.1062)	0.699 (0.1716)	0.0021 (0.1291)	0.0056 (0.1838
	5	0.5549 (0.1241)	0.548 (0.127)	0.5593 (0.3807)	0.6182 (0.331)	-0.3507 (0.2264)	-0.4051 (0.1806)	0.1495 (0.0442)	0.1351 (0.0435)	-0.327 (0.4217)	-0.3214 (0.2881)	-2.175 (2.7708)	-1.9933 (2.4338)	-2.5821 (2.9199)	-2.4964 (3.0858)	0.4351 (0.2347)	0.3927 (0.17)	-0.3128 (0.1974)	-0.336 (0.2175

[a] *Dataset I* comprises of the financial variables and technical indicators while *Dataset II* contains both *Dataset I* and the candlestick chart (price) patterns as described in Table S.V.1. *MRAR* stands for Morningstar's risk-adjusted rating.
[b] The initial endowment is $1000.
[c] The IQR values are shown in parentheses.
[d] There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

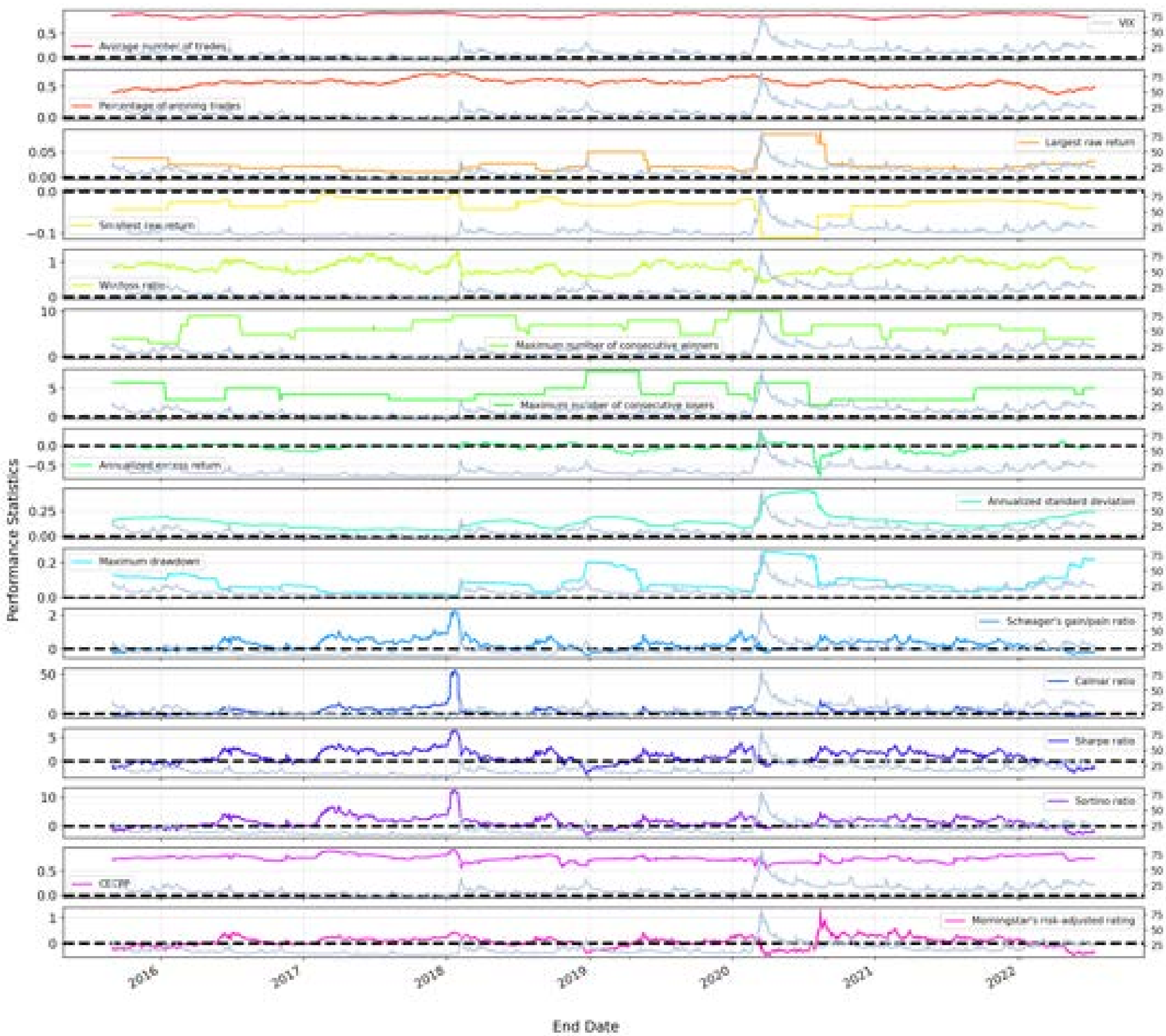
Figure S.IX.5: The performance statistics of 1718 trading sessions with the fixed transaction cost strategy investing in *SPY* for 100 days based on one-day ahead forecasts by a LGBM model (trained with the *As1* loss function, then cross-validated with the *Sharpe ratio* scoring function) using the data on financial variables and technical indicators (*Dataset I*) described in Table S.V.1



[a] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

[b] The horizonal axis contains the end dates of the trading sessions.

Figure S.IX.6: The performance statistics of 1718 trading sessions with the fixed transaction cost strategy investing in *SPY* for 100 days based on one-day ahead forecasts by a LGBM model (trained with the *As1* loss function, then cross-validated with the *Sharpe ratio* scoring function) using the data on financial variables, technical indicators, and candlestick chart patterns (*Dataset II*) described in Table S.V.1



[a] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

[b] The horizonal axis contains the end dates of the trading sessions.

Table S.IX.2: The median (and IQR) of all the performance metrics of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the LGBM model trained with the *As1* loss function, then cross-validated with the *Sharpe ratio* scoring function for different amounts of transaction cost across various holding periods (in days)

Holding period	Trans. cost	Percent. of win. trades		Win/loss ratio		Annualized excess ret.		Annualized std. dev.		Schwager's gain/pain ratio		Sharpe ratio		Sortino ratio		CECPP		MRAR	
		Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset
100	0.05	0.5618 (0.1248)	0.561 (0.0814)	0.8842 (0.2005)	0.849 (0.2197)	-0.0278 (0.1028)	-0.0268 (0.0717)	0.1303 (0.0659)	0.1319 (0.0705)	0.267 (0.4829)	0.2074 (0.4305)	1.2038 (2.1317)	0.9203 (1.9225)	1.4166 (2.6951)	1.0977 (2.2837)	0.7624 (0.0867)	0.7585 (0.0645)	0.1152 (0.2657)	0.0918 (0.2401
	0.1	0.5618 (0.1248)	0.5604 (0.0814)	0.8817 (0.2015)	0.8454 (0.2169)	-0.0298 (0.1022)	-0.0293 (0.0733)	0.1303 (0.0659)	0.1319 (0.0705)	0.2613 (0.4776)	0.2022 (0.4267)	1.1807 (2.1182)	0.8972 (1.9172)	1.386 (2.6903)	1.067 (2.2828)	0.7611 (0.0874)	0.7577 (0.0643)	0.1126 (0.2649)	0.0887 (0.2408
	0.5	0.5568 (0.1301)	0.5542 (0.0803)	0.8614 (0.2254)	0.845 (0.2242)	-0.051 (0.1002)	-0.055 (0.0822)	0.1304 (0.0658)	0.1321 (0.07)	0.2172 (0.4373)	0.1596 (0.3995)	0.9925 (2.0198)	0.7054 (1.8518)	1.166 (2.6017)	0.8532 (2.2317)	0.7503 (0.0946)	0.7479 (0.0641)	0.0908 (0.2584)	0.0652 (0.2394
	1	0.5506 (0.1281)	0.5506 (0.0833)	0.8497 (0.2343)	0.8286 (0.2296)	-0.0742 (0.1021)	-0.079 (0.0916)	0.1307 (0.0656)	0.1325 (0.0689)	0.1653 (0.3858)	0.1123 (0.3659)	0.7527 (1.8657)	0.4875 (1.7687)	0.9064 (2.4276)	0.5969 (2.1499)	0.7251 (0.1077)	0.7301 (0.0707)	0.0613 (0.249)	0.0359 (0.2335
	5	0.5169 (0.1191)	0.5181 (0.0693)	0.7106 (0.2162)	0.7045 (0.1969)	-0.2629 (0.1566)	-0.2616 (0.1707)	0.1417 (0.0622)	0.1424 (0.0661)	-0.1839 (0.2022)	-0.1831 (0.2075)	-1.1471 (1.3507)	-1.1751 (1.3952)	-1.3964 (1.5621)	-1.4076 (1.8382)	0.3873 (0.2648)	0.3907 (0.3094)	-0.1589 (0.1887)	-0.1873 (0.2201
200	0.05	0.5738 (0.1135)	0.5681 (0.0725)	0.8735 (0.2006)	0.8274 (0.2016)	-0.0225 (0.0614)	-0.0323 (0.0646)	0.1397 (0.0518)	0.1382 (0.0516)	0.2392 (0.3272)	0.2142 (0.2726)	1.0366 (1.4298)	0.9677 (1.245)	1.2135 (1.7746)	1.1372 (1.5133)	0.7606 (0.0606)	0.7356 (0.0582)	0.116 (0.1878)	0.0901 (0.1796
	0.1	0.5738 (0.1135)	0.5667 (0.0705)	0.8707 (0.1984)	0.8251 (0.1945)	-0.0253 (0.0615)	-0.0351 (0.0654)	0.1398 (0.0519)	0.1382 (0.0517)	0.2334 (0.3244)	0.2086 (0.27)	1.0246 (1.4231)	0.9455 (1.24)	1.2016 (1.7619)	1.1132 (1.4962)	0.7597 (0.0606)	0.7352 (0.0582)	0.113 (0.1853)	0.0868 (0.179)
	0.5	0.5683 (0.1154)	0.5614 (0.07)	0.8635 (0.214)	0.8218 (0.1952)	-0.0459 (0.0652)	-0.0563 (0.0675)	0.14 (0.052)	0.1382 (0.0514)	0.1965 (0.3078)	0.1579 (0.2527)	0.8878 (1.3561)	0.7307 (1.1774)	1.0402 (1.6867)	0.9009 (1.4158)	0.7497 (0.0674)	0.7296 (0.0581)	0.0911 (0.1758)	0.0619 (0.1773
	1	0.5637 (0.1129)	0.5556 (0.0763)	0.8498 (0.202)	0.8189 (0.2076)	-0.0716 (0.0695)	-0.0822 (0.0655)	0.1403 (0.0521)	0.1381 (0.0511)	0.1474 (0.2863)	0.1082 (0.2304)	0.6874 (1.2916)	0.487 (1.1318)	0.815 (1.5768)	0.5885 (1.3282)	0.729 (0.0814)	0.7185 (0.0478)	0.0624 (0.1682)	0.0341 (0.1752
	5	0.5249 (0.1092)	0.5233 (0.061)	0.7345 (0.1654)	0.6628 (0.1597)	-0.2658 (0.1259)	-0.29 (0.0962)	0.1499 (0.0492)	0.1491 (0.0523)	-0.1666 (0.144)	-0.1781 (0.1763)	-1.0233 (0.9011)	-1.0979 (1.1878)	-1.2138 (1.0077)	-1.3148 (1.5427)	0.4285 (0.2497)	0.4346 (0.2297)	-0.1677 (0.1471)	-0.2013 (0.1563

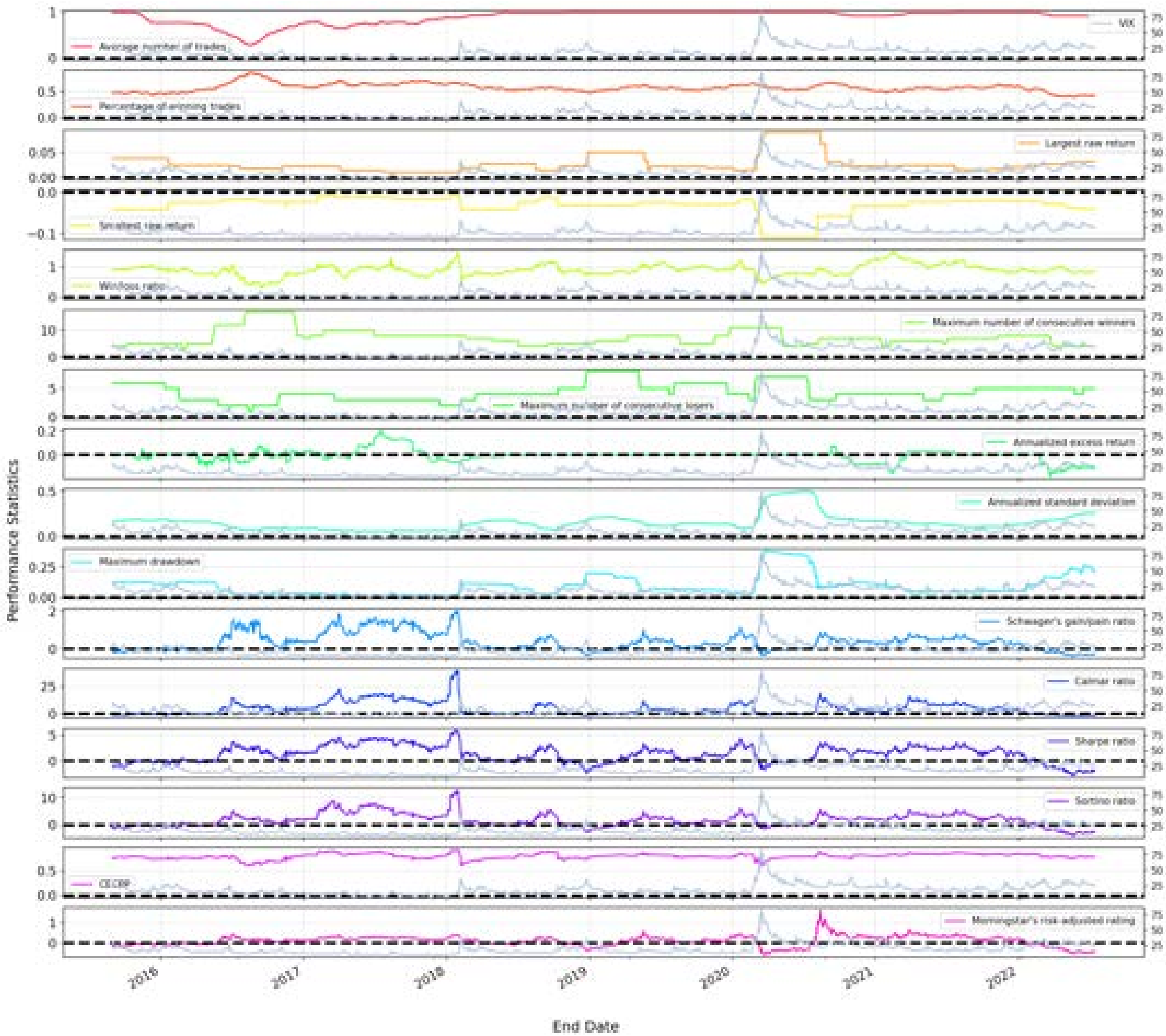
[a] *Dataset I* comprises of the financial variables and technical indicators while *Dataset II* contains both *Dataset I* and the candlestick chart (price) patterns as described in Table S.V.1. *MRAR* stands for Morningstar's risk-adjusted rating.

[b] The initial endowment is $1000.

[c] The IQR values are shown in parentheses.

[d] There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

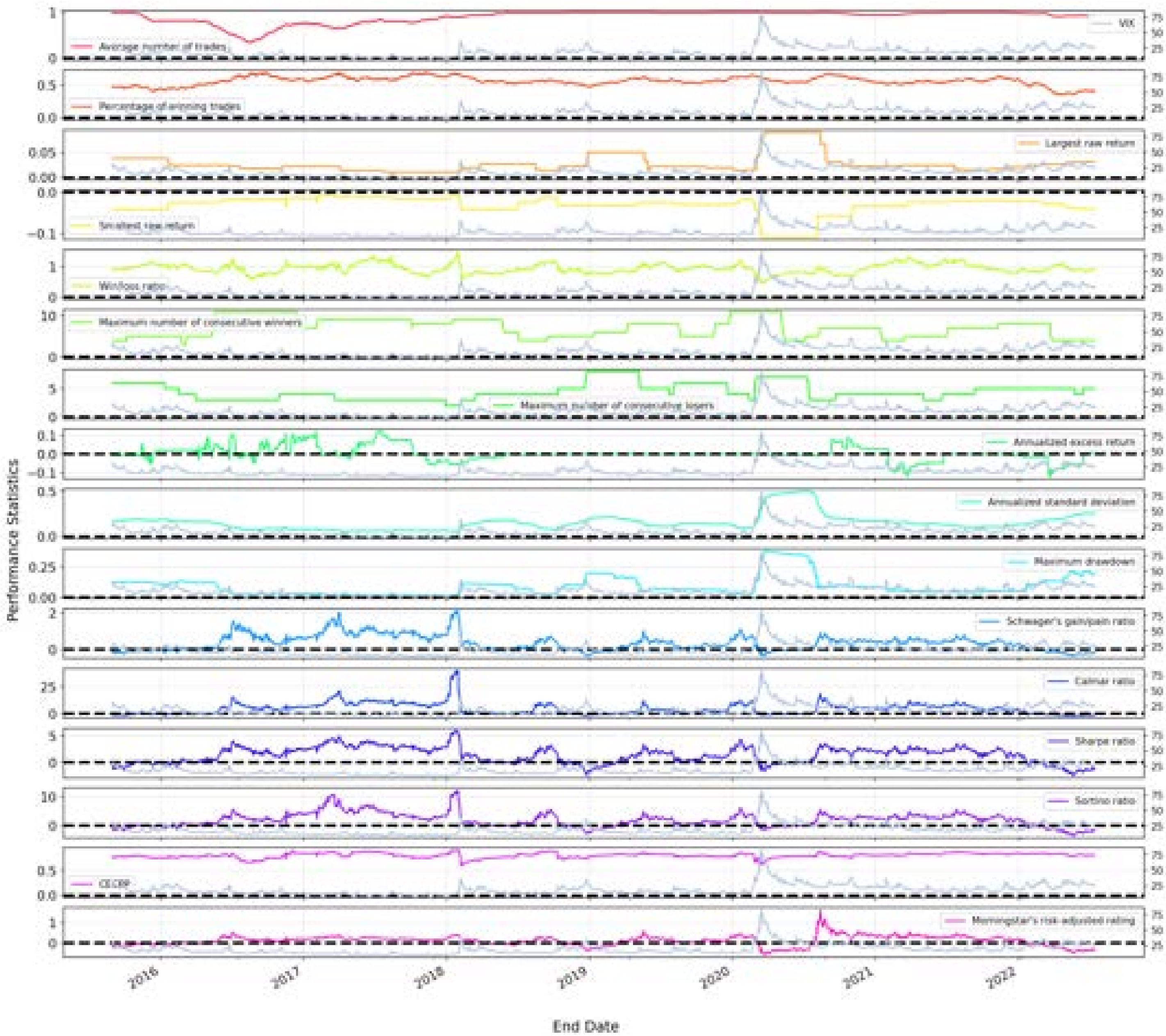
Figure S.IX.7: The performance statistics of 1718 trading sessions with the fixed transaction cost strategy investing in *SPY* for 100 days based on one-day ahead forecasts by a RF model (cross-validated with the *Sharpe ratio* scoring function) using the data on financial variables and technical indicators (*Dataset I*) described in Table S.V.1



[a] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

[b] The horizonal axis contains the end dates of the trading sessions.

Figure S.IX.8: The performance statistics of 1718 trading sessions with the fixed transaction cost strategy investing in *SPY* for 100 days based on one-day ahead forecasts by a RF model (cross-validated with the *Sharpe ratio* scoring function) using the data on financial variables, technical indicators, and candlestick chart patterns (*Dataset II*) described in Table S.V.1



[a] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

[b] The horizonal axis contains the end dates of the trading sessions.

Table S.IX.3: The median (and IQR) of all the performance metrics of trading sessions with the fixed transaction cost strategy investing in *SPY* based on one-day ahead forecasts by the RF model cross-validated with the *Sharpe ratio* scoring function for different amounts of transaction cost across various holding periods (in days)

Holding period	Trans. cost	Percent. of win. trades		Win/loss ratio		Annualized excess ret.		Annualized std. dev.		Schwager's gain/pain ratio		Sharpe ratio		Sortino ratio		CECPP		MRAR	
		Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset
100	0.05	0.567 (0.0918)	0.567 (0.0758)	0.899 (0.2127)	0.9265 (0.2145)	-0.0 (0.0156)	-0.0 (0.0186)	0.1401 (0.0815)	0.1397 (0.0833)	0.3102 (0.5742)	0.3362 (0.5717)	1.3977 (2.4593)	1.535 (2.3288)	1.6326 (3.1202)	1.8695 (3.108)	0.7933 (0.0615)	0.8016 (0.0585)	0.1462 (0.3033)	0.1565 (0.2787
	0.1	0.567 (0.0918)	0.567 (0.0758)	0.897 (0.2086)	0.9258 (0.2132)	-0.0 (0.0174)	-0.0 (0.0183)	0.1401 (0.0814)	0.1397 (0.0833)	0.3069 (0.5734)	0.3352 (0.5623)	1.3694 (2.4405)	1.51 (2.3199)	1.6011 (3.1102)	1.8585 (3.0969)	0.7922 (0.0612)	0.8007 (0.058)	0.145 (0.3004)	0.1546 (0.2774
	0.5	0.5612 (0.087)	0.5625 (0.0763)	0.8913 (0.2025)	0.9205 (0.2237)	-0.0 (0.0395)	-0.0039 (0.0276)	0.1401 (0.0812)	0.1397 (0.0833)	0.2776 (0.5501)	0.3139 (0.5152)	1.2448 (2.2974)	1.3724 (2.2058)	1.4435 (2.9823)	1.6775 (2.9267)	0.7848 (0.0576)	0.7932 (0.0558)	0.1229 (0.283)	0.1334 (0.2726
	1	0.5612 (0.0805)	0.5591 (0.0716)	0.889 (0.2038)	0.9098 (0.209)	-0.006 (0.0661)	-0.0098 (0.0479)	0.1403 (0.0801)	0.1399 (0.083)	0.231 (0.4941)	0.2707 (0.4634)	1.039 (2.1669)	1.1949 (2.0998)	1.2159 (2.7864)	1.4122 (2.7638)	0.7683 (0.0581)	0.7806 (0.0529)	0.0984 (0.27)	0.1054 (0.2675
	5	0.551 (0.0713)	0.5417 (0.064)	0.8411 (0.2569)	0.8626 (0.2571)	-0.0365 (0.2723)	-0.0709 (0.1929)	0.1438 (0.0721)	0.1429 (0.0759)	-0.0377 (0.4528)	-0.0391 (0.4598)	-0.2484 (2.4719)	-0.2585 (2.5271)	-0.3056 (3.013)	-0.3339 (3.2527)	0.4627 (0.2682)	0.4587 (0.2874)	-0.0906 (0.3188)	-0.0838 (0.3443
200	0.05	0.5729 (0.064)	0.5736 (0.0494)	0.8618 (0.1855)	0.8965 (0.1524)	-0.0 (0.0234)	-0.0006 (0.0252)	0.1377 (0.0465)	0.1383 (0.0469)	0.2217 (0.4063)	0.2249 (0.4053)	0.9458 (1.8189)	0.9759 (1.8913)	1.0736 (2.2719)	1.1277 (2.5637)	0.7772 (0.0578)	0.7881 (0.0522)	0.1079 (0.2332)	0.1112 (0.2199
	0.1	0.5707 (0.0646)	0.5736 (0.0494)	0.8608 (0.1844)	0.8937 (0.1513)	-0.0 (0.0242)	-0.0007 (0.0241)	0.1378 (0.0465)	0.1383 (0.0469)	0.2193 (0.4036)	0.2235 (0.4043)	0.9365 (1.7926)	0.9696 (1.8832)	1.0687 (2.2528)	1.1167 (2.5548)	0.7765 (0.0566)	0.7875 (0.0519)	0.1065 (0.2304)	0.1101 (0.216)
	0.5	0.5707 (0.0635)	0.5685 (0.0463)	0.8585 (0.1745)	0.887 (0.1535)	-0.0028 (0.0379)	-0.0041 (0.0179)	0.138 (0.0465)	0.1385 (0.047)	0.2022 (0.3732)	0.2066 (0.3856)	0.8859 (1.6337)	0.8951 (1.7668)	0.9925 (2.0459)	1.0365 (2.3883)	0.771 (0.0493)	0.7803 (0.0503)	0.0961 (0.2069)	0.0988 (0.1956
	1	0.5707 (0.0579)	0.5685 (0.0444)	0.8607 (0.1825)	0.8815 (0.157)	-0.0235 (0.0522)	-0.0179 (0.0344)	0.1385 (0.0464)	0.1384 (0.047)	0.1821 (0.3226)	0.1846 (0.3424)	0.7981 (1.5078)	0.814 (1.5847)	0.9085 (1.9246)	0.9334 (2.1666)	0.7566 (0.0446)	0.7661 (0.0428)	0.0788 (0.188)	0.085 (0.1803
	5	0.5645 (0.0442)	0.5533 (0.0381)	0.828 (0.2831)	0.8381 (0.2301)	-0.1125 (0.2866)	-0.0976 (0.2352)	0.1459 (0.0469)	0.1431 (0.0492)	0.0154 (0.341)	0.0148 (0.3078)	-0.0067 (1.872)	-0.0022 (1.6402)	-0.0074 (2.255)	-0.003 (2.0716)	0.4863 (0.2136)	0.4929 (0.2161)	-0.0676 (0.2386)	-0.063 (0.2176

[a] *Dataset I* comprises of the financial variables and technical indicators while *Dataset II* contains both *Dataset I* and the candlestick chart (price) patterns as described in Table S.V.1.

[b] The initial endowment is $1000.

[c] The IQR values are shown in parentheses.

[d] There are 1718 possible trading sessions of length 100 days or 1618 possible trading sessions of length 200 days.

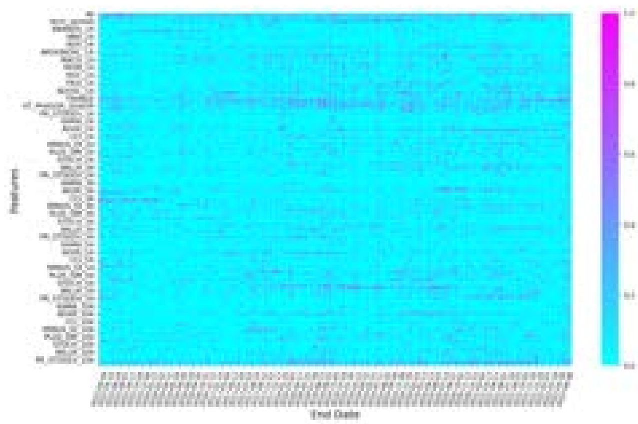
S.X Important Predictors

We study the contribution of each predictor to the performance of our trading strategies across trading sessions. Figures S.X.1 and S.X.3 show the heat maps (over all rolling windows) of the global Shapley values of top predictors from Dataset I [which are used to produce one-day ahead forecasts of the moving directions of *SPY* prices] selected by a LGBM model. This model is trained with either the *Brier* or *As1* loss function and cross-validated with one of the following four scoring functions: *AUC*, *Schwager's gain/pain ratio*, *Sharpe ratio*, and *CECPP* (where the first is a statistical scoring function while the last three are economic scoring functions). The top ten predictors are: 1) dividend/price ratio (*dp*), 2) term spread (*term_spread*), 3) Bollinger bands with 14-day MAs (*BBANDS_14*), 4) 14-day simple MA (*SMA_14*), 5) 14-day average directional movement index (*ADX_14*), 6) 14-day Aroon oscillator (*AROONOSC_14*), 7) 14-day MA convergence/divergence (*MACD_14*), 8) 14-day momentum (*MOM_14*), 9) 14-day rate of change (*ROC_14*), 10) the 1-day ROC of a triple smooth 14-day exponential MA (*TRIX_14*) and the Chaikin accumulation/distribution line oscillator based on the last 14 days (*ADOSC_14*). There is a significant amount of evidence documenting the predictive power of *dp* and *term_spread* in the existing literature [see, e.g., Rapach and Zhou (2022)]. While existing works primarily focus on technical indicators based on MAs of historical prices or trading volumes [e.g., Brock et al. (1992); Lin (2018)], we find that other technical indicators, such as *MOM* or *ADX*, also have a strong predictability for the direction of aggregate market price movements. Technical indicators constructed from observations of past 14 days are stronger predictors than those using observations longer than 14 days (i.e., recent histories are more relevant).

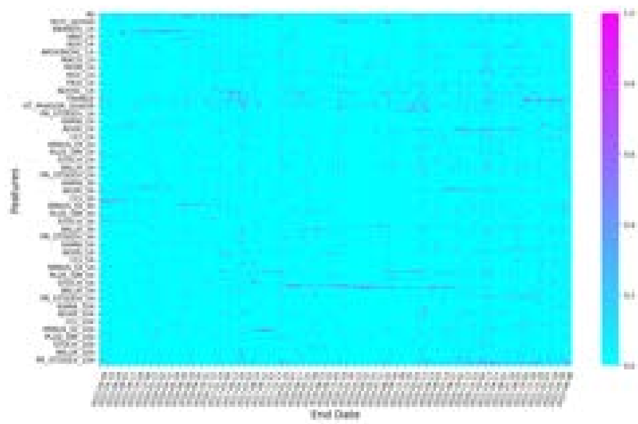
Figures S.X.2 and S.X.4 show the heat maps (over all rolling windows) of the global Shapley values of top predictors from Dataset II [which are used to produce one-day ahead forecasts of the moving directions of *SPY* prices] selected by a LGBM model. This model is trained with either the *Brier* or *As1* loss function and cross-validated with one of the following four scoring functions: *AUC*, *Schwager's gain/pain ratio*, *Sharpe ratio*, and *CECPP*. The top ten predictors are: 1) dividend/price ratio (*dp*), 2) earning per share (*EPS*), 3) 14-day exponential MA (*EMA_14*), 4) 14-day parabolic stop and reverse (*SAR_14*), 5) 14-day Aroon oscillator (*AROONOSC_14*), 6) money flow index over the last 14 days (*MFI_14*), 7) the plus directional movement index over the last 14 days (*PLUS_DM_14*), 8) the stochastic relative strength index based on the last 14 days (*STOCHRSI_14*), 9) the Chaikin accumulation/distribution line oscillator based on the last 14 days (*ADOSC_14*), 10) the Hilbert transform - dominant cycle period (*HT_DCPERIOD*) and the Hilbert transform - sinewave (*HT_SINE_UPPER*). Candlestick chart patterns that have some predictability for the price moving direction include: 1) the three advancing white soldiers pattern (*3WHITESOLDIERS*), 2) the closing Marubozu pattern (*CLOSING-MARUBOZU*), 3) the Doji Star pattern (*DOJISTAR*), 4) the Up/Down-gap Side-by-Side White Lines pattern (*GAPSIDESIDEWHITE*), 5) The Harami Cross pattern (*HARAMICROSS*), 6) the Identical Three Crows pattern (*IDENTICAL3CROWS*), 7) the Ladder Bottom pattern (*LADDERBOT-*

TOM), 8) the Mat Hold pattern (*MATHOLD*), 9) the Rickshaw Man pattern (*RICKSHAWMAN*), 10) the Stalled pattern (*STALLEDPATTERN*), 11) the Tristar pattern (*TRISTAR*). As suggested in Table S.V.1, *MATHOLD* (with the performance rank 86 out of 103) and *STALLEDPATTERN* (with the performance rank 93 out of 103) are among the best performing candlestick chart patterns suggested in the literature. We have also obtained same set of top predictors by applying RF [which we do not report here].

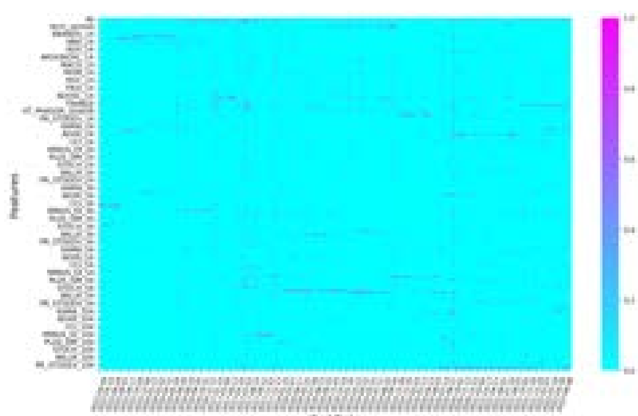
Figure S.X.1: The heat map of the global Shapley values of top predictors (used to forecast *one-day ahead* the moving direction of the *SPY* price) selected by LGBM models trained with the *Brier* loss function and cross-validated with four different scoring functions [based on the fixed transaction cost strategy] for the data on financial variables and technical indicators (*Dataset I*) described in Table S.V.1



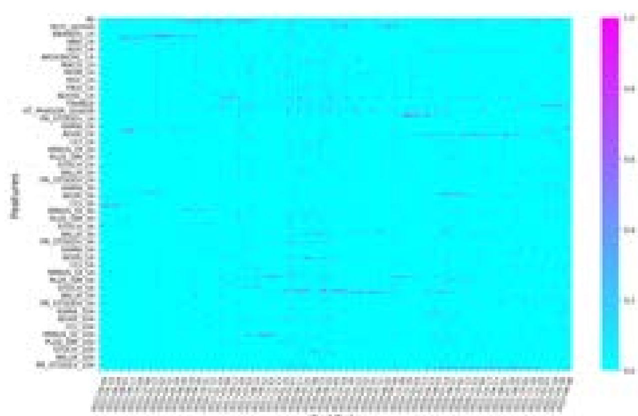
(a) *AUC* score



(b) *Schwager's (2012) gain to pain ratio* score



(c) *Sharpe ratio* score



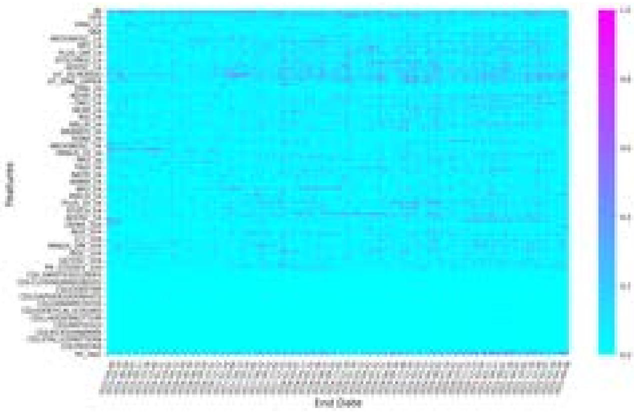
(d) *CECPP* score

[a] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

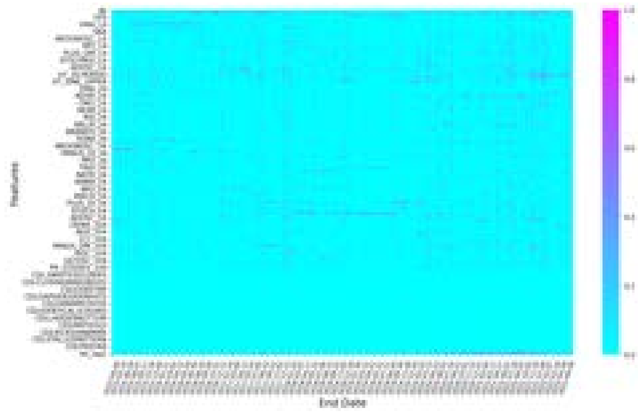
[b] The stratified K-Folds cross-validation with the number of folds = 5 is used to cross-validate a LGBM model.

[c] The horizonal axis contains the end dates of the rolling windows.

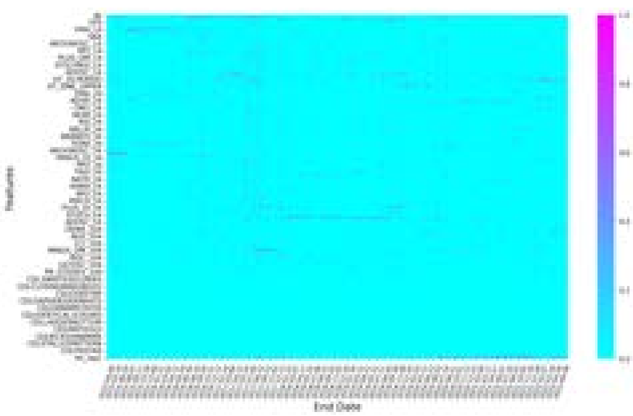
Figure S.X.2: The heat map of the global Shapley values of top predictors (used to forecast *one-day ahead* the moving direction of the *SPY* price) selected by LGBM models trained with the *Brier* loss function and cross-validated with four different scoring functions [based on the fixed transaction cost strategy] for the data on financial variables, technical indicators, and candlestick chart patterns (*Dataset II*) described in Table S.V.1



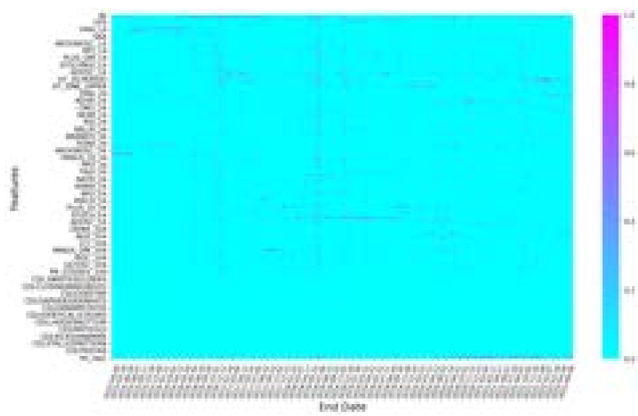
(a) *AUC* score



(b) *Schwager's (2012) gain to pain ratio* score



(c) *Sharpe ratio* score



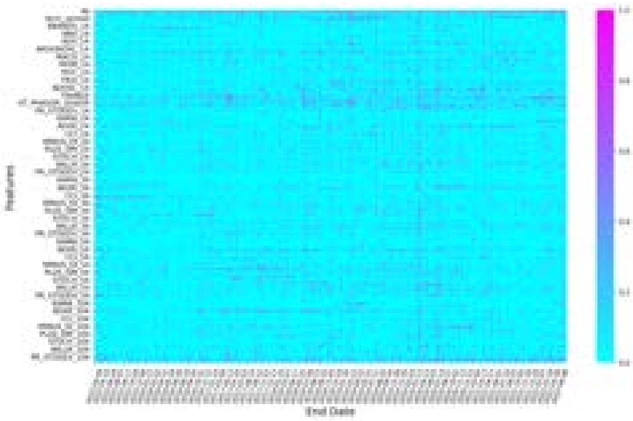
(d) *CECPP* score

[a] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

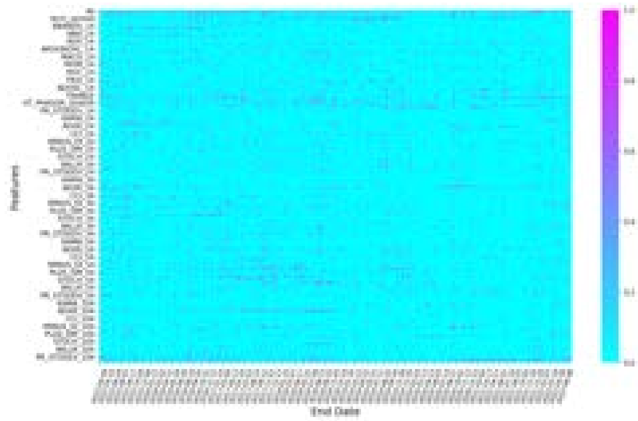
[b] The stratified K-Folds cross-validation with the number of folds = 5 is used to cross-validate a LGBM model.

[c] The horizonal axis contains the end dates of the rolling windows.

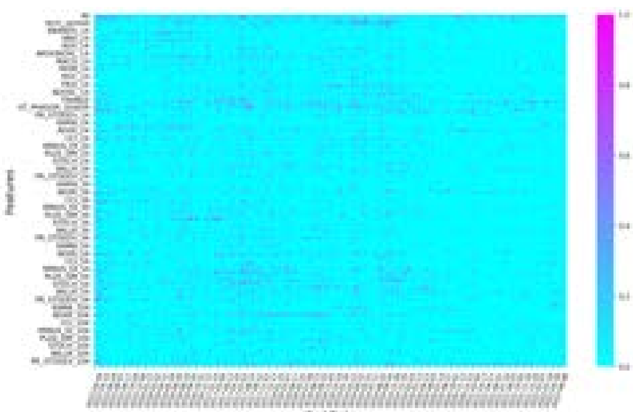
Figure S.X.3: The heat map of the global Shapley values of top predictors (used to forecast *one-day ahead* the moving direction of the *SPY* price) selected by LGBM models trained with the *As1* loss function and cross-validated with four different scoring functions [based on the fixed transaction cost strategy] for the data on financial variables and technical indicators (*Dataset I*) described in Table S.V.1



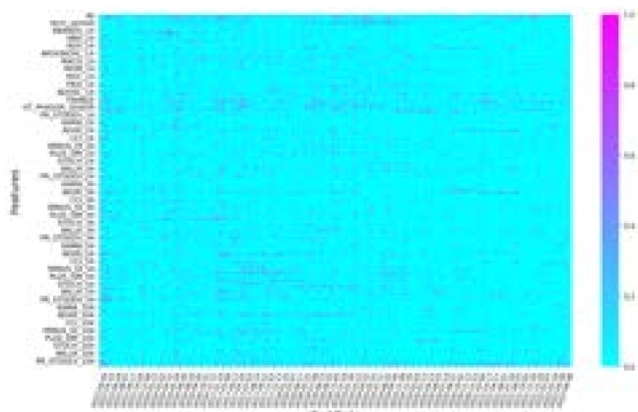
(a) *AUC* score



(b) *Schwager's (2012) gain to pain ratio* score



(c) *Sharpe ratio* score



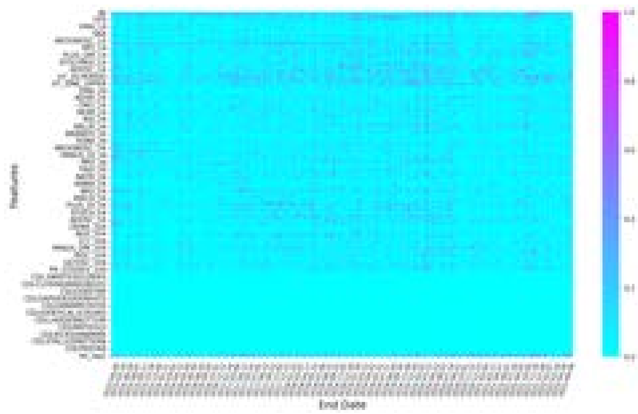
(d) *CECPP* score

[a] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

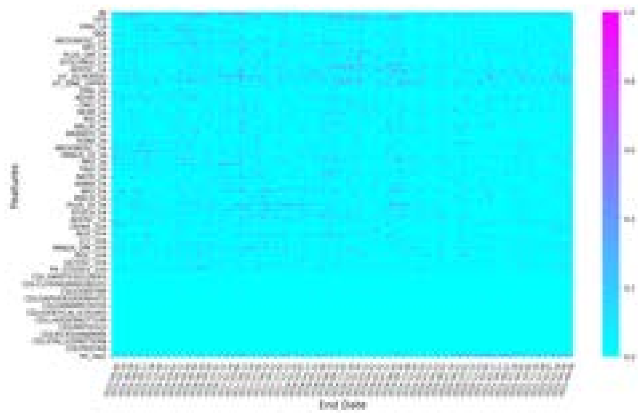
[b] The stratified K-Folds cross-validation with the number of folds = 5 is used to cross-validate a LGBM model.

[c] The horizonal axis contains the end dates of the rolling windows.

Figure S.X.4: The heat map of the global Shapley values of top predictors (used to forecast *one-day ahead* the moving direction of the *SPY* price) selected by LGBM models trained with the *As1* loss function and cross-validated with four different scoring functions [based on the fixed transaction cost strategy] for the data on financial variables, technical indicators, and candlestick chart patterns (*Dataset II*) described in Table S.V.1

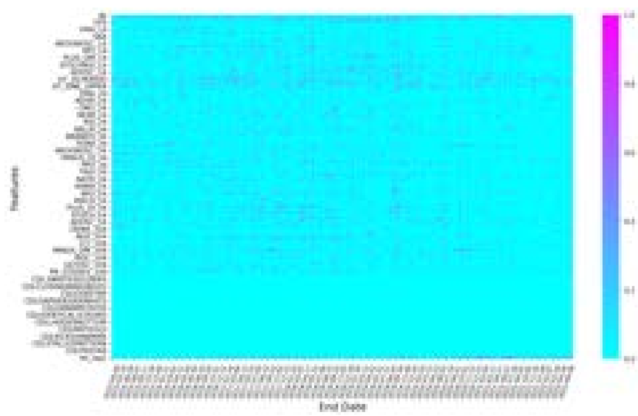


(a) *AUC* score



(b) *Schwager's (2012) gain to pain ratio* score

(c) *Sharpe ratio* score



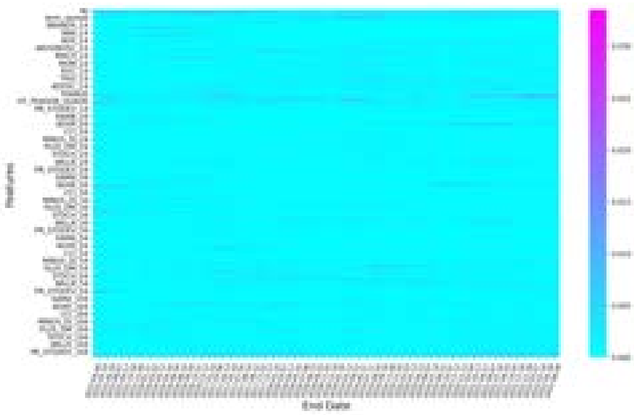
(d) *CECPP* score

[a] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

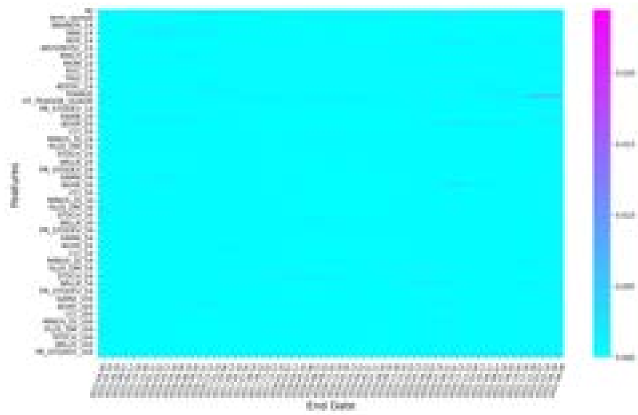
[b] The stratified K-Folds cross-validation with the number of folds = 5 is used to cross-validate a LGBM model.

[c] The horizonal axis contains the end dates of the rolling windows.

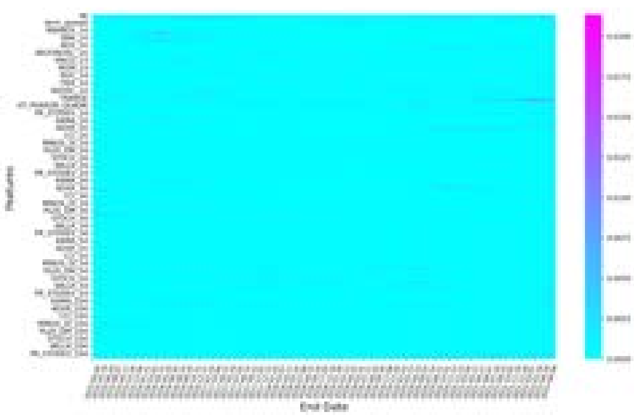
Figure S.X.5: The heat map of the global Shapley values of top predictors (used to forecast *one-day ahead* the moving direction of the *SPY* price) selected by RF models cross-validated with four different scoring functions [based on the fixed transaction cost strategy] for the data on financial variables and technical indicators (*Dataset I*) described in Table S.V.1



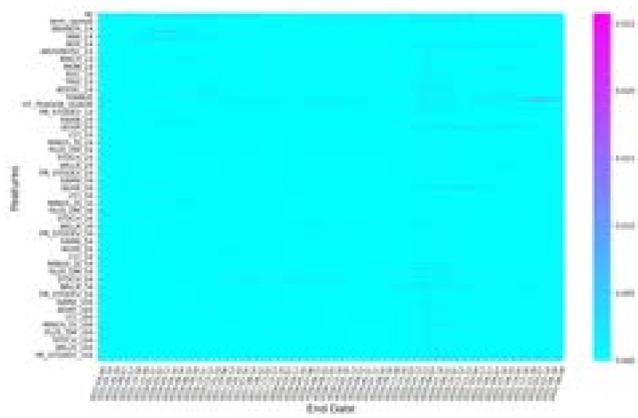
(a) *AUC* score



(b) *Schwager's (2012) gain to pain ratio* score



(c) *Sharpe ratio* score



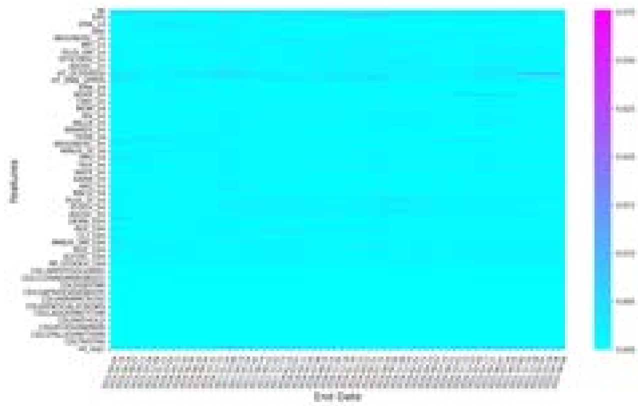
(d) *CECPP* score

[a] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

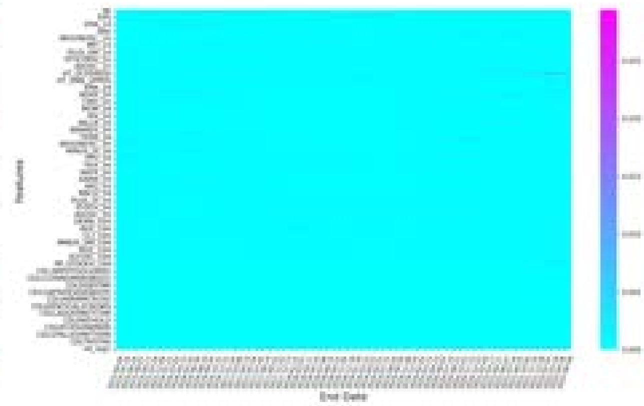
[b] The stratified K-Folds cross-validation with the number of folds = 5 is used to cross-validate a LGBM model.

[c] The horizonal axis contains the end dates of the rolling windows.

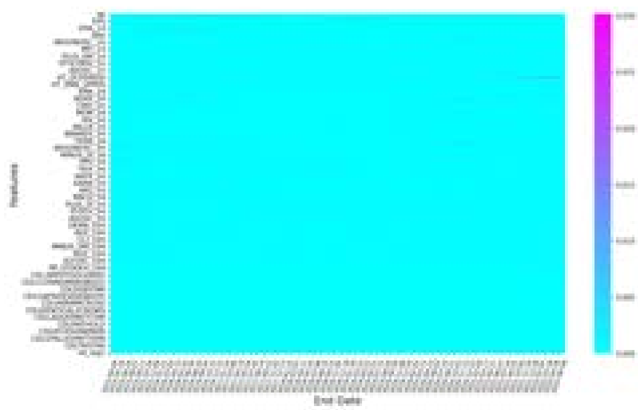
Figure S.X.6: The heat map of the global Shapley values of top predictors (used to forecast *one-day ahead* the moving direction of the *SPY* price) selected by RF models cross-validated with four different scoring functions [based on the fixed transaction cost strategy] for the data on financial variables, technical indicators, and candlestick chart patterns (*Dataset II*) described in Table S.V.1



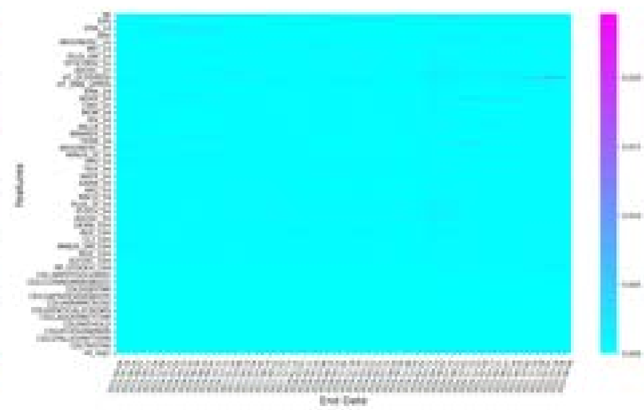
(a) *AUC* score



(b) *Schwager's (2012) gain to pain ratio* score



(c) *Sharpe ratio* score



(d) *CECPP* score

[a] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

[b] The stratified K-Folds cross-validation with the number of folds = 5 is used to cross-validate a LGBM model.

[c] The horizonal axis contains the end dates of the rolling windows.

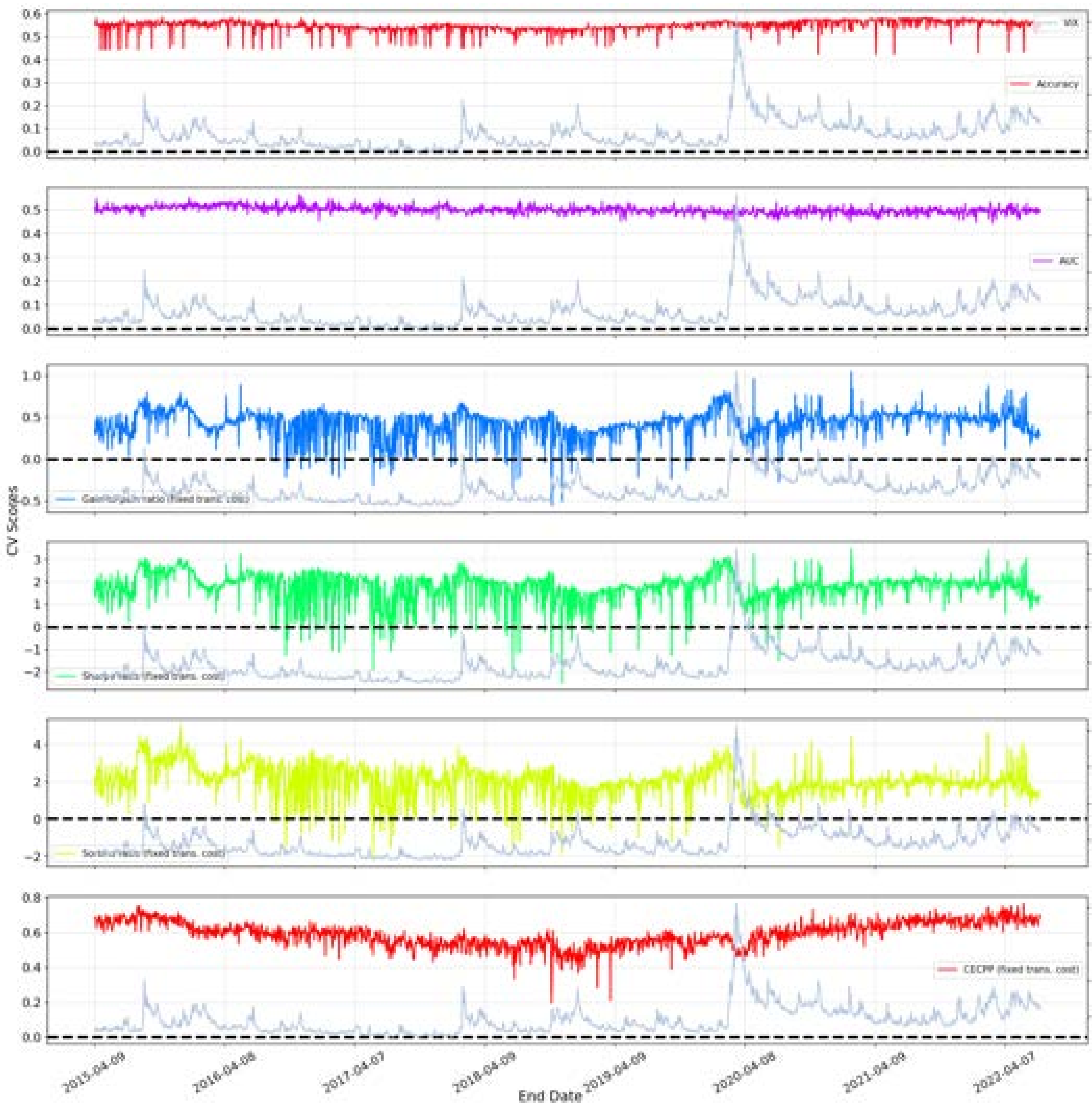
S.XI Cross-Validation Scores

To examine whether or not the ML algorithms used to predict the moving direction of future prices overfit on in-sample data, we also report the average values of the scoring functions evaluated with validation subsamples in every rolling window. A model is overfitting when the average value of a scoring function evaluated with validation samples is much greater than the value of this scoring function evaluated with test data and OoS forecasts.

Figures S.XI.1 and S.XI.2 line-plot the average scoring functions evaluated with validation sub-samples (namely, the in-sample average scoring function) by a LGBM model trained with the *Brier* loss function and cross-validated with the *CECPP* scoring function (based on the fixed transaction cost strategy) for Dataset I and Dataset II respectively. First of all, we can notice that the *Accuracy* and *AUC* scores are a little above 50% (which are too low for a model to be considered to overfit the training data). The in-sample average *gain/pain ratio* values are well below 2.0, which are comparable with the OoS values reported in Figures S.IX.3 and S.IX.4. The in-sample average *Sharpe ratio* and *Sortino ratio* values are well below 4.0, which are comparable with the OoS *Sharpe ratio* and *Sortino ratio* values. The in-sample average *CECPP* values are mostly somewhere in between 0.4 and 0.8, which are also comparable with the OoS *CECPP* values.

Reviewing the line plots of the in-sample average scoring functions calculated from RF models cross-validated with the *CECPP* scoring function for Dataset I and Dataset II (i.e., Figures S.XI.3 and S.XI.4), we find no evidence of overfitting for the RF models used to forecast the moving direction of future price in every trading session.

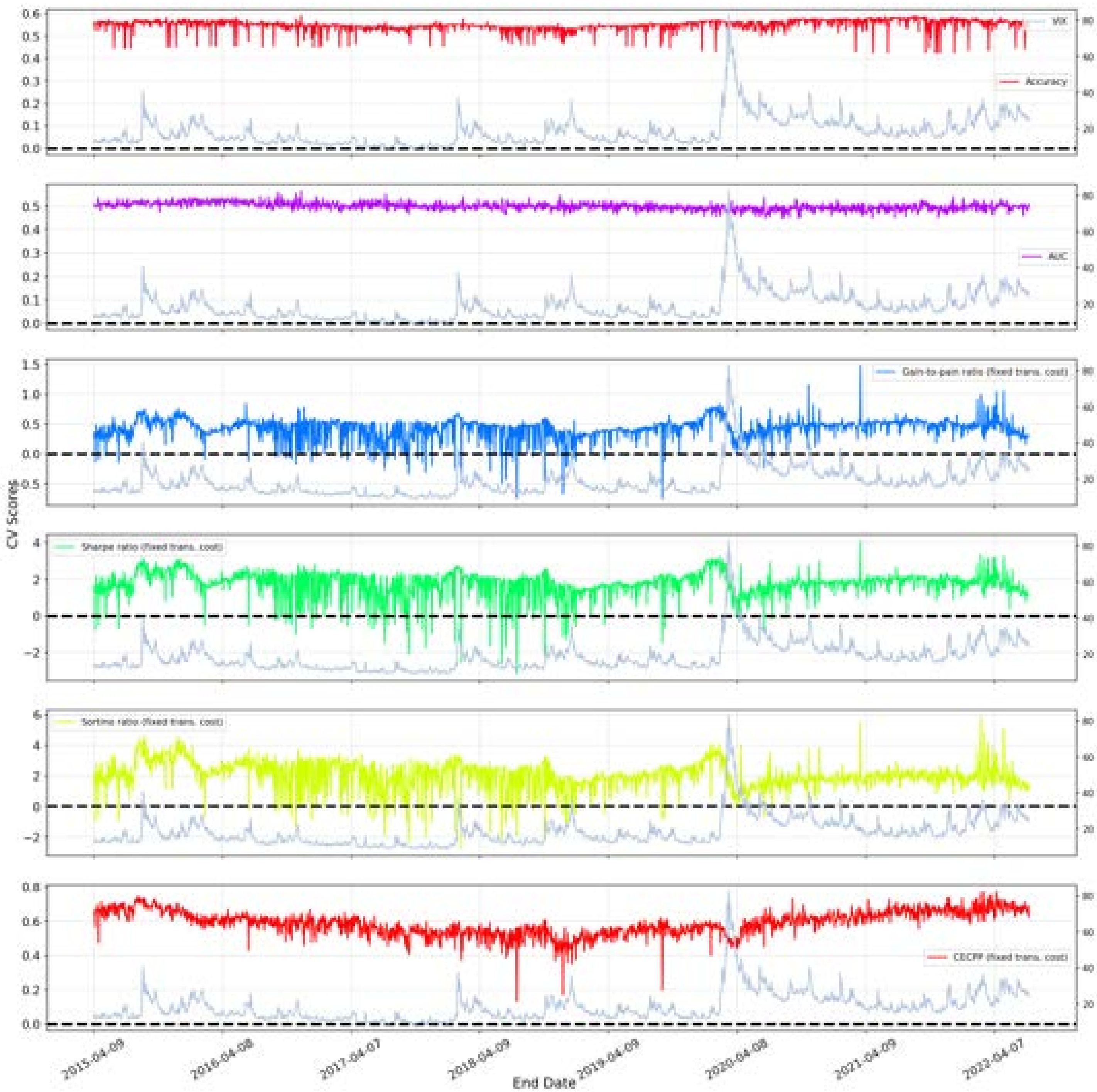
Figure S.XI.1: The average scoring functions (based on the fixed transaction cost strategy investing in *SPY*) evaluated on forecasts of validation samples made by a LGBM model trained with the *Brier* loss function and cross-validated with the *CECPP* scoring function using the data on financial variables and technical indicators (*Dataset I*) described in Table S.V.1



[a] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

[b] The horizonal axis contains the end dates of rolling windows.

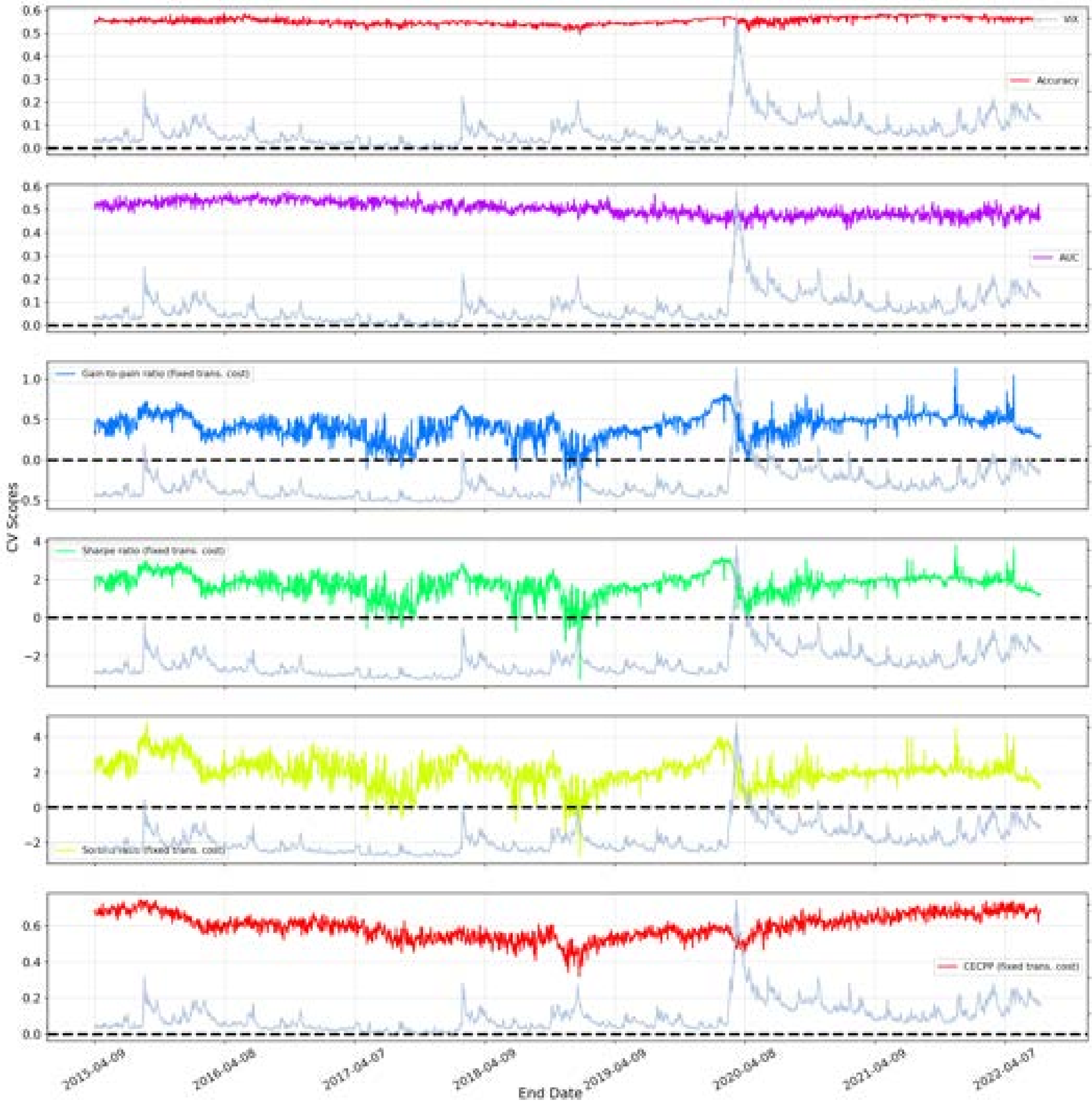
Figure S.XI.2: The average scoring functions (based on the fixed transaction cost strategy investing in *SPY*) evaluated on forecasts of validation samples made by a LGBM model trained with the *Brier* loss function and cross-validated with the *CECPP* scoring function using the data on financial variables, technical indicators, and candlestick chart patterns (*Dataset II*) described in Table S.V.1



[a] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

[b] The horizonal axis contains the end dates of rolling windows.

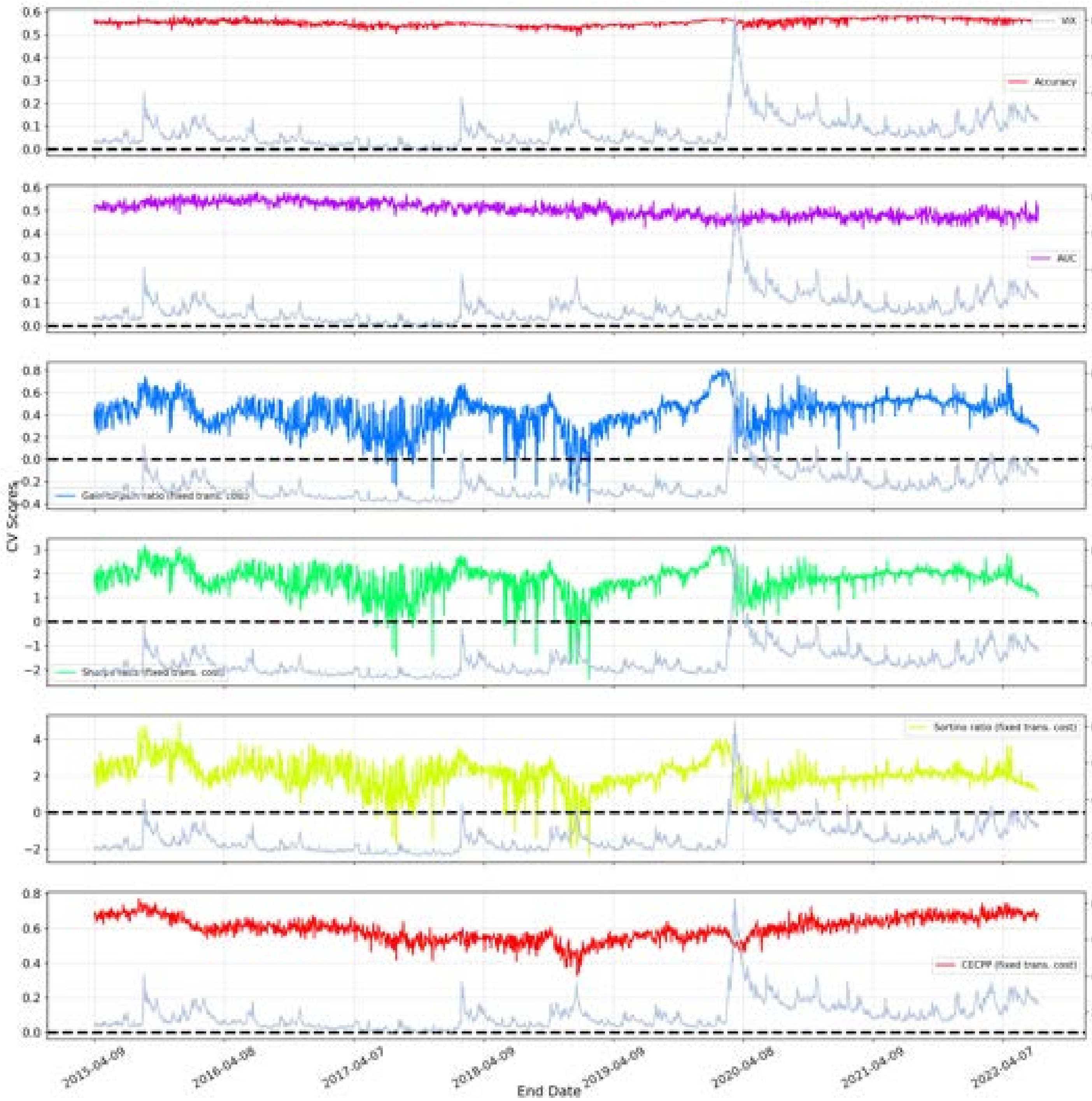
Figure S.XI.3: The average scoring functions (based on the fixed transaction cost strategy investing in *SPY*) evaluated on forecasts of validation samples made by a RF model cross-validated with the *CECPP* scoring function using the data on financial variables and technical indicators (*Dataset I*) described in Table S.V.1



[a] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

[b] The horizonal axis contains the end dates of rolling windows.

Figure S.XI.4: The average scoring functions (based on the fixed transaction cost strategy investing in *SPY*) evaluated on forecasts of validation samples made by a RF model cross-validated with the *CECPP* scoring function using the data on financial variables, technical indicators, and candlestick chart patterns (*Dataset II*) described in Table S.V.1

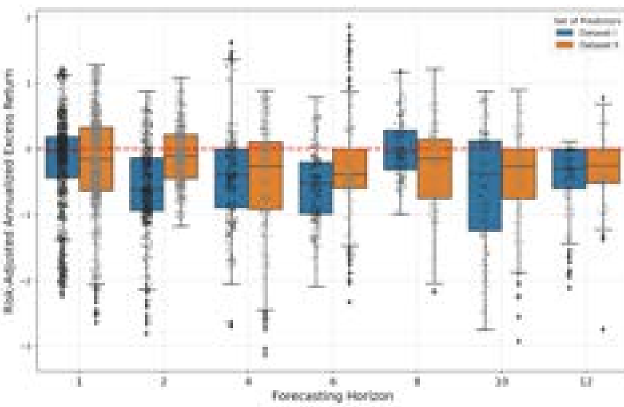


[a] The initial endowment is \$1000, and the amount of fixed transaction cost $c = \$0.05$.

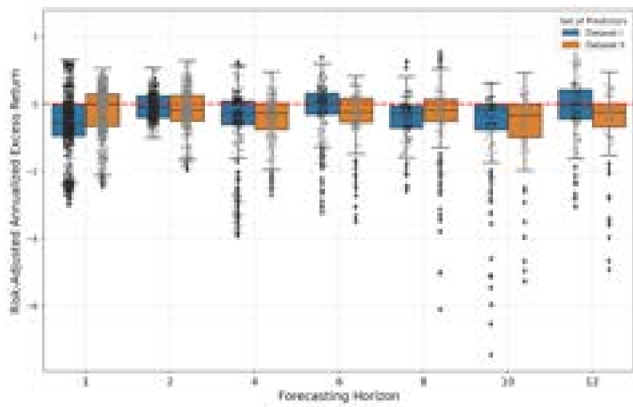
[b] The horizonal axis contains the end dates of rolling windows.

S.XII Risk-Adjusted Annualized Excess Returns across Forecasting Horizons

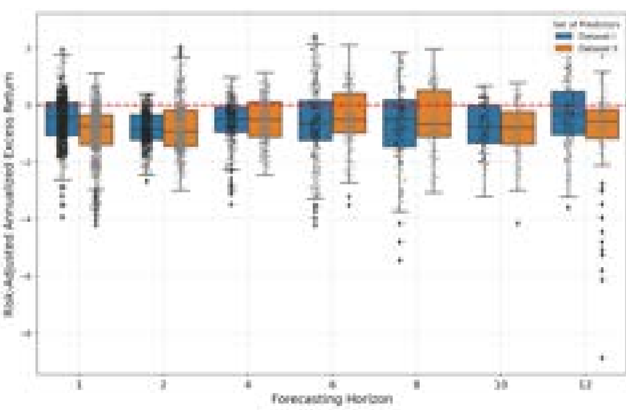
Figure S.XII.1: The annualized excess returns of the fixed transaction cost strategy investing in *SPY* for *200 days* based on multi-horizon forecasts [of price moving directions] by LGBM models trained with the *Brier* or *As2* loss function and cross-validated with the *CECPP* or *Sharpe ratio* scoring functions



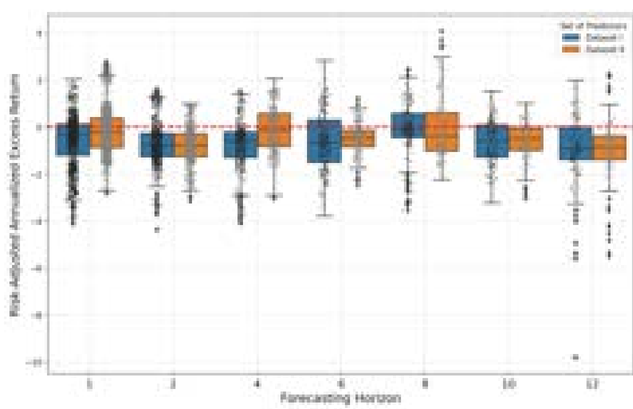
(a) *Brier* loss and *CECPP* score



(b) *Brier* loss and *Sharpe ratio* score



(c) *As2* loss and *CECPP* score



(d) *As2* loss and *Sharpe ratio* score

[a] The initial endowment is $1000, and the amount of fixed transaction cost $c = \$0.05$.

[b] Two sets of predictors are used: *Dataset I* (including the financial variables and technical indicators) and *Dataset II* (including *Dataset I* and candlestick chart patterns as described in Table S.V.1).

References

Achelis, S. B. (2001). *Technical Analysis from A to Z*. McGraw-Hill.

Anghel, D. G. (2021). Data snooping bias in tests of the relative performance of multiple forecasting models. *Journal of Banking and Finance 126*, 106112–106157.

Avramov, D., T. Chordia, and A. Goyal (2006). Liquidity and autocorrelation in individual stock returns. *Journal of Finance 61*(5), 2365–2394.

Baele, L., G. Bekaert, and K. Inghelbrecht (2010). The determinants of stock and bond return comovements. *Review of Financial Studies 23*(6), 2374–2428.

Bekaert, G. and E. Engstrom (2010). Inflation and the stock market: understanding the 'FED model'. *Journal of Monetary Economics 57*(3), 278–294.

Bessembinder, H. (2018). Do stocks outperform treasury bills. *Journal of Financial Economics 129*(3), 440–457.

Bollinger, J. A. (2001). *Bollinger on Bollinger Bands* (First ed.). McGraw Hill.

Brock, W., J. Lakonishok, and B. LeBaron (1992). Simple technical trading rules and the stochastic properties of stock returns. *Journal of Finance 47*(5), 1731–1764.

Bulkowski, T. N. (2008). *Encyclopedia of Candlestick Charts* (First ed.). John Wiley & Sons.

Campbell, J. Y., , and J. Ammer (1993). What moves the stock and bond markets? a variance decomposition for long-term asset returns. *Journal of Finance 48*(1), 3–37.

Campbell, J. Y. and R. J. Shiller (1988a). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies 1*(3), 195–228.

Campbell, J. Y. and R. J. Shiller (1988b). Stock prices, earnings, and expected dividends. *Journal of Finance 43*(3), 661–676.

Chande, T. S. (1995). A time price oscillator. *Technical Analysis of Stocks & Commodities 13*, 369–374.

Chande, T. S. and S. Kroll (1994). *The New Technical Trader: Boost Your Profit by Plugging into the Latest Indicators* (First ed.). John Wiley & Sons.

Chesler, D. (2004a). Quantifying market deception with the Hikkake pattern. *The Technical Analyst (November/December)*, 23–27.

Chesler, D. (2004b). Trading false moves with the Hikkake pattern. *Active Trader (April)*, 42–46.

Colby, R. W. (2002). *The Encyclopedia Of Technical Market Indicators* (Second ed.). McGraw-Hill.

Connolly, R., C. Stivers, and L. Sun (2005). Stock market uncertainty and the stock-bond return relation. *Journal of Financial and Quantitative Analysis 40*(1), 161–194.

Conrad, J., A. Hameed, and C. Niden (1994). Volume and autocovariances in short-horizon individual security returns. *Journal of Finance 49*(4), 1305–1329.

Ehlers, J. F. (2001). *Rocket Science for Traders: Digital Signal Processing Applications* (First ed.). John Wiley & Sons.

Fama, E. (1981). Stock returns, real activity, inflation, and money. *American Economic Review 71*(4), 545–565.

Fama, E. F. and K. R. French (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics 25*(1), 23–49.

Gerritsen, D. F., E. Bouri, E. Ramezanifar, and D. Roubaud (2020). The profitability of technical trading rules in the bitcoin market. *Finance Research Letters 34*, 101–263.

Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica 74*(6), 1545–1578.

Granville, J. E. (1976). *New Strategy of Daily Stock Market Timing for Maximum Profit*. Prentice-Hall.

Gustafson, G. (2001). Which volatility measure? *Technical Analysis of Stocks & Commodities 19*(6), 46–50.

Hsu, P.-H., M. P. Taylor, and Z. Wang (2016). Technical trading: Is it still beating the foreign exchange market? *Journal of International Economics 102*, 188–208.

Hutson, J. K. (1983). Good TRIX - triple exponential smoothing oscillator. *Technical Analysis of Stocks & Commodities 2*(3), 91–93.

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

Kablan, A. and W. L. Ng (2010). High frequency trading strategy using the hilbert transform. *The 6th International Conference on Networked Computing and Advanced Information Management*, 466–471.

Kaufman, P. J. (2019). *Trading Systems and Methods* (Six ed.). John Wiley & Sons.

Lamont, O. (1998). Earnings and expected returns. *Journal of Finance 53*(5), 1563–1587.

Lane, G. C. (1984). Lane's stochastics. *Technical Analysis of Stocks & Commodities 2*(3), 87–90.

Levy, R. A. (1967). Relative strength as a criterion for investment selection. *Journal of Finance 22*(4), 595–610.

Li, L., K. Jamieson, G. DeSalvo, A. Rostamizadeh, and A. Talwalkar (2018). Hyperband: A novel bandit-based approach to hyperparameter optimization. *Journal of Machine Learning Research 18*, 1–52.

Lin, Q. (2018). Technical analysis and stock return predictability: An aligned approach. *Journal of Financial Markets 38*, 103–123.

Meliones, A. and G. Makrides (2019). Automated stock price motion prediction using technical analysis datasets and machine learning. In G. A. Tsihrintzis, M. Virvou, E. Sakkopoulos, and L. C. Jain (Eds.), *Machine Learning Paradigms: Applications of Learning and Analytics in Intelligent Systems*, Volume 1 of *Learning and Analytics in Intelligent Systems*, pp. 207–228. Springer.

Mulloy, P. G. (1994). Smoothing data with faster moving averages. *Technical Analysis of Stocks & Commodities 12*(1), 11–19.

Politis, D. N. and J. P. Romano (1994). The stationary bootstrap. *Journal of the American Statistical Association 89*(428), 1303–1313.

Rapach, D. and Q. Zhou (2022). Asset pricing: time-series predictability. In J. H. Hamilton (Ed.), *Oxford Research Encyclopedia of Economics and Finance*, Oxford. Oxford University Press.

Schlossberg, B. (2006). *Technical Analysis of the Currency Market: Classic Techniques for Profiting from Market Swings and Trader Sentiment*. John Wiley & Sons.

Schwager, J. D. (2012). *Hedge Fund Market Wizards: How Winning Traders Win* (1 ed.). New Jersey: John Wiley & Sons.

Spiro, P. S. (1990). The impact of interest rate changes on stock prices volatility. *Journal of Portfolio Management 16*(2), 63–68.

Wang, G. B. X. (2010). Inflation risk and the inflation risk premium. *Economic Policy 25*(64), 755–806.

White, H. (2000). A reality check for data snooping. *Econometrica 68*(5), 1097–1126.

Wilder, J. W. (1978). *New Concepts in Technical Trading Systems* (First ed.). Trend Research.

Williams, L. (1985). The ultimate oscillator. *Technical Analysis of Stocks & Commodities 3*(4), 140–141.

Worden, D. (1950). The evolution of the cumulative money stream and balance of power. https://www.worden.com/TeleChartHelp/Content/EvolutionofBOPMoneyStream/EvolutionOfMSandBOP.htm.

Yen, S. M. F. and Y.-L. Hsu (2010). Profitability of technical analysis in financial and commodity futures markets - a reality check. *Decision Support Systems 50*, 128–139.