

Article

Does Machine Learning Outperform Simple Investment Rules? Comparative Performance and Strategy Robustness in European Equity Markets

Flavia Mirela Barna ^{1,*} , Anca Țăranu ¹ and Grațiela Georgiana Noja ² 

¹ Department of Finance, Information Systems and Business Modeling, West University of Timișoara, 300115 Timișoara, Romania; anca.taranu91@e-uvt.ro

² Department of Marketing, International Business and Economics, West University of Timișoara, 300115 Timișoara, Romania; gratiela.noja@e-uvt.ro

* Correspondence: flavia.barna@e-uvt.ro

Highlights

Please indicate how your work links to systems science via your contributions to systems practice, theory, and/or methodology.

- It conceptualizes machine-learning-based portfolio selection as a decision-making subsystem integrating data, predictive models, portfolio-construction rules, implementation constraints, risk assessment, and institutional governance.
- It evaluates statistical and implementation robustness through out-of-sample performance, ranking diagnostics, Newey–West inference, turnover, transaction-cost sensitivity, and alternative selection thresholds, while discussing operational and institutional resilience in relation to model complexity, interpretability, monitoring, data integrity, and implementation intensity.

What are the main findings and/or the implications of the main findings?

- Logistic Regression achieved the strongest classification performance and the highest gross cumulative portfolio return, whereas the transparent momentum strategy recorded the highest observed Sharpe ratio and substantially lower target-weight turnover than Logistic Regression and the three-model ensemble. Sharpe-ratio comparisons are descriptive.
- The three-model ensemble underperformed both Logistic Regression and the momentum rule, indicating that additional algorithmic complexity did not generate incremental investment value under the study's restricted technical information set and common portfolio-construction framework



Academic Editors: Mara Madaleno and João Jungo

Received: 6 August 2026

Revised: 2 September 2026

Accepted: 9 September 2026

Published: 11 September 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Abstract

Artificial intelligence is increasingly used in asset management, although evidence that greater model complexity consistently improves investment performance remains limited. This study examines whether machine-learning methods generate incremental value in European equity selection beyond transparent investment rules. The analysis used 562 eligible monthly price series drawn from the March 2026 STOXX Europe 600 constituents and applied retrospectively as a common ex post reference universe. The models were trained on observations from January 2011 to December 2020 and evaluated out of sample using signals formed from January 2021 to March 2026, with corresponding portfolio returns realized from February 2021 to April 2026. Logistic regression, random forest, XGBoost, and an equal-probability ensemble are compared with momentum, low-volatility, and

equal-weight strategies under common portfolio-construction rules. Logistic regression recorded the strongest classification results and the highest gross cumulative portfolio return. The transparent momentum strategy recorded the highest observed Sharpe ratio and substantially lower target-weight turnover, while the three-model ensemble underperformed both logistic regression and momentum. Newey–West inference did not establish a statistically significant difference in mean monthly returns between logistic regression and momentum. The results indicate that additional algorithmic complexity did not generate incremental investment value under the restricted technical information set and portfolio framework considered. The findings support evaluating model complexity jointly through predictive quality, gross performance, statistical uncertainty, implementation sensitivity, interpretability, and governance requirements.

Keywords: machine learning; European equity markets; complex financial systems; stock selection; strategy resilience

1. Introduction

Artificial intelligence (AI) and machine learning (ML) have moved rapidly from experimental applications to established components of the asset-management value chain. They are used for data extraction, research support, signal generation, portfolio construction, risk monitoring, compliance, and trading. Yet, adoption has advanced faster than evidence of investment performance. A 2026 Mercer survey of 131 global asset managers reported that 55% had integrated AI into at least one investment process and 91% planned to expand its use within twelve months, while only 8% reported measurable improvements in investment returns [1]. The contrast between adoption and realized investment benefits makes the evaluation of model complexity an important empirical and institutional question.

Financial markets are complex adaptive systems in which heterogeneous investors, institutions, technologies, and regulations interact under changing conditions [2–4]. A predictive strategy is not an isolated algorithm. It is one component of a broader decision-making framework encompassing data sources, feature engineering, model estimation, portfolio rules, execution, risk management, governance, and regulatory oversight. A model can therefore be statistically more sophisticated without improving the overall resilience of the investment system. Complexity can capture nonlinear relationships, but it may also amplify estimation error, obscure underlying economic mechanisms, increase dependence on data and software providers, and make validation more demanding.

The empirical finance literature provides strong evidence that ML methods can identify nonlinear relationships in large predictor sets [5]. However, the magnitude and stability of this advantage vary with the market, frequency, predictor set, and validation design. Evidence from U.S. equities cannot be transferred automatically to European markets, which span multiple national exchanges and operate under a distinct institutional framework. Moreover, much of the practical value attributed to ML may originate from well-known return regularities—particularly, momentum—that can also be implemented through transparent decision rules.

Recent evidence reinforces the view that the investment value of machine learning is highly design-dependent. Interpretable deep-learning models have generated economically meaningful portfolio outcomes that remain robust to transaction costs [6], while explainable reinforcement-learning approaches have produced out-of-sample gains relative to equal-weight portfolios in European equity applications [7]. These findings demonstrate that sophisticated architectures can generate investment value in suitable informational

and portfolio-design settings. They also sharpen the question addressed in this study, i.e., whether additional algorithmic complexity creates incremental value when competing methods are deliberately restricted to the same technical information set and evaluated under common portfolio-construction rules. This study investigates that question using a common ex post reference universe drawn from the March 2026 STOXX Europe 600 constituents. Logistic regression provided an interpretable linear specification, while random forest and XGBoost represent nonlinear machine-learning approaches based on bagging and sequential boosting, respectively. All three models use the same 14 technical indicators derived from historical prices. Their predicted probabilities are evaluated individually and through an equal-probability ensemble. The resulting portfolios are compared with an internally constructed equal-weight comparator and with transparent strategies based on 12-1 momentum and trailing volatility.

The study addresses three research questions, as follows:

RQ1. Do technical indicators derived from historical prices contain out-of-sample information about the relative next-month performance of European equities?

RQ2. Do the nonlinear random forest and XGBoost models, or their combination with logistic regression in an equal-probability ensemble, improve predictive and portfolio performance relative to the linear logistic regression specification?

RQ3. Do the evaluated machine-learning specifications generate incremental portfolio value relative to transparent momentum and low-volatility strategies and to the internally constructed equal-weight comparator?

The contribution of this study is fourfold. First, it provides a controlled comparison of linear and nonlinear machine-learning specifications, an ensemble, and transparent rule-based strategies with regard to European equities, with all approaches evaluated using the same ex post reference universe, test period, rebalancing frequency, selection thresholds, and weighting rule. Second, it separates predictive complexity from economic value by comparing machine-learning portfolios directly with momentum and low-volatility strategies derived from the same restricted technical information set. Third, it evaluates statistical and implementation robustness through ranking diagnostics, Newey–West inference for mean monthly return differences, target-weight turnover, transaction-cost sensitivity, and alternative portfolio-selection thresholds. Fourth, it links the empirical findings to systems and institutional considerations while distinguishing directly measured statistical and implementation outcomes from broader governance implications.

The central finding is deliberately conditional. All three machine-learning models exhibit only modest out-of-sample discriminatory performance. Logistic regression recorded the strongest classification results and the highest gross cumulative portfolio return, whereas the transparent momentum strategy recorded the highest observed Sharpe ratio and substantially lower target-weight turnover. The three-model ensemble underperformed both logistic regression and momentum. Newey–West inference did not establish a statistically significant difference in mean monthly returns between logistic regression and momentum. The findings therefore do not imply that machine learning is generally ineffective. Rather, they indicate that additional algorithmic complexity did not generate incremental investment value under the restricted technical information set and common portfolio framework examined in this study.

The rest of the paper is organized as follows. Section 2 presents the theoretical background concerning complex adaptive financial systems, documented return regularities, machine learning, strategy resilience, and European model governance. Section 3 describes the investment universe, data, feature engineering, model specifications, portfolio-construction rules, statistical inference, target-weight turnover, and robustness procedures. Section 4 reports the empirical results. Section 5 discusses their systems, implementation,

and institutional implications and presents the study's limitations. Section 6 concludes this research endeavor.

2. Theoretical Background

2.1. Financial Markets as Complex Adaptive Systems

The efficient market hypothesis (EMH) is the standard benchmark in the assessment of return predictability. The weak form of the EMH says that current prices reflect information contained in past prices and this limits the ability to earn systematic excess returns by using technical analysis [8]. Empirical evidence, however, documents recurring return patterns. Their persistence need not necessarily imply simple market inefficiency but can also be a matter of risk compensation, behavioral biases, institutional constraints, limits to arbitrage, and changing interactions between market participants [9].

The complex-systems viewpoint is an attractive supplement to this reading. Markets consist of adaptive agents who learn from common information and from each other's behavior. A profitable pattern attracts capital and the resulting trading can weaken, change, or reinforce the pattern temporarily. Thus, predictor–return relationships are not immutable physical laws. They are the contingent products of an evolving system. The adaptive markets hypothesis (AMH) considers market efficiency as a function of context, which is a function of competition, innovation, and changes in the environment [3].

This view is particularly important in algorithmic investment. Individual optimization can lead to collective effects when many institutions use similar data, software libraries, features, and models. Crowded signals can lead to correlated trades, liquidity pressure, and herding, mechanisms that have also been discussed in the literature on high-speed and algorithmic trading [10,11]. A strategy that seems robust in an isolated backtest can therefore become fragile when deployed at scale. The durability of the system depends on the diversity of models, controls, and behavior, and on the ability of the institutions to detect degradation of models and to adapt decision rules.

This study treats machine-learning-based portfolio selection as a decision-making subsystem embedded in a broader financial and institutional environment. The relevant question is not simply whether a predictive model produces a marginal improvement in classification performance, but whether the resulting stock-selection process generates economically meaningful value relative to simpler alternatives after statistical uncertainty and implementation sensitivity are considered. The systems perspective therefore requires predictive models to be evaluated jointly with their information inputs, portfolio-construction rules, implementation demands, interpretability, and governance requirements.

2.2. Technical Indicators and Documented Return Anomalies

One of the most robust empirical regularities in asset pricing is momentum. Jegadeesh and Titman [12] find that stocks with high returns over the prior 3–12 months tend to continue outperforming over intermediate horizons. Carhart [13] includes momentum in a four-factor model, while Fama and French [14] document its effect in international markets, including Europe. The standard 12-1 momentum construction omits the most recent month to reduce potential contamination from short-term reversal.

The low-volatility anomaly is a challenge to the classical positive risk-return relation. Ang et al. [15] find that stocks with high idiosyncratic volatility earn lower future average returns than low-volatility stocks. Possible explanations are leverage constraints, benchmark incentives, preference for lottery-like payoffs, and limits to arbitrage.

Other price-based signals capture relative strength and reversal. George and Hwang [16] find that distance from the 52-week high predicts future returns net of conventional momentum. Long-horizon reversal is found to be consistent with overreaction by De

Bondt and Thaler [17], but the horizon of this effect is different from intermediate-term momentum. These mechanisms can be captured by a systematic feature set of moving-average deviations and trend-acceleration measures.

These indicators provide economically interpretable inputs for machine-learning models and therefore create a useful controlled comparison. If a complex model relies primarily on combinations or reweightings of momentum-related variables, its predictive architecture may be more sophisticated while its underlying economic exposure remains similar to that of a transparent momentum rule. This possibility cannot be established from model architecture alone and must be evaluated by comparing predictive rankings, portfolio outcomes, and implementation characteristics.

2.3. Machine Learning, Ensembles, and Parsimony

Gu, Kelly, and Xiu [5] show that ML can improve cross-sectional return prediction by capturing nonlinearities and interactions among a large set of firm characteristics. Krauss et al. [18] reported profitable statistical-arbitrage results for deep neural networks, gradient boosting, and random forest on S&P 500 data, but also showed that performance degraded over time. These results indicate the potential of nonlinear methods and the adaptive nature of markets to signals that are broadly exploited.

More recent research shows that this potential extends to different financial prediction and portfolio-design settings. Li et al. [19] found XGBoost to be the strongest model among the evaluated machine-learning specifications for stock-price crash-risk prediction when using heterogeneous firm- and CEO-level characteristics. Recent work published in *Systems* has also extended portfolio selection toward deep reinforcement-learning architectures [20]. These findings illustrate the capacity of nonlinear learners to exploit rich information structures, while suggesting that their relative advantage may depend importantly on the information set available to the model.

Technical indicators have also been embedded in more sophisticated AI-based trading systems. Yu et al. [21] integrated technical-analysis inputs with deep-learning models in a lightweight automated stock-trading framework and evaluated the resulting backtested returns. Cheng et al. [22] similarly combined deep learning with indicator selection and explicitly assessed both stock-price forecasting and trading profits. This literature is particularly relevant because it highlights the distinction between statistical prediction and economic performance; superior predictive metrics do not necessarily translate into superior portfolio outcomes once ranking, portfolio construction, risk, target-weight turnover, and implementation constraints are taken into account.

Model flexibility comes with risks. Financial observations are noisy, non-stationary, and cross-sectionally dependent. Backtests may overstate performance through look-ahead bias, data snooping, repeated model selection, survivorship bias, and inappropriate validation [23–25].

These concerns are consistent with the broader asset-pricing literature. Bagnara [26], in a critical review of machine learning in asset pricing, emphasizes both the flexibility and predictive capacity of modern ML approaches and the methodological care required when translating predictive improvements into economically interpretable conclusions. Ensemble learning combines predictions from multiple models with the objective of producing more stable or accurate results than any individual specification. Different ensemble approaches achieve this in different ways. Bagging reduces prediction variance by averaging models trained on modified samples of the data [27], while gradient boosting builds models sequentially, with each new model attempting to correct previous errors [28]. In this study, the ensemble is constructed as the equal-weighted average of the predicted probabilities generated by logistic regression, random forest, and XGBoost. Averaging may improve

predictive stability when the component models capture sufficiently different relationships or make weakly correlated errors. However, when the models rely on the same observations and closely related technical indicators, their predictions may be too similar for combination to generate an incremental performance benefit. The value of the ensemble must therefore be assessed through its observed out-of-sample ranking and portfolio outcomes rather than inferred from its greater architectural complexity.

2.4. Strategy Resilience and Institutional Governance

Resilience in an investment system can be considered across predictive, financial, implementation, operational, and institutional dimensions. Predictive resilience concerns the persistence of ranking quality outside the estimation sample. Financial resilience concerns the stability of portfolio outcomes, including return variability and drawdown. Implementation resilience concerns whether observed gross performance remains economically meaningful after accounting for trading intensity and implementation frictions. Operational resilience relates to data continuity, system dependencies, execution processes, and ongoing model monitoring. Institutional resilience relates to whether an organization can explain, validate, challenge, oversee, and govern the investment system.

The present study directly evaluates only selected predictive, financial, and implementation dimensions through out-of-sample classification and ranking results, portfolio performance, drawdown, target-weight turnover, transaction-cost sensitivity, calendar-period comparisons, and alternative portfolio-selection thresholds. Operational and institutional resilience are not measured empirically. They are discussed as implications of the observed differences in model complexity, interpretability, data requirements, monitoring demands, and implementation intensity.

This multidimensional interpretation is consistent with the growing literature on explainable artificial intelligence in finance. Černevičienė and Kabašinskas [29], in a systematic review of 138 studies, show that explainability has become increasingly important in financial applications because opaque algorithmic decisions create challenges for risk assessment, trust, and institutional decision-making. Their findings connect explainability with the development of more trustworthy financial systems. Accordingly, interpretability is treated in this study as a model characteristic relevant to institutional resilience, rather than as direct empirical evidence that a particular strategy is institutionally resilient.

Performance evaluation should therefore combine predictive diagnostics, portfolio outcomes, statistical uncertainty, risk indicators, and implementation sensitivity. At the same time, governance implications should be presented proportionately and should not be treated as directly measured institutional outcomes.

These trade-offs are especially relevant when considering European regulation. ESMA noted that the MiFID II requirements are still applicable to investment firms using AI, including requirements on organizational arrangements, conduct of business, client interests, records, controls, and governance [30]. The EU AI Act implements a horizontal risk-based approach, but an AI model that is used for stock selection is not automatically considered a high-risk system simply because it is used in finance [31]. The regulatory relevance of this study is hence not so much in claiming a universal statutory duty for explainability as in the practical necessity for traceability, oversight by humans, data governance, and model-risk controls.

Evidence from European investment funds also lends itself to a conservative interpretation of AI adoption. Bagattini and Piazza [32] identify 145 EU funds that reported the use of AI or ML in their investment strategies, representing approximately 0.1% of assets under management for UCITS. Their net performance was not very different from comparable actively managed funds. This evidence from investment funds motivates

the present strategy-level analysis, which investigates whether machine-learning models based on technical indicators generate additional investment value or simply reproduce exposures already captured by established strategies such as momentum.

Taken together, recent evidence does not support a simple conclusion that either complex or parsimonious models are universally superior. Sophisticated machine-learning architectures can generate substantial predictive and economic value when they exploit rich information sets, portfolio-specific objectives, or advanced learning mechanisms [6,7,19–22]. At the same time, the literature increasingly emphasizes interpretability, validation, and the distinction between predictive accuracy and economically implementable performance [7,26,29]. It remains less explored whether additional algorithmic complexity creates incremental investment value when competing methods use the same restricted technical information set and are evaluated under identical portfolio-construction rules. The present study addresses this gap through a controlled comparison of linear and nonlinear machine-learning models, an ensemble, and transparent momentum and low-volatility strategies applied with regard to European equities.

3. Materials and Methods

3.1. Investment Universe and Data Collection

The analysis begins with the STOXX Europe 600 constituent list available in March 2026. The source list contained 600 constituent records corresponding to 599 unique tickers. Two legacy constituent records, Novozymes and Chr. Hansen, correspond to the same post-merger issuer, Novonesis, following the completion of their statutory combination in January 2024. The resulting B share trades on Nasdaq Copenhagen under the ticker NSIS B, represented as NSIS-B.CO in the Yahoo Finance ticker convention [33,34]. The STOXX Europe 600 includes large-, mid-, and small-cap companies across 17 developed European countries and represents approximately 90% of the region's free-float market capitalization [35].

Monthly auto-adjusted closing prices for the period from January 2010 to April 2026 were obtained from Yahoo Finance, using the Python yfinance library (version 1.4.1). Exchange-specific ticker suffixes and ticker histories were reconciled manually where necessary. The raw data were retrieved on 31 May 2026 using Google Colab (Python 3.12) and yfinance 1.4.1, with interval = "1mo" and auto_adjust = True. The requested download window was January 2010–April 2026, with the API end date set to 1 May 2026 in the yfinance API call. Exchange-specific Yahoo Finance ticker mappings were manually reconciled with the constituent list. The reproducibility package contains the constituent mapping, processed price and feature datasets, filtering rules, and the source code used to reconstruct the reported results.

Companies were required to have at least 60 monthly price observations. After accounting for duplicate legacy mapping, 37 unique tickers were excluded because of recent listings, corporate reorganizations, acquisitions, ticker-history discontinuities, or unavailable price data. The final price database therefore contains 562 eligible company series.

Feature construction initially produced 96,131 usable company-month observations after the initialization of the rolling windows. A predefined data-quality screen identified 27 observations with forward monthly returns above 100% or below −90%. These observations represented approximately 0.03% of the initial feature sample and included 17 tickers. Six observations related to 3i Group (III.L) and included visible adjustment discontinuities. The remaining observations may reflect corporate actions, genuine extreme returns, ticker-history discontinuities, or source-data problems. Because point-in-time corporate-action validation was unavailable for every observation, all 27 observations were

excluded using the same predefined rule. The resulting analytical dataset contained 96,104 company-month observations.

The cross-sectional mean forward return and the binary prediction target were recomputed after applying the extreme-return screen. Consequently, the 27 excluded observations do not affect the target labels of the remaining observations through the cross-sectional benchmark. Relative to the initially retained feature dataset, this procedure changes 1560 target labels across 121 signal-formation months. All model estimation, classification evaluation, portfolio construction, and robustness tests reported below use the post-screen target definition.

The March 2026 constituent snapshot was used because it was the index composition available when the research database was assembled. Its retrospective application means that the sample is a common ex post reference universe rather than a point-in-time reconstruction of historical STOXX Europe 600 membership. Firms removed from the index before March 2026 are absent from the sample, while investors operating in earlier periods could not have known the March 2026 composition. The design therefore introduces both survivorship bias and look-ahead bias.

The use of the same fixed universe for all machine-learning models, rule-based strategies, and the internally constructed equal-weight comparator improves internal comparability because the evaluated approaches share the same eligible observations, test period, rebalancing frequency, selection thresholds, and weighting rule. However, this common-universe design does not eliminate survivorship or look-ahead bias. The biases may overstate absolute portfolio performance and may also affect the strategies differently if their signals have unequal exposure to firms that subsequently exit the index. Accordingly, the results should be interpreted as a controlled comparison of stock-selection methods within a shared ex post universe and not as the historical performance of an investable STOXX Europe 600 strategy.

The analysis relies on auto-adjusted prices supplied by a third-party public data source. Differences in currency denomination, corporate-action adjustments, missing observations, and ticker histories may introduce measurement inconsistencies across national markets. The extreme-return screen reduces the influence of visible discontinuities but cannot ensure that all data-quality problems have been identified. The reported returns should therefore be interpreted as comparative backtest results rather than as directly realizable investment performance.

Prices were retained in their local quotation currencies and were not converted into a common currency. Consequently, the cross-sectional target and the portfolio returns combine local-currency returns and do not incorporate exchange-rate gains or losses from the perspective of a euro-based investor. The reported results do not represent the performance of a directly investable euro-denominated portfolio.

3.2. Feature Engineering and Target Definition

Fourteen technical predictors were constructed separately for each security using price and return information available through the end of signal-formation month t . No information from month $t + 1$ was used in predictor construction. Table 1 summarizes the variables.

The variable names are retained to preserve exact correspondence with the processed feature dataset and Python code. In particular, `vol_1m` is a short-horizon volatility indicator calculated as the rolling standard deviation of the latest three monthly returns; it is not the standard deviation of a single monthly observation. Similarly, the `dist_52w_high` and `dist_52w_low` labels refer to the highest and lowest prices observed over the latest 12 monthly observations, which approximate a 52-week horizon in monthly data.

Table 1. Predictive variables.

Category	Variable	Definition
Momentum	mom_1m	$(P_t/P_{t-1} - 1)$
Momentum	mom_3m	$(P_t/P_{t-3} - 1)$
Momentum	mom_6m	$(P_t/P_{t-6} - 1)$
Momentum	mom_12m	$(P_t/P_{t-12} - 1)$
Momentum	mom_12_1	$(P_{t-1}/P_{t-12} - 1)$
Volatility	vol_1m	Rolling standard deviation of the latest three monthly returns
Volatility	vol_6m	Standard deviation of the latest six monthly returns
Volatility	vol_12m	Standard deviation of the latest twelve monthly returns
Mean reversion	mean_rev_3m	$(P_t/SMA_3(t) - 1)$
Mean reversion	mean_rev_6m	$(P_t/SMA_6(t) - 1)$
Mean reversion	mean_rev_12m	$(P_t/SMA_{12}(t) - 1)$
Relative strength	dist_52w_high	$(P_t/\max(P_{t-11}, \dots, P_t) - 1)$
Relative strength	dist_52w_low	$(P_t/\min(P_{t-11}, \dots, P_t) - 1)$
Trend	trend_3_12	mom_3m – mom_12m

The prediction target is a binary cross-sectional indicator. For security i in signal-formation month t , the target equals one if the security's simple return in month $t + 1$ exceeds the cross-sectional mean return of all eligible securities available for that realized-return month, and zero otherwise: $y_{i,t} = 1 \{r_{i,t+1} > \bar{r}_{t+1}\}$

where $1\{\cdot\}$ is the indicator function, $r_{i,t+1}$ denotes the security's forward one-month return, and $\bar{r}_{t+1}$ denotes the cross-sectional mean forward return of the eligible analytical universe. Thus, $y_{i,t} = 1$ if $r_{i,t+1} > \bar{r}_{t+1}$ and $y_{i,t} = 0$ otherwise.

After the extreme-return screen and the subsequent recomputation of the cross-sectional benchmark and target, the complete analytical sample contained approximately 48.84% positive observations and 51.16% negative observations. In the out-of-sample test period, the corresponding proportions were approximately 49.10% positive and 50.90% negative. Consequently, 50.90% is the relevant majority-class accuracy benchmark for evaluating the test-period classification results.

3.3. Training Design and Models

The analytical sample was divided chronologically into a training period and an out-of-sample test period. After the extreme-return screen, the training sample contained 60,909 company-month observations from January 2011 to December 2020. The test sample contained 35,195 company-month observations and 63 signal-formation months from January 2021 to March 2026. Because the target is defined using the following month's return, the corresponding realized portfolio returns extend from February 2021 to April 2026. No random train–test split was used.

The three classifiers were selected to represent distinct levels and forms of algorithmic complexity. Logistic regression provides a regularized and interpretable linear benchmark. Random forest represents nonlinear bagging and can capture interactions while limiting model variance through tree averaging. XGBoost represents sequential gradient boosting and provides a more flexible nonlinear specification. The equal-probability ensemble tested

whether combining predictions from these structurally different learners generated value beyond the individual model specifications. The momentum and low-volatility portfolios provide transparent economic comparators directly related to predictors included in the common technical information set.

Three classifiers were estimated, as follows:

- Logistic regression, using an L2-regularized probabilistic classifier estimated after standardizing the 14 predictors with StandardScaler ($C = 1.0$, solver = “lbfgs”, tol = 0.0001, fit_intercept = True, max_iter = 1000, random_state = 42);
- Random forest, using 200 trees, criterion = “gini”, max_depth = 6, min_samples_split = 2, min_samples_leaf = 50, max_features = “sqrt”, bootstrap = True, class_weight = None, random_state = 42, and n_jobs = 1;
- XGBoost, using 200 estimators, objective = “binary:logistic”, booster = “gbtree”, tree_method = “hist”, max_depth = 4, learning_rate = 0.05, subsample = 0.8, colsample_bytree = 0.8, min_child_weight = 1.0, gamma = 0.0, reg_alpha = 0.0, reg_lambda = 1.0, eval_metric = “logloss”, random_state = 42, and n_jobs = 1.

The logistic regression standardization parameters were estimated using only the 2011–2020 training observations and were then applied unchanged to the 2021–2026 test observations. Random forest and XGBoost were estimated using the unstandardized predictor values. Model estimation used the complete training sample, and the fitted models were held fixed throughout the out-of-sample test period; no rolling or expanding-window re-estimation was performed.

The model classes and hyperparameters were fixed before examining portfolio-performance results and were not optimized on the January 2021–March 2026 test sample. They were selected as parsimonious, regularized specifications intended to limit model complexity rather than through an exhaustive hyperparameter-tuning exercise. Consequently, this study compares fixed model specifications and does not claim that each algorithm is optimally tuned. Hyperparameters not listed above retained the defaults of the software versions reported in the reproducibility materials.

Classification performance was evaluated using accuracy and the area under the receiver operating characteristic curve (AUC-ROC). Accuracy evaluates classifications produced at the default probability threshold, whereas AUC-ROC evaluates the ranking of positive observations across all possible classification thresholds. The metrics are interpreted descriptively. In the out-of-sample test sample, the negative class represented approximately 50.90% of observations; therefore, 50.90% is the relevant majority-class accuracy benchmark. An AUC-ROC of 0.5000 represents random ranking. Small differences relative to these benchmarks do not by themselves establish economically meaningful predictability.

For each stock-month, the ensemble score is the arithmetic mean of the three predicted probabilities. Individual-model and ensemble results are reported separately to distinguish the performance of each standalone classifier from that of the combined prediction system.

3.4. Portfolio Construction and Comparators

At the end of each month, stocks were ranked according to the predicted probability generated by logistic regression or according to the equal-probability ensemble score. The top 20% of stocks (approximately 112 companies) were selected to form the portfolio for the following month. The selected stocks received equal weights, and the portfolios were rebalanced monthly. An equal-weight portfolio constructed internally from all eligible securities in the common analytical sample was used as the primary comparator. This portfolio is distinct from the official STOXX Europe 600 index. This approach made it

possible to isolate the differences generated by the selection rules while keeping the sample and weighting method constant.

Signals are formed at the end of month t using information available through that month, and the selected portfolio earns the forward return in month $t + 1$. The 63 signal-formation dates extended from January 2021 to March 2026, while the corresponding realized returns extended from February 2021 to April 2026. All annual performance results were assigned to the calendar month in which the return is realized, rather than to the month in which the signal was formed.

The following comparators were used:

- a simple momentum portfolio that selected the top 20% of stocks according to mom_12_1;
- a simple low-volatility portfolio that selected the bottom 20% of stocks according to vol_12m;
- the internally constructed equal-weight portfolio, which includes all eligible securities available in the common analytical sample for the corresponding month.

Cumulative return [36–38] is calculated as follows:

$$R_{cum} = \prod_{t=1}^T (1 + r_t) - 1$$

The geometric annualized return, which is used in the principal performance tables, is:

$$R_{ann} = (1 + R_{cum})^{12/T} - 1$$

For transparency, arithmetic annualization based on the mean monthly return was calculated separately but is not used as the geometric annualized performance measure reported in the principal tables. Annualized volatility is calculated as follows:

$$\sigma_{ann} = sd(r_t) \sqrt{12}$$

The annual risk-free rate of 3% is converted to the monthly equivalent as follows:

$$r_{f,m} = (1 + 0.03)^{1/12} - 1$$

and the Sharpe ratio is calculated as follows:

$$SR = \sqrt{12} \frac{\overline{r_t - r_{f,m}}}{s(r_t - r_{f,m})}$$

Sharpe ratios are used as descriptive measures of observed risk-adjusted performance. This study does not conduct formal inference on differences between Sharpe ratios. Accordingly, differences between the reported Sharpe ratios should not be interpreted as statistically significant differences.

Statistical inference for mean return differences is based on Newey–West heteroskedasticity- and autocorrelation-consistent standard errors with three monthly lags.

Portfolio trading intensity was approximated using one-way target-weight turnover reconstructed from the security-level portfolio selections generated at each monthly rebalancing. For strategy s , target-weight turnover in month t was calculated as follows:

$$\text{TargetTurnover}_{s,t} = \frac{1}{2} \sum_{i=1}^N |w_{i,s,t}^* - w_{i,s,t-1}^*|$$

where $w_{i,s,t}^*$ and $w_{i,s,t-1}^*$ denote the target portfolio weights assigned to security i at two consecutive monthly portfolio-formation dates. A weight is set equal to zero when a security is not included in the corresponding portfolio. The factor of one half prevents purchases and sales associated with the same portfolio replacement from being counted twice. Because the portfolios are equal-weighted, the measure primarily captures changes in portfolio membership and small changes in target weights caused by variation in the number of eligible securities.

This measure should be interpreted as target-weight or selection turnover rather than as exact realized turnover. It compares consecutive target allocations and does not account for the drift in security weights generated by returns between rebalancing dates. Consequently, it may understate the amount of trading required to restore equal portfolio weights, particularly for the equal-weight comparator and during periods of high cross-sectional return dispersion.

Net portfolio returns were evaluated under proportional transaction-cost assumptions of 10, 25, and 50 basis points per unit of one-way target-weight turnover:

$$r_{s,t}^{net} = r_{s,t}^{gross} - c \times TargetTurnover_{s,t}$$

where c denotes the assumed proportional transaction cost per unit of one-way target-weight turnover. The resulting net returns represent stylized cost-sensitivity estimates based on changes in portfolio target allocations. They do not represent exact realized net returns because the turnover proxy excludes return-induced weight drift, security-specific bid–ask spreads, taxes, commissions, market impact, implementation shortfall, and capacity constraints. The first portfolio-formation month is excluded from the average target-weight turnover calculation because no preceding target portfolio is available for comparison.

3.5. Statistical Inference, Transaction Costs, and Robustness Tests

Statistical inference focuses on differences in mean monthly portfolio returns. For each pair of strategies, the monthly return of the second strategy was subtracted from that of the first strategy, and the mean difference was evaluated using Newey–West heteroskedasticity- and autocorrelation-consistent standard errors with three monthly lags. These tests account for heteroskedasticity and serial dependence in the monthly return-difference series. The Newey–West tests do not evaluate differences in Sharpe ratios or establish the statistical significance of the observed risk-adjusted-performance rankings. Sharpe ratios are therefore reported and discussed descriptively. Formal inference on Sharpe-ratio differences is outside the scope of the present study.

Within the scope of the available database, strategy robustness was assessed through (i) out-of-sample classification and ranking performance; (ii) comparison of the standalone logistic regression model, the ensemble, and the transparent strategies; (iii) performance across realized-return calendar periods; (iv) drawdown and descriptive risk-adjusted indicators; (v) Newey–West inference for mean monthly return differences; (vi) target-weight turnover and transaction-cost sensitivity; and (vii) alternative portfolio-selection thresholds. Operational and institutional resilience are discussed as implications rather than directly measured outcomes.

4. Results

4.1. Classification Performance

Table 2 presents the out-of-sample classification results. Logistic regression achieved the highest accuracy and AUC-ROC among the three evaluated models, although its accuracy exceeded the test-period majority-class benchmark by only approximately 0.19

percentage points. As illustrated in Figure 1, the performance of all three machine-learning models remains close to the majority-class accuracy benchmark and the random-ranking AUC-ROC reference, indicating limited out-of-sample discriminatory ability.

Table 2. Out-of-sample classification performance, January 2021–March 2026.

Model	Accuracy	AUC-ROC
Logistic regression	51.09%	0.5103
Random forest	50.33%	0.5045
XGBoost	50.26%	0.5020
Majority-class reference	50.90%	0.5000

Note: The majority-class reference assigns all test observations to the most frequent test-period class. Its accuracy is 50.90%, reflecting the class distribution obtained after the extreme-return screen and the subsequent recomputation of the cross-sectional target. An AUC-ROC of 0.5000 represents random ranking performance.

Figure 1 provides a visual comparison of model accuracy and AUC-ROC against their respective reference benchmarks.

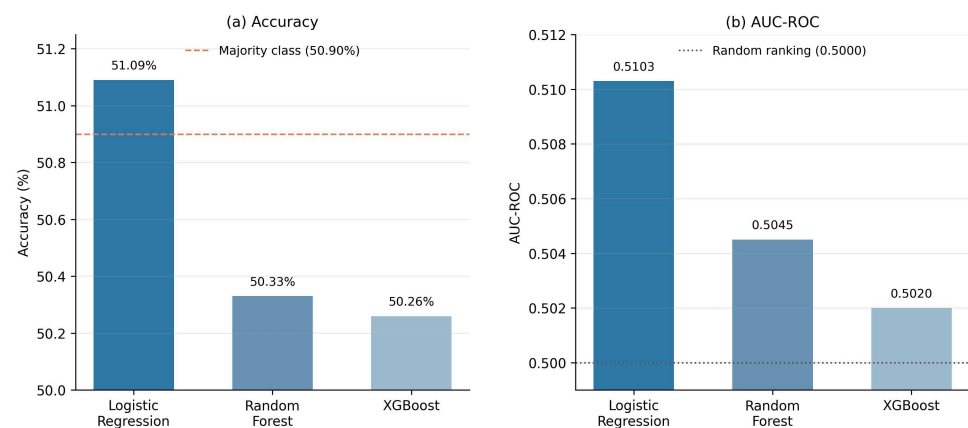


Figure 1. Out-of-sample classification performance by model. Note: Panel (a) reports classification accuracy, with the dashed line indicating the 50.90% test-period majority-class benchmark. Panel (b) reports AUC-ROC, with the dotted line indicating the random-ranking reference of 0.5000. Source: Authors' calculations.

The results indicate weak out-of-sample discriminatory performance rather than strong predictability. Logistic regression achieved the highest accuracy (51.09%) and AUC-ROC (0.5103) among the three models. Its accuracy exceeded the 50.90% test-period majority-class benchmark by approximately 0.19 percentage points, while its AUC-ROC remained only modestly above the random-ranking value of 0.5000. Random forest and XGBoost produced accuracies below the majority-class benchmark and AUC-ROC values close to 0.5000. Thus, logistic regression provides only a very small improvement over the majority-class benchmark at the default classification threshold, while the ranking performance of all three models remains weak. Overall, the classification results provide only limited evidence that the selected technical indicators can predict the relative performance of stocks in the following month. Logistic regression performed better than random forest and XGBoost, but the differences are small and do not indicate strong predictive ability. The results therefore do not support the assumption that more complex nonlinear models necessarily improve prediction when all models rely exclusively on technical indicators derived from historical prices. A possible explanation is that the relationships captured by these indicators are sufficiently simple and stable to be represented by a regularized linear

model. By contrast, the more flexible nonlinear models may identify complex relationships that were specific to the 2011–2020 training period but did not remain sufficiently stable during the 2021–2026 test period.

Classification performance and portfolio performance capture different aspects of the models and should therefore be interpreted separately. Even modest ranking information reflected in AUC-ROC may produce a meaningful difference in portfolio returns when predicted probabilities are used to rank stocks and the portfolio includes only those with the highest scores. At the same time, better classification results do not necessarily lead to higher returns or stronger risk-adjusted performance. For this reason, the economic value of the models is assessed through the portfolio comparisons presented in the following sections, rather than through classification metrics alone.

The model-specific built-in feature-importance outputs indicate that momentum-related variables remain important for both nonlinear models. For Random Forest, the most prominent variables are *mom_1m*, *mean_rev_3m*, *mom_12_1*, *vol_6m*, *trend_3_12*, and *mom_12m*. For XGBoost, the leading variables are *mom_1m*, *mom_12_1*, *mom_12m*, *mean_rev_3m*, *trend_3_12*, and *vol_12m*. These outputs are used only as descriptive diagnostics: importance values are model-dependent, are not directly comparable across algorithms, and do not identify causal effects. Because several predictors are correlated, the importance assigned to an individual variable may also be distributed across related indicators. The results therefore support only the limited conclusion that the models draw substantially on momentum-related information, which provides the economic justification for comparison with the transparent 12–1 momentum strategy.

4.2. Logistic Regression Portfolio Versus the Equal-Weight Comparator

Table 3 reports the performance of the standalone logistic regression portfolio, distinct from the ensemble results presented in Section 4.3.

Table 3. Performance of the logistic regression portfolio and the internally constructed equal-weight comparator.

Metric	LR Portfolio	Equal-Weight Comparator
Geometric annualized return	15.22%	9.96%
Annualized volatility	17.15%	13.60%
Sharpe ratio	0.743	0.550
Maximum drawdown	−23.83%	−21.96%
Calmar ratio	0.638	0.454
Win rate	65.1%	61.9%
Information Ratio	0.818	—

The LR portfolio combines a higher annualized return with higher volatility and a slightly deeper maximum drawdown. Its Sharpe and Calmar ratios are higher than those of the equal-weight comparator in the gross backtest. These gross performance indicators should be interpreted jointly with target-weight turnover and transaction-cost sensitivity results reported in Section 4.4.

Newey–West heteroskedasticity- and autocorrelation-consistent inference with three monthly lags indicates that the LR portfolio generated an average monthly return differential of 0.437 percentage points relative to the equal-weight comparator. The corresponding HAC statistic is 2.020 and the *p*-value is 0.043. The estimated mean-return advantage is

therefore statistically significant at the 5% level, although the result remains subject to the short out-of-sample period and the limitations of the fixed ex post universe.

4.3. Individual, Ensemble, and Rule-Based Strategy Comparison

Table 4 compares the gross cumulative and risk-adjusted performance of the logistic regression portfolio, the three-model ensemble, the momentum and low-volatility strategies, and the internally constructed equal-weight comparator. Implementation sensitivity after turnover-based transaction-cost deductions is examined separately in Section 4.4.

Table 4. Out-of-sample gross portfolio performance.

Strategy	Cumulative Return	Geometric Annualized Return	Sharpe Ratio
Logistic regression	110.36%	15.22%	0.743
Simple momentum (mom_12_1)	100.63%	14.18%	0.837
ML ensemble (LR + RF + XGBoost)	80.32%	11.88%	0.582
Simple low volatility	67.10%	10.27%	0.671
Equal-weight comparator	64.61%	9.96%	0.550

Note: Performance based on signals formed from January 2021 to March 2026 and returns realized from February 2021 to April 2026. Sharpe ratios are rounded to three decimal places.

Figure 2 illustrates the cumulative gross performance paths of the Logistic Regression portfolio, the momentum and low-volatility strategies, and the equal-weight comparator over the out-of-sample period.

LR produced the highest cumulative return, exceeding the simple momentum strategy by approximately 9.73 percentage points. Momentum achieved the highest observed Sharpe ratio in the sample, at 0.837 compared with 0.743 for LR. Because no formal test of Sharpe-ratio differences was conducted, this ranking should be interpreted descriptively and does not establish a statistically significant risk-adjusted advantage. Momentum exceeds the ML ensemble by approximately 20.31 percentage points in cumulative return. The low-volatility strategy also achieved a higher Sharpe ratio than the ensemble despite a lower cumulative return. The reported differences represent gross portfolio outcomes before transaction-cost deductions.

These comparisons produce a more nuanced interpretation. The strongest individual ML specification generates the highest gross cumulative return, but combining the three classifiers reduces performance. Relative to a rule constructed from one of the dominant predictors, the ensemble's incremental value is not evident. Greater model complexity therefore does not generate a monotonic performance improvement.

This finding illustrates a systems trade-off between complexity and parsimony. The ensemble integrates more information and reduces dependence on one algorithm, but integration does not guarantee that the combined decision rule preserves the strongest economic signal. A transparent rule can be more resilient when its mechanism remains stable and when added features contribute noise rather than diversification.

To assess the statistical uncertainty surrounding the observed performance differences, Table 5 reports Newey–West heteroskedasticity- and autocorrelation-consistent (HAC) inference for pairwise differences in mean monthly portfolio returns.

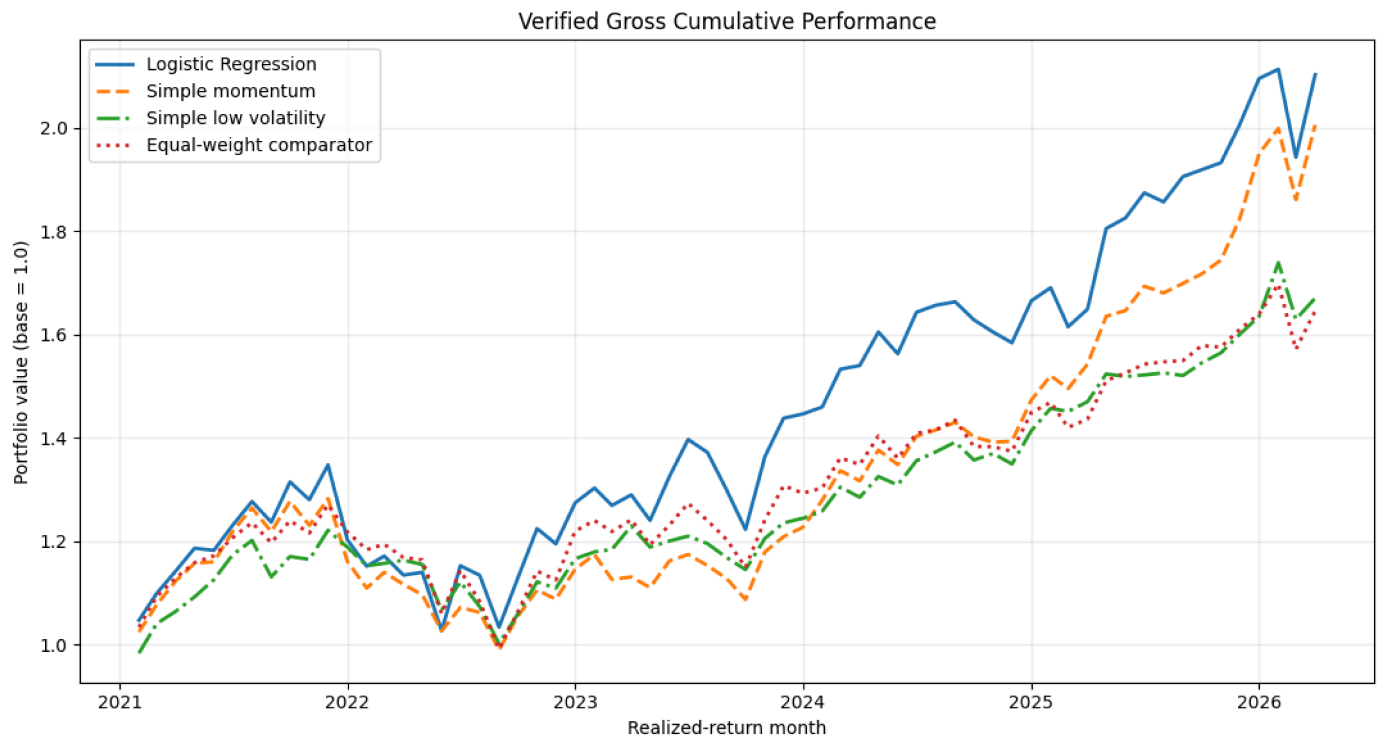


Figure 2. Cumulative gross performance of the logistic regression portfolio, simple momentum, simple low-volatility, and the internally constructed equal-weight comparator. Note: Portfolio values were indexed to 1.0 at the beginning of the out-of-sample evaluation period. Signals were formed from January 2021 to March 2026, and the corresponding forward portfolio returns were realized from February 2021 to April 2026. The horizontal axis reports realized-return months. Returns are gross of transaction costs. Source: Authors' calculations.

Table 5. Newey–West HAC inference for pairwise differences in mean monthly portfolio returns.

Comparison	Mean Monthly Difference	HAC Standard Error	HAC Statistic	p-Value	95% Confidence Interval
LR – Equal-weight	0.437%	0.216%	2.020	0.043	[0.013%, 0.862%]
LR – Momentum	0.122%	0.205%	0.594	0.552	[−0.280%, 0.523%]
LR – Ensemble	0.254%	0.115%	2.217	0.027	[0.029%, 0.479%]
Momentum – Ensemble	0.133%	0.223%	0.595	0.552	[−0.304%, 0.570%]
Momentum – Equal-weight	0.316%	0.231%	1.369	0.171	[−0.136%, 0.768%]
Ensemble – Equal-weight	0.183%	0.173%	1.055	0.291	[−0.157%, 0.523%]
Low volatility – Equal-weight	−0.001%	0.176%	−0.005	0.996	[−0.345%, 0.343%]

Note: Differences are calculated as the monthly return of the first strategy minus the monthly return of the second strategy. Newey–West heteroskedasticity- and autocorrelation-consistent standard errors use three monthly lags. Mean differences, standard errors, and confidence intervals are expressed in monthly percentage terms. The tests evaluate whether the mean monthly return difference is statistically different from zero. They do not test differences between Sharpe ratios. The reported Sharpe ratios are descriptive estimates of observed risk-adjusted performance.

The pairwise HAC tests did not establish that LR generated a higher mean monthly return than momentum or that momentum generated a higher mean monthly return than the ensemble. LR nevertheless generated a statistically significant higher mean monthly

return than both the equal-weight comparator ($p = 0.043$) and the ensemble ($p = 0.027$) at the 5% level. The remaining pairwise mean-return differences did not reach conventional statistical significance thresholds. Consequently, the cumulative-return rankings should still be interpreted as sample-specific estimates rather than as evidence of universal strategy superiority. These tests concern only differences in mean monthly return and do not establish whether the observed differences between Sharpe ratios are statistically significant. Momentum achieved the highest observed Sharpe ratio, but its apparent risk-adjusted advantage should therefore be interpreted descriptively rather than inferentially.

4.4. Target-Weight Turnover and Transaction-Cost Sensitivity

Gross returns do not fully capture the investment value of a strategy when portfolio rankings generate different trading intensities. Table 6 reports the average one-way target-weight turnover of each strategy and the geometric annualized return obtained after applying proportional transaction costs of 10, 25, and 50 basis points per unit of target-weight turnover. The indicator captures changes between consecutive target portfolio allocations and should not be interpreted as exact realized trading turnover.

Table 6. Average target-weight turnover and net performance under alternative transaction-cost assumptions.

Strategy	Average Monthly Target-Weight Turnover	Annualized Target-Weight Turnover	Gross Geometric Annualized Return	Net Geometric Annualized Return: 10 bps	Net Geometric Annualized Return: 25 bps	Net Geometric Annualized Return: 50 bps
Logistic Regression	50.23%	602.74%	15.22%	14.54%	13.54%	11.88%
Momentum	24.42%	293.08%	14.18%	13.86%	13.37%	12.56%
ML ensemble	54.58%	654.98%	11.88%	11.17%	10.11%	8.36%
Low volatility	14.52%	174.29%	10.27%	10.09%	9.81%	9.34%
Equal-weight comparator	0.06%	0.76%	9.96%	≈9.96%	≈9.96%	≈9.95%

Notes: Target-weight turnover is defined as one half of the sum of the absolute differences between portfolio target weights at two consecutive monthly portfolio-formation dates. Annualized target-weight turnover equals average monthly target-weight turnover multiplied by 12. Net returns are geometric annualized returns obtained after deducting the assumed proportional cost from each monthly gross return according to the corresponding target-weight turnover. The measure captures changes in target portfolio allocations but does not incorporate return-induced weight drift between rebalancing dates. It should therefore be interpreted as a selection-based trading-intensity proxy rather than as exact realized turnover. The cost scenarios do not include security-specific bid-ask spreads, taxes, commissions, market impact, capacity constraints, or implementation shortfall. The initial portfolio-formation month is excluded because no preceding target portfolio is available for comparison.

The target-weight turnover results provide an additional indication of the implementation implications of model parsimony. The ensemble generated the highest selection-based trading intensity, while logistic regression required considerably more frequent changes in target holdings than the momentum strategy. Under the stylized cost assumptions, LR retains a small annualized-return advantage at 10 basis points, this advantage nearly disappears at 25 basis points, and momentum generates the higher net annualized return at 50 basis points. These results indicate that the strongest gross-return strategy is not necessarily the strongest strategy after applying costs to changes in target allocations. However, because the turnover measure excludes return-induced weight drift, the resulting net returns should be interpreted as sensitivity estimates rather than as precise achievable performance.

These findings also clarify the meaning of incremental investment value in the present study. A model adds implementable value only if its predictive contribution is sufficiently

large and stable to compensate for the trading required to exploit changes in its rankings. Greater architectural complexity may therefore impose both governance costs and direct portfolio-implementation costs without necessarily producing a commensurate improvement in net performance.

The ensemble generated the highest average monthly target-weight turnover, at 54.58%, followed by Logistic Regression at 50.23%. Momentum required substantially fewer changes in target holdings, with average monthly target-weight turnover of 24.42%, while low volatility remained the most stable active selection strategy, at 14.52%. Under the 10-basis-point scenario, LR retains a net annualized-return advantage over momentum, generating 14.54% compared with 13.86%. At 25 basis points, LR generates 13.54%, compared with 13.37% for momentum, leaving only a small difference. At 50 basis points, momentum generates the higher net annualized return, 12.56% compared with 11.88% for LR. Therefore, LR's higher gross cumulative return does not imply an unambiguous implementation advantage.

The ensemble remains below both LR and momentum under all three cost assumptions. Its higher trading intensity further weakens the case that additional architectural complexity generates incremental implementable value. These results reinforce the distinction between gross predictive performance and net investment value.

4.5. Logistic Regression Ranking Diagnostics

To assess whether the logistic regression scores contain economically meaningful cross-sectional ranking information beyond classification accuracy, Table 7 reports forward returns by predicted-probability quintile and the monthly Spearman information coefficient.

Table 7. Forward returns by predicted-probability quintile and Spearman Information Coefficient.

Panel/Diagnostic	Estimate	HAC Standard Error	HAC Statistic	p-Value	95% Confidence Interval
Panel A. Mean monthly forward returns by LR-score quintile					
Q1 – lowest scores	0.727%	—	—	—	—
Q2	0.573%	—	—	—	—
Q3	0.794%	—	—	—	—
Q4	0.950%	—	—	—	—
Q5 – highest scores	1.307%	—	—	—	—
Q5 – Q1 spread	0.580%	0.351%	1.655	0.098	[−0.107%, 1.267%]
Panel B. Monthly Spearman Information Coefficient					
Mean IC	0.0257	0.0145	1.767	0.077	[−0.0028, 0.0541]
Standard deviation of monthly IC	0.1258	—	—	—	—
Months with positive IC	57.1%	—	—	—	—

Note: Stocks are sorted each month into quintiles using the logistic regression-predicted probability, from Q1 (lowest score) to Q5 (highest score). The Q5–Q1 spread and the mean monthly Spearman information coefficient are evaluated using Newey–West HAC standard errors with three monthly lags. Quintile returns, the spread, its standard error, and its confidence interval are expressed in monthly percentage terms.

Although the average return of Q2 is lower than that of Q1, forward returns subsequently increase across the LR score distribution, reaching 1.307% per month in the

highest-score quintile. The Q5 – Q1 spread equals 0.580% per month. Newey–West inference indicates that this spread is marginally significant at the 10% level but not at the 5% level (HAC statistic = 1.655, $p = 0.098$). The mean monthly Spearman information coefficient between the LR score and the subsequent return is 0.0257 and is positive in 57.1% of the test months. Its Newey–West p -value is 0.077. These results provide limited evidence that LR contains economically relevant ranking information, but they do not support claims of strong or uniformly stable predictability.

4.6. Annual Performance and Regime Sensitivity

Table 8 reports the annual gross performance of the Logistic Regression portfolio and the equal-weight comparator by realized-return calendar year. The Logistic Regression portfolio records a higher compounded return than the equal-weight comparator in every reported calendar period. As illustrated in Figure 3, this relative performance is observed in both positive- and negative-return periods. The 2021 result covers February–December, whereas the 2026 result covers January–April. These partial-period observations should not be interpreted as full-year returns.

Table 8. Annual gross performance by realized-return year.

Realized-Return Year	LR Portfolio	Equal-Weight Comparator	Arithmetic Difference
2021 *	34.79%	27.05%	+7.74 pp
2022	−11.35%	−11.45%	+0.10 pp
2023	20.36%	16.23%	+4.13 pp
2024	10.16%	5.08%	+5.08 pp
2025	26.68%	17.18%	+9.49 pp
2026 *	4.81%	2.23%	+2.58 pp

Note: Returns are assigned to the calendar month in which they are realized, rather than to the signal-formation month. * Partial calendar years: 2021 covers February–December 2021; 2026 covers January–April 2026. The arithmetic difference is the LR portfolio return minus the equal-weight comparator return, expressed in percentage points. Differences are calculated using unrounded return values; therefore, they may differ slightly from differences computed directly from the rounded figures displayed in the table.

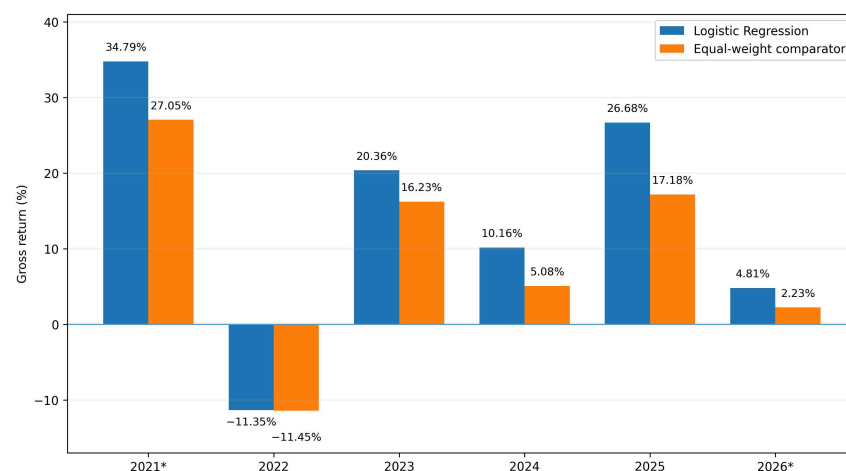


Figure 3. Annual gross portfolio returns: Logistic Regression versus equal-weight comparator. Note: Annual returns are compounded from monthly realized portfolio returns and are reported by realized-return calendar year. The asterisk (*) indicates a partial calendar year: 2021 covers February–December, whereas 2026 covers January–April. Returns are gross of transaction costs. Source: Authors' calculations.

The annual pattern is directionally favorable but heterogeneous. The LR portfolio exceeds the equal-weight comparator in every realized-return calendar period, although the difference in 2022 is only 0.10 percentage points. The arithmetic return difference ranges from 0.10 percentage points in 2022 to 9.49 percentage points in 2025. This variation is consistent with regime-dependent predictability and shows why cumulative outperformance should not be interpreted as evidence of a constant structural premium.

These annual results remain conditional on the fixed March 2026 constituent universe. They exclude firms that left the index before that date and therefore cannot be interpreted as the realized historical performance of an investable STOXX Europe 600 strategy. Because the annual comparisons are based on local-currency gross returns, they describe relative differences between selection methods within the shared ex post sample. Transaction-cost sensitivity is evaluated separately at the full-period strategy level in Section 4.4.

4.7. Robustness to Portfolio-Selection Thresholds

To assess whether the baseline results depend on the proportion of stocks included in each portfolio, Table 9 compares strategy performance across alternative selection thresholds of 10%, 20%, and 30%.

Table 9. Strategy performance across alternative portfolio-selection thresholds.

Selection Threshold	Strategy	Cumulative Return	Geometric Annualized Return	Annualized Volatility	Sharpe Ratio
10%	Logistic Regression	145.02%	18.61%	19.30%	0.832
10%	Momentum	114.78%	15.67%	15.06%	0.851
10%	ML ensemble	76.34%	11.41%	18.51%	0.517
10%	Low volatility	64.42%	9.93%	10.71%	0.665
20% (baseline)	Logistic Regression	110.36%	15.22%	17.15%	0.743
20% (baseline)	Momentum	100.63%	14.18%	13.47%	0.837
20% (baseline)	ML ensemble	80.32%	11.88%	16.64%	0.582
20% (baseline)	Low volatility	67.10%	10.27%	11.14%	0.671
30%	Logistic Regression	100.58%	14.18%	15.80%	0.734
30%	Momentum	90.44%	13.05%	13.13%	0.779
30%	ML ensemble	77.93%	11.60%	15.48%	0.597
30%	Low volatility	64.95%	10.00%	11.50%	0.632

Note: LR, momentum, and ensemble portfolios select the top 10%, 20%, or 30% of eligible stocks according to their respective ranking signals; the low-volatility strategy selects the corresponding bottom fraction by trailing volatility. Portfolios are equal-weighted and rebalanced monthly. Results are gross of transaction costs. The 20% threshold is the baseline specification used in the main analysis.

The main qualitative result is robust to alternative selection thresholds. LR generates the highest geometric annualized return at the 10%, 20%, and 30% thresholds, while momentum records the highest observed Sharpe ratio at each threshold. The ensemble does not outperform either LR or momentum in terms of geometric annualized return or Sharpe ratio at any of the evaluated thresholds. Thus, the non-monotonic relationship between model complexity and portfolio performance is not driven solely by the baseline 20% selection rule.

5. Discussion

5.1. Complexity Does Not Guarantee Incremental Value

The empirical results obtained in this study reveal a non-monotonic relationship between model complexity and investment performance. Among the evaluated strategies, logistic regression generated the highest gross cumulative return, while the transparent momentum rule achieved the highest observed Sharpe ratio. The ensemble combining logistic regression, random forest, and XGBoost underperformed both approaches. These findings show that increasing model complexity does not necessarily improve portfolio outcomes. In this context, the present study does not seek to determine whether machine learning is universally effective, but rather whether more complex models provide additional value beyond transparent investment signals already included in the feature set.

The ensemble result also shows that predictive diversification differs from portfolio diversification. Averaging LR, RF, and XGBoost probabilities reduces dependence on a single learner, but the component models share the same observations and correlated technical indicators. Their errors may therefore be too similar for averaging to create a strong ensemble gain.

The evidence also shows that the relative attractiveness of the strategies depends on the performance dimension considered. Logistic regression provides the best classification results among the three evaluated models and the highest gross cumulative return, whereas the transparent momentum strategy achieves the highest observed Sharpe ratio and substantially lower target-weight turnover. Once stylized proportional transaction costs are introduced, the gross return advantage of logistic regression narrows and eventually reverses under the highest cost scenario considered. Model complexity should therefore be evaluated jointly in terms of predictive performance, gross portfolio outcomes, risk-adjusted performance, and implementation intensity rather than against a single performance measure.

5.2. Implications for Financial-System Resilience

The results indicate that strategy robustness cannot be assessed solely through cumulative returns. Logistic regression generated the highest gross cumulative return, whereas the simple momentum strategy achieved the highest observed Sharpe ratio and the three-model ensemble underperformed both approaches. The realized-return calendar analysis further shows that the LR portfolio exceeded the equal-weight comparator in each reported calendar period, although the magnitude of this advantage varies substantially over time. The arithmetic difference ranges from 0.10 percentage points in 2022 to 9.49 percentage points in 2025. This temporal variation is consistent with regime-dependent predictability and indicates that cumulative outperformance should not be interpreted as evidence of a constant structural premium. Moreover, the relatively short out-of-sample period does not cover a sufficiently broad range of market regimes to establish the long-term robustness of any strategy.

Implementation robustness provides an additional dimension to this comparison. Logistic regression and the ML ensemble generated substantially higher target-weight turnover than the transparent momentum and low-volatility strategies. Consequently, transaction costs affect the relative economic attractiveness of the strategies differently. Logistic regression retains a small annualized-return advantage over momentum under the 10-basis-point cost scenario, the two strategies become economically similar at 25 basis points, and momentum generates the higher net annualized return at 50 basis points. These results show that a model that dominates in gross cumulative performance need not dominate after implementation frictions are considered. In this sense, parsimony

may contribute to investment-system resilience not only through interpretability and governance, but also through lower implementation intensity.

The analysis also shows that data quality can materially influence portfolio results. The screening procedure identified a small number of extreme observations associated with adjustment discontinuities, corporate actions, ticker-history problems, or potentially genuine extreme returns. Removing these observations improved the consistency of the analytical dataset, but the use of publicly available adjusted prices does not guarantee that all data-quality problems were identified. Moreover, the retrospective use of the March 2026 STOXX Europe 600 constituents introduced survivorship bias. Consequently, the reported results should be interpreted as a controlled comparison of selection methods within a common ex post sample rather than as evidence of historically investable performance.

The prominence of momentum-related variables in the machine-learning models and the strong performance of the transparent momentum strategy suggest that the evaluated models may rely on a common underlying market signal. This finding raises a potential robustness concern; models with different computational structures may appear diversified while remaining economically exposed to the same return pattern. Nevertheless, the present study does not directly measure market-wide model adoption, correlated trading, crowding, or systemic risk. Any implications concerning herding and market-level resilience should therefore be treated as plausible mechanisms consistent with the findings and concerns identified by asset managers [1], rather than as effects demonstrated empirically by this analysis.

At the institutional level, the comparison provides support for selecting model complexity in relation to demonstrated incremental value. In this study, the three-model ensemble did not outperform either logistic regression or the transparent momentum rule, despite its greater computational complexity. A simpler approach may therefore offer advantages in interpretation, validation, monitoring, and communication without necessarily sacrificing performance. This conclusion nevertheless remains conditional on the restricted technical information set and the simplified implementation framework. Although target-weight turnover is reconstructed and proportional transaction-cost scenarios are evaluated, these scenarios do not capture security-specific spreads, market impact, taxes, commissions, liquidity constraints, or implementation shortfall. The results therefore support a proportionate approach to model complexity, while stopping short of establishing the precise net performance that could be achieved in institutional implementation. Because the target-weight turnover measure does not incorporate return-induced drift in portfolio weights, these comparisons capture differences in selection-driven implementation intensity rather than the total amount of trading that would be required in live portfolio management.

5.3. European Governance Implications

The governance implications of this study arise directly from the differences observed among the evaluated models. Logistic regression achieved the best classification performance among the three evaluated models and the highest gross cumulative return, while the more complex three-model ensemble did not produce an additional performance advantage. These results indicate that greater computational complexity may increase validation and monitoring requirements without necessarily improving portfolio outcomes. They do not imply that investment-selection models are automatically classified as high-risk systems under the EU AI Act. Rather, they show why the use of machine learning in investment processes should remain subject to the organizational, conduct, record-keeping, conflict-management, and client-interest requirements applicable to investment firms [30].

The empirical analysis also identifies specific governance issues relevant to institutional implementation. First, the extreme observations detected in the public price data

demonstrate the importance of documenting data sources, corporate-action adjustments, ticker changes, and data-cleaning decisions. Second, the modest predictive out-of-sample performance of all three models highlights the need to define validation boundaries and monitoring criteria for model deterioration. Third, the stronger performance of logistic regression relative to random forest, XGBoost, and the ensemble shows that interpretability may be achieved without necessarily accepting lower performance. Because the coefficients of logistic regression can be examined directly, institutions can more easily identify the contribution of individual predictors and explain changes in stock rankings. Tree-based models and multi-model ensembles may require additional interpretation, validation, and monitoring tools.

The prominence of momentum-related variables and the highest observed Sharpe ratio achieved with the transparent momentum strategy also provide a direct link between the empirical findings and model governance. Although the machine-learning models combine several technical indicators, their predictions may still depend largely on an established momentum effect. Institutions should therefore assess not only the technical architecture of a model but also its underlying economic exposures. A model that is computationally complex may offer limited diversification if it primarily reproduces a familiar investment signal. The target-weight turnover and transaction-cost sensitivity results reinforce this governance argument; additional model complexity should be justified not only by improvements in predictive metrics or gross returns, but also by incremental performance that remains economically meaningful after accounting for the changes in portfolio target allocations generated by the model. Because the turnover measure excludes return-induced weight drift and other implementation frictions, these results should be interpreted as comparative sensitivity evidence rather than as precise estimates of realized cost.

5.4. Limitations

Several limitations require the results to be interpreted with caution. Although the analysis includes HAC inference, target-weight turnover, transaction-cost sensitivity, and alternative portfolio-selection thresholds, these procedures cannot eliminate the limitations associated with the ex post universe, the relatively short evaluation period, and the simplified representation of real-world implementation conditions.

First, the sample was constructed using the companies included in the STOXX Europe 600 in March 2026 and was applied retrospectively to the entire analysis period. Consequently, companies removed from the index before that date were not represented, which introduced survivorship bias. This bias may overstate absolute portfolio returns and may affect the relative performance of the evaluated strategies differently.

Second, the test period contained only 63 monthly forward-return observations associated with signals formed from January 2021 to March 2026, which is too short to represent the full range of market regimes. Relatedly, the limited number of out-of-sample monthly observations constrains statistical power. Although Newey–West HAC inference accounts for heteroskedasticity and serial dependence in monthly return differences, several economically meaningful estimates do not reach conventional statistical significance thresholds. The portfolio rankings should therefore be interpreted primarily as sample-specific comparative evidence rather than as estimates of persistent population-level superiority.

Third, this study reconstructs target-weight turnover from consecutive security-level portfolio selections and evaluates net performance under proportional transaction-cost assumptions of 10, 25, and 50 basis points per unit of one-way target-weight turnover. This measure captures changes in selected securities and target allocations but does not account for the drift in portfolio weights caused by security returns between rebalancing dates. It may therefore understate the total amount of trading required to restore equal

portfolio weights, particularly during periods of high cross-sectional return dispersion. The assumed costs are also applied uniformly across securities, strategies, countries, trading venues, and market conditions and do not account separately for security-specific bid–ask spreads, brokerage commissions, taxes, market impact, implementation shortfall, liquidity constraints, or portfolio capacity. Accordingly, the reported net returns should be interpreted as selection-based transaction-cost sensitivity estimates rather than as precise measures of institutionally achievable performance.

Fourth, the feature set contained technical price indicators only. Fundamental, macroeconomic, sentiment, liquidity, and ESG variables could provide information that better exploits nonlinear models.

Fifth, public-source adjusted prices, local-currency quotations, and ticker mapping introduce measurement risk. No conversion into a common currency was performed, so the results exclude exchange-rate effects and do not represent the performance of a euro-denominated portfolio. The extreme-return filter may remove both data errors and genuine extreme observations, while also failing to detect less visible discontinuities.

Sixth, the models were evaluated under a fixed temporal split. Relationships can change as market participants adapt. Rolling or expanding estimation would provide a stronger evaluation of adaptation, but this is outside the scope of the present study.

6. Conclusions

This study examines whether increasing machine-learning complexity generates additional investment value in European equity selection. All three evaluated models produced AUC-ROC values slightly above the random-ranking reference, but their overall predictive performance remained weak. The linear logistic regression model achieved the best classification results among the three evaluated models and generated the highest gross cumulative portfolio return. At the same time, the transparent 12–1 momentum strategy achieved the highest observed Sharpe ratio and outperformed the ensemble combining all three models. These results therefore identify a non-monotonic relationship between model complexity and investment performance; more sophisticated architectures do not automatically produce superior portfolio outcomes.

The main contribution of the study is to distinguish the ability of machine-learning models to organize and rank technical signals from their capacity to generate value beyond transparent investment rules. The prominence of momentum-related variables and the performance of the simple momentum strategy suggest that the evaluated models draw substantially on economically familiar momentum-related information. These findings suggest that additional machine-learning complexity is most likely to create incremental value when it captures economically relevant relationships that are not already represented by simpler and more interpretable strategies.

From a systems perspective, the findings show that architectural diversification does not necessarily produce economic diversification. Although the three-model ensemble combines linear and nonlinear approaches, all component models use the same observations and closely related technical indicators. Their shared exposure to momentum-related information may help explain the limited diversification benefit observed when their predictions are combined. This result extends the portfolio comparison beyond a simple ranking of strategies and demonstrates that the resilience of an investment system depends on the diversity of its underlying economic signals, not only on the number or complexity of its models.

The implementation analysis further qualifies the gross-performance results. Logistic regression achieved the highest gross cumulative return, but its higher target-weight turnover makes its relative advantage sensitive to proportional costs applied to changes in

target portfolio allocations. Because the turnover proxy excludes return-induced portfolio-weight drift, the cost analysis provides a comparative measure of selection-driven implementation sensitivity rather than an estimate of exact realized trading costs. Momentum combined the highest observed Sharpe ratio with substantially lower target-weight turnover and generated the higher net annualized return under the highest proportional-cost scenario considered. At the same time, Newey–West inference does not establish a statistically significant mean-return advantage of logistic regression over momentum. Taken together, these results indicate that incremental investment value should be assessed jointly through predictive quality, gross performance, risk-adjusted outcomes, statistical uncertainty, and implementation costs.

This study also highlights the institutional relevance of model parsimony. In the evaluated setting, logistic regression provides stronger predictive and portfolio results than the nonlinear alternatives while remaining easier to interpret, validate, and monitor. From an institutional perspective, additional model complexity is most readily justified when its incremental predictive or economic contribution is sufficient to compensate for the additional requirements associated with data management, validation, monitoring, implementation, and governance. This conclusion links the empirical findings directly to the design of resilient and institutionally feasible investment systems.

The results should be interpreted within the study’s methodological boundaries. The sample is based on the March 2026 STOXX Europe 600 constituents and therefore contains survivorship and look-ahead bias. Firms removed before that date are absent, and the reported outcomes are not the realized historical returns of an investable index strategy. The evaluation period covers a limited number of market regimes, and the analysis relies exclusively on technical price indicators. Implementation costs are represented through stylized proportional transaction-cost scenarios rather than security-specific realized execution costs. In addition, securities are evaluated in their local quotation currencies without conversion into a common currency. These constraints limit generalization and direct implementation. The findings should therefore be interpreted as comparative evidence on portfolio-selection methods within a shared ex post universe rather than as precise estimates of institutionally achievable, euro-denominated investment performance.

Future research should test whether the identified relationship between model complexity and investment performance persists when using point-in-time index membership, longer evaluation periods, rolling or expanding estimation windows, security-specific transaction-cost and liquidity measures, common-currency portfolio returns, and broader fundamental, macroeconomic, liquidity, sentiment, and ESG information. Particular attention should be given to whether nonlinear models and ensembles generate stronger incremental value when the information set contains economically diverse predictors rather than predominantly technical signals. Future work could also examine whether combining machine learning with economically distinct signals, portfolio constraints, and human oversight produces a more robust balance between predictive performance, implementation efficiency, interpretability, and institutional resilience.

Author Contributions: Conceptualization, F.M.B. and A.Ț.; methodology, F.M.B., A.Ț. and G.G.N.; software, A.Ț.; formal analysis, A.Ț.; investigation, A.Ț.; data curation, A.Ț.; writing-original draft preparation, F.M.B., A.Ț. and G.G.N.; writing-review and editing, F.M.B. and G.G.N.; supervision, F.M.B. All authors have read and agreed to the published version of the manuscript.

Funding: West University of Timișoara supported the publication of this article with funding from the CNFIS-FDI-2026-F-0369 project.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The analysis uses monthly auto-adjusted price data obtained from Yahoo Finance through the yfinance Python library. The data supporting the conclusions of this article can be made available by the corresponding author upon request.

Acknowledgments: The authors acknowledge the institutional support of West University of Timișoara.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Mercer. *How Artificial Intelligence Is Shaping Asset Management*; Mercer: New York, NY, USA, 2026. Available online: <https://www.mercer.com/en-us/about/newsroom/how-artificial-intelligence-is-shaping-asset-management/> (accessed on 3 August 2026).
2. Arthur, W.B. Complexity and the economy. *Science* **1999**, *284*, 107–109. [CrossRef]
3. Lo, A.W. The adaptive markets hypothesis. *J. Portf. Manag.* **2004**, *30*, 15–29. [CrossRef]
4. Farmer, J.D.; Lo, A.W. Frontiers of finance: Evolution and efficient markets. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9991–9992. [CrossRef]
5. Gu, S.; Kelly, B.; Xiu, D. Empirical asset pricing via machine learning. *Rev. Financ. Stud.* **2020**, *33*, 2223–2273. [CrossRef]
6. Zografopoulos, L.; Iannino, M.C.; Psaradellis, I.; Sermpinis, G. Industry return prediction via interpretable deep learning. *Eur. J. Oper. Res.* **2025**, *321*, 257–268. [CrossRef]
7. González Cortés, D.; Onieva, E.; Pastor, I.; Trinchera, L.; Wu, J. Portfolio construction using explainable reinforcement learning. *Expert Syst.* **2024**, *41*, e13667. [CrossRef]
8. Fama, E.F. Efficient capital markets: A review of theory and empirical work. *J. Financ.* **1970**, *25*, 383–417. [CrossRef]
9. Shleifer, A.; Vishny, R.W. The limits of arbitrage. *J. Financ.* **1997**, *52*, 35–55. [CrossRef]
10. Biais, B.; Foucault, T.; Moinas, S. Equilibrium fast trading. *J. Financ. Econ.* **2015**, *116*, 292–313. [CrossRef]
11. Kirilenko, A.; Lo, A.W. Moore’s law versus Murphy’s law: Algorithmic trading and its discontents. *J. Econ. Perspect.* **2013**, *27*, 51–72. [CrossRef]
12. Jegadeesh, N.; Titman, S. Returns to buying winners and selling losers: Implications for stock market efficiency. *J. Financ.* **1993**, *48*, 65–91. [CrossRef]
13. Carhart, M.M. On persistence in mutual fund performance. *J. Financ.* **1997**, *52*, 57–82. [CrossRef]
14. Fama, E.F.; French, K.R. Size, value, and momentum in international stock returns. *J. Financ. Econ.* **2012**, *105*, 457–472. [CrossRef]
15. Ang, A.; Hodrick, R.J.; Xing, Y.; Zhang, X. The cross-section of volatility and expected returns. *J. Financ.* **2006**, *61*, 259–299. [CrossRef]
16. George, T.J.; Hwang, C.Y. The 52-week high and momentum investing. *J. Financ.* **2004**, *59*, 2145–2176. [CrossRef]
17. De Bondt, W.F.M.; Thaler, R. Does the stock market overreact? *J. Financ.* **1985**, *40*, 793–805. [CrossRef]
18. Krauss, C.; Do, X.A.; Huck, N. Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *Eur. J. Oper. Res.* **2017**, *259*, 689–702. [CrossRef]
19. Li, Y.; Xue, H.; Wei, S.; Wang, R.; Liu, F. A machine learning approach for investigating the determinants of stock price crash risk: Exploiting firm and CEO characteristics. *Systems* **2024**, *12*, 143. [CrossRef]
20. Gao, N.; Liu, Y.; He, Y.; Zhang, J.; Zhang, L. A novel portfolio selection method via deep reinforcement learning. *Systems* **2026**, *14*, 292. [CrossRef]
21. Yu, S.; Yang, S.-B.; Yoon, S.-H. The design of an intelligent lightweight stock trading system using deep learning models: Employing technical analysis methods. *Systems* **2023**, *11*, 470. [CrossRef]
22. Cheng, C.-H.; Tsai, M.-C.; Chang, C. A time series model based on deep learning and integrated indicator selection method for forecasting stock prices and evaluating trading profits. *Systems* **2022**, *10*, 243. [CrossRef]
23. López de Prado, M. *Advances in Financial Machine Learning*; Wiley: Hoboken, NJ, USA, 2018.
24. Bailey, D.H.; Borwein, J.M.; López de Prado, M.; Zhu, Q.J. The probability of backtest overfitting. *J. Comput. Financ.* **2017**, *20*, 39–69. [CrossRef]
25. Harvey, C.R.; Liu, Y.; Zhu, H. . . . and the cross-section of expected returns. *Rev. Financ. Stud.* **2016**, *29*, 5–68. [CrossRef]
26. Bagnara, M. Asset pricing and machine learning: A critical review. *J. Econ. Surv.* **2024**, *38*, 27–56. [CrossRef]
27. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [CrossRef]
28. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016*; ACM: New York, NY, USA, 2016; pp. 785–794. [CrossRef]
29. Černevičienė, J.; Kabašinskas, A. Explainable artificial intelligence (XAI) in finance: A systematic literature review. *Artif. Intell. Rev.* **2024**, *57*, 216. [CrossRef]

30. European Securities and Markets Authority. *Public Statement on the Use of Artificial Intelligence in the Provision of Retail Investment Services*; ESMA35-335435667-5924; ESMA: Paris, France, 2024. Available online: https://www.esma.europa.eu/sites/default/files/2024-05/ESMA35-335435667-5924_Public_Statement_on_AI_and_investment_services.pdf (accessed on 3 August 2026).
31. European Parliament and Council. Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence. *Off. J. Eur. Union* **2024**, 2024, 1689. Available online: <https://eur-lex.europa.eu/eli/reg/2024/1689/oj> (accessed on 3 August 2026).
32. Bagattini, G.; Piazza, F. *Artificial Intelligence in EU Investment Funds: Adoption, Strategies and Portfolio Exposures*; ESMA TRV Risk Analysis; ESMA50-43599798-9923; European Securities and Markets Authority: Paris, France, 2025. Available online: https://www.esma.europa.eu/sites/default/files/2025-02/ESMA50-43599798-9923_TRV_Article_Artificial_intelligence_in_EU_investment_funds.pdf (accessed on 3 August 2026). [[CrossRef](#)]
33. Novonesis. *The Combination of Novozymes and Chr. Hansen Is Now Successfully Completed, Creating Novonesis—A Leading Global Biosolutions Partner*; Novonesis: Copenhagen, Denmark, 2024. Available online: <https://www.novonesis.com/en/news/combination-of-novozymes-and-chr-hansen-now-successfully-completed> (accessed on 31 August 2026).
34. Nasdaq Derivatives Markets. *Anticipated Change of Name and Ticker Code for Novozymes*; Product Information 67/24; Nasdaq: Stockholm, Sweden, 2024. Available online: <https://view.news.eu.nasdaq.com/view?id=b6e45c347e97ae345487178dcb750f544&lang=en> (accessed on 31 August 2026).
35. STOXX Ltd. *STOXX Europe 600 Index*; STOXX Ltd.: Zug, Switzerland, 2026. Available online: <https://www.stoxx.com/index-details?symbol=SXXP> (accessed on 3 August 2026).
36. Grinold, R.C.; Kahn, R.N. *Active Portfolio Management*, 2nd ed.; McGraw-Hill: New York, NY, USA, 1999.
37. Markowitz, H. Portfolio selection. *J. Financ.* **1952**, 7, 77–91. [[CrossRef](#)]
38. Sharpe, W.F. Capital asset prices: A theory of market equilibrium under conditions of risk. *J. Financ.* **1964**, 19, 425–442. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.