

Machine Learning, Valuation Signals, and Portfolio Return Forecasting: Evidence from the Pakistan Stock Exchange



Sabahat Mateen ^a Muhammad Naveed Jan ^b Abdul Ghafoor Khan ^c

Abstract: *This study aims to analyze whether the use of machine-learning models provides better out-of-sample forecasts for the returns on characteristic-sorted portfolios on the Pakistan Stock Exchange (PSX) and to determine whether the E/P ratio and P/D ratio offer incremental predictability beyond the Fama-French five-factor (FF5) model. This study uses 264 months of data (July 2003-June 2025) on non-financial KSE-100 constituents to benchmark FF5/OLS regression models against the XGBoost and LightGBM gradient-boosting models in 18 size-by-B/M portfolios using a 22-variable feature set (FF5 factors and lags, macroeconomic variables, and valuation ratios) under five-fold walk-forward validation approach. XGBoost and LightGBM generate significantly lower out-of-sample forecast errors relative to the FF5/OLS benchmark model; Diebold-Mariano tests support significant improvement in 14 of 18 portfolios for XGBoost and 15 of 18 for LightGBM. FF5 captures a small amount of variation in return (average $R^2 = 24.1\%$). However, including E/P and P/D increases R^2 to 24.8%, where E/P is the second most significant explanatory variable in the XGBoost regression model. This study demonstrates that machine learning algorithms can supplement traditional asset pricing models based on factors in an emerging market economy and that valuation variables still have some predictive power.*

Keywords: Machine Learning, Asset Pricing, Portfolio Return Forecasting, XGBoost, LightGBM, Fama-French Five-Factor Model, Earnings Yield, Price-To-Dividend Ratio, Pakistan Stock Exchange, Emerging Markets

Introduction

Portfolio return forecasting is important in the process of investment decisions, risk management, and capital allocation, and it has played a key role in the development of asset pricing research since the CAPM and the multifactor models that followed it (Sharpe, 1964; Ross, 1976; Merton, 1973). The FF5 by Fama & French (2015) continues to be used as the standard linear model for the cross-section of expected returns after the prior three-factor version (Fama & French, 1993) and Carhart's (1997) momentum model. These models rely on the fact that a limited number of linear risk factors sufficiently explain return variation, which has become more and more controversial given increasing evidence of nonlinearity, interaction effects, and time-varying risk premia that could be effectively captured using flexible modeling techniques (Gu et al., 2020; Chen et al., 2021). Recent studies have revealed that gradient boosting models like XGBoost and LightGBM can enhance return prediction performance out of sample through modeling the complex relationship and interaction effects between firm features, factor exposure, and macroeconomic variables that could not be captured by linear regressions (Gu et al., 2020; Bryzgalova et al., 2023; Kelly et al., 2019).

Whether such benefits hold in an environment such as that on the Pakistan Stock Exchange (PSX) remains an unanswered and highly relevant question. Indeed, emerging market environments exhibit a combination of lower trading volume, incomplete information disclosure, and high volatility in macroeconomic variables

^a Research Scholar, Department of Management Sciences, COMSATS University Islamabad, Abbottabad Campus, Khyber Pakhtunkhwa, Pakistan.

^b Assistant Professor, Department of Management Sciences, COMSATS University Islamabad, Abbottabad Campus, Khyber Pakhtunkhwa, Pakistan.

^c Lecturer, Department of Economics, COMSATS University Islamabad, Abbottabad Campus, Khyber Pakhtunkhwa, Pakistan.

such as inflation, interest rate, exchange rate, and money supply growth, along with the identical factor structure as assumed in models built for developed markets, so that a linear model calibrated based on relatively stable factor return relationships could be ill-suited for a situation in which factor return relationships vary with macroeconomic regimes and interact with the characteristics of the firm (Feng et al., 2020; Kozak et al., 2020). If there exist thresholds, interactions, or time-varying loadings within the return-generating process on the PSX that cannot be captured by a linear FF5 regression model, the forecast errors would be systematically larger compared to those under a nonlinear model.

PSX provides a valuable context within which this issue can be studied. First, it is a medium-liquidity emerging stock market that is prone to episodes of currency devaluation, inflation, and changes in monetary policies, and secondly, its non-financial listed companies are characterized by a great variety of value regimes throughout more than two decades, including the global financial crisis in 2008 and further inclusion in MSCI Frontier and Emerging Markets indices. Thus, the PSX provides a challenging field of testing linear factor models on one hand, and an opportunity to see whether there still remains some superiority of machine learning techniques under the conditions of macroeconomic uncertainty and scarcity of information, on the other hand.

Existing PSX-focused research provides only partial evidence on this question. Several studies have applied machine-learning and related methods to the PSX, but the emphasis has typically fallen on predicting the direction of the KSE-100 index or of individual share prices rather than on the returns of characteristic-sorted portfolios (Ali et al., 2021; Munir et al., 2024; Jadoon et al., 2024; Zafar, 2025). Other work has compared machine-learning forecasts against factor benchmarks for individual PSX securities (Raza et al., 2024), extended the FF5 framework with a human-capital factor estimated with machine-learning methods (Siddiqui et al., 2025; Zada et al., 2025), or compared artificial-neural-network and FF5 forecasts for Pakistani equities (Jan & Ayub, 2019). What remains largely untested is the joint predictive contribution of valuation characteristics — specifically the earnings-to-price ratio (E/P) and the price-to-dividend ratio (P/D) — together with firm characteristics, Fama-French factors and macroeconomic variables, to one-month-ahead, portfolio-level return forecasts evaluated with time-ordered, walk-forward validation and both statistical and economic performance criteria. The present study examines two research questions as follows. RQ1 asks how effectively XGBoost and LightGBM predict monthly portfolio returns on the PSX relative to an FF5/OLS benchmark, using firm-level, factor and macroeconomic information. RQ2 asks whether E/P and P/D, used both as portfolio-sorting characteristics and as predictive features, provide statistically significant incremental information for subsequent monthly portfolio returns beyond the FF5 factors. Consistent with value-investing theory and the Gordon growth model, the associated hypothesis is that a higher E/P (a lower price relative to earnings) and a lower P/D (a higher dividend yield) are associated with higher subsequent returns, although the machine-learning specifications allow this relationship to be nonlinear and conditional on the prevailing macroeconomic environment.

The study makes four contributions. Empirically, it provides portfolio-level, rather than stock- or index-level, evidence on return predictability from an under-studied frontier/emerging market over a 22-year sample (July 2003-June 2025). Methodologically, it benchmarks two gradient-boosting algorithms, XGBoost and LightGBM, against an FF5/OLS specification using identical, walk-forward out-of-sample validation rather than in-sample fit. On the information side, it isolates the incremental explanatory and predictive contribution of E/P and P/D once standard risk factors, their lags, and macroeconomic variables are already included. Finally, it integrates statistical forecast-accuracy testing (mean absolute error and the Diebold-Mariano test) with economic-performance measures (the Sharpe ratio and FF5 alpha), a combination that is necessary because a statistically more accurate forecast need not translate into an economically exploitable trading advantage.

The remainder of the article is organized as follows. Section 1 reviews the relevant asset-pricing and machine-learning literature and develops the study's hypotheses. Section 2 describes the data, sample,

portfolio construction and empirical methodology. Section 3 reports the results. Section 4 discusses the findings, their theoretical and practical implications, and the distinction between statistical and economic significance. The final section concludes and notes the study's limitations.

Literature Review and Theoretical Development

Traditional Asset Pricing

Return prediction begins with Sharpe's (1964) capital asset pricing model, whereby securities are priced based on their covariance with respect to the market portfolio only. Arbitrage pricing theory, as suggested by Ross (1976), is an extension of CAPM that relaxes the assumption of CAPM by including several unspecified risk factors. The intertemporal version of CAPM by Merton (1973) adds to CAPM state variables representing changes in the investment opportunity set; this is important especially when applied to stock markets like PSX. Fama and French (1993) operationalized a multifactor structure empirically with size (SMB) and value (HML) factors added to the market factor, later extended by Carhart (1997) to include momentum. Fama and French (2015) proposed the five-factor model used as the benchmark in this study, adding profitability (RMW) and investment (CMA) factors motivated by the dividend-discount model. These models remain widely used because they are parsimonious, economically interpretable, and estimable by ordinary least squares (OLS), but they impose a linear, time-invariant relationship between factor exposures and expected returns. The growing number of proposed factors (Feng et al., 2020) and the shrinkage methods used to select among them (Kozak et al., 2020) point to unresolved model uncertainty even in developed markets, motivating the search for more flexible specifications.

Machine Learning in Asset Pricing

Machine-learning methods relax the linearity and additivity assumptions of factor models. Gu, Kelly, and Xiu (2018, 2020) provide a benchmark comparison of machine-learning methods, including gradient-boosted trees and neural networks, against OLS for equity return prediction, finding that the more flexible methods achieve materially larger out-of-sample R^2 by capturing nonlinear characteristic interactions. Embedding the no-arbitrage assumption within the deep learning algorithm is Chen et al. (2021), building test asset portfolios from cross-sectional data of characteristics using decision trees is Bryzgalova et al. (2023), and modelling factor loadings as a function of characteristics is Kelly et al. (2019). The use of tree-based ensembles in financial data with structured and tabular nature makes gradient boosting especially useful: XGBoost and LightGBM train a sequence of small trees on the residuals of previously trained trees using regularization techniques to avoid overfitting. Moreover, the leaf-wise growing approach of LightGBM and its histogram-based splitting make LightGBM computationally efficient for a high number of features (Guennioui et al., 2024; Dezhkam et al., 2023). Comparative studies show that gradient boosting models significantly outperform linear models as well as other types of machine learning models (support vector machines (Huang et al., 2005; Ballings, Van den Poel et al., 2015) and LSTM networks (Fischer & Krauss, 2018)) in prediction tasks of returns on stocks and portfolios (Sak et al., 2024; Sarioglu Duran et al., 2025; Manogna & Kulkarni, 2025; Zhu, 2025; Wang, Stefanidis, Jiang, & Su, 2025; Pagliaro, The consistent theme across this literature is not that machine learning replaces factor models, but that it can recover a portion of the return variation a linear factor specification leaves unexplained (Fang & Taylor, 2021; Fang et al., 2023; Zhang et al., 2025).

Valuation Features

Earnings yield (E/P) and the price-dividend ratio (P/D) occupy a dual role in this literature: they are simultaneously portfolio-sorting characteristics, in the spirit of Fama and French's value factor, and candidate predictive features usable inside a forecasting model. Value-investing theory and the Gordon growth model imply that, holding growth and risk constant, a higher earnings yield signals relative undervaluation and should be associated with higher subsequent returns, while a higher P/D ratio (equivalently, a lower dividend yield) implies investors are paying a higher premium for a given stream of cash dividends and should be

associated with lower subsequent returns. Because E/P and P/D are constructed from annual accounting and market data, their relationship with monthly returns need not be linear or stable: the value effect can interact with size, profitability and macroeconomic conditions — precisely the kind of conditional relationship that tree-based ensembles can represent but that a single OLS coefficient cannot. This motivates testing E/P and P/D in two complementary specifications: as additional linear regressors in an augmented FF5 regression, where coefficient sign and incremental R^2 can be evaluated directly, and as inputs to the machine-learning feature set, where their contribution is assessed through feature-importance rankings and out-of-sample forecast accuracy rather than a single linear coefficient.

Macroeconomic Variables

Firm-level and factor-based predictors do not operate in isolation from the macroeconomic environment, particularly in an emerging market subject to periodic currency depreciation and inflationary pressure. Consumer price inflation, policy interest rate, PKR/USD exchange rate, and money supply (M2) growth influence the discount rates, input cost, and risk-taking behavior of the investors, respectively, and have been found influential for stock return variability in South Asian stock markets (Poudel, 2025; Shakya, 2024). Since these variables vary with the same periodicity as the monthly portfolio returns, and since the effect of these variables may also depend on the states of the economy – such as when there is currency depreciation or monetary tightening – they make for suitable features for building a machine-learning algorithm.

Research Gap and Hypothesis Development

Taken together, the literature supports three propositions that motivate the present design. First, linear factor models such as FF5 are likely to leave a meaningful share of PSX portfolio-return variation unexplained, given the market's macroeconomic volatility and the well-documented tendency of linear models to underfit non-linear return processes (Gu et al., 2020). Second, gradient-boosting methods, which have consistently outperformed linear and alternative machine-learning benchmarks in other markets, are plausible candidates for closing part of this gap on the PSX (Guennioui et al., 2024; Iftikhar et al., 2024). Third, E/P and P/D are likely to retain incremental predictive content once standard factors are controlled for, although the linear evidence may understate their importance if the underlying relationship is nonlinear. Formally:

H0 (RQ2): E/P and P/D contain no statistically significant incremental information for subsequent monthly portfolio returns, net of the FF5 factors.

H1 (RQ2): E/P and/or P/D provide statistically significant incremental information, with higher E/P and lower P/D associated with higher subsequent returns.

Existing PSX evidence, reviewed above, has concentrated on index direction, individual-stock returns or classification tasks (Ali et al., 2021; Munir et al., 2024; Zafar, 2025; Jadoon et al., 2024) rather than on the one-month-ahead performance of characteristic-sorted portfolios benchmarked jointly against FF5/OLS with time-ordered validation, statistical forecast-comparison tests and economic performance measures. This combination — portfolio-level forecasting, an explicit FF5 benchmark, two independently validated gradient-boosting algorithms, and both statistical and economic evaluation criteria — defines the gap the present study addresses.

Methodology

Research Design

This is a quantitative, empirical study designed around out-of-sample forecasting rather than in-sample explanation. It combines a linear asset-pricing benchmark (FF5/OLS) with two supervised machine-learning algorithms (XGBoost and LightGBM) to forecast one-month-ahead returns of characteristic-sorted portfolios, and evaluates all three models using identical, time-ordered validation and a common held-out test design, applying both statistical (MAE, Diebold-Mariano) and economic (Sharpe ratio, FF5 alpha) performance criteria.

Population and Sample

The target population comprises all non-financial companies listed on the Pakistan Stock Exchange between 2003 and 2025; banks, insurance companies and investment companies are excluded because their regulatory capital requirements, balance-sheet composition and reporting practices differ systematically from those of non-financial firms. The sampling frame will be that of the KSE-100 Index, which is PSX's benchmark index comprising the largest and most liquid companies on the exchange, ensuring that the analytical sample contains firms with sufficient trading and information for portfolio analysis. Firms will qualify to be part of the analytical sample if they have at least three years of monthly returns and annual financial statements; otherwise, they are filtered out. The resulting return series covers the period July 2003 to June 2025 (264 months), with the five FF5 factors available from September 2003 (262 monthly observations); the underlying panel of firm-year/firm-month observations totals 3,911.

Monthly, rather than daily, data are used for four reasons. First, the accounting-based characteristics used in portfolio formation and as predictors — E/P, P/D, book-to-market, operating profitability and asset growth — update annually and change far more slowly than daily prices, so daily returns would repeatedly pair unchanged accounting information with many price observations without adding genuine information. Second, the macroeconomic predictors (inflation, the policy rate, money supply and the exchange rate) are available or economically interpreted at monthly frequency. Third, monthly returns mitigate thin trading, bid-ask bounce and non-synchronous pricing, which are more pronounced at daily frequency in an emerging market. Fourth, a one-month-ahead horizon is consistent with the annual FF5 factor-construction and rebalancing convention used throughout the study.

Data and Variables

The dependent variable is the monthly return of each characteristic-sorted portfolio. Predictors fall into four groups: firm-level characteristics (market capitalization, book-to-market, operating profitability and asset growth), valuation ratios (E/P and P/D, measured at the KSE-100 market level from annual data and matched to the monthly return series), macroeconomic variables (CPI inflation, the SBP policy rate, the PKR/USD exchange rate, M2 money supply and the trade balance), and the five contemporaneous and two-period-lagged Fama-French factors (MKT, SMB, HML, RMW and CMA). Table 1 summarizes the definitions, measurement, and sources.

Table 1

Variables, Measurement and Data Sources

Variable	Definition / measurement	Source
Portfolio return (dependent)	Monthly return of each of the 18 size-by-B/M portfolios	PSX; own calculation
Market capitalisation (size)	Price $\times$ shares outstanding	PSX
Earnings-to-price (E/P)	Earnings per share $\div$ market price (1/P–E)	Company annual reports; PSX
Book-to-market (B/M)	Book value of equity $\div$ market value of equity	Annual reports
Price-dividend (P/D)	Market price $\div$ dividend per share (100 / dividend yield)	PSX; annual reports
Asset growth (AG)	(Total assets _t – total assets _{t-1}) $\div$ total assets _{t-1}	Annual reports
Operating profitability (OP)	Operating income $\div$ total assets	Financial statements
Trade balance	Exports – imports	Pakistan Bureau of Statistics

Variable	Definition / measurement	Source
CPI inflation	Monthly consumer price index	Pakistan Bureau of Statistics
Policy rate	SBP target policy rate (monthly)	State Bank of Pakistan (SBP)
Exchange rate	Monthly average PKR per USD	SBP
Money supply (M2)	Monthly M2 level / growth	SBP
MKT	Value-weighted market return minus the risk-free rate (T-bill)	PSX; SBP
SMB	Size factor: average small-firm return minus average large-firm return	Constructed (Fama & French, 2015)
HML	Value factor: high-B/M return minus low-B/M return	Constructed
RMW	Profitability factor: robust minus weak operating profitability	Constructed
CMA	Investment factor: conservative minus aggressive asset growth	Constructed

Note: FF5 factor construction follows Fama and French (2015). All accounting variables are lagged to the most recently available annual statement prior to the portfolio-formation date.

Portfolio Formation

Following standard Fama-French sorting logic, at each annual formation date eligible firms are independently sorted into three size groups (market-capitalization terciles) and six book-to-market groups (sextile breakpoints), producing 18 size-by-book-to-market test portfolios, labelled (size group, B/M group) from (1,1) to (3,6). The SMB, HML, RMW and CMA factors are separately constructed from parallel 2×3 independent sorts on size, book-to-market, operating profitability and asset growth, following Fama and French (2015). Portfolio membership is fixed for the subsequent 12-month holding period, and monthly constituent returns are used to compute both factor and test-portfolio returns; the FF5 factor portfolios are value-weighted, consistent with standard factor-construction practice, while the 18 test portfolios used for forecasting are equally weighted.

To avoid look-ahead bias, only accounting information publicly available before the formation date enters the feature set: market capitalization is measured at the formation date itself, and annual accounting variables are taken from the most recently available financial statements and lagged accordingly. A firm-year data point with non-positive book equity is not considered in the book-to-market portfolio sorting (and therefore not part of HML) during the particular year instead of being provided with an arbitrary figure; firm-year data points with a non-positive denominator resulting in undefined ratios such as E/P and P/D are considered not available for that particular characteristic instead of being arbitrarily assigned a value. E/P and P/D, calculated annually at the market level of the KSE-100, are included as additional valuation features of the portfolio in both the augmented regression analysis and machine learning, together with the size-by-book-to-market sorting characteristics.

Fama-French Five-Factor Benchmark

For each of the 18 portfolios, excess returns are regressed on the five contemporaneous FF5 factors:

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_1 \text{MKT}_t + \beta_2 \text{SMB}_t + \beta_3 \text{HML}_t + \beta_4 \text{RMW}_t + \beta_5 \text{CMA}_t + \varepsilon_{i,t}$$

where $R_{f,t}$ is the monthly Pakistan T-bill rate and α_i (Jensen's alpha) measures the average monthly abnormal return left unexplained by the five factors; the null hypothesis $\alpha_i = 0$ is tested for each portfolio. An augmented specification adds the annual E/P and P/D ratios as additional regressors:

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_1 \text{MKT}_t + \dots + \beta_5 \text{CMA}_t + \beta_6 \text{EP}_t + \beta_7 \text{PD}_t + \varepsilon_{i,t}$$

A positive β_6 is consistent with value-investing theory (higher earnings yield, higher subsequent return); the incremental R^2 between the baseline and augmented specifications indicates how much additional explanatory power the valuation ratios contribute beyond the five standard factors.

XGBoost and LightGBM

XGBoost and LightGBM are gradient-boosting ensembles that build an additive sequence of shallow regression trees, each fitted to the gradient of the loss function with respect to the current ensemble's predictions, with explicit regularization of tree complexity to control overfitting in a relatively short (264-month) sample. XGBoost uses a level-wise tree-growth strategy and second-order gradient information within a regularized objective; LightGBM instead grows trees leaf-wise and uses histogram-based feature binning, which improves computational efficiency and lets it handle a larger feature set without a proportional increase in training time (Guennioui et al., 2024). Both models are trained on a 22-feature vector comprising the five contemporaneous FF5 factors, ten lagged factor values (two lags each), four macroeconomic variables, and three valuation-ratio variables (E/P, P/D and earnings yield); estimators are set to 150 trees, maximum depth 4, learning rate 0.05 and subsample rate 0.8 for both models. Results are reported both with and without the valuation-ratio block to isolate its marginal contribution to forecast accuracy.

Walk-Forward Forecasting

All three models are evaluated using time-ordered, walk-forward validation rather than randomized cross-validation, which would otherwise allow information from future periods to influence training and bias accuracy estimates upward. The sample is partitioned into five sequential folds; in each fold the model is trained only on data preceding the test period and then used to forecast returns in the subsequent, unseen period, after which the training window advances and the process repeats. This approach is consistent with how a forecaster would actually employ the model and guarantees that the accuracy reported truly is out-of-sample accuracy. The procedure is conducted in the same way for the OLS, XGBoost, and LightGBM models in order to avoid any biases in the comparison.

Evaluation Metrics

Forecast accuracy is measured with the mean absolute error, $MAE = (1/T) \sum |R_t - \hat{R}_t|$, which expresses the average absolute deviation between forecast and realized portfolio returns in the original return units and is robust to the sign changes and near-zero values that make percentage-based metrics such as MAPE difficult to interpret for monthly equity returns. Statistical differences in forecast accuracy between model pairs are tested with the Diebold-Mariano (DM) test, based on the loss differential $d_t = L(e_{1,t}) - L(e_{2,t})$ between two models' squared forecast errors, with heteroskedasticity- and autocorrelation-consistent (Newey-West / HAC) standard errors; a significantly negative DM statistic indicates the machine-learning model produces significantly lower forecast error than the OLS/FF5 benchmark. Because a lower forecast error need not imply a superior investment strategy, economic performance is assessed separately using the annualised Sharpe ratio — computed from the portfolio's mean monthly return in excess of a 0.50%-per-month (6% per annum) risk-free rate, scaled by its monthly standard deviation and annualised by the square root of 12 — together with the FF5 alpha estimated for each portfolio. Feature importance for the extended machine-learning models is reported as XGBoost's gain-based importance and LightGBM's normalized split-count importance, averaged across all 18 portfolios.

Results

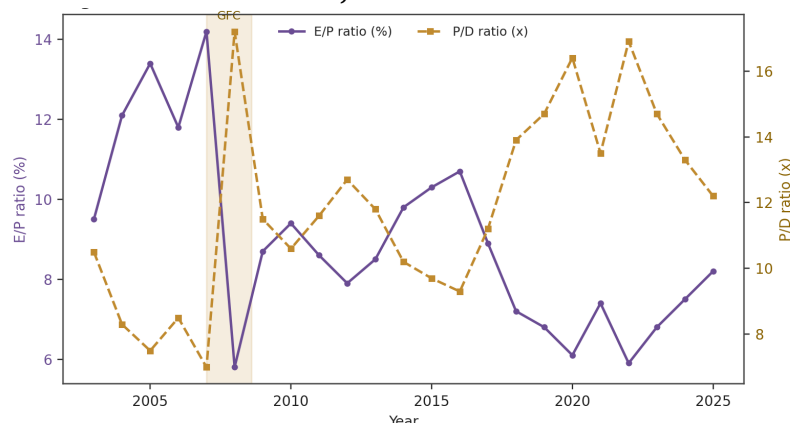
Descriptive Evidence

KSE-100 market-level valuation ratios display a pronounced cyclical pattern over the sample period (Figure 2). The E/P ratio (earnings yield) reached a sample maximum of 14.2% in 2007, ahead of the 2008 global financial crisis, before falling sharply to a sample minimum of 5.8% in 2008 as the crisis compressed

valuations; the P/D ratio moved inversely over the same period, rising from 7.0x in 2007 to 17.2x in 2008. E/P falls from 2016 forward while P/D rises, which is to be expected given the steady deterioration of the PKR as reflected by this data set. Indeed, another local minimum for E/P is achieved in 2020 (E/P = 6.1%) and maximum for P/D is achieved in 2020 (P/D = 16.4x). Overall, in the sample we find the mean E/P ratio to be 8.9% (SD = 2.3) and mean P/D to be 11.9x (SD = 2.9).

Figure 1

KSE-100 E/P and P/D Ratios, 2003-2025



Valuation Ratios and Correlations

Correlations between the E/P and P/D ratios of the annual year and the monthly returns of the portfolios (Table 2; $N = 262$ for each portfolio) suggest that there is a clear, albeit not consistently statistically significant pattern. The ratio of E/P correlates positively with returns in 15 out of 18 portfolios; it is significant ($p < 0.05$ or $p < 0.10$) in five of those 15 portfolios. Similarly, the ratio of P/D correlates negatively with returns in 16 out of 18 portfolios; it is significant in seven portfolios. The strongest correlations seem to be found among portfolios like (1,1), (1,4), (2,3), and (3,1), where correlations are significant and of the right sign as expected by the theory.

Table 2

Pearson Correlations Between Valuation Ratios and Portfolio Returns (Selected Portfolios)

Portfolio	corr(E/P, Ret)	p-value	corr(P/D, Ret)	p-value	N
1,1	0.1927**	0.002	-0.2207**	0.000	262
1,4	0.1783**	0.004	-0.1816**	0.003	262
2,3	0.1529**	0.013	-0.1639**	0.008	262
2,5	0.0921	0.137	-0.1212*	0.050	262
3,1	0.0765	0.217	-0.1426**	0.021	262
3,5	0.1291**	0.037	-0.1499**	0.015	262
3,6	0.1052*	0.089	-0.1622**	0.009	262

Note: * $p < 0.10$, ** $p < 0.05$, two-tailed. Full 18-portfolio results are available from the authors. Portfolio labelled (size group, B/M group), size 1=small-3=large, B/M 1=low-6=high.

FF5 and Augmented FF5 Results

The R^2 for the FF5/OLS regression is fairly low among the 18 test portfolios, varying from 0.063 (portfolio 1,2) to 0.426 (portfolio 3,1) with an average of 0.241. By including the valuation ratios, E/P, and P/D in the regression, the R^2 becomes 0.248, which is an increase of 0.7 percentage points, and while small in absolute magnitude, the increase in the R^2 is observed in all of the 18 portfolios (see Table 3 & Figure 3). The sign of the E/P and P/D coefficients in the augmented regression is not uniform across portfolios: E/P is negatively signed in 14 of 18 portfolios and P/D in 12 of 18, which at first appears inconsistent with the positive-E/P, negative-P/D pattern found in the simple correlations. This is not necessarily a contradiction: the augmented

regression estimates a contemporaneous partial relationship after controlling for five correlated risk factors, whereas the correlation and machine-learning evidence evaluate E/P and P/D as predictive, lagged characteristics; a single linear coefficient, estimated jointly with five collinear factors, is a considerably less complete summary of the valuation-return relationship than the combination of correlation, incremental R^2 and feature importance considered together. Statistically meaningful positive alpha remains in 8 of the 18 portfolios even after the five factors are included, indicating the linear specification leaves an economically non-trivial share of average returns unexplained; this is best read not as evidence that FF5 is invalid, but as establishing a relatively undemanding linear benchmark against which the nonlinear machine-learning models can be evaluated.

Table 3

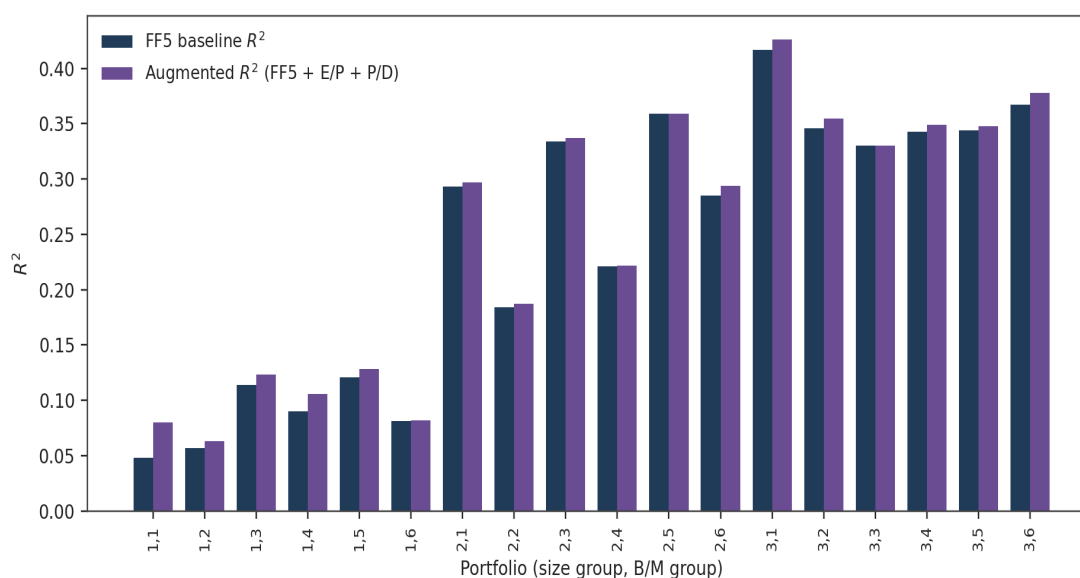
FF5 Baseline vs Augmented (FF5 + E/P + P/D) OLS Regression Results, 18 Portfolios (Selected)

Portfolio	Alpha	t(α)	b EP	t(EP)	b PD	t(PD)	R^2 (aug.)	ΔR^2
1,1	0.1425	1.25	-0.0034	-0.56	-0.0181	-1.60	0.080	+0.032
2,1	0.0131	0.14	-0.0015	-0.30	0.0014	0.15	0.297	+0.004
2,5	0.0050	0.05	-0.0001	-0.02	0.0008	0.08	0.359	+0.000
3,1	0.2636	2.16	-0.0123	-1.90	-0.0244	-2.01	0.426	+0.009
3,4	0.1998	1.45	-0.0102	-1.39	-0.0149	-1.09	0.349	+0.006
3,6	0.3351	2.13	-0.0150	-1.80	-0.0328	-2.10	0.378	+0.011
Average (18)	0.0952	—	-0.0062	—	-0.0094	—	0.248	+0.007

Note: Portfolio labelled (size group [1=small-3=large], B/M group [1=low-6=high]). ΔR^2 = gain over FF5 baseline. Full 18-portfolio table available from the authors.

Figure 2

FF5 Baseline vs. Augmented (FF5 + E/P + P/D) R^2 . 18 Portfolios



Machine-Learning Forecast Accuracy

Table 4 reports the out-of-sample MAE of the XGBoost and LightGBM specifications, estimated with and without the E/P, P/D and earnings-yield block, under five-fold walk-forward validation. Average MAE is 0.0648 for XGBoost and 0.0651 for LightGBM using the macro-and-factor feature set; adding the valuation-ratio block leaves average MAE for both models essentially unchanged (0.0655 for XGBoost, 0.0652 for LightGBM), with the effect varying by portfolio — MAE falls in some portfolios and rises marginally in others once the valuation ratios are added, most consistently among mid- and large-capitalisation portfolios. This indicates that the marginal contribution of E/P and P/D to point-forecast accuracy, measured by MAE alone, is modest and portfolio-specific; their more consequential contribution, operates through feature importance

and the augmented-regression R^2 rather than through a uniform reduction in absolute forecast error. The comparison against the FF5/OLS benchmark itself is evaluated formally through the Diebold-Mariano test a directly comparable, portfolio-by-portfolio point-estimate MAE series for the OLS benchmark was not available in the underlying project documentation, so the DM test — which compares squared forecast errors between model pairs directly — provides the primary and most reliable statistical evidence on the OLS-versus-machine-learning comparison.

Table 4

Out-of-Sample MAE, XGBoost and LightGBM, With and Without Valuation Features (Selected Portfolios)

Portfolio	XGB (macro)	XGB (+E/P)	Δ XGB	LGB (macro)	LGB (+E/P)	Δ LGB
1,1	0.05798	0.05934	−2.34%	0.06066	0.06135	−1.14%
1,6	0.05237	0.05513	−5.28%	0.05536	0.05604	−1.22%
2,4	0.05128	0.05165	−0.71%	0.04780	0.04719	+1.27%
3,1	0.06740	0.06916	−2.62%	0.06924	0.07008	−1.22%
3,4	0.08585	0.08540	+0.52%	0.08255	0.08320	−0.79%
Average (18)	0.06478	0.06554	−1.26%	0.06512	0.06519	−0.10%

Note: 'macro' = FF5 factors + lags + macroeconomic variables; '+E/P' = macro set plus E/P, P/D and earnings yield. Positive Δ = MAE reduced (improvement) after adding valuation features. Five-fold walk-forward cross-validation. Full 18-portfolio table available from the authors.

Diebold-Mariano Tests

The Diebold-Mariano test compares the OLS/FF5 benchmark against each machine-learning model using HAC-corrected squared-error loss differentials (Table 5; Figure 4). XGBoost produces a significantly negative DM statistic — indicating significantly lower forecast error than OLS — in 14 of 18 portfolios (77.8%) at the 5% or 10% level, and LightGBM in 15 of 18 (83.3%). The largest improvements occur in portfolios (2,5) and (3,4), with DM statistics of −4.112 (XGBoost) and −4.399 (LightGBM) respectively, both significant at $p < 0.001$. Only portfolios (1,2), (1,4) and (1,6) fail to show a significant improvement for either model, and these are concentrated among small-capitalization portfolios, where returns are noisier and sample information more limited. Because the DM test is conducted within the same walk-forward, out-of-sample design used for the MAE comparison, and because the improvement is concentrated in a large majority of portfolios rather than a small subset, the result is unlikely to reflect chance variation in a single evaluation window.

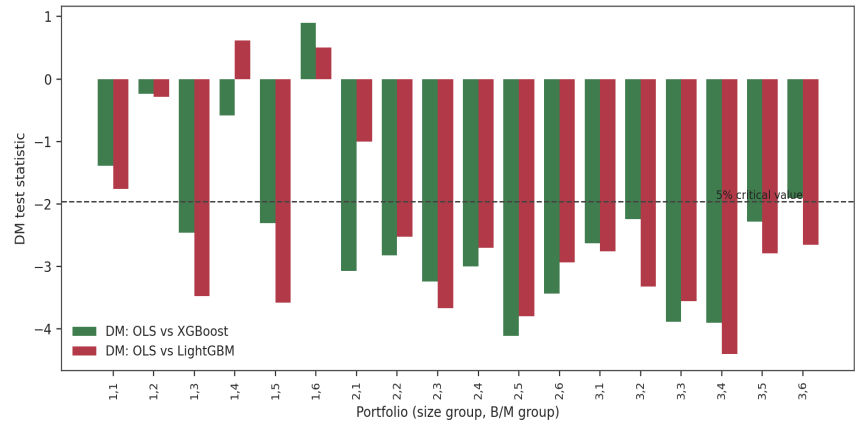
Table 5

Diebold-Mariano Test Statistics, OLS vs XGBoost and OLS vs LightGBM (Selected Portfolios)

Portfolio	DM (OLS vs XGB)	p-value	Sig.	DM (OLS vs LGB)	p-value	Sig.
1,3	−2.459	0.014	**	−3.474	0.001	**
2,1	−3.074	0.002	**	−1.005	0.315	
2,5	−4.112	0.000	**	−3.796	0.000	**
3,1	−2.633	0.008	**	−2.758	0.006	**
3,4	−3.905	0.000	**	−4.399	0.000	**
3,6	−1.906	0.057	*	−2.656	0.008	**

Note: negative DM $\Rightarrow$ machine-learning model more accurate than OLS. HAC (Newey-West) variance correction. * $p < 0.10$, ** $p < 0.05$. Full 18-portfolio table available from the authors.

Figure 3
Diebold-Mariano Test Statistics, OLS v XGBoots/LightGBM (18 Portfolios)



Economic Performance

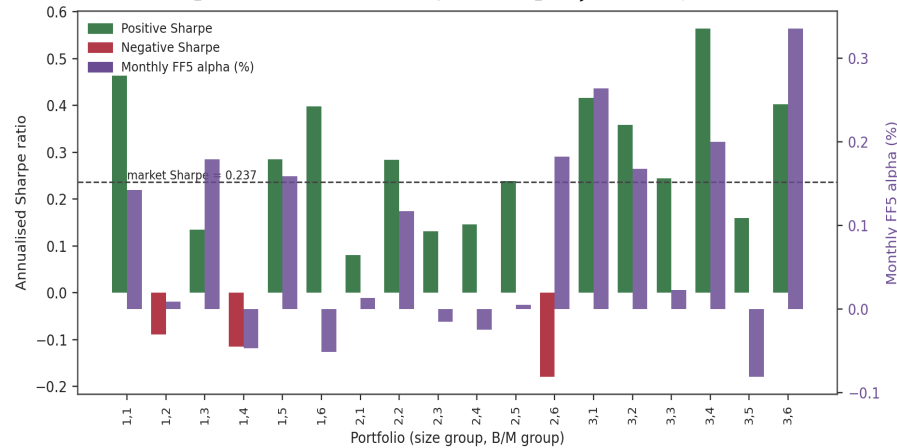
Table 6 and Figure 5 report annualised Sharpe ratios and monthly FF5 alphas for each of the 18 portfolios (risk-free rate = 0.50% per month). Risk-adjusted performance varies substantially: portfolio (3,4) achieves the highest annualised Sharpe ratio (0.564), exceeding the market Sharpe ratio of 0.237, while three portfolios (1,2), (1,4) and (2,6) — record negative Sharpe ratios over the sample period. There is no clear association between a portfolio's Sharpe ratio and its FF5 alpha: portfolios with strong risk-adjusted performance do not systematically also carry large abnormal returns once the five factors are controlled for, suggesting the two measures capture different, largely independent dimensions of performance. Neither the statistically superior machine-learning forecasts nor a higher Sharpe ratio should be read as direct evidence of an implementable abnormal-profit opportunity, particularly once transaction costs and market-impact costs — not incorporated into this analysis — are considered.

Table 6
Annualised Sharpe Ratios and Monthly FF5 Alpha, 18 Portfolios (Selected; July 2003–June 2025)

Portfolio	Mean monthly return	Monthly SD	Sharpe ratio (ann.)	FF5 alpha (monthly, %)
1,1	0.0163	0.0842	0.4639	0.1425
1,2	0.0025	0.0953	−0.0891	0.0091
2,6	0.0006	0.0849	−0.1792	0.1820
3,1	0.0187	0.1143	0.4164	0.2636
3,4	0.0248	0.1215	0.5642	0.1998
MKT	0.0099	0.0720	0.2370	—
Mean (18)	0.0121	0.1001	0.2181	0.0952

Note: Sharpe ratio annualised by multiplying by $\sqrt{12}$. Full 18-portfolio table available from the authors.

Figure 4
Annualised Sharpe Ratio and Monthly FF5 Alpha, 18 Portfolios



Feature Importance

Feature-importance scores, averaged across the 18 portfolios (Table 7; Figure 6), show the market factor (MKT) is the single most important predictor in both models. E/P is the second most important feature in the XGBoost model (importance = 0.0613), ranking ahead of every macroeconomic variable, and P/D ranks seventh (0.0482); together with a related valuation feature, the valuation variables account for approximately 15.6% of total XGBoost feature importance. Money supply (M2), the exchange rate, the policy rate and CPI inflation all rank among the ten most important XGBoost features, confirming that macroeconomic conditions carry meaningful predictive content alongside firm-level and factor-based information. The LightGBM ranking broadly corroborates this pattern for the market factor and several FF5-factor lags, although the relative importance attached to specific macroeconomic and valuation features differs between the two algorithms — an expected consequence of their different splitting and importance-normalization conventions (gain-based for XGBoost, split-count-based for LightGBM) rather than evidence of disagreement about which broad categories of information matter.

Table 7

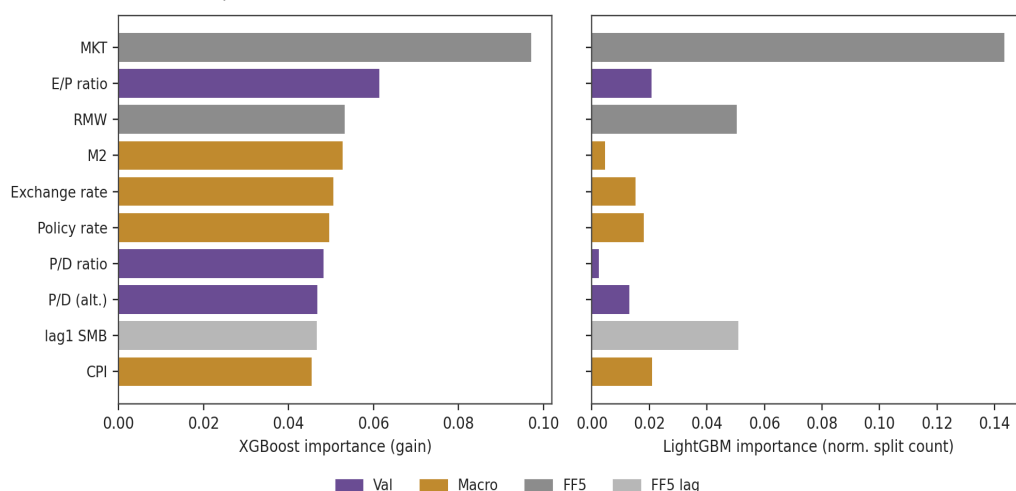
Feature Importance, Extended Machine-Learning Model (22 Features, Top 10, Averaged Across 18 Portfolios)

Feature	XGBoost importance	LightGBM importance	XGBoost rank	Category
MKT	0.0971	0.1435	1	FF5 factor
E/P ratio	0.0613	0.0208	2	Valuation
RMW	0.0532	0.0504	3	FF5 factor
M2	0.0527	0.0046	4	Macro
Exchange rate	0.0506	0.0152	5	Macro
Policy rate	0.0496	0.0180	6	Macro
P/D ratio	0.0482	0.0024	7	Valuation
P/D (earnings-yield form)	0.0468	0.0130	8	Valuation
Lag-1 SMB	0.0467	0.0510	9	FF5 lag
CPI	0.0454	0.0209	10	Macro

Note: XGBoost = gain-based importance; LightGBM = split-count importance normalised to sum to 1, averaged across 18 portfolios. Full 22-feature table available from the authors.

Figure 5

Feature Importance, Top 10 to 22 Features (Averaged across 18 Portfolios)



Discussion

Main Findings

Two consistent findings emerge. First, both XGBoost and LightGBM produce significantly lower out-of-sample forecast error than the linear FF5/OLS benchmark in a clear majority of the 18 test portfolios, a result confirmed by the Diebold-Mariano test rather than by in-sample fit alone — the more demanding and more decision-relevant standard of evidence. Second, the explanatory power offered by E/P and P/D is marginal

but still adds to the augmented linear regression equation, and they are some of the top predictors in the machine-learning feature-importance ranking, indicating that the characteristics that make up valuation still have predictive power for the PSX portfolio returns.

Theoretical Contribution

The findings above can be viewed not as a counterargument but rather in complement to the factor-based approach. Low R^2 (24.1%) achieved in the FF5 model together with the continued existence of significant alpha in nearly half of the test portfolios demonstrate that there is some unexplained variance of returns in the case of a linear and time-invariant specification of factors; this justifies, instead of refuting, the adoption of machine learning methods that are capable of modeling nonlinear thresholds and interactions between characteristics. The paper contributes to empirics and methodology (provides an example of where and by how much a nonlinear specification performs better compared to a linear one) rather than theory.

Emerging-Market Implications

The structural characteristics of the PSX – lower volume trading, incomplete information disclosure, higher sensitivity to currency and inflation effects, and limited ability for arbitrage compared to the advanced markets – provide a reasonable explanation as to why the linear model does not perform well and why macroeconomic factors have predictive value. This is consistent with the broader emerging-market literature linking macroeconomic volatility to stock-return variation (Poudel, 2025; Shakya, 2024) and suggests that machine-learning forecasts calibrated for one emerging market should not be assumed to transfer directly to another without re-estimation, given that the specific macroeconomic drivers and their relative importance are themselves market-specific.

Managerial and Investment Implications

For portfolio managers and analysts, the results suggest gradient-boosting forecasts can usefully supplement, rather than replace, factor-based analysis, particularly for portfolios where FF5 alpha remains large and unexplained. Valuation screens based on E/P and P/D remain a reasonable input to portfolio selection, but the modest incremental R^2 and portfolio-specific MAE results indicate such screens are more informative when combined with macroeconomic context than when used in isolation. For institutions considering machine-learning-based forecasting more broadly, the results point to the importance of walk-forward (rather than randomized) validation, benchmark comparison against a transparent linear model, and periodic monitoring of feature importance and forecast error as governance safeguards against overfitting and unstable, black-box predictions. The findings also point, more tentatively, to the value of timelier and more standardized disclosure of the earnings, dividend and macroeconomic data on which such valuation and forecasting signals depend.

Statistical versus Economic Significance

A central caution is that statistical forecast superiority, economic performance and investment profitability are three distinct constructs. A significantly negative DM statistic establishes that a model's squared forecast error is lower than the benchmark's; it does not by itself establish that the improvement is large enough, or stable enough, to generate an exploitable trading strategy net of transaction costs. Similarly, a portfolio's high Sharpe ratio in this sample does not establish that a live, implementable strategy built on the machine-learning forecasts would achieve the same risk-adjusted performance once transaction costs, market impact and portfolio-turnover constraints — none of which are incorporated here — are taken into account.

Limitations and Future Research

Several limitations qualify these findings. E/P and P/D are measured at annual frequency, limiting the variation available to a monthly forecasting model; the PSX-listed non-financial universe is comparatively small relative to developed markets, which can introduce construction noise into factor and portfolio returns;

the 264-month walk-forward sample is modest for training gradient-boosting models, so results may be sensitive to hyper-parameter choices; portfolios are equally, rather than value, weighted; transaction costs, market impact and trading restrictions are not incorporated into the Sharpe-ratio analysis; and the construction does not explicitly correct for survivorship bias, which could overstate realised portfolio returns if delisted firms are removed from the sample non-randomly. Some suggestions for future work include constructing high frequency (quarterly and trailing twelve-month) valuation ratios, studying returns at the stock level and double sorting for trading purposes, investigating differences in predictability before and after the MSCI reclassification in 2017, considering value-weighted portfolios and transaction cost implications, and using survivorship bias adjustment and valuation metrics by sectors.

Conclusions

In this study, we investigate whether machine learning algorithms perform better than OLS in out-of-sample prediction of characteristic-sorted portfolio returns on the Pakistan Stock Exchange (PSE). Whether the two ratios earnings yield (E/P) and price dividend (P/D) contribute to the explanatory power of Fama French five-factor model. The analysis is conducted using 264 months of data from July 2003 to June 2025 for nonfinancial firms of KSE 100 index firms sorted into 18 size book-to-market portfolios.

The results support two main conclusions. First, both XGBoost and LightGBM produce statistically significantly lower forecast error than the FF5/OLS benchmark, confirmed by Diebold-Mariano tests in 14 of 18 portfolios for XGBoost and 15 of 18 for LightGBM, indicating that the return-generating process on the PSX contains nonlinearities and characteristic interactions that a linear factor model does not fully capture. Second, E/P and P/D contain incremental information for portfolio returns: they raise the average explanatory power of the augmented FF5 regression from 24.1% to 24.8%, and E/P ranks as the second most important predictor, after the market factor, in the XGBoost feature-importance results, ahead of every macroeconomic variable considered.

The study's contribution is empirical and methodological rather than theoretical: it does not propose a new asset-pricing theory, and it does not establish that FF5 is invalid, but it provides portfolio-level evidence, from an under-studied frontier/emerging market, that gradient-boosting machine-learning methods can usefully complement conventional factor-based asset pricing when forecasting accuracy, rather than in-sample explanatory power, is the criterion of interest. For portfolio managers, risk managers and financial-market participants, the results support incorporating valuation information into portfolio decisions, developing governance frameworks for machine-learning-based forecasting (validation, benchmarking, monitoring) and improving the timeliness and standardization of the earnings, dividend and macroeconomic disclosures on which such forecasts depend.

Several qualifications should guide the interpretation of these findings. The relationships documented here are predictive, not causal: the models identify information that improves forecasts within this study's design, not mechanisms by which changes in valuation ratios or macroeconomic variables produce changes in future returns. A statistically significant forecast-accuracy improvement is not equivalent to a profitable, implementable trading strategy, since transaction costs, market impact and portfolio turnover are not incorporated into the analysis. The sample's annual valuation-ratio frequency, moderate size, equally weighted portfolio construction and absence of an explicit survivorship-bias correction all constrain the generalizability of the results beyond the specific sample and period studied. Within these bounds, the evidence indicates that machine learning is best understood as a complement to, rather than an automatic replacement for, conventional asset-pricing models in the Pakistan Stock Exchange.

References

- Ali, M., Khan, D. M., Aamir, M., Ali, A., & Ahmad, Z. (2021). Predicting the direction movement of financial time series using artificial neural network and support vector machine. *Complexity*, 2021, Article 2906463. <https://doi.org/10.1155/2021/2906463>
- Ballings, M., Van den Poel, D., Hespeels, N., & Gryp, R. (2015). Evaluating multiple classifiers for stock price direction prediction. *Expert Systems with Applications*, 42(20), 7046–7056. <https://doi.org/10.1016/j.eswa.2015.05.013>
- Bryzgalova, S., Pelger, M., & Zhu, J. (2023). Forest through the trees: Building cross-sections of stock returns. *The Journal of Finance*, 78(5), 2835–2883. <https://doi.org/10.1111/jofi.13477>
- Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, 52(1), 57–82. <https://doi.org/10.1111/j.1540-6261.1997.tb03808.x>
- Chen, L., Pelger, M., & Zhu, J. (2021). Deep learning in asset pricing. *Management Science*, 67(12), 7149–7176. <https://doi.org/10.1287/mnsc.2023.4695>
- Dezhkam, A., et al. (2023). Forecasting stock market for an efficient portfolio by combining XGBoost and Hilbert–Huang transform. *Engineering Applications of Artificial Intelligence*, 118, Article 105616. <https://doi.org/10.1016/j.engappai.2022.105616>
- Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1), 3–56. [https://doi.org/10.1016/0304-405X\(93\)90023-5](https://doi.org/10.1016/0304-405X(93)90023-5)
- Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1), 1–22. <https://doi.org/10.1016/j.jfineco.2014.10.010>
- Fang, M., & Taylor, S. (2021). A machine learning based asset pricing factor model comparison on anomaly portfolios. *Economics Letters*, 204, Article 109919. <https://doi.org/10.1016/j.econlet.2021.109919>
- Fang, Y., Chen, Y., & Ren, H. (2023). A factor pricing model based on machine learning algorithm. *International Review of Economics & Finance*, 88, 280–297. <https://doi.org/10.1016/j.iref.2023.06.012>
- Feng, G., Giglio, S., & Xiu, D. (2020). Taming the factor zoo: A test of new factors. *The Journal of Finance*, 75(3), 1327–1370. <https://doi.org/10.1111/jofi.12883>
- Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2), 654–669. <https://doi.org/10.1016/j.ejor.2017.11.054>
- Gu, S., Kelly, B., & Xiu, D. (2018). Empirical asset pricing via machine learning. *National Bureau of Economic Research Working Paper Series*, Article 25398. <https://doi.org/10.3386/w25398>
- Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5), 2223–2273. <https://doi.org/10.1093/rfs/hhaa009>
- Guennioui, O., et al. (2024). Improving global stock market prediction with XGBoost and LightGBM machine learning models. *Review of Economics and Finance*, 21, 2603–2610. <https://doi.org/10.55365/1923.x2023.21.278>
- Huang, W., Nakamori, Y., & Wang, S.-Y. (2005). Forecasting stock market movement direction with support vector machine. *Computers & Operations Research*, 32(10), 2513–2522. <https://doi.org/10.1016/j.cor.2004.03.016>
- Iftikhar, H., Khan, M., Turpo-Chaparro, J. E., Canas Rodrigues, P., & López-Gonzales, J. L. (2024). Forecasting stock prices using a novel filtering-combination technique: Application to the Pakistan stock exchange. *AIMS Mathematics*, 9(2), 3264–3288. <https://doi.org/10.3934/math.2024159>
- Jadoon, A. K., Mahmood, T., Sarwar, A., Javaid, M. F., & Iqbal, M. (2024). Prediction of stock market movement using Long Short-Term Memory (LSTM) artificial neural network: Analysis of KSE 100 index. *Pakistan Journal of Life and Social Sciences*, 22(1), 107–119. <https://doi.org/10.57239/PJLSS-2024-22.1.009>
- Jan, M. N., & Ayub, U. (2019). Do the Fama and French five-factor model forecast well using ANN? *Journal of Business Economics and Management*, 20(1), 168–191. <https://doi.org/10.3846/jbem.2019.8250>

- Kelly, B., Pruitt, S., & Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, 134(3), 501–524. <https://doi.org/10.1016/j.jfineco.2019.05.001>
- Kozak, S., Nagel, S., & Santosh, S. (2020). Shrinking the cross-section. *Journal of Financial Economics*, 135(2), 271–292. <https://doi.org/10.1016/j.jfineco.2019.06.008>
- Manogna, R. L., & Kulkarni, N. (2025). Portfolio optimization model for stock price prediction using machine learning. *Journal of Statistical Theory and Applications*, 24, 1091–1108. <https://doi.org/10.1007/s44199-025-00140-z>
- Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica*, 41(5), 867–887. <https://doi.org/10.2307/1913811>
- Munir, T., Mamlook, R. E. A., Rahman, A. R., Alrashidi, A., & Yaseen, A. M. (2024). COVID-19's influence on Karachi Stock Exchange: A comparative machine learning algorithms study for forecasting. *Heliyon*, 10(7), Article e33190. <https://doi.org/10.1016/j.heliyon.2024.e33190>
- Pagliaro, A. (2026). Regime-aware LightGBM for stock market forecasting: A validated walk-forward framework with statistical rigor and explainable AI analysis. *Electronics*, 15(6), Article 1334. <https://doi.org/10.3390/electronics15061334>
- Poudel, S. R. (2025). Macroeconomic variables and common stock returns in Nepalese capital market. *The Journal of Business and Management*, 9(1), 73–87. <https://doi.org/10.3126/jbm.v9i1.81186>
- Raza, H., & Akhtar, Z. (2024). Predicting stock prices in the Pakistan market using machine learning and technical indicators. *Modern Finance*, 2(2), 46–63. <https://doi.org/10.61351/mf.v2i2.167>
- Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3), 341–360. [https://doi.org/10.1016/0022-0531\(76\)90046-6](https://doi.org/10.1016/0022-0531(76)90046-6)
- Sak, H., Huang, T., & Chng, M. T. (2024). Exploring the factor zoo with a machine-learning portfolio. *International Review of Financial Analysis*, 96, Article 103599. <https://doi.org/10.1016/j.irfa.2024.103599>
- Sarioğlu Duran, E., Korkmaz, T., & Ersöz Kaya, I. (2025). Machine learning for out-of-sample prediction of industry portfolio returns within multi-factor asset pricing models. *Applied Sciences*, 15(24), Article 12866. <https://doi.org/10.3390/app152412866>
- Shakya, L. (2024). Impact of macroeconomic factors and firm characteristics on stock returns in Nepalese insurance companies. *Nepalese Journal of Economics*, 8(4), 232–249. <https://doi.org/10.3126/nje.v8i4.79758>
- Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3), 425–442. <https://doi.org/10.1111/j.1540-6261.1964.tb02865.x>
- Siddiqui, O., Khan, N., & Bhatti, A. A. (2025). Human capital in asset pricing: A machine learning perspective on the six-factor model for Pakistan's equity market. *Modern Finance*, 3(4), 96–117. <https://doi.org/10.61351/mf.v3i4.416>
- Wang, H., Stefanidis, A., Jiang, Z., Su, J., et al. (2025). SSCL-GBM: A semi-supervised stock prediction approach with custom loss function. *Journal of Mathematics*, 2025, Article 5991303. <https://doi.org/10.1155/jom/5991303>
- Zada, H., Khan, N., Woo, K.-Y., & Chortane, S. G. (2025). Asset pricing: A comparative analysis of Fama-French five-factor with human capital-based six-factor model. *Advances in Decision Sciences*, 29(4), 1–37. <https://doi.org/10.47654/v29y2025i4p1-37>
- Zafar, Q., Farid, S., Ashraf, M. A., & Ashraf, Z. (2025). Forecasting stock index movement using tree based classifiers: Evidence from an emerging economy. *Review of Applied Management and Social Sciences*, 8(2), 847–871. <https://doi.org/10.47067/ramss.v8i2.522>
- Zhang, A., Pan, M., & Zhang, X. (2025). The pricing ability of factor model based on machine learning: Evidence from high-frequency data in China. *International Review of Economics & Finance*, 101, Article 104153. <https://doi.org/10.1016/j.iref.2025.104153>
- Zhu, H. (2025). LGBM on stock returns prediction and portfolio construction. In *Proceedings of the 2nd International Conference on Data Science and Engineering (ICDSE 2025)* (pp. 521–526). SCITEPRESS. <https://doi.org/10.5220/0013s701300004670>