Proceedings of the 5th International Conference on Financial Technology and Business Analysis (ICFTBA 2026)
DOI: 10.54254/2754-1169/2026.39392

Can Machine Learning Improve Corporate Earnings Forecasts and Valuation Inputs? ROE Forecasting and Residual Income Analysis of Shanghai and Shenzhen A-Share Manufacturing Firms

Fangjian Zhang

Central University of Finance and Economics, Beijing, China
2024311732@email.cufe.edu.cn

Abstract. Earnings forecasts are a foundation for valuation and fundamental analysis. Whether the greater flexibility of machine learning produces a stable out-of-sample forecasting advantage, and whether such an advantage improves valuation inputs, remains an empirical question. Using Shanghai and Shenzhen A-share manufacturing firms, this study sets June 30 of each year as the forecast origin. Under a strict temporal out-of-sample design, it compares three naive benchmarks, a mean-reversion model, Elastic Net, random forest, and XGBoost in forecasting return on equity (ROE) at several horizons. Ablation tests and SHAP analysis assess the incremental predictive value of information groups and how the models use them. With common assumptions for growth, cost of equity, and the equity path, the study then tests how ROE forecast errors pass through to finite-horizon residual-income errors. The results show a horizon boundary. Random forest shows a relatively consistent advantage at h1 under the main specification. Its average error remains low at h2, but the advantage weakens. Mean reversion performs better in the exploratory h3 results. Earnings history dominates the information used by short-horizon tree models, while the share of financial information rises at the medium horizon. Ablation results show that the incremental value of financial information depends on the model and horizon. Under the finite-horizon residual-income mapping, the short- and medium-horizon advantage of random forest translates into a lower two-period input error. The evidence supports a conditional, rather than universal, role for machine learning in improving earnings forecasts and valuation inputs.

Keywords: machine learning, earnings forecasting, return on equity, residual income, out-of-sample forecasting

1. Introduction

Corporate earnings forecasts are central to fundamental valuation. Return on equity (ROE) is a key measure of profitability. By relating net income to shareholders' equity, ROE indicates how efficiently a firm uses its equity capital. Yet annual ROE may be affected by financial condition,

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2754-1169/2026.39392

cash-flow quality, or industry shocks. Future ROE therefore cannot simply be extrapolated from current ROE. More accurate ROE forecasts may improve expectations about firms' financial condition and, in turn, the quality of inputs to fundamental valuation models.

Fama and French find persistence and mean reversion in profitability [1]. Cao and You show that financial-statement information helps predict future earnings. Binz et al. use margins, turnover, and other financial ratios to improve out-of-sample profitability forecasts [2, 3]. Dessaint et al. show that short-term-oriented alternative data affect forecast informativeness differently across horizons [4]. Hunt, Myers, and Myers apply machine learning to earnings prediction. Chen et al. relate the predictive gains of machine-learning models to richer financial information and nonlinear interactions [5, 6]. These findings do not establish a stable advantage at every horizon. Ohlson's residual income valuation (RIV) framework links earnings measures to intrinsic value. Dechow, Hutton, and Sloan find that Ohlson's model offers only limited improvements over existing implementation methods [7, 8]. Evidence on how improved earnings forecasts pass through to valuation errors for Chinese A-share firms and specified horizons remains limited. The study addresses three research questions. RQ1: Which forecasting methods predict ROE most accurately under strict temporal out-of-sample conditions? RQ2: Which information groups add predictive value, and what is the structure of the information used by the models? RQ3: Does a reduction in ROE forecast error translate into lower finite-horizon RI error?

This study therefore compares seven methods for forecasting the profitability of Shanghai and Shenzhen A-share manufacturing firms. It evaluates ROE forecast accuracy, the incremental value and model use of different information groups, and the transmission of ROE forecast errors to errors in the present value of finite-horizon residual income. It provides evidence for earnings forecasting and fundamental valuation in the Chinese market.

2. Theoretical background

2.1. The basis for ROE forecasting

ROE measures a firm's profitability relative to shareholders' equity. Other earnings measures include net income and return on assets (ROA). This study compares profitability across firms of different sizes and uses a residual-income model based on equity rather than total assets. It therefore uses an ROE measure. ROE can be defined in several ways, with different economic meanings. Here ROE is defined as

$$ROE_{i,t} = \frac{NI_{i,t}}{B_{i,t}^{\text{open}}} \tag{1}$$

In equation (1), NI is net income attributable to the parent, B^{open} is opening equity attributable to the parent, i indexes firms, and t denotes the fiscal year.

Liu et al. show that short-term profitability changes contain information about future profitability in the Chinese stock market [9]. Li provides evidence of mean reversion in Chinese listed firms from a residual-income perspective [10]. Easton et al. forecast earnings from recent earnings histories [11]. Esplin shows that the relative accuracy of industry- and firm-level long-term earnings-growth forecasts varies with firm characteristics [12]. Profitability may persist in the short run because of competitive advantages and operating inertia. Over longer horizons, capital flows and industry competition may induce mean reversion. The industry median, the recent three-year mean, and the latest ROE thus represent industry reversion, firm-level reversion, and persistence, respectively.

DOI: 10.54254/2754-1169/2026.39392

Firms with the same ROE may have different earnings and operating structures. Relevant measures include profit margin, asset turnover, the equity multiplier, cash-flow quality, and leverage. Industry may also matter. These financial and industry features can therefore be used to capture relationships that may improve forecasts. A linear model makes sensitivity to individual features easier to see, but may miss nonlinear interactions. A nonlinear model can represent such interactions more flexibly, while increasing the risk of fitting noise. Whether machine learning outperforms traditional methods is therefore tested under specified conditions.

2.2. Residual income and the transmission of ROE error

Zhang et al. connect accounting information with residual-income valuation in the Chinese market [13]. Residual income is net income less the return required by shareholders. It measures earnings in excess of the opportunity cost of equity. ROE forecasts have valuation relevance because for a given equity base, ROE determines forecast net income, residual income (RI), and ultimately residual income valuation (RIV). Factoring out the equity base gives

$$RI_j = (r_j - k)B_{j-1} \tag{2}$$

where

$$B_{j-1} = B_0 q_{j-1}, q_{j-1} = (1+g)^{j-1} \tag{3}$$

In the full RIV framework, intrinsic value equals current book value plus the present value of future residual income. This study uses finite-horizon RI to test whether lower ROE forecast error translates into a more accurate valuation input. It does not test whether the full valuation estimate improves.

The forecast present value of finite-horizon residual income is defined as

$$\hat{PVRI}_H^{\,m} = B_0 \sum_{j=1}^{H} \frac{q_{j-1}(\hat{r}_j^{\,\mathrm{m}} - k)}{(1+k)^j} \tag{4}$$

The corresponding realized present value is

$$PVRI_H^{\,realized} = B_0 \sum_{j=1}^{H} \frac{q_{j-1}(r_j - k)}{(1+k)^j} \tag{5}$$

Under the main scenario, $g = 0$, so the period-j equity multiplier $q_{j-1} = 1$. Economically, the equity base is assumed not to grow over the finite horizon. The required equity return is $k = 10\%$, yielding the forecast-actual difference. This difference is divided by initial equity to compare errors across firms of different sizes, giving the standardized error:

$$E_H^{\,m} = \frac{\hat{PVRI}_H^{\,\mathrm{m}} - PVRI_H^{\,realized}}{B_0} \tag{6}$$

Substitution yields

$$E_H^{\,m} = \sum_{j=1}^{H} w_j e_j^{\,\mathrm{m}}, w_j = \frac{q_{j-1}}{(1+k)^j}, e_j^{\,\mathrm{m}} = \hat{r}_{j\,m} - r_j \tag{7}$$

DOI: 10.54254/2754-1169/2026.39392

When $H = 1$, $E_1 = w_1 e_1$. This standardized error is a single-period measure: it multiplies ROE error by a constant and cannot establish an improvement in accumulated multi-period error. When $H \geq 2$, errors across periods may offset one another. Their combination shows whether improved ROE forecasts can translate into lower finite-horizon residual-income error.

Squaring this standardized error gives

$$(E_H{}^{\mathrm{m}})^2 = \sum_{j=1}^{H} w_{j_2} (e_j{}^{\mathrm{m}})^2 + 2 \sum_{j<\ell}^{H} w_j w_\ell e_{jm} e_\ell{}^{\mathrm{m}} \tag{8}$$

The squared terms capture the contribution of each period to total squared error. Cross terms capture whether errors across periods move in the same direction. Positive cross terms indicate accumulation in the same direction, while negative terms indicate offsetting errors. The net cross term reflects their combined effect.

In equations (2)-(8), j indexes future fiscal years and H denotes the finite forecast horizon. The symbol r denotes actual ROE, and r with a hat denotes the forecast from model m. B_0 is common initial parent-attributable equity, g is equity-path growth, k is the cost of equity, and q is the equity-path coefficient. The symbol e denotes predicted minus actual ROE, w is the transmission weight, and E is standardized finite-horizon RI transmission error.

3. Research design

3.1. Sample selection and data sources

Sample. The sample comprises Shanghai and Shenzhen A-share manufacturing firms. At each forecast origin, the population includes firms listed at that time. Firms are not excluded because they delist subsequently. Manufacturing status is determined from historical industry records for each year, rather than inferred backward from current status. Forecast origins fall on June 30 each year. The sample covers 2013–2025.

Data sources. Annual financial statements mainly come from Eastmoney's structured balance sheets, income statements, and cash-flow statements. Historical industry records combine dated information from the China Securities Regulatory Commission, the China Association for Public Companies, and CNINFO. Key fields and unusual observations were checked against available original annual reports.

Historical availability. Predictors for forecast year Y use only annual financial information from $Y - 1$ or earlier. When a reliable annual-report release date is available, that date is used. Otherwise, the date is inferred from other available financial statements for the firm. Identical data vintages available at the historical date cannot be guaranteed, and later revisions may be present. This is therefore a retrospective out-of-sample study with temporal availability constraints, rather than a full reconstruction of the real-time information set.

Sample screening. Opening equity attributable to the parent must be positive. Latest total assets and closing equity attributable to the parent at the forecast origin must also be positive. Observations whose core accounting measures cannot be recovered are excluded. Missing ordinary features and future delisting do not cause observations to be removed. Historical label availability is assessed separately for h1, h2, and h3.

DOI: 10.54254/2754-1169/2026.39392

3.2. Forecast targets and input features

At the June 30 forecast origin in year Y, h1 targets ROE for fiscal year Y, h2 for $Y+1$, and h3 for $Y+2$. These targets correspond to fiscal years, not to the 12, 24, or 36 months following the observation date. Separate models are fitted for each horizon.

Equation (9) shows how each forecast horizon maps to its target.

$$y_{i,Y,h} = ROE_{i,Y+h-1}, \quad h \in \{1,2,3\} \tag{9}$$

Input features are prespecified in four groups: P, profitability history; S, firm state; F, financial information; and I, industry information. Table 1 defines the main features and their construction.

Table 1. Information groups and core feature definitions

Group	Variable	Construction
P	Latest and lagged ROE	NI / opening parent-attributable equity for $Y-1$ and $Y-2$
P	Historical mean, change, and loss	Mean valid ROE over the latest three calendar years; adjacent ROE difference; latest $NI < 0$
S	Listing age and equity/assets	Time since listing at origin; latest closing parent-attributable equity / total assets
F	Net and gross margins	Consolidated net income / revenue; (revenue − cost of sales) / revenue
F	Turnover and leverage	Revenue / average total assets; total liabilities / total assets
F	Cash flow and accruals/assets	Net operating cash flow / total assets; (consolidated net income − net operating cash flow) / total assets
F	Revenue and asset growth	Year-on-year growth in revenue and total assets
F	Size and asset composition	Natural log of total assets; cash, receivables, and inventory / total assets
I	Peer ROE and industry	Leave-one-out industry median (all-manufacturing fallback below 20 peers); training-fold industry coding

Note: NI in target ROE refers to net income attributable to the parent. Profit margin and accrual measures use consolidated net income. All inputs follow the forecast-year and release-date rules. Source: Annual financial data compiled for this study.

The peer-industry ROE measure excludes the focal firm. If fewer than 20 peers have valid ROE, the leave-one-out median for all manufacturing firms is used. Industry status is based on the classification for that year. Net income for profit margin differs from the parent-attributable net income used for ROE.

The main training specification winsorizes extreme numeric features at the 1st and 99th percentiles within each training fold. Missing values are replaced with the training-fold median, and a binary missingness indicator is added. Industry is one-hot encoded. Elastic Net additionally standardizes numeric features to align their scales; random forest and XGBoost do not. All medians, percentiles, scaling parameters, and industry categories are estimated only within the training fold. ROE is not winsorized in the main model's training data or in the validation and test data. Missing target ROE is not imputed.

3.3. Model construction and comparison

3.3.1. Benchmarks and forecasting models

The study evaluates seven methods: three naive benchmarks, one mean-reversion model, and three machine-learning models.

DOI: 10.54254/2754-1169/2026.39392

The naive benchmarks are Latest, which carries the latest ROE forward; Historical, which uses the mean ROE over the most recent three years; and Industry, which uses the leave-one-out median ROE for the firm's industry. The historical benchmark requires at least two valid years; otherwise it is missing.

The mean-reversion speed is estimated from the historical training sample available at each origin rather than imposed ex ante. The model uses the latest ROE, median peer ROE, and a recent-loss indicator. The indicator equals one when the latest parent-attributable net income is negative.

Equation (10) gives the mean-reversion forecast.

$$\hat{ROE}^{MR}_{i,Y,h} = \beta_{0,h} + \beta_{1,h} ROE_{i,Y-1} + \beta_{2,h} PeerROE_{i,Y-1} + \beta_{3,h} Loss_{i,Y-1} \tag{10}$$

Here, $\hat{ROE}^{MR}_{i,Y,h}$ is the mean-reversion forecast for firm i at origin Y and horizon h; $ROE_{i,Y-1}$ is the latest available ROE before the origin; $PeerROE_{i,Y-1}$ is the peer ROE median excluding the focal firm; $Loss_{i,Y-1}$ is the latest-loss indicator. The coefficients $\beta_{0,h}, \beta_{1,h}, \beta_{2,h}, \beta_{3,h}$ are estimated from the eligible historical training sample.

Elastic Net (EN) is a regularized linear model and serves as a linear comparator using the same information [14]. Random forest (RF) averages predictions from randomized decision trees to capture nonlinear relationships [15]. XGBoost is a boosted-tree model whose later trees correct earlier prediction errors and capture nonlinear interactions [16].

The candidate hyperparameter grid for the machine-learning models is shown in Table 2.

Table 2. Models and candidate hyperparameters

Model	Candidate hyperparameters	Fixed settings	Combinations
Elastic Net	Penalty α: 0.1, 0.01, 0.001, 0.0001; L1 ratio: 0.9, 0.5, 0.1	Standardize numeric features within training fold	12
Random forest	Maximum depth: 4, 8; minimum leaf size: 50, 20	500 trees; feature fraction per split = 0.8	4
XGBoost	Maximum depth: 2, 4; trees: 200, 500; minimum child weight: 30, 10	Learning rate = 0.05; row sample = 0.8; column sample = 0.8; L2 penalty = 1	8

Note: Candidate settings are selected using inner historical validation. Ties follow the prespecified grid order. The random seed is 20260906.

3.3.2. Temporal validation and model evaluation

For each forecast origin, historically eligible training observations are first identified. After specifying the parameter grid, the two latest historical forecast origins with legally available labels in the outer training set serve as validation origins. Models are trained on earlier eligible observations and evaluated at these origins using mean squared error (MSE). The two yearly MSE values receive equal weight for each parameter setting. The lowest-error setting is selected, then the model is refitted on all historical data available at the current origin.

Formal outer test windows are 2019–2025 for h1, 2020–2024 for h2, and 2022–2023 for h3. At least three training-origin years with labels are required. Because only two test years are available for h3 under the strict temporal design, it is exploratory and cannot establish robustness over longer test histories.

DOI: 10.54254/2754-1169/2026.39392

Formal model rankings use observations with predictions and actual values for all seven methods. This common sample contains 20,582, 14,784, and 6,255 observations for h1, h2, and h3. Annual stability, error quantiles, and paired uncertainty comparisons of RF with mean reversion and XGBoost use a five-model common sample: Latest, mean reversion, EN, RF, and XGBoost. It contains 20,601, 14,800, and 6,263 observations, respectively. This avoids losing valid observations solely because the other two benchmarks are unavailable. Metrics from the two samples are not directly comparable. Information-group analyses use a separate machine-learning evaluation sample.

The primary metrics are equal-year-weighted root mean squared error (RMSE) and mean absolute error (MAE). Equal weighting prevents later years with more listed firms from receiving more weight automatically. Equal-year-weighted RMSE is the square root of the mean of yearly MSE values. Specifically,

$$RMSE_{EY} = \sqrt{\frac{1}{T}\sum_{y=1}^{T}\frac{1}{n_y}\sum_{i=1}^{n_y}{e_{i,y}}^2} \tag{11}$$

$$MAE_{EY} = \frac{1}{T}\sum_{y=1}^{T}\frac{1}{n_y}\sum_{i=1}^{n_y}|e_{i,y}| \tag{12}$$

In equations (11)-(12), T is the number of test years, n_y is the common-sample size in year y, and $e_{i,y}$ is firm i's ROE forecast error in year y.

RMSE summarizes squared error but does not describe the entire error distribution. Additional measures include pooled RMSE and MAE, median absolute error, signed bias, the 90th and 95th percentiles of absolute error (P90 and P95), and annual wins.

To measure cross-sectional sampling uncertainty within test years, the study uses firm-cluster bootstrap [17]. Predictions are held fixed; models are not retrained. Firms in the test sample are resampled with replacement 2,000 times to form 95% intervals for the error metrics. These intervals describe uncertainty across firms within the fixed test years.

3.3.3. Incremental information and model explanation

Ablation tests evaluate the incremental value of information groups for RF and XGBoost at h1 and h2. Six specifications are considered: $P+S$, $P+S+F$, the full set, and sets excluding P, F, or I. Each specification is tuned separately on its own data set. The tests ask whether a group improves out-of-sample error relative to a specified baseline.

SHAP describes how fitted models use information [18, 19]. Contributions to an individual ROE forecast are first calculated by feature, then aggregated to P, S, F, and I. A variable's contribution is combined with its missingness indicator. Signed contributions are summed within each group, their absolute values are taken, and values are aggregated within years before equal weighting across years. SHAP describes model use of information [19].

3.4. Descriptive statistics

The eligible firm-origin panel has 31,510 observations from 3,800 firms. These observations represent the full eligible panel before preprocessing and therefore differ from the formal h1, h2, and h3 test samples.

Table 3 reports descriptive statistics for the core input features.

DOI: 10.54254/2754-1169/2026.39392

Table 3. Summary of core input features

Variable	N	P25	Median	P75
Latest ROE	31380	0.025	0.075	0.142
Historical mean ROE	31488	0.034	0.092	0.167
Parent equity/assets	31510	0.441	0.599	0.754
Net margin	31504	0.022	0.070	0.134
Liabilities/assets	31510	0.231	0.379	0.531
Operating cash flow/assets	31510	0.011	0.048	0.089
Total asset turnover	31480	0.402	0.571	0.793

Note: Statistics use nonmissing values from the 31,510 eligible firm-origin observations. Raw ratios are in decimal units; total asset turnover is measured in times. No training-fold preprocessing is applied. Source: Final feature table.

4. Empirical results and analysis

4.1. Out-of-sample ROE forecasts

Table 4 compares error metrics for all candidate methods on the seven-model common sample. At h1, RF has an equal-year-weighted RMSE of 140.28 percentage points, below 147.78 for mean reversion and 144.40 for XGBoost. Its MAE is 13.54 percentage points, below 16.96 and 16.00, respectively. Against mean reversion, RF reduces RMSE by 7.50 percentage points, or about 5.08%, and MAE by 3.42 percentage points. This indicates a clear short-horizon out-of-sample advantage. Firms with very small ROE denominators can produce extreme ratios, to which squared loss assigns heavy weight. RMSE must therefore be read alongside MAE, error quantiles, and the behavior of extreme errors rather than treated alone as typical-firm accuracy. At h2, RF again has the lowest equal-year-weighted RMSE and MAE, at 82.88 and 13.51 percentage points. Relative to mean reversion, however, the gains are only 2.15 and 0.16 percentage points. The average advantage is much smaller. Model complexity does not have a monotonic relation with forecast accuracy; performance depends on the setting.

At h3, mean reversion has an RMSE of 70.68 and an MAE of 13.28 percentage points. Both are below EN (70.95 and 16.13), RF (72.94 and 14.59), and XGBoost (74.18 and 16.33). These longer-horizon results do not support a general machine-learning advantage. Yet h3 has only two test origins, so this exploratory result does not establish that any method dominates in all longer-horizon settings.

Table 4. Overall forecast comparison on the seven-model common sample

Horizon	Model	N	Equal-year RMSE	Equal-year MAE	Pooled RMSE	Pooled MAE	Bias
h1	Latest	20582	185.24	16.85	181.38	16.29	4.75
h1	Historical	20582	160.39	18.75	156.55	18.46	6.88
h1	Industry	20582	148.78	15.93	143.57	15.55	6.40
h1	Mean reversion	20582	147.78	16.96	142.57	16.66	1.03
h1	EN	20582	147.10	16.95	141.89	16.74	2.20
h1	RF	20582	140.28	13.54	134.85	13.10	2.16
h1	XGB	20582	144.40	16.00	139.47	15.66	2.51
h2	Latest	14784	160.72	18.28	159.53	17.96	5.25

DOI: 10.54254/2754-1169/2026.39392

Table 4. (continued)

h2	Historical	14784	112.70	19.36	112.27	19.30	7.30
h2	Industry	14784	85.76	14.40	83.70	14.19	6.47
h2	Mean reversion	14784	85.03	13.67	82.93	13.42	1.54
h2	EN	14784	84.48	14.98	82.39	14.75	2.43
h2	RF	14784	82.88	13.51	80.93	13.26	2.30
h2	XGB	14784	86.69	16.54	85.05	16.20	1.99
h3	Latest	6255	167.58	19.14	169.66	19.16	9.44
h3	Historical	6255	110.91	20.88	111.41	20.91	10.80
h3	Industry	6255	72.21	14.11	72.29	14.09	9.47
h3	Mean reversion	6255	70.68	13.28	70.77	13.29	1.10
h3	EN	6255	70.95	16.13	71.03	16.09	1.96
h3	RF	6255	72.94	14.59	73.02	14.57	2.60
h3	XGB	6255	74.18	16.33	74.25	16.30	1.92

Note: Values are percentage points. The seven-model common sample is used for overall ranking. Bias is the pooled mean of predicted minus actual ROE. Equal-year-weighted RMSE is the square root of mean annual MSE. Metrics from the seven-model and five-model samples are not directly comparable. Source: Common predictions for all seven models.

The preprocessing sensitivity results in Table 5 also show that some model rankings change with preprocessing. Claims about RF's advantage are therefore limited to the main specification and its evaluation conditions.

Table 5. RMSE and MAE under alternative preprocessing specifications

Horizon	Specification	EN	RF	XGB
h1	Main	151.36 / 17.75	143.76 / 14.26	147.41 / 16.83
h1	No feature winsorization	153.18 / 17.93	141.63 / 13.98	140.52 / 15.43
h1	Training-target winsorization	151.90 / 14.24	151.72 / 13.87	151.60 / 13.91
h2	Main	84.55 / 15.14	83.14 / 13.80	87.36 / 16.96
h2	No feature winsorization	84.75 / 14.67	84.20 / 15.57	88.09 / 17.94
h2	Training-target winsorization	84.79 / 13.21	84.62 / 12.78	84.22 / 13.05
h3	Main	70.83 / 16.19	72.82 / 14.63	74.06 / 16.40
h3	No feature winsorization	939.94 / 26.65	74.32 / 16.15	76.39 / 17.30
h3	Training-target winsorization	70.79 / 12.50	70.70 / 12.32	70.81 / 12.40

Note: Each cell reports equal-year-weighted RMSE followed by MAE in percentage points for the machine-learning sample. Validation and test targets are never winsorized. Source: Existing preprocessing sensitivity results.

4.2. Error distribution and annual stability

At h1, RF has the lowest median absolute error, P90, and P95 among the five models: 3.71, 17.38, and 30.29 percentage points, respectively. At h2, its median absolute error remains lowest, but its P90 and P95 exceed those of mean reversion. Thus its medium-horizon advantage does not extend across the entire error distribution. At h3, mean reversion has the lowest median absolute error and P95.

In MAE and MSE comparisons with mean reversion and XGBoost, RF is better on nearly all h1 measures. The upper bounds of the 95% intervals for both loss differences against mean reversion

DOI: 10.54254/2754-1169/2026.39392

are below zero, giving relatively consistent evidence of a short-horizon advantage. At h2, both intervals against mean reversion include zero, and its annual wins are fewer.

Table 6 reports paired loss differences and firm-cluster intervals.

Table 6. Paired RF loss differences and firm-cluster intervals

Horizon	Comparator	Loss	RF − comparator	Lower 95% CI	Upper 95% CI	Years won
h1	Mean reversion	MAE	-3.47	-3.90	-3.07	7/7
h1	Mean reversion	MSE	-2361.62	-4671.31	-634.38	7/7
h1	XGB	MAE	-2.46	-2.96	-1.97	7/7
h1	XGB	MSE	-1121.62	-2319.46	258.03	6/7
h2	Mean reversion	MAE	-0.15	-0.47	0.16	3/5
h2	Mean reversion	MSE	-360.17	-875.18	17.25	3/5
h2	XGB	MAE	-3.03	-3.52	-2.57	5/5
h2	XGB	MSE	-646.24	-1323.90	24.17	4/5

Note: MAE is in percentage points and MSE in squared percentage points. Negative values favor RF. The 95% intervals use 2,000 firm-cluster bootstrap resamples, with model predictions and test years held fixed. They reflect cross-sectional sampling uncertainty within the given test years.

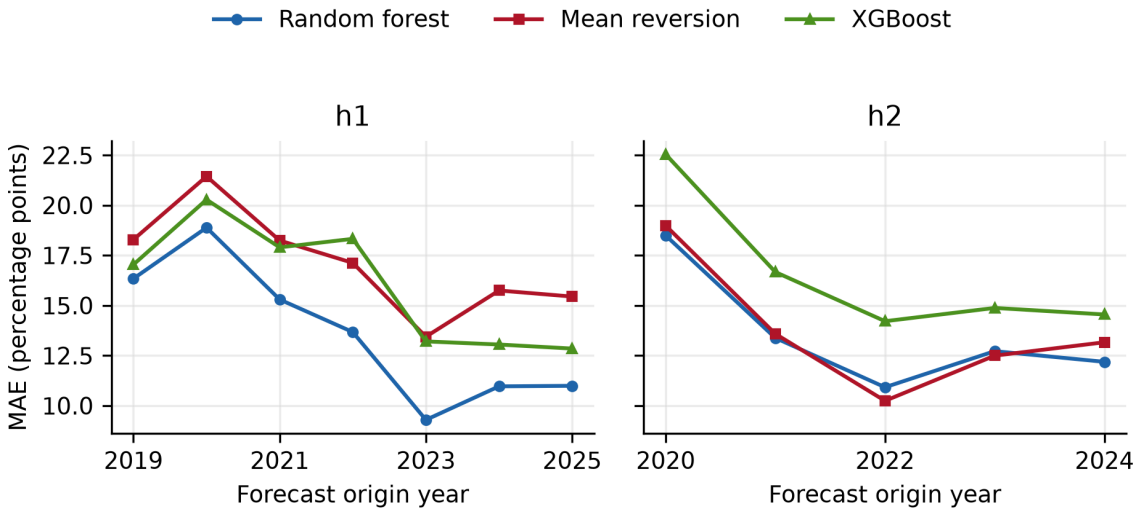

Figure 1. Annual MAE on the common sample

Note: The figure reports RF, mean reversion, and XGBoost results for the five-model common sample. Source: Annual error table.

Figure 1 shows the annual distribution of errors. At h1, RF outperforms mean reversion on MAE and MSE in all seven years. Against XGBoost, it wins on MAE in all seven years and on MSE in six. The annual pattern weakens at h2. RF beats mean reversion on both losses in only three years; both losses are higher in 2022 and 2023.

4.3. Incremental information and model explanation

Ablation tests for RF and XGBoost examine the incremental value of information groups relative to a baseline. Table 7 shows that adding F to $P + S$ at h1 improves RF's MAE by 0.38, with a small RMSE gain. Removing F from the full set slightly improves RMSE by 0.11 but worsens MAE by

DOI: 10.54254/2754-1169/2026.39392

0.33. Financial information thus does not provide a consistent gain on every metric for RF at the short horizon. At h2, adding F improves MAE by 0.77 and RMSE by 0.82. Removing F from the full set worsens both metrics. Financial information thus has relatively consistent incremental value for RF at the medium horizon. For XGBoost, financial information worsens results in most cases. Its incremental value depends on the model, horizon, and loss metric.

Table 7. Equal-year-weighted errors by information-group specification

Horizon	Specification	RF RMSE	RF MAE	XGB RMSE	XGB MAE
h1	PS	144.49	14.68	145.75	16.45
h1	PSF	144.07	14.30	147.63	16.85
h1	full	143.76	14.26	147.41	16.83
h1	noP	144.84	14.30	149.01	16.72
h1	noF	143.65	14.59	145.98	16.75
h1	noI	144.07	14.30	147.63	16.85
h2	PS	84.05	14.50	86.50	16.18
h2	PSF	83.23	13.73	87.20	16.80
h2	full	83.14	13.80	87.36	16.96
h2	noP	83.62	13.64	86.97	16.31
h2	noF	83.99	14.56	87.28	17.06
h2	noI	83.23	13.73	87.20	16.80

Note: Values are percentage points. The machine-learning sample has 20,669 h1 and 14,850 h2 observations. Full is $P + S + F + I$; noP, noF, and noI omit the named group. PSF is equivalent to noI. Each specification is tuned separately. Source: Information-group forecast results.

SHAP analysis describes how the fitted models use information. Figure 2 shows that profitability history (P) is the most heavily used group in both RF and XGBoost at the short horizon. At the medium horizon, the share attributed to financial information rises. Short-horizon ROE forecasts rely more on profitability history. The share of financial information rises at the medium horizon, while profitability history remains important.

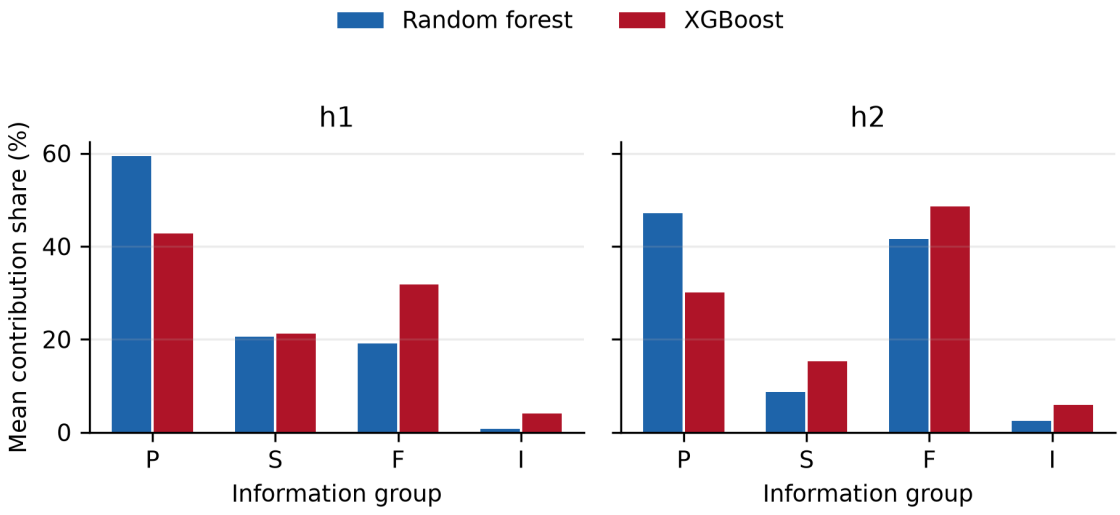

Figure 2. Information-group contribution shares in tree models

DOI: 10.54254/2754-1169/2026.39392

4.4. Transmission of ROE forecast errors to finite-horizon RI

In the main scenario, $H = 2$, $g = 0$, and the required equity return is $k = 10\%$. RF has lower RMSE and MAE than the other models. Its short- and medium-horizon ROE forecast advantage thus translates into lower two-period RI input error under these assumptions. Mean reversion is better than RF for h3 alone. When $H = 3$, however, RF still has the lowest finite-horizon RI error on both metrics. The ranking reverses relative to the single h3 forecast. A single-horizon forecast ranking therefore need not determine the ranking for a multi-period residual-income combination.

Table 8 reports finite-horizon residual-income transmission errors under the main parameters.

Table 8. Finite-horizon residual-income error under main parameters

Periods	Model	N	Equal-year RMSE	Equal-year MAE	Pooled RMSE	Pooled MAE
2	Latest	14799	280.44	27.68	278.59	27.21
2	Mean reversion	14799	151.70	22.75	148.48	22.49
2	Elastic Net	14799	150.83	23.70	147.62	23.54
2	Random forest	14799	143.17	19.53	139.58	19.19
2	XGBoost	14799	151.46	22.87	147.78	22.51
3	Latest	6247	408.77	37.87	413.26	37.90
3	Mean reversion	6247	160.26	27.58	158.22	27.56
3	Elastic Net	6247	159.58	30.92	157.56	30.91
3	Random forest	6247	148.28	25.00	146.47	24.94
3	XGBoost	6247	157.43	28.75	155.50	28.69

Note: Values are percentages of initial parent-attributable equity B_0. Here $g = 0$ and $k = 10\%$. Results use the five-model common multi-horizon sample. $H = 2$ is the main test and $H = 3$ is exploratory. Source: Finite-horizon transmission calculations from ROE forecasts.

At $H = 2$, the cross and within-period squared terms show how errors combine across horizons. In Figure 3, the net cross term is positive for every model. It accounts for 4.07% of RF's error, 8.29% of mean reversion's, and 37.78% of the Latest benchmark's. Overall, RF has a smaller net cross-period contribution than mean reversion.

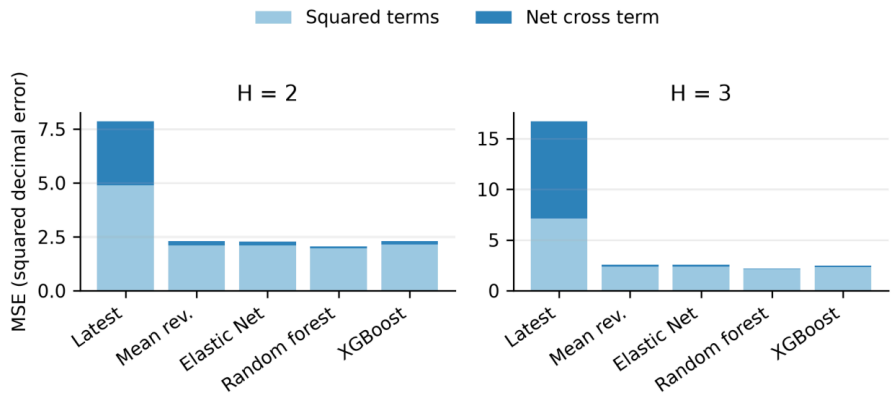

Figure 3. Components of finite-horizon transmission MSE

Note: $g = 0$ and $k = 10\%$. The units are squared decimal errors. The squared terms and net cross term sum to total MSE.

DOI: 10.54254/2754-1169/2026.39392

5. Discussion

5.1. Horizon limits of forecasting advantages

One possible explanation is that recent ROE, profitability history, and current financial conditions remain closely associated with near-term profitability because operating performance adjusts gradually. At longer horizons, new operating shocks, industry changes, and macroeconomic factors reduce the inertia of historical information. Historical financial features gradually lose relevance. Capital flows may also intensify competition and pull excess profitability toward industry norms. A simple mean-reversion model may then win because its simple structure fits less noise. The short-horizon result is consistent with earlier evidence that machine learning can use complex financial features to improve earnings forecasts [20]. Related evidence shows that alternative data affect analysts' forecast informativeness differently at short and long horizons. The h3 result also accords with evidence of long-run profitability mean reversion.

5.2. Model use and incremental information value

A high SHAP share indicates that a model relies heavily on an information group. It does not establish that the group adds predictive value relative to a baseline. Ablation instead asks how out-of-sample forecasts change when a group is added or removed. Model use and incremental predictive value are therefore distinct questions. The results show that financial information has different incremental value across models and horizons; its usefulness must be evaluated in the specific forecasting setting.

5.3. Finite-horizon residual-income transmission

Mean reversion outperforms RF for the single h3 forecast, yet RF performs better when errors across horizons are combined. RF is better in the first two periods, while mean reversion wins only in the last. This ranking difference may reflect errors at each horizon, common discount weights, and the way their errors combine across periods. A single-horizon ranking cannot directly determine the ranking of finite-horizon RI input errors.

The aggregate cross term is positive for every model, and the improvement reflects both smaller within-period squared-error contributions and a smaller net cross-period contribution. The evidence shows that, under shared assumptions for growth, the required equity return, and the equity path, better forecasts can produce lower finite-horizon RI input error.

6. Conclusions and future work

6.1. Main findings

For RQ1, the relative performance of forecasting methods depends on the horizon. RF has a relatively stable advantage at h1. It still leads on average at h2, but its advantage weakens across years and under bootstrap uncertainty. Mean reversion performs better at h3, although only two test years make that result exploratory. For RQ2, earnings history makes the largest SHAP contribution at the short horizon, while the contribution of financial information rises at the medium horizon. Ablation tests show a relatively consistent contribution from financial information to RF at h2, but not across all models or horizons. For RQ3, better ROE forecasts can reduce finite-horizon RI input error. A single-period error ranking need not match the multi-period RI ranking. RF has lower finite-

DOI: 10.54254/2754-1169/2026.39392

horizon RI error in the main $H = 2$ scenario. Overall, the evidence does not support a universal claim that machine learning improves corporate earnings forecasts and valuation inputs. In this A-share manufacturing sample and strict temporal out-of-sample setting, some models improve ROE forecasts at certain horizons. Under common parameters, those gains can reach finite-horizon residual-income input errors. The advantage depends on horizon, model, and information set.

6.2. Implications

First, model complexity does not guarantee out-of-sample accuracy. The ability of machine learning to capture nonlinear relationships and interactions does not establish superiority over simpler methods. Forecasting methods should be evaluated through a combination of economic reasoning and strict temporal tests, the forecast horizon, and the error distribution. Simple mean reversion may remain competitive at longer horizons.

Second, information value depends on the model and horizon. Its predictive value should be assessed against a specified model, horizon, and baseline information set.

Third, model selection for residual-income analysis should consider combined forecast errors across horizons rather than accuracy at a single horizon. The weights and interactions of errors across horizons also matter.

6.3. Limitations and future research

First, the data do not fully recover the originally released versions of firms' financial records; later revisions may be present. Future work could use commercial databases or complete historical disclosures to reconstruct the information genuinely available at each date.

Second, the valuation mapping assumes common growth and required equity returns, fixes the equity path, and does not implement the full residual-income model. Future work could add firm-specific costs of equity, equity paths, and terminal values once the sources of forecast error can be clearly attributed. It could then test whether better ROE forecasts improve complete firm-value estimates.

References

[1] Fama, E. F., & French, K. R. (2000). Forecasting profitability and earnings. *The Journal of Business, 73*(2), 161–175. https://doi.org/10.1086/209638

[2] Cao, K., & You, H. (2024). Fundamental analysis via machine learning. *Financial Analysts Journal, 80*(2), 74–98. https://doi.org/10.1080/0015198X.2024.2313692

[3] Binz, O., Schipper, K., & Standridge, K. R. (2025). Estimating profitability decomposition frameworks via machine learning: Implications for earnings forecasting and financial statement analysis. *Journal of Accounting and Economics, 80*(2–3), 101805. https://doi.org/10.1016/j.jacceco.2025.101805

[4] Dessaint, O., Foucault, T., & Fresard, L. (2024). Does alternative data improve financial forecasting? The horizon effect. *The Journal of Finance, 79*(3), 2237–2287. https://doi.org/10.1111/jofi.13323

[5] Hunt, J. O. S., Myers, J. N., & Myers, L. A. (2022). Improving earnings predictions and abnormal returns with machine learning. *Accounting Horizons, 36*(1), 131–149. https://doi.org/10.2308/HORIZONS-19-125

[6] Chen, X., Cho, Y. H., Dou, Y., & Lev, B. (2022). Predicting future earnings changes using machine learning and detailed financial data. *Journal of Accounting Research, 60*(2), 467–515. https://doi.org/10.1111/1475-679X.12429

[7] Ohlson, J. A. (1995). Earnings, book values, and dividends in equity valuation. *Contemporary Accounting Research, 11*(2), 661–687. https://doi.org/10.1111/j.1911-3846.1995.tb00461.x

[8] Dechow, P. M., Hutton, A. P., & Sloan, R. G. (1999). An empirical assessment of the residual income valuation model. *Journal of Accounting and Economics, 26*(1–3), 1–34. https://doi.org/10.1016/S0165-4101(98)00049-4

DOI: 10.54254/2754-1169/2026.39392

[9] Liu, Y., Yang, B., & Su, Y. (2023). Changes in firm profitability, heterogeneous investor beliefs, and stock returns. *Journal of Management Science and Engineering, 8*(2), 258–272. https://doi.org/10.1016/j.jmse.2022.10.004

[10] Li, H. F. (2012). Quality and value analysis of Chinese listed companies during 1994—2010—Based on residual income model. *Journal of Xihua University (Philosophy & Social Sciences), 31*(5), 81–89. https://aj.xhu.edu.cn/xhdxxbzskb/cn/article/id/381ff429-0dde-4b29-ae99-36d421b81fca

[11] Easton, P. D., Kapons, M. M., Monahan, S. J., Schütt, H. H., & Weisbrod, E. H. (2024). Forecasting earnings using k-nearest neighbors. *The Accounting Review, 99*(3), 115–140. https://doi.org/10.2308/TAR-2021-0478

[12] Esplin, A. (2022). Industry-level versus firm-level forecasts of long-term earnings growth. *Finance Research Letters, 47*, 102516. https://doi.org/10.1016/j.frl.2021.102516

[13] Zhang, R. J., Liu, H., & Hu, X. B. (2002). A corporate valuation model using accounting information: Development and application of the RIR model. *Journal of Finance and Economics, 28*(7), 69–75.

[14] Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B, 67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

[15] Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32. https://doi.org/10.1023/A:1010933404324

[16] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785

[17] Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics, 7*(1), 1–26. https://doi.org/10.1214/aos/1176344552

[18] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In I. Guyon, U. von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems 30* (pp. 4765–4774). Curran Associates.

[19] Li, M., Sun, H., Huang, Y., & Chen, H. (2024). Shapley value: From cooperative game to explainable artificial intelligence. *Autonomous Intelligent Systems, 4*, Article 2. https://doi.org/10.1007/s43684-023-00060-8

[20] Jones, S., Moser, W. J., & Wieland, M. M. (2023). Machine learning and the prediction of changes in profitability. *Contemporary Accounting Research, 40*(4), 2643–2672. https://doi.org/10.1111/1911-3846.12888