Accounting Research Letters

Article in Press

Accounting Research Letters

Volume X, Issue X s

202X

pp. XX-XX

The Accounting and Economic Association of Japan

Machine Learning for Forecast Accuracy and Value Relevance

Minjun Jeong

Declaration of Interest

The authors declare no conflicts of interest.

Declaration of Generative AI

This manuscript was proofread and partially edited using ChatGPT and Claude.

Data Availability Statement

The data used in this study are available upon request.

Funding Information

This work was supported by the Japan Science and Technology Agency (JST) through its *Support for Pioneering Research Initiated by the Next Generation (SPRING)* program (Grant No. JPMJSP2139) and by Osaka Metropolitan University.

Acknowledgments

I would like to thank Professor Hiroyuki Ishikawa for valuable feedback on research design, Professor Jumpei Hamamura (Editor), and an anonymous reviewer for their helpful comments. The usual disclaimers apply.

Accounting Research Letters

Article in Press

Accounting Research Letters

Volume X, Issue X s

202X

pp. XX-XX

The Accounting and Economic Association of Japan

Machine Learning for Forecast Accuracy and Value Relevance

Minjun Jeong

Osaka Metropolitan University, Graduate School of Business

Submitted: March 18, 2025; Revised: July 30, 2025; Accepted: August 16, 2025;

Available online: XXXXX XX, 202X

Abstract

This study investigates the accuracy and value relevance of machine learning-based earnings forecasts in the Japanese equity market. Using XGBoost and a dataset of 50,568 firm-year observations from 2002 to 2022, the analysis reveals that machine learning-based forecasts show mixed effects on bias correction, with improvements in mean bias but deterioration in median forecasts, while maintaining comparable overall accuracy. The effectiveness of these forecasts, however, varies across firm size; even so, they provide incremental informational value. These findings suggest ML techniques offer potential for enhancing financial disclosure usefulness and equity valuation, though their overall impact is modest and effectiveness varies across contexts.

Keywords: machine learning; management forecast; forecast accuracy; value relevance

Corresponding Author: Minjun Jeong, Osaka Metropolitan University, Graduate School of Business, 3-3-138 Sugimoto, Sumiyoshi-ku, Osaka, 558-8585, Japan

Email: sw23662q@st.omu.ac.jp / dlrlwnsh@icloud.com

1. Introduction

This study investigates whether machine learning (ML) improves the forecast accuracy and value relevance of earnings forecasts. Using XGBoost, a gradient-boosting algorithm noted for strong predictive performance, this study generates machine-learning-based earnings forecasts (MLEF) and embeds them in the Ohlson (2001) valuation framework to test their incremental value relevance.

In the Japanese equity market, management earnings forecasts (MEF) play a significant role in corporate financial disclosure practices, even though issuing such forecasts is voluntary. Ota (2010) documents that MEF have the strongest association with stock prices among accounting variables, underscoring their importance in market valuation.

Nevertheless, MEF often exhibit optimistic bias that can erode decision usefulness (Ota, 2006). Traditional linear forecasting models are likewise constrained by limited variable sets and by difficulty in capturing the nonlinear interactions inherent in financial data.

ML methods provide a flexible alternative. Tree-based algorithms such as gradient boosting accommodate high-dimensional inputs and recognize nonlinearities, interaction effects, and asymmetric patterns without explicit functional-form assumptions[1].

Accordingly, this study pursues three objectives. First, it tests whether MLEF deliver higher forecast accuracy than MEF across firm characteristics (e.g., firm size and firm maturity). Second, it evaluates the value relevance of MLEF within the Ohlson framework. Third, it examines whether MLEF convey incremental information beyond MEF.

2. Machine Learning Method

This study uses financial and stock data from the NEEDS-FinancialQUEST database, for firms listed on the Tokyo Stock Exchange between 2002 and 2022. Firms were excluded from the dataset if they operated in the financial sector, had a fiscal year that is not equal to 12 months, or met exclusion criteria (e.g., negative shareholders' equity, delayed disclosure of MEF or financial statements beyond 61 days after the fiscal period end, or missing values). After applying these filters, the dataset consisted of 50,568 firm-year observations.

Following Campbell et al. (2024), the machine learning model employs an indirect approach where the target variable is the systematic bias component of management forecasts (defined as

[1] Gu et al. (2020) find that machine learning models outperform traditional linear methods in measuring asset risk premiums, and that the resulting estimates lead to investment strategies with higher Sharpe ratios, demonstrating both statistical and economic significance.

MEF minus actual earnings). The predicted bias estimates are then subtracted from original management forecasts to generate bias-adjusted earnings forecasts, thereby preserving management's informational content while mitigating systematic distortions.

The analysis uses 111 variables based on Cao and You (2024), including current-period values, lagged values, and first-order differences[2]. This study utilizes raw financial variables as input features, leveraging the tree-based algorithms' inherent robustness to outliers due to their reliance on relative value rankings for split decisions[3].

To respect the time-series nature of forecasting, the model is trained with a 10-year (120-month) rolling-window, validated on the subsequent 12 months for hyper-parameter tuning, and then applied three months after each fiscal year-end when both annual statements and initial MEF become available[4]. Both initial MEF and MLEF were constructed strictly using information available at the estimation point, without incorporating any future data such as revised MEF or actual earnings[5].

This study applies XGBoost, a gradient boosting algorithm chosen for its ability to handle missing data and control model complexity. The model utilizes mean absolute error as the loss function for its robustness[6].

3. Research Design

This study evaluates the forecast accuracy and value relevance of MLEF compared with MEF. The research design comprises two primary components: (1) forecast accuracy evaluation and (2) value relevance assessment.

3.1. Forecast Accuracy

Forecast accuracy is evaluated using two metrics: optimistic bias and absolute forecast error:

$$MFbias_{i,t+1} = (MEF_{i,t+1} - Actual\_E_{i,t+1}) / MVE_{i,t}$$

$$MLbias_{i,t+1} = (MLEF_{i,t+1} - Actual\_E_{i,t+1}) / MVE_{i,t}$$

$$MFAFE_{i,t+1} = |MFbias_{i,t+1}|$$

$$MLAFE_{i,t+1} = |MLbias_{i,t+1}|$$

2 In addition to the variables used by Cao and You (2024), the model also incorporates gross profit, operating profit, ordinary profit, financial and investment cash flows, MEFs, the bias in MEFs, and its absolute value.

3 Anand et al. (2019) demonstrate that tree-based earnings prediction models achieve similar accuracy whether using winsorized or non-winsorized data, confirming that their method is not sensitive to outliers.

4 Hyperparameters were optimized using scikit-learn's XGBRegressor with values close to defaults: n_estimators [100, 500, 1000], subsample [0.5, 1.0], and colsample_bytree [0.5, 1.0].

5 For example, in forecasting FY2014 earnings for a firm with a March fiscal year-end, the model was trained on data from March 2002 to February 2012, validated on data from March 2012 to February 2013, and tested on March 2013 data.

6 This study employs mean absolute error rather than mean squared error as the loss function, following recent studies that demonstrate MAE's superior robustness to outliers in earnings forecasting applications (Campbell et al., 2024; Gu et al., 2020).

where $MEF_{i,t+1}$ and $MLEF_{i,t+1}$ represent earnings forecasts for period $t+1$ by management and machine learning respectively, $Actual\_E_{i,t+1}$ is actual earnings for period $t+1$, and $MVE_{i,t}$ denotes the market value of equity at the fiscal year-end[7]. Differences in forecast bias and accuracy between MEF and MLEF are tested using paired t-tests for mean differences and Wilcoxon signed-rank tests for median differences, as each firm maintains both management and machine learning forecasts for the same period. Forecast accuracy is examined by comparing the differences between MEF and MLEF across three firm characteristics.

First, firm size is analyzed. Firms with smaller market capitalization are more likely to exhibit higher information asymmetry, which leads to optimistic bias in MEF (Kato et al., 2009). Forecast accuracy may improve if machine learning models can effectively identify and correct such biases.

Second, firm maturity is considered. Growth firms, characterized by lower retained earnings-to-total assets ratios, tend to issue more optimistic MEF. While the relationship between firm maturity and forecast accuracy has been documented for analyst forecasts (DeAngelo et al., 2006), the extent to which this applies to MEF warrants empirical investigation[8].

Third, prior MEF errors are examined. Firms with historically large forecast errors may benefit more from MLEF, whereas firms with consistently accurate past forecasts demonstrate smaller differences between MEF and MLEF. Ota (2006) found that firms tend to maintain their directional bias over time, meaning that firms with previously optimistic (or pessimistic) forecasts tend to exhibit similar tendencies in subsequent forecasts[9].

3.2. Value Relevance

This study examines the value relevance of MLEF within a modified Ohlson (2001) framework through three complementary analyses.

3.2.1 Basic Valuation Framework

For the valuation analysis, this study employs deflated versions of the earnings forecasts, denoted as $MF_{i,t+1}$ and $ML_{i,t+1}$, where:

$$MF_{i,t+1}=MEF_{i,t+1}/MVE_{i,t-1}$$
$$ML_{i,t+1}=MLEF_{i,t+1}/MVE_{i,t-1}$$

The market value of equity is modeled as follows:

$$MVE_{i,t}=\alpha_0+\beta_1 B_{i,t}+\beta_2 E_{i,t}+\beta_3 MF_{i,t+1}+Year+Ind+\varepsilon_{i,t}$$
$$MVE_{i,t}=\alpha_0+\beta_1 B_{i,t}+\beta_2 E_{i,t}+\beta_3 ML_{i,t+1}+Year+Ind+\varepsilon_{i,t}$$

---

7 Unlike the $MVE_{i,t}$ used in the forecasting accuracy analysis, the $MVE_{i,t}$ in the value relevance analysis is based on the market capitalization as of the end of June.

8 Firm maturity is grouped into quintiles based on the ratio of retained earnings to total assets at year-end.

9 To evaluate the impact of prior forecast accuracy, firms are classified into quintiles based on the three-year average of *MFAFE*, including the current fiscal year.

where $MVE_{i,t}$ is the market value of equity three months after the fiscal year-end (i.e., at the end of June), $B_{i,t}$ is the book value of equity for period t, $E_{i,t}$ denotes actual earnings for period t, and $MF_{i,t+1}$ and $ML_{i,t+1}$ refer to MEF and MLEF for period t+1, respectively[10]. i represents the company, and *Year* is the fiscal year dummy variable, *Ind* is the industry dummy variable[11]. A significant β_3 indicates that either MEF or MLEF contributes to explaining equity market value. The value relevance analysis is conducted exclusively on firms with a March fiscal year-end (t), to ensure consistency in financial reporting periods and comparability across observations.

3.2.2 Incremental Information

To evaluate whether MLEF provides additional information given MEF, the difference between the two forecasts is included:

$$diffML_{i,t+1}=(MLEF_{i,t+1}-MEF_{i,t+1})/MVE_{i,t-1}$$

$$MVE_{i,t}=\alpha_0+\beta_1 B_{i,t}+\beta_2 E_{i,t}+\beta_3 MF_{i,t+1}+\beta_4 diffML_{i,t+1}+Year+Ind+\varepsilon_{i,t}$$

where $diffML_{i,t+1}$ represents the differential information between MEF and MLEF. A significant β_4 suggests that MLEF adjusts for biases in MEF. If $\beta_4>0$, it indicates that upward machine learning adjustments relative to management forecasts ($diffML > 0$) are associated with higher firm valuations, while downward adjustments ($diffML < 0$) correspond to lower firm valuations.

3.2.3 Incremental Information by Firm Size

This study examines whether the incremental information content of *diffML* varies across firm sizes. This analysis contributes to the literature by examining three considerations specific to machine learning applications in earnings forecasting.

First, machine learning models process financial data systematically across all firm sizes, potentially offering different information patterns compared to analyst forecasts that exhibit preferential coverage for larger firms[12]. Second, larger firms tend to demonstrate higher financial reporting quality (Dechow and Dichev, 2002) and given that machine learning performance is likely to depend on input data quality, *diffML* effects vary across different reporting environments. Third, smaller firms often exhibit higher information asymmetry and greater management forecast bias, which may create more opportunities for machine learning bias correction.

If *diffML* exhibits significance for both large and small firms, it could suggest that machine

[10] These variables are normalized by dividing them by the market capitalization three months after the fiscal year-end t-1 (i.e., the end of June).

[11] Using the 33-industry sector classification provided by the Tokyo Stock Exchange.

[12] These expectations differ from Nara and Noma (2013), who found that analyst forecast adjustments provide value relevance primarily for large firms due to analysts' preferential information production. In their study, immediately after management forecast releases, the difference between management and analyst forecasts shows no information content for larger firms, while a few weeks later, information content emerges only for larger firms. Unlike analysts who face resource constraints leading to firm-size preferences, machine learning processes available data systematically, allowing this study to test whether *diffML* provides incremental information more uniformly across firm sizes.

learning adjustments may be valued regardless of firm size. If significance appears only for large firms, this suggests that market recognition depends on information quality similar to analyst forecasts. Alternatively, stronger effects for small firms could suggest that machine learning provides incremental value relevance in information-sparse environments.

4. Results

After excluding outliers in forecast bias measures (top and bottom 0.5%), the analysis sample comprised 24,143 firm-year observations from 2013-2022.

4.1. Forecast Accuracy

Table 1 Descriptive Statistics and Correlations for Forecast Accuracy Analysis

Panel A: Descriptive statistics

	N	Mean	SD	Min	P25	Median	P75	Max
$Actual\_E_{i,t+1}$	24,143	9,320	53,820	-671,216	305	1,149	4,420	2,493,983
$MEF_{i,t+1}$	24,143	9,162	48,112	-85,000	350	1,140	4,300	2,250,000
$MLEF_{i,t+1}$	24,143	8,875	46,338	-96,168	318	1,121	4,261	2,316,489
$MFbias_{i,t+1}$	24,143	0.432	8.799	-260.551	-1.883	-0.273	1.367	167.788
$MLbias_{i,t+1}$	24,143	0.146	8.662	-264.667	-2.086	-0.269	1.404	150.984
$MFAFE_{i,t+1}$	24,143	3.888	7.906	0.000	0.615	1.668	4.037	260.551
$MLAFE_{i,t+1}$	24,143	3.949	7.711	0.000	0.671	1.761	4.185	264.667

Panel B: Correlation coefficients

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) $Actual\_E_{i,t+1}$		0.950	0.925	-0.068	-0.066	-0.066	-0.061
(2) $MEF_{i,t+1}$	0.891		0.984	-0.010	-0.009	-0.041	-0.037
(3) $MLEF_{i,t+1}$	0.879	0.968		-0.011	-0.002	-0.042	-0.042
(4) $MFbias_{i,t+1}$	-0.353	-0.048	-0.070		0.965	0.473	0.420
(5) $MLbias_{i,t+1}$	-0.286	0.001	0.039	0.870		0.431	0.370
(6) $MFAFE_{i,t+1}$	-0.300	-0.250	-0.257	-0.057	-0.029		0.963
(7) $MLAFE_{i,t+1}$	-0.307	-0.266	-0.295	-0.043	-0.094	0.801	

The values of bias and AFE in Panel A are multiplied by 100 and presented as percentages. The upper triangular matrix in Panel B represents Pearson's product-moment correlation coefficients, while the lower triangular matrix represents Spearman's rank correlation coefficients. Earnings forecasts ($MEF_{i,t+1}$ and $MLEF_{i,t+1}$) and actual earnings ($Actual\_E_{i,t+1}$) are presented as undeflated values to show the magnitude and distribution of earnings and forecast levels, while bias and accuracy measures are deflated by the market capitalization three months after the fiscal year-end t-1 (i.e., the end of June).

Table 1 summarizes forecast accuracy. Panel A presents descriptive statistics, showing that MLEF exhibits lower optimistic bias and comparable accuracy to MEF. Panel B reports a positive correlation between the two forecast types, suggesting similar trend capture despite differing bias.

Table 2 Forecast Bias and Accuracy by Firm Characteristics

Panel A: By firm size

		mean			median			mean			median		
Size	N	*MF bias*	*ML bias*	*difference*	*MF bias*	*ML bias*	*difference*	*MF AFE*	*ML AFE*	*difference*	*MF AFE*	*ML AFE*	*difference*
1 (small)	4,833	1.803	0.879	0.923*** (18.635)	−0.068	−0.326	0.258*** (16.514)	6.910	6.933	−0.023 (−0.488)	3.276	3.418	−0.142
2	4,831	0.305	0.099	0.206*** (6.476)	−0.309	−0.361	0.052*** (4.201)	4.216	4.257	−0.041 (−1.395)	2.152	2.244	−0.092 (0.209)
3	4,829	−0.006	−0.143	0.137*** (4.437)	−0.356	−0.328	−0.028 (0.121)	3.455	3.513	−0.057** (−2.121)	1.699	1.754	−0.055 (−0.839)
4	4,826	0.086	0.046	0.040* (1.741)	−0.283	−0.224	−0.060*** (−4.008)	2.841	2.895	−0.054*** (−2.639)	1.275	1.324	−0.049 (0.014)
5 (large)	4,824	−0.030	−0.150	0.120*** (5.324)	−0.258	−0.223	−0.035 (−1.550)	2.013	2.142	−0.129*** (−6.137)	0.864	0.961	−0.097*** (−4.384)

Panel B: By firm maturity

		mean			median			mean			median		
RE/AT	N	*MF bias*	*ML bias*	*difference*	*MF bias*	*ML bias*	*difference*	*MF AFE*	*ML AFE*	*difference*	*MF AFE*	*ML AFE*	*difference*
1 (growth)	4,829	2.187	1.083	1.104*** (21.958)	−0.038	−0.287	0.249*** (22.568)	6.702	6.763	−0.061 (−1.290)	2.767	3.093	−0.326 (−0.900)
2	4,829	0.435	0.146	0.289*** (7.913)	−0.307	−0.341	0.034*** (4.520)	4.393	4.500	−0.106*** (−3.142)	2.085	2.193	−0.108** (−2.224)
3	4,829	0.049	−0.053	0.102*** (3.718)	−0.341	−0.297	−0.044 (−0.153)	3.530	3.602	−0.072*** (−2.883)	1.720	1.833	−0.112 (−1.406)
4	4,828	−0.203	−0.185	−0.019 (−0.858)	−0.354	−0.303	−0.051*** (−5.018)	2.703	2.735	−0.031 (−1.596)	1.386	1.437	−0.051 (0.268)
5 (mature)	4,828	−0.309	−0.260	−0.049*** (−2.876)	−0.258	−0.184	−0.074*** (−8.149)	2.111	2.144	−0.033** (−2.086)	1.042	1.085	−0.044 (−0.235)

Panel C: By prior management forecast errors

		mean			median			mean			median		
Errors	N	*MF bias*	*ML bias*	*difference*	*MF bias*	*ML bias*	*difference*	*MF AFE*	*ML AFE*	*difference*	*MF AFE*	*ML AFE*	*difference*
1 (small)	4,833	0.014	−0.081	0.096*** (8.578)	−0.155	−0.151	−0.004** (2.287)	1.350	1.433	−0.083*** (−8.176)	0.639	0.711	−0.072*** (−7.474)
2	4,831	0.002	−0.096	0.098*** (5.895)	−0.304	−0.287	−0.016 (0.861)	2.069	2.154	−0.085*** (−5.853)	1.176	1.272	−0.095*** (−3.619)
3	4,829	0.044	−0.111	0.155*** (6.695)	−0.341	−0.347	0.006* (1.833)	2.891	2.970	−0.080*** (−3.800)	1.732	1.780	−0.048 (−0.684)
4	4,826	0.004	−0.227	0.231*** (7.307)	−0.515	−0.494	−0.021** (2.556)	4.390	4.445	−0.056* (−1.942)	2.482	2.566	−0.084 (0.261)
5 (large)	4,824	2.097	1.249	0.848*** (14.210)	−0.374	−0.468	0.095*** (11.438)	8.749	8.750	0.000 (−0.006)	4.472	4.637	−0.165** (2.041)

***, **, * Denote statistical significance at the 1%, 5%, and 10% levels, respectively. All bias and accuracy measures presented in this table use the same deflated values as defined in Table 1. Test statistics are reported in parentheses (t-statistics from paired t-tests for mean differences, z-statistics from Wilcoxon signed-rank tests for median differences).

Table 2 provides a comparative analysis of forecast accuracy across firm characteristics.

Panel A, focusing on firm size, reveals varied patterns based on statistical measures. For smaller firms, mean values indicate MLEF mitigates optimistic bias in MEF (1.803 vs 0.879), while median values suggest MLEF exacerbates pessimistic tendencies (-0.068 vs -0.326). This divergence suggests that machine learning corrects extremely optimistic outliers but overcorrects for typical small firms. For larger firms, mean forecasts become more pessimistic with MLEF (-0.030 vs -0.150), while median forecasts improve slightly (-0.258 vs -0.223). This aligns with evidence that larger firms issue less biased forecasts (Kato et al., 2009), providing less room for improvement. Despite these mixed effects, accuracy improvements remain modest across all firm size categories.

Panel B analyzes firm maturity and reveals similar patterns. For growth firms, mean values demonstrate MLEF mitigates optimistic bias in MEF (2.187 vs 1.083), while median values show MLEF exacerbating pessimistic tendencies (-0.038 vs -0.287). This suggests growth firms benefit from bias correction for extreme cases but experience overcorrection for typical forecasts. Mature firms (quintile 5) show minimal changes in both mean and median bias. The limited impact on mature firms reflects their stable operating environments and accurate forecasting practices, offering less opportunity for enhancement.

Panel C examines prior forecast error histories and displays the most notable mean-median divergence. For firms with large historical errors, mean values reveal that MLEF reduces optimistic bias in MEF (2.097 vs 1.249), while median values show MLEF worsening already pessimistic forecasts (-0.374 vs -0.468). In contrast, firms with smaller historical errors (quintile 1) demonstrate that MLEF worsens mean bias (0.014 vs -0.081) while modestly improving median performance (-0.155 vs -0.151), showing mixed effects across statistical measures.

The analysis reveals a mean-median divergence where machine learning adjustments show contrasting effects depending on the statistical measure examined. While MLEF reduces extreme optimistic bias (improving mean values), it shifts median forecasts toward greater pessimism. This pattern appears across firm size, maturity, and prior forecast error analyses, indicating that the practical benefits of machine learning bias correction are limited and context dependent. Although mean bias reduction represents an improvement, the concurrent deterioration in median forecasts suggests that the overall enhancement in forecast quality is modest.

4.2. Value Relevance

Table 3 Descriptive Statistics and Correlations for Value Relevance Analysis

Panel A: Descriptive statistics

	N	Mean	SD	Min	P25	Median	P75	Max
$MVE_{i,t}$	16,794	1.113	0.348	0.449	0.894	1.056	1.256	3.357
$B_{i,t}$	16,794	1.285	0.748	0.084	0.726	1.150	1.707	4.173
$E_{i,t}$	16,794	0.070	0.075	−0.522	0.040	0.069	0.105	0.385
$MF_{i,t+1}$	16,794	0.080	0.046	−0.075	0.047	0.074	0.105	0.270
$ML_{i,t+1}$	16,794	0.078	0.054	−0.080	0.041	0.070	0.107	0.271
$diffML_{i,t+1}$	16,794	−0.002	0.021	−0.219	−0.008	0.000	0.007	0.186

Panel B: Correlation coefficients

	(1)	(2)	(3)	(4)	(5)	(6)
(1) $MVEi,t$		0.104	0.205	0.338	0.344	0.127
(2) Bi,t	0.148		0.218	0.370	0.336	0.036
(3) Ei,t	0.266	0.355		0.610	0.657	0.317
(4) $MF_{i,t+1}$	0.343	0.410	0.772		0.918	0.123
(5) $ML_{i,t+1}$	0.366	0.377	0.795	0.923		0.506
(6) $diffML_{i,t+1}$	0.198	0.078	0.328	0.173	0.465	

The upper triangular matrix in Panel B represents Pearson's product-moment correlation coefficients, while the lower triangular matrix represents Spearman's rank correlation coefficients. All variables are deflated by market capitalization three months after fiscal year-end t-1 (i.e., the end of June). $MF_{i,t+1}$ and $ML_{i,t+1}$ represent deflated versions of the earnings forecasts $MEF_{i,t+1}$ and $MLEF_{i,t+1}$ from the forecast accuracy analysis.

Following the forecast accuracy analysis, this study examines the value relevance of *diffML*, investigating whether forecast differences provide informational content for equity valuation. The value relevance of MLEF is examined within the Ohlson (2001) valuation framework, with preliminary analysis in Table 3 and key regression results in Table 4.

Table 3 presents descriptive statistics and correlation coefficients. Panel A shows that the difference between MLEF and MEF (*diffML*) has a modest negative mean (-0.002) with a median of zero, capturing deviations between forecast types across firms. Panel B reports the correlation matrix, highlighting a high positive correlation between MEF and MLEF[13].

Table 4 evaluates the value relevance of different forecasting models. Panel A shows that models incorporating *diffML* better explain firm value compared to those using only MF or ML, with significant coefficients. Panel B presents Vuong test results, demonstrating that models including *diffML* perform better than those without it, supporting its contribution to value relevance.

These findings partially align with Ota (2010), who found management forecasts positively associated with firm value. However, this study's negative coefficient on book value contrasts with Ota's positive result, demonstrating how forward-looking variables now more comprehensively capture value-relevance expectations in recent years[14].

Table 5 investigates how firm size affects the value relevance of *diffML*. Panels A and B analyze

[13] Correlation coefficients and VIF tests (all values < 2) indicate minimal multicollinearity concerns.

[14] Ota (2010) analyzed 1979-1999 data, while this study covers 2013-2022. Year-by-year tests found positive book value coefficients only in 2013.

Table 4 Value Relevance Regression Results: Comparison of Forecast Models

Panel A: *ML* and *MF*, *diffML*

	Intercept	*B*	*E*	*MF*	*ML*	*diffML*	*Adj R^2*	*N*	*AIC*
MF_model	1.088***	−0.021***	0.032	2.463***			0.342	15,353	5,782
	(24.213)	(−5.517)	(0.800)	(34.970)					
ML_model	1.125***	−0.017***	−0.111***		2.273***		0.346	15,353	5,693
	(25.147)	(−4.625)	(−2.626)		(36.319)				
diff_model	1.103***	−0.021***	−0.121***	2.519***		1.476***	0.348	15,353	5,636
	(24.652)	(−5.574)	(−2.869)	(35.861)		(12.172)			

Panel B: Vuong test

Model to compare	z value by Vuong test
MF_model vs ML_model	−1.838**
MF_model vs diff_model	−4.792***
ML_model vs diff_model	−3.005***

***, **, * Denote statistical significance at the 1%, 5%, and 10% levels, respectively. The t-statistics are reported in the parentheses. All variables are deflated by market capitalization three months after fiscal year-end t-1(i.e., the end of June).

Table 5 Value Relevance of diffML by Firm Size

Panel A: By small firms

	Intercept	*B*	*E*	*MF*	*ML*	*diffML*	*Adj R^2*	*N*	*AIC*
MF_model	1.119***	0.017***	0.142**	1.392***			0.359	2,927	423.4
	(14.619)	(2.696)	(2.465)	(12.106)					
ML_model	1.141***	0.020***	0.094		1.156***		0.356	2,927	435.6
	(14.887)	(3.026)	(1.537)		(11.575)				
diff_model	1.125***	0.017***	0.091	1.409***		0.488***	0.360	2,927	418.2
	(14.712)	(2.603)	(1.498)	(12.245)		(2.675)			

Panel B: Vuong test by small firms

Model to compare	z value by Vuong test
MF_model vs ML_model	0.767
MF_model vs diff_model	−1.074
ML_model vs diff_model	−0.712

Panel C: By large firms

	Intercept	*B*	*E*	*MF*	*ML*	*diffML*	*Adj R^2*	*N*	*AIC*
MF_model	1.089***	−0.026*	−0.259**	3.803***			0.385	3,157	883.9
	(12.398)	(−1.864)	(−2.016)	(17.517)					
ML_model	1.136***	−0.019	−0.376***		3.554***		0.395	3,157	834.3
	(13.110)	(−1.438)	(−2.936)		(19.001)				
diff_model	1.108***	−0.026*	−0.435***	3.873***		2.673***	0.396	3,157	827.3
	(12.730)	(−1.883)	(−3.360)	(17.986)		(7.636)			

Panel D: Vuong test by large firms

Model to compare	z value by Vuong test
MF_model vs ML_model	−1.697***
MF_model vs diff_model	−2.729***
ML_model vs diff_model	−3.228***

***, **, * Denote statistical significance at the 1%, 5%, and 10% levels, respectively. The t-statistics are reported in the parentheses. All variables are deflated by market capitalization three months after fiscal year-end t-1(i.e., the end of June). Panels A and B analyze small firms, which represent the bottom quintile based on fiscal year-end market capitalization within each year. Panels C and D analyze large firms, which represent the top quintile based on fiscal year-end market capitalization within each year.

small firms, where *diffML* provides incremental information with performance comparable to both MF and ML models. Vuong test results indicate that while *diffML* has a statistical advantage over ML, its improvement relative to the MF model is limited for small firms.

Panels C and D examine large firms, where *diffML* shows positive influence on firm valuation and added explanatory power compared to alternative models. Vuong test results indicate the *diffML* model has a statistical advantage over both MF and ML in this context. However, the adjusted R^2 improves only modestly (0.011 increase), which appears economically modest.

Notably, this pattern appears to contrast with the forecast accuracy results: while smaller firms exhibit larger bias corrections (Table 2), larger firms demonstrate greater value relevance for *diffML*. This finding demonstrates that the magnitude of bias correction may not directly correspond to market recognition, though the underlying mechanisms warrant further investigation.

While machine learning reduces forecast bias for small firms, the value relevance of *diffML* appears more evident for large firms. This pattern may reflect differences in financial reporting quality across firm sizes: although small firms appear to benefit more from bias correction, large firms' potentially higher-quality financial disclosures facilitate market participants' ability to better interpret and value the corrected information, possibly leading to greater incremental value relevance despite smaller bias corrections[15].

5. Conclusion

This study investigates whether earnings forecasts from machine learning (ML) models provide incremental value relevance beyond management forecasts in the Japanese equity market. The analysis demonstrates that ML forecasts reduce optimistic bias but do not improve accuracy. The difference between ML and management forecasts (diffML) carries incremental information valued by the market, although its overall impact is modest. This effect varies across firm sizes, with stronger results for larger firms. These findings demonstrate that ML forecasts complement management forecasts and contribute to earnings forecasting literature, although they exhibit mixed effectiveness across different measures and firm characteristics. However, the magnitude of the incremental information's contribution appears limited, and the mechanisms through which investors incorporate ML-based insights require further investigation.

[15] Chen et al. (2022) further show that machine learning model performance in earnings forecasting depends critically on financial data quality, with better results for higher-quality financial reporting implementations.

References

Anand, V., Brunner, R., Ikegwu, K., and Sougiannis, T. 2019. Predicting profitability using machine learning. Available at SSRN: http://dx.doi.org/10.2139/ssrn.3466478

Campbell, J. L., Ham, H., Lu, Z., and Wood, K. 2024. Expectations matter: When (not) to use machine learning earnings forecasts. Available at SSRN: http://dx.doi.org/10.2139/ssrn.4495297

Cao, K., and You, H. 2024. Fundamental analysis via machine learning. *Financial Analysts Journal*, 80(2): 74-98.

Chen, X., Cho, Y. H., Dou, Y., and Lev, B. 2022. Predicting future earnings changes using machine learning and detailed financial data. *Journal of Accounting Research*, 60(2): 467-515.

DeAngelo, H., DeAngelo, L., and Stulz, R. M. 2006. Dividend policy and the earned/contributed capital mix: a test of the life-cycle theory. *Journal of Financial Economics*, 81(2): 227-254.

Dechow, P. M., and Dichev, I. D. 2002. The quality of accruals and earnings: The role of accrual estimation errors. *The Accounting Review*, 77(s-1): 35-59.

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

Kato, K., Skinner, D. J., and Kunimura, M. 2009. Management forecasts in Japan: An empirical study of forecasts that are effectively mandated. *The Accounting Review*, 84(5): 1575-1606.

Nara, S., and Noma, M. 2013. Do analysts add value after management forecast revision? *Securities Analysts Journal*, 51: 88-98.

Ohlson, J. A. 2001. Earnings, book values, and dividends in equity valuation: An empirical perspective. *Contemporary Accounting Research*, 18(1): 107-120.

Ota, K. 2006. Determinants of bias in management earnings forecasts: Empirical evidence from Japan. *International accounting: Standards, regulations, and financial reporting*, 267-294.

Ota, K. 2010. The value relevance of MEF and their impact on analysts' forecasts: Empirical evidence from Japan. *Abacus*, 46(1): 28-59.