Incremental Predictive Value of Management Forward - Looking Disclosures with Respect to Subsequent Fiscal - Year Profitability: Evidence from Chinese A - Share Listed Companies

Ye Wu

Business School, Sichuan University, Chengdu, China
Email: wuye031026@163.com

Abstract This study checks if management forward - looking disclosure (FLD) brings any additional predictive power with regard to subsequent fiscal-year profitability—beyond that of ordinary accounting data—for Chinese A-share non-financial firms. FLD is recognized at the sentence level via a TF - IDF logistic classifier; it is measured as the fraction of forward-looking sentences in MD&A. A fixed stratified sample is taken in chronological order and split into training, validation, and test periods. Ridge and XGBoost models compare a purely financial specification against one of identical form but with added FLDIntensity. With regard to the central ROA test and to robustness analyses involving ROE, rolling test years, or non - ST/*ST firms, FLDIntensity fails to yield a consistent improvement in forecast accuracy out-of-sample. That result does not prove that managerial narratives is of no value; it allows for the possibility that a crude prevalence count omits tone or detail present in more complex text forms.

Keywords Forward-looking disclosure(FLD); MD&A; Profitability forecasting; Machine learning; A-share listed firms;

1. Introduction

Financial statements give an account of actual economic results, whereas the Management Discussion and Analysis (MD&A) portion of a report gives room to discuss operations, strategy, planned investments, risks, and expectations about the future. With regard to the firm's condition, managers have access to data not entirely available to outside investors — hence the value of narrative disclosure[1][2]. Forward - looking disclosure (FLD) is possibly helpful in predicting later profitability. But it is not enough merely to ask whether forward - looking language has content. If current accounting indicators already capture the basic signals in question, extra forward - looking language may not greatly improve forecasts.

Earlier work has indicated that corporate narratives may have economic value as sources of information [3]–[5], while studies of MD&A features show that richer textual measures can also help explain or predict firm outcomes[7]–[9]. Still such findings do not justify treating a straightforward measure of how much of an MD&A is forward-looking as a source of added predictive value over standard financial facts. That distinction is significant since association and incremental prediction are separate empirical aims[13]. One textual variable may be connected to future profitability without helping to lower forecast error—once the overlapping information of current profitability, cash flow, growth, leverage, and such accounting items is known. The research gap is not wide: it is now a matter of deciding whether the frequency of forward-looking sentences yields reliable incremental out-of-sample predictive value.

With an eye toward this question, the present study uses data about Shanghai - and Shenzhen - listed A - share non - financial companies. It constructs a text measure based on a reproducible year - by - industry proportional stratified sample of 6000 firm - years (2015–2024). Forward - looking sentences are marked by means of a sentence - level machine - learning classifier, and FLDIntensity is taken to be the proportion of valid MD&A sentences regarded as forward - looking. Profitability data pertain to 5,

356 firm - years between 2015 and 2023. With regard to Ridge regression and to XGBoost, a purely financial model (M1) is compared with one of the same structures that also has FLDIntensity (M2). The same data set and same preprocessing steps are used throughout, with prior years assigned to model building and later years kept aside for out-of-sample testing. Incremental value is measured directly by the question of whether M2 cuts forecast error when compared to M1.

The results do not confirm a constant positive incremental role for FLDIntensity. With respect to the main next-fiscal-year ROA target, inclusion of FLDIntensity tends to make Ridge forecasts only slightly less accurate (with negligible economic consequences), whereas XGBoost brings no marked improvement — nor one that is reliably statistical. The same basic conclusion holds when ROE is substituted as a target, when the out-of-sample test moves forward annually, and when ST/*ST firm - years are deleted. Therefore, the data do not back up the idea that FLD prevalence yields a consistent advantage in predicting profitability over the information available from standard financial variables.

This paper presents an analysis that distinguishes between the overall information content of managerial narratives and the incremental predictive value of one textual measure. It avoids the question of whether MD&A text is generally informative and instead asks if a clear prevalence measure offorward - looking language gives added value to a given accounting information set under time - ordered out-of-sample testing. It also identifies a limit for simple disclosure counts: the volume of forward - looking language may not help much with prediction—even if the narrative's tone or its specificity or credibility or semantic content are all useful.

2. Literature Review and Hypothesis Development

Financial statements are mainly concerned with past or realized results, whereas the Management Discussion and Analysis (MD&A) permits managers to refer to current operating facts, to strategic choices, to planned investments, to risks, and to expectations which do not yet show up wholly in accounting data. With private knowledge of upcoming developments, managers are in a position to make useful narrative disclosures concerning information asymmetry between firms and outside investors[1][2]. Forward - looking statements may be of help in forecasting, since they refer to planned actions, to expectations, to dangers, or to other unrealized facts of the situation. Still, one must ask whether such statements provide — separately from known facts — any new information that bears on future earnings and is not already embodied in standard financial variables.

Available texts on analysis suggest that corporate narratives can be sources of economically useful information, though they also make clear that outcomes vary with text representation. Earlier work has dealt with readability, forward-looking tone, and finance-related language [3][5][6]. Other studies examine shifts in MD&A wording and forward-looking disclosure [7][8]. Such measures cover (in part) separate aspects of managerial communication—not only the quantity of forward-looking text—seems important. With regard to the forecasting question at hand, Bochkay and Levine[9]claim that richer MD&A narrative characteristics help to predict out-of-sample profitability when joined with standard financial data. Evidence drawn from Chinese companies shows a similar link between MD&A tone or forward - looking disclosure and later value for the firm, stock liquidity, or earnings[10]-[12]. Taken together, these findings point to managerial narratives as possible sources of forward - looking information; they also note dependence of textual value on the summary feature used.

The present study deals with a deliberately narrow textual construct. FLDIntensity is given as the share of valid MD&A sentences assigned to the category of forward - looking statements. It is an indicator of the frequency of statements directed toward the future, but not of optimism or pessimism, of specificity or vagueness, of plausibility or formula - talk, nor of particular topics (e.g., sales). Under this definition, evidence that tone or semantic features carry information does not imply that FLD prevalence improves forecast performance. What is being asked here is not an association but an increment: if one already has current ROA, ROE, cash flow figures, growth, leverage, turnover, size, or firm age, is FLDIntensity helpful in reducing forecast error for later (unseen) cases? Such a division corresponds to the wider dichotomy of explanation versus prediction noted by Shmueli [13].

To have incremental predictive value, FLDIntensity requires fulfillment of two conditions. Forward-looking sentences need to give some account of later operations, and the relevant part of that account should not be fully captured by the accounting facts used in the baseline model. With both conditions satisfied, the addition of FLDIntensity to a forecasting model (otherwise unaltered) is likely to help predict profitability out-of-sample. The present study proceeds to test the hypothesis:

H1: Adding management forward-looking disclosure intensity to conventional financial information improves out-of-sample prediction of subsequent fiscal-year profitability.

The hypothesis is not about the general informational value of managerial narrative but specifically about the incremental value of FLD prevalence.

ISSN 3106-5783 (Print) ISSN 3106-7042 (Online)

3. Data, Text Measurement, and Research Design

3.1 Population and sample construction

Financial variables are obtained from CSMAR. The study covers Shanghai- and Shenzhen-listed A-share non-financial firms from 2015 to 2024. After financial and listing-status screening, 38,487 firm-year observations can be matched to full annual reports. From this population, a fixed formal text sample of 6,000 firm-years is drawn without replacement, with 600 observations from each fiscal year and industry-proportional stratification within each year. Requiring a valid FLDIntensity measure and subsequent profitability data yields the final supervised modeling sample of 5,356 firm-years from 2015 to 2023.

Table 1. Sample construction and final prediction sample

Stage	Years / rule	N	Comment
Matched annual-report population	2015–2024	38,487	Shanghai/Shenzhen A-share non-financial firm-years
Formal text sample	600 per year; industry-proportional stratification	6,000	Fixed sample drawn without replacement
Final supervised modeling sample	2015–2023	5,356	Valid FLDIntensity and subsequent profitability data

3.2 Financial predictors, targets, and timing

The main profitability outcome is ROA for fiscal year t+1, while the financial predictors and FLDIntensity are indexed to fiscal year t. The baseline financial information set includes current ROA, current ROE, revenue growth, leverage, total-asset turnover, operating cash flow divided by total assets, firm size, and firm age. ROE(t+1) is used as an alternative profitability outcome. Outcomes are matched to predictors by firm and fiscal year.

An important timing limitation must be acknowledged. For calendar-year firms, the annual report for fiscal year t is normally released during calendar year t+1[18]. Consequently, ROA(t+1) measures profitability over the same fiscal year in which the year-t annual report is disclosed, rather than over a complete twelve-month period beginning after the disclosure date. The design should therefore be interpreted as a fiscal-year-aligned incremental predictive comparison, not as a strict real-time one-year-ahead forecast from the filing date. Because this timing convention is identical for M1 and M2, it still permits a controlled comparison of whether FLDIntensity adds predictive content beyond the financial information set, but it limits stronger claims about forecasting future profitability.

Table 2. Prediction variables

Variable	Definition / role
ROA(t+1)	Subsequent fiscal-year return on assets; main target
ROA(t)	Current return on assets
ROE(t)	Current return on equity
RevenueGrowth(t)	Revenue growth rate
Leverage(t)	Asset-liability ratio
AssetTurnover(t)	Total-asset turnover
OCF/Assets(t)	Net operating cash flow divided by total assets
Size(t)	Natural logarithm of total assets
FirmAge(t)	Fiscal year minus listing year plus one
FLDIntensity(t)	Forward-looking MD&A sentences / total valid MD&A sentences

3.3 Annual-report retrieval and MD&A extraction

Historical disclosure records identify the relevant annual report. When direct CNINFO retrieval was throttled, the same reports were obtained from public mirrors and retained only after PDF structure and content checks. Only the management narrative is extracted. The parser recognizes historical variants of MD&A headings and uses major formal headings as section boundaries; older board-report structures are used only when no formal MD&A heading is found. For dual mainland/Hong Kong reporting

formats, only the mainland-rule MD&A is retained. The pipeline was stress-tested on 100 reports before the formal run. In the formal sample, unresolved extraction failures are retained as missing rather than eliminated by progressively relaxing the extraction rules.

3.4 Reference annotation and sentence classifier

After non-substantive fragments are removed, MD&A text is segmented using Chinese sentence-ending punctuation. A sentence is labeled as forward-looking when it contains substantive discussion of future plans, strategic actions, targets, expectations, investment arrangements, risks, uncertainty, or conditional future outcomes. Purely historical events, descriptions of the current state, and generic statements without a substantive future implication are labeled non-forward-looking. Sentences combining historical information with substantive future content are classified as forward-looking.

The classifier is developed on an annotated corpus of 1,190 sentences drawn from 60 annual reports, including 450 FLD and 740 non-FLD sentences. Character-level and word-level TF-IDF representations are evaluated with logistic regression and linear SVM under report-grouped five-fold cross-validation, so sentences from the same annual report remain in the same fold. The fused character-plus-word TF-IDF logistic regression model is selected and frozen before the final evaluation.

Final evaluation uses a held-out set of 300 sentences from 20 annual reports that were not used in model development, comprising 96 FLD and 204 non-FLD sentences. The frozen classifier achieves 84.33% accuracy, 75.79% precision, 75.00% recall, and a 75.39% F1-score for the FLD class. Because the held-out set is limited in size and the reference labels were not independently double-coded, these statistics should be interpreted cautiously. They provide a workable operational basis for constructing firm-year FLDIntensity rather than evidence of definitive classification validity.

Table 3. Report-grouped five-fold classifier development results

Text representation / classifier	Accuracy	Precision	Recall	F1
Character TF-IDF + Logistic	0.6815	0.5741	0.6111	0.5920
Character TF-IDF + Linear SVM	0.6748	0.5793	0.5111	0.5431
Character + Word TF-IDF + Logistic	0.7008	0.6000	0.6267	0.6130
Character + Word TF-IDF + Linear SVM	0.6933	0.6019	0.5578	0.5790

Note: Precision, recall, and F1 refer to the FLD class.

3.5 Construction and interpretation of FLDIntensity

$$\text{FLDIntensity}(i,t)=\text{ForwardLookingSentences}(i,t)/\text{TotalValidMDASentences}(i,t) \quad (1)$$

FLDIntensity is a prevalence measure only. It captures how much of the MD&A is forward-looking, not whether the statements are positive or negative, specific or vague, credible or boilerplate.

3.6 Incremental prediction framework

The forecasting exercise is predictive rather than causal. Within each algorithm, M1 and M2 use the same firm-years, preprocessing, temporal split, tuning procedure, and evaluation rules; the only information-set change is the addition of FLDIntensity.

$$M1: ROA(i,t+1) = f\left[FinancialInformation(i,t)\right] \quad (2)$$

$$M2: ROA(i,t+1) = f[FinancialInformation(i,t), FLDIntensity(i,t)] \quad (3)$$

Ridge is used as a regularized linear benchmark[14] and XGBoost as a nonlinear comparison[15]. Persistence and historical-mean forecasts provide simple benchmarks. The main split is 2015-2019 for training, 2020-2021 for validation, and 2022-2023 for final testing. Hyperparameters are selected on validation data, after which the model is refit on train plus validation and evaluated once on the final test set.

Preprocessing uses training-period information only. Financial-predictor winsorization cutoffs are estimated from Train and carried forward unchanged; Ridge standardizes predictors within its pipeline, while XGBoost uses the clipped values directly.

3.7 Forecast evaluation and paired inference

Performance is evaluated with MAE, RMSE, test-set R^2, and out-of-sample R^2 relative to the historical-mean benchmark. Incremental value is measured by the M1-M2 error difference, where a positive value favors M2.

$$\Delta MAE = MAE(M1) - MAE(M2); \Delta RMSE = RMSE(M1) - RMSE(M2) \quad (4)$$

A positive difference means that M2 has the lower error. Uncertainty is evaluated with a paired cluster bootstrap at the firm level[16][17]. Each replication resamples firms as whole clusters and applies the same resampled firms to the M1 and M2 forecasts, preserving both the pairing of the two model errors and within-firm dependence when a firm appears more than once in the test data. The main comparison uses 2,000 bootstrap replications. Coefficients and feature-importance values are reported only as diagnostics because the study is designed for prediction rather

ISSN 3106-5783 (Print) ISSN 3106-7042 (Online)

than causal parameter estimation.

4. Validation of the Text Measure and Descriptive Evidence

4.1 Final held-out classification performance

Table 4. Final held-out FLD-classification performance

Metric	Value
Accuracy	0.8433
Precision (FLD)	0.7579
Recall (FLD)	0.7500
F1-score (FLD)	0.7539
Test sentences	300
FLD / non-FLD	96 / 204
Confusion matrix	[[181, 23], [24, 72]]

The held-out evaluation indicates workable sentence-level classification performance, with 84.33% accuracy and an FLD-class F1-score of 75.39%. Because the reference labels were not independently double-coded, these results support the operational use of the classifier for constructing FLDIntensity but should not be interpreted as definitive classification validity.

4.2 Formal-sample text coverage and FLD distribution

Table 5. Formal-sample FLD coverage and distribution by fiscal year

Year	Formal N	Valid FLD	Missing	Coverage	Mean FLD	Median FLD
2015	600	598	2	99.67%	0.3099	0.3081
2016	600	598	2	99.67%	0.3141	0.3086
2017	600	597	3	99.50%	0.3063	0.3036
2018	600	599	1	99.83%	0.3077	0.3046
2019	600	594	6	99.00%	0.3150	0.3120
2020	600	588	12	98.00%	0.3137	0.3132
2021	600	593	7	98.83%	0.3049	0.3042
2022	600	593	7	98.83%	0.3070	0.3017
2023	600	596	4	99.33%	0.3196	0.3176
2024	600	595	5	99.17%	0.3237	0.3202
Total	6,000	5,951	49	99.18%	0.3122	0.3093

Valid FLDIntensity is obtained for 5,951 of the 6,000 sampled firm-years, giving 99.18% overall text-measure coverage. Annual mean FLDIntensity remains close to 0.31 throughout the sample, indicating that the prevalence measure is broadly stable across years rather than being driven by a single period.

4.3 Descriptive statistics for the main modeling sample

Table 6. Descriptive statistics: 2015–2023 main modeling sample

Variable	N	Mean	SD	Min	P25	Median	P75	Max
ROA(t)	5,356	0.0375	0.0717	-0.2779	0.0120	0.0381	0.0733	0.2170
ROE(t)	5,356	0.0513	0.1598	-0.8050	0.0249	0.0701	0.1210	0.3925
RevenueGrowth(t)	5,356	0.1617	0.4576	-0.5899	-0.0447	0.0889	0.2527	3.1133
Leverage(t)	5,356	0.4130	0.2073	0.0658	0.2440	0.4010	0.5631	0.9101

Variable	N	Mean	SD	Min	P25	Median	P75	Max
AssetTurnover(t)	5,356	0.6198	0.4231	0.0496	0.3491	0.5302	0.7664	2.6074
OCF/Assets(t)	5,356	0.0490	0.0680	-0.1637	0.0111	0.0476	0.0892	0.2397
Size(t)	5,356	22.2570	1.2830	19.7970	21.3637	22.0584	22.9877	26.1132
FirmAge(t)	5,356	11.4341	8.1855	1.0000	4.0000	9.0000	19.0000	27.0000
FLDIntensity(t)	5,356	0.3109	0.0830	0.0000	0.2549	0.3084	0.3647	0.6279
ROA(t+1)	5,356	0.0288	0.0918	-1.3330	0.0091	0.0338	0.0667	1.2848

Note: Financial predictors reflect the final model-input convention: Train-only 1%/99% winsorization thresholds are applied to all splits. FLDIntensity and ROA(t+1) are not winsorized.

The 5,356-observation modeling sample retains an FLD-Intensity distribution very similar to that of the full valid-text sample, so the supervised-sample restriction does not materially alter the text measure. ROA(t+1) is left un-winsorized, preserving extreme realized profitability outcomes for the main forecasting test.

4.4 Pearson correlations

Table 7. Selected Pearson correlations

Variable	Corr. with ROA(t+1)	Corr. with FLDIntensity
ROA(t)	0.5155	0.0107
ROE(t)	0.4403	-0.0029
RevenueGrowth(t)	0.1166	-0.0220
Leverage(t)	-0.2270	-0.0134
AssetTurnover(t)	0.1616	0.0282
OCF/Assets(t)	0.3386	0.0148
Size(t)	0.0082	-0.0397
FirmAge(t)	-0.1373	-0.0374
FLDIntensity(t)	0.0060	1.0000

Note: The complete pairwise matrix is reported in Appendix D.

Current ROA shows the strongest simple association with subsequent ROA, whereas FLDIntensity is essentially uncorrelated with ROA(t+1). This near-zero unconditional correlation does not rule out conditional or nonlinear incremental value, which is evaluated directly through the out-of-sample M1-M2 comparisons in Section 5.

5. Main Out-of-Sample Results

5.1 Absolute forecasting performance

Table 8. Main ROA prediction results: 2022–2023 final test set

Model	Specification	MAE	RMSE	Test R^2	OOS R^2
Benchmark	Persistence: ROA(t)	0.037284	0.068166	0.161937	0.196927
Benchmark	Historical mean	0.048758	0.076065	-0.043570	0.000000
Ridge	M1: Financial	0.034353	0.060888	0.331341	0.359258
Ridge	M2: Financial + FLD	0.034379	0.060914	0.330755	0.358697
XGBoost	M1: Financial	0.034889	0.063196	0.279672	0.309746
XGBoost	M2: Financial + FLD	0.034867	0.063206	0.279441	0.309526

The financial information set provides useful out-of-sample forecasts, as both Ridge and XGBoost outperform the two simple benchmarks. The key comparison, however, is between M1 and M2 within each algorithm. Adding FLD-Intensity does not produce a clear improvement: Ridge performance becomes marginally worse, while XGBoost changes only trivially.

ISSN 3106-5783 (Print) ISSN 3106-7042 (Online)

5.2 Incremental predictive value of FLDIntensity

Table 9. Main incremental predictive-value test

Model	ΔMAE	MAE change	95% CI	p	ΔRMSE	RMSE change	95% CI	p	ΔR²
Ridge	-0.000027	-0.0779%	[-0.000054, -0.000002]	0.028	-0.000027	-0.0438%	[-0.000048, -0.000004]	0.024	-0.000586
XGBoost	0.000022	0.0618%	[-0.000222, 0.000274]	0.865	-0.000010	-0.0160%	[-0.000480, 0.000496]	0.967	-0.000230

Note: ΔMAE = MAE(M1) − MAE(M2) and ΔRMSE = RMSE(M1) − RMSE(M2).

Positive values favor M2. CIs and p-values are based on 2,000 firm-clustered paired bootstrap replications. The incremental tests provide no evidence of a stable forecasting gain from FLDIntensity. For Ridge, the direction is negative: adding FLDIntensity slightly increases both MAE and RMSE. Although the paired bootstrap detects this deterioration statistically, the magnitude is economically negligible, at well below 0.1%. For XGBoost, the changes are close to zero and statistically indistinguishable from no difference. Taken together, M2 does not stably reduce out-of-sample error relative to M1, so H1 is not supported.

5.3 Model diagnostics: use of FLD versus incremental value

Model diagnostics do not alter this conclusion. FLDIntensity receives a very small standardized coefficient in Ridge and non-zero feature importance in XGBoost, showing that the models can use the variable. These quantities do not demonstrate incremental predictive value, which is determined by the out-of-sample M1-M2 error comparison above.

6. Robustness and Stability Analyses

6.1 Alternative profitability target: ROE(t+1)

ROE creates a more difficult target because a small or negative equity denominator can generate extreme values. In the raw diagnostic, training-set ROE(t+1) ranges from -65.6367 to 8.7150, which makes squared-loss tree forecasts unstable. For this robustness test, target winsorization is therefore added without using future data: the 1st and 99th percentile cutoffs are estimated from the 2015-2019 training target only (-0.823607 and 0.353994) and then applied unchanged to Train, Validation, and Test. The adjustment affects only about 2% of observations in each split—60 of 2,973 in Train, 24 of 1,164 in Validation, and 23 of 1,178 in Test.

Table 10. ROE(t+1) robustness after Train-only target winsorization

Model	Specification	MAE	RMSE	Test R²	OOS R²
Benchmark	Persistence: ROE(t)	0.078773	0.156626	0.198978	0.240405
Benchmark	Historical mean	0.098275	0.179711	-0.054537	0.000000
Ridge	M1: Financial	0.076157	0.143990	0.323011	0.358023
Ridge	M2: Financial + FLD	0.076225	0.144092	0.322053	0.357115
XGBoost	M1: Financial	0.075047	0.138967	0.369428	0.402039
XGBoost	M2: Financial + FLD	0.074774	0.138469	0.373936	0.406314

With the winsorized ROE target, Ridge M2 remains slightly worse than M1, while XGBoost M2 improves MAE and RMSE by less than 0.4%. The paired bootstrap intervals include zero, so the ROE exercise does not provide stable support for H1.

6.2 Rolling annual out-of-sample forecasts

Rolling tests evaluate one test year at a time from 2019 to 2023, with preprocessing and hyperparameters re-estimated within each historical window. This checks whether the main result depends on a single 2022-2023 test period.

Table 11. Year-by-year rolling incremental predictive value

Test year	Model	ΔMAE	p(MAE)	ΔRMSE	p(RMSE)	ΔR²
2019	Ridge	0.000008	0.590	-0.000002	0.804	-0.000043
2019	XGBoost	0.001447	0.120	0.001245	0.624	0.025557

Test year	Model	ΔMAE	p(MAE)	ΔRMSE	p(RMSE)	ΔR²
2020	Ridge	0.000010	0.080	-0.000001	0.924	-0.000013
2020	XGBoost	0.000544	0.332	0.000789	0.752	0.017694
2021	Ridge	0.000002	0.862	0.000010	0.346	0.000151
2021	XGBoost	-0.000172	0.598	-0.000798	0.170	-0.013975
2022	Ridge	-0.000028	0.166	-0.000010	0.638	-0.000230
2022	XGBoost	-0.000219	0.302	-0.000457	0.324	-0.011346
2023	Ridge	-0.000017	0.114	-0.000025	0.012	-0.000515
2023	XGBoost	0.000111	0.360	0.000584	0.076	0.012223

Note: Positive Δ error favors M2. No annual positive improvement has a 95% bootstrap interval entirely above zero.

Ridge remains close to zero in every test year. XGBoost changes sign across years, and no positive annual improvement is statistically significant at the 5% level.

Pooling the rolling predictions produces the same conclusion: neither Ridge nor XGBoost shows a statistically stable reduction in MAE or RMSE. The pattern is more consistent with small, period-specific nonlinear signals than with persistent incremental forecasting value.

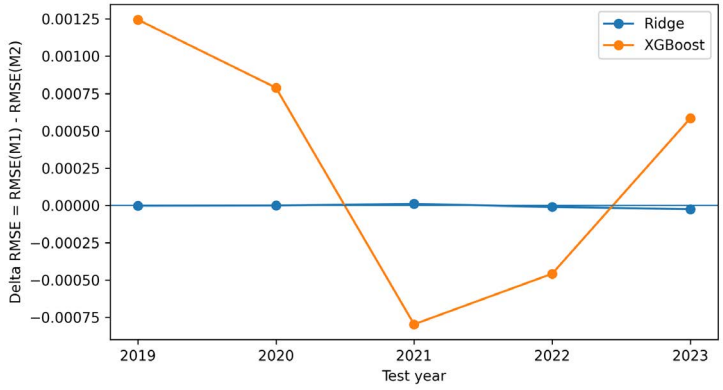


Figure 1. Year-by-year ΔRMSE = RMSE(M1) − RMSE(M2). Positive values favor the FLD-augmented model.

6.3 Excluding ST/*ST firm-years

Removing 124 ST/*ST firm-years leaves 5,232 observations. Preprocessing thresholds are re-estimated from the non-ST training sample before refitting the models.

Table 12. ROA robustness after excluding ST/*ST firm-years

Model	Specification	MAE	RMSE	Test R²	OOS R²
Benchmark	Persistence: ROA(t)	0.035715	0.065836	0.181194	0.220227
Benchmark	Historical mean	0.047945	0.074555	-0.050056	0.000000
Ridge	M1: Financial	0.033655	0.059917	0.321797	0.354127
Ridge	M2: Financial + FLD	0.033702	0.059972	0.320564	0.352953
XGBoost	M1: Financial	0.033486	0.060151	0.316487	0.349070
XGBoost	M2: Financial + FLD	0.033467	0.060308	0.312924	0.345677

The substantive result remains unchanged after excluding ST/*ST observations. Ridge M2 is slightly worse, and XGBoost shows only a trivial MAE improvement alongside a worse RMSE and R^2. The null incremental result is therefore not driven by distressed firm-years.

7. Discussion

7.1 Interpreting the null incremental result

The null incremental result is narrower than a claim that managerial narratives is uninformative. It shows that, conditional on the conventional financial predictors in M1, the share of MD&A sentences classified as forward-look-

ISSN 3106-5783 (Print) ISSN 3106-7042 (Online)

ing does not reliably reduce out-of-sample profitability forecast error. The same broad inference appears in the main ROA test and remains after using ROE, rolling the test year, and excluding ST/*ST observations. The evidence therefore identifies a boundary of the FLDIntensity measure rather than a general absence of information in forward-looking disclosure.

7.2 Why the result can coexist with prior positive textual evidence

This finding does not contradict prior studies reporting informative corporate text because those studies often examine different textual constructs or outcomes. Research using tone, semantic features, changes in MD&A wording, or richer narrative representations preserves information about what managers discuss and how they express it [5][9] -[12]. FLDIntensity, by contrast, retains only prevalence: how much of the MD&A is classified as forward-looking. In addition, an association between a textual measure and future firm outcomes is not equivalent to an improvement in out-of-sample prediction once an existing financial information set is held fixed. The present result is therefore compatible with positive findings for richer textual measures.

7.3 Why FLDIntensity may add little beyond accounting information

Two related explanations are plausible. First, FLDIntensity compresses an entire MD&A into a single ratio and therefore discards direction, numerical specificity, magnitude, topic, uncertainty, and credibility. Forward-looking sentences with very different economic implications receive the same weight in the measure, so potentially useful semantic information may be lost in aggregation. Second, some forward-looking language may be redundant with contemporaneous accounting signals. Current profitability, cash flow, growth, leverage, and turnover already summarize important aspects of the firm's operating position, and managers' forward-looking statements may respond to those same conditions. The small Ridge coefficient and the absence of lower XGBoost test error are consistent with these explanations, although they do not establish either mechanism causally.

A natural next step is therefore to enrich the text measure rather than simply increase model complexity. Future work could separate positive plans from downside risks, identify numerical targets and explicit horizons, distinguish topics, or use semantic representations that preserve more of the content discarded by a prevalence ratio.

8. Conclusion

This study examines whether the prevalence of management forward-looking disclosure in MD&A adds predictive value beyond conventional financial information for subsequent corporate profitability. In the Chinese A-share sample, the evidence does not show a stable positive incremental contribution from FLDIntensity. Adding the measure slightly worsens Ridge performance and leaves XGBoost performance essentially unchanged. The same broad conclusion remains when ROE is used as the profitability outcome, when test years are rolled forward, and when ST/*ST firm-years are excluded. H1 is therefore not supported.

This conclusion is specific to the prevalence measure used in the study. FLDIntensity records how much of the MD&A is forward-looking, but it does not preserve tone, specificity, credibility, topic, magnitude, or other semantic content. Accordingly, the findings do not imply that managerial narrative is uninformative. Rather, they identify a boundary of using a one-dimensional disclosure ratio to obtain incremental out-of-sample forecasting gains beyond accounting information.

References

[1] P. M. Healy and K. G. Palepu, "Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature," Journal of Accounting and Economics, vol. 31, nos. 1–3, pp. 405–440, 2001. doi: 10.1016/S0165-4101(01)00018-0.

[2] R. E. Verrecchia, "Essays on disclosure," Journal of Accounting and Economics, vol. 32, nos. 1–3, pp. 97–180, 2001. doi: 10.1016/S0165-4101(01)00025-8.

[3] F. Li, "Annual report readability, current earnings, and earnings persistence," Journal of Accounting and Economics, vol. 45, nos. 2–3, pp. 221–247, 2008. doi: 10.1016/j.jacceco.2008.02.003.

[4] P. C. Tetlock, M. Saar-Tsechansky, and S. Macskassy, "More than words: Quantifying language to measure firms' fundamentals," The Journal of Finance, vol. 63, no. 3, pp. 1437–1467, 2008. doi: 10.1111/j.1540-6261.2008.01362.x.

[5] F. Li, "The information content of forward-looking statements in corporate filings—A naïve Bayesian machine learning approach," Journal of Accounting Research, vol. 48, no. 5, pp. 1049–1102, 2010. doi: 10.1111/j.1475-679X.2010.00382.x.

[6] T. Loughran and B. McDonald, "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks," The Journal of Finance, vol. 66, no. 1, pp. 35–65, 2011. doi: 10.1111/j.1540-6261.2010.01625.x.

[7] S. V. Brown and J. W. Tucker, "Large-sample evidence on firms' year-over-year MD&A modifications," Journal of Accounting Research, vol. 49, no. 2, pp. 309–346, 2011. doi: 10.1111/j.1475-679X.2010.00396.x.

[8] V. Muslu, S. Radhakrishnan, K. R. Subramanyam, and D. Lim, "Forward-looking MD&A disclosures and the information environment," Management Science, vol. 61, no. 5, pp. 931–948, 2015. doi: 10.1287/mnsc.2014.1921.

[9] K. Bochkay and C. B. Levine, "Using MD&A to improve earnings forecasts," Journal of Accounting, Auditing & Finance, vol. 34, no. 3, pp. 458–482, 2019. doi: 10.1177/0148558X17722919.

[10] D. X. Wu, X. Yao, and J. L. Guo, "Is textual tone informative or inflated for firm's future value? Evidence

from Chinese listed firms," Economic Modelling, vol. 94, pp. 513–525, 2021. doi: 10.1016/j.econmod.2020.02.027.

[11] M. Z. Abedin, M. A. Goldstein, Q. Huang, and H. Zeng, "Forward-looking disclosure effects on stock liquidity in China: Evidence from MD&A text analysis," International Review of Financial Analysis, vol. 95, Part B, art. 103484, 2024. doi: 10.1016/j.irfa.2024.103484.

[12] S. Li, D. Wang, and L. Jiang, "Can the MD&A tone of listed firms' annual reports predict their future performance? Empirical evidence based on machine learning text analysis," Asia-Pacific Journal of Financial Studies, vol. 54, no. 1, pp. 54–81, 2025. doi: 10.1111/ajfs.12500.

[13] G. Shmueli, "To explain or to predict?" Statistical Science, vol. 25, no. 3, pp. 289–310, 2010. doi: 10.1214/10-STS330.

[14] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," Technometrics, vol. 12, no. 1, pp. 55–67, 1970. doi: 10.1080/00401706.1970.10488634.

[15] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

[16] B. Efron and R. J. Tibshirani, An Introduction to the Bootstrap. New York, NY, USA: Chapman & Hall, 1993.

[17] J. G. MacKinnon, M. Ø. Nielsen, and M. D. Webb, "Cluster-robust inference: A guide to empirical practice," Journal of Econometrics, vol. 232, no. 2, pp. 272–299, 2023. doi: 10.1016/j.jeconom.2022.04.001.

[18] China Securities Regulatory Commission (CSRC), Measures for the Administration of Information Disclosure by Listed Companies, CSRC Order No. 40, Jan. 30, 2007, revised by CSRC Order No. 182, Mar. 18, 2021.