

Can Digital-Transformation Disclosure Improve Following-Year ROA Estimation? A Small-Sample Machine-Learning Study of Chinese Listed Restaurant Firms

Honto Zhou*

Hangzhou HGW High School, Hangzhou, China

*Corresponding author: Honto Zhou.

Abstract

This small-sample study tests whether an annual-report disclosure signal improves following-year return-on-assets (ROA) estimation in Chinese listed restaurant firms. The panel contains three firms, 20 transitions, 11 chronological training observations, and nine holdout observations. A frozen 12-term Chinese dictionary measures digital-transformation disclosure density. We compare historical-mean and current-ROA persistence references with controls-only, controls-plus-disclosure, and controls-plus-year versions of fixed Ridge, random-forest, and gradient-boosting regressors. Disclosure lowers holdout ROA mean absolute error (MAE) relative to controls in all three model families: 0.1027 to 0.0615 for Ridge, 0.0754 to 0.0672 for random forest, and 0.0746 to 0.0633 for gradient boosting. Only Ridge is lower than the persistence MAE of 0.0626, and disclosure density is correlated with feature year in the training panel (Spearman 0.787). Results are descriptive, without a hypothesis test. Disclosure density is therefore an auxiliary feature for longer-panel validation, not evidence of stable independent, causal, or generalizable forecasting value.

Keywords

digital-transformation disclosure, annual reports, ROA estimation, restaurant firms, machine learning

1. Introduction

Annual reports combine accounting statements with narrative disclosure about strategy and operations. Financial-text research treats that narrative as a possible complement to ratios, while warning that dictionary choice, wording, and document processing affect measurement [1]. Digital-transformation research has developed broad conceptual accounts of organisational change [2, 3] and used annual-report language to construct firm-level indicators linked to financial allocation, liquidity, or performance [4, 5]. Most such evidence uses broad listed-company samples and association-oriented designs. The narrower question here is whether a fixed disclosure count reduces following-year ROA estimation error beyond prior-year financial controls in listed restaurant firms.

The industry restriction is important. Restaurant companies face demand shocks, substantial fixed operating costs, and different degrees of online sales or digital service adoption. Evidence from listed food-and-beverage companies has connected digital transformation with financial performance, but the setting, measurement, and country differ from the Chinese restaurant-only panel used here [6]. The present task therefore sits between disclosure measurement and simple financial prediction without treating either literature as direct proof of the study hypothesis.

Small samples make evaluation design more important than model complexity. A chronological holdout protects later rows from model fitting, and fixed configurations reduce choices made after inspecting results. Forecast-evaluation guidance emphasises genuine out-of-sample assessment, and time-series validation research cautions that temporal dependence changes what a validation result can support [7, 8]. We consequently use only Ridge regression, random forest, and gradient boosting, together with two simple reference forecasts. The core comparison is a feature ablation: each model uses accounting controls alone and then the same controls plus one frozen disclosure-density feature.

Observed holdout error, rather than algorithmic novelty, defines the study question. The strict CSRC H62 restaurant universe yields three firms and 20 transitions, with 11 observations in the earlier training period and nine in the later holdout. Disclosure lowers error relative to controls alone in the three fixed comparisons, but it does not consistently beat current-ROA persistence and is correlated with calendar time. Accordingly, the paper does not treat model families, seeds, firms, or years as independent draws for a population test. Its contribution is an auditable and reproducible forecasting framework with an explicit evidence boundary, not evidence that disclosure causes financial performance.

Figure 1 previews the complete information flow from public annual reports to the three model specifications, fixed regressors, reference forecasts, and chronological evaluation. This design makes the time split and the limited interpretation visible before any performance result is considered.

2. Related Work

Financial disclosure text can be represented by dictionaries, linguistic features, or learned text models. Accounting and finance research describes the role of unstructured disclosure alongside conventional ratios [1, 9]. Domain dictionaries are useful when their vocabulary, matching rule, and denominator are explicitly reported. More complex text classifiers may improve contextual detection, whereas a fixed literal dictionary is easier to audit in a small panel. Our density feature is therefore a measurement choice, not a claim that dictionary counting is the best text-analysis method.

Research on enterprise digital transformation spans several outcomes and empirical strategies. Studies of Chinese listed firms have examined financial performance and stock liquidity [4, 5], while sector-specific work has examined listed food-and-beverage firms in a different national setting [6]. The results and identification strategies are not interchangeable, and they do not justify a causal statement in the present setting. We borrow only the broad measurement idea that annual-report language can form a transparent digital-transformation indicator.

Sector-specific evidence remains limited. A study of Vietnamese listed food-and-beverage companies connects digital transformation and financial performance using financial reports and survey measurement [6]. Neither line evaluates a Chinese A-share restaurant-only, following-year profitability estimate with an untouched chronological holdout. Our distinction is narrow: we test descriptive incremental information within that precise sample.

Ridge regression supplies a regularised linear learner, while random forests and gradient boosting provide limited non-linearity [10–12]. These models are familiar reference learners rather than competing algorithmic contributions. Changing conditions and small evaluation samples can destabilise apparent gains irrespective of the model family [7, 8]. This motivates fixed configurations, simple reference forecasts, and descriptive reporting rather than model ranking.

3. Method

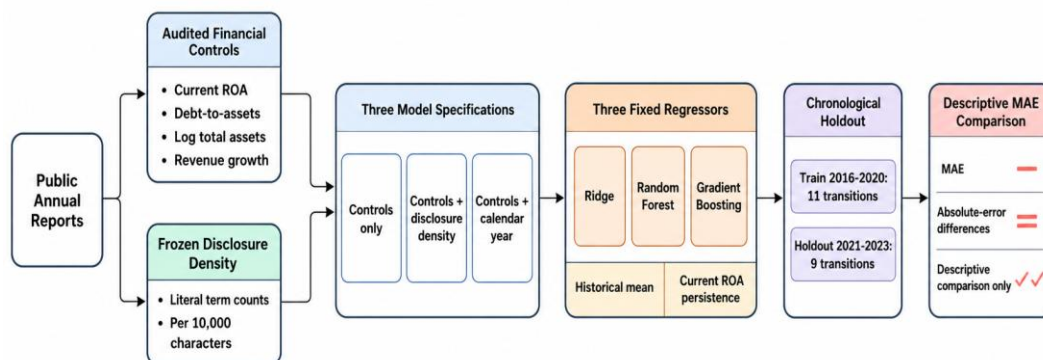
3.1 Study Design, Data, and Prediction Time

The sample comprises three Chinese A-share firms classified in CSRC H62: Xi'an Catering (000721), Quanjude (002186), and Tongqinglou (605108). Public annual reports were downloaded from CNINFO [13]. Each retained report has a source URL, release date, and SHA-256 hash; the reconstructed panel retains the report page used for each financial value. Audited consolidated statements supplied attributable net profit, total assets, attributable equity, total liabilities, and operating revenue. A row was excluded when a required value could not be traced, and no values were imputed.

Each row links features from fiscal year t to the same firm's ROA in fiscal year $t + 1$. The t annual report becomes public in March or April of $t + 1$, after the following fiscal year has begun. The 20 release dates range from 24 March 2017 to 25 April 2024, and the reproducibility record lists one release date for every row. The target is therefore following-year profitability estimation after the feature report is released, rather than a forecast available at the start of $t + 1$.

The pre-analysis CNINFO archive covers 2016–2024: reports from 2016–2023 provide feature years, and the three 2024 reports provide outcomes for 2023 rows. Thus, 2016 is a collection, not listing-date, boundary for Xi'an Catering or Quanjude; an earlier extension requires a new uniformly audited archive. The fixed chronological split assigns 2016–2020 to training (11 rows) and 2021–2023 to an untouched holdout (nine rows). Xi'an Catering and Quanjude contribute five training and three holdout rows each; Tongqinglou contributes one training and three holdout rows. Figure 1 shows the information flow and the time split.

Figure 1: Source-traceable feature-ablation workflow. Public annual reports provide financial controls and a frozen disclosure-density measure. Year- t features enter controls-only, controls-plus-disclosure, and calendar-year diagnostic versions of three fixed regressors. The chronological holdout contains feature years 2021–2023, and evaluation concerns following-year ROA after the year- t report is released



3.2 Frozen Disclosure Feature

Table 1 reports the complete frozen dictionary. The original Chinese strings are counted as literal, case-sensitive matches; no synonyms, stemming, weighting, or post-result term selection is used. Text extraction retained Chinese characters, Latin letters, and digits for the denominator. Counts are summed and divided by retained character count, then reported as mentions per 10,000 retained characters. All text emitted by the PDF layout extractor is included, including body prose, tables, and any extractable headers or footers. An extraction exception or fewer than 1,000 retained characters would fail the pipeline; no feature value would be imputed. No report triggered that rule.

Table 1: Frozen Chinese dictionary. Literal terms are shown exactly as matched; synonyms were not included.

Category	Literal Chinese terms
Digital framing	数字化; 数字转型; 信息化
Data and cloud	大数据; 云计算
Intelligent systems	人工智能; 智能化
Digital channels	线上; 电子商务; 数字营销; 移动支付; 智慧餐饮

All 283 literal-match contexts were retained with the study records for traceability. The literal count was not altered by a contextual classification step. The resulting density should therefore be read as a transparent disclosure-intensity signal, not as a verified measure of technology investment or implementation quality.

The financial controls are current ROA, debt-to-assets ratio, log total assets, and revenue growth. Current ROA is calculated as attributable net profit divided by closing total assets. This simplified definition is applied consistently across all firms and years; therefore, the reported ROA results use closing rather than average total assets. Debt-to-assets divides closing liabilities by closing assets; log total assets applies the natural logarithm; and revenue growth compares years t and $t - 1$. Following-year ROE is a secondary descriptive outcome using year-end attributable equity. Table 2 gives panel-level descriptive statistics.

Table 2: Panel descriptive statistics for 20 transitions

Variable	Mean	SD	Min.	Max.
Following-year ROA	-0.026	0.078	-0.184	0.076
Following-year ROE	-0.085	0.189	-0.423	0.136
Disclosure density	1.148	0.675	0.269	2.877
Debt-to-assets	0.395	0.168	0.153	0.724
Log total assets	21.279	0.349	20.772	22.105
Revenue growth	0.077	0.323	-0.500	0.993

3.3 Fixed Models and Descriptive Evaluation

Two simple reference forecasts are included: the training-set historical mean and persistence. Let R denote ROA; for persistence the next-period estimate, $\hat{R}_{t+1}$, equals the current value, R_t . Each learned model is fitted in three versions: controls only, controls plus disclosure density, and controls plus a numeric calendar-year index ($t - 2016$). The year index is used unscaled alongside the other raw inputs for the tree models; for Ridge it enters the same training-only standardisation pipeline as the other predictors. It is a diagnostic comparator, not a proposed predictor. Ridge provides a regularized linear benchmark, while shallow random forest and gradient boosting probe limited nonlinear relationships without extensive tuning under the small-sample setting. Ridge uses penalty 1. Random forest uses 100 trees, maximum depth 2, and minimum leaf size 2. Gradient boosting uses 30 estimators, learning rate 0.05, depth 1, and minimum leaf size 2. No hyperparameter search was performed. The two stochastic learners use five fixed seeds (7, 13, 29, 43, and 71); their per-row predictions are averaged before calculating a single MAE. All implementation uses scikit-learn [14], and Ridge preprocessing is fitted only on the relevant training rows.

The primary measure is mean absolute error (MAE),

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (1)$$

Here, n denotes the number of holdout observations, y_i the observed following-year ROA, and $\hat{y}_i$ its estimate. We report MAE changes, row-level differences, grouped results, and an expanding-window backtest; descriptive ROA RMSE, ROE metrics, and tree-model seed ranges are also reported. All results remain descriptive because fixed model families, seeds, firms, and years are not independent random draws from a population.

4. Experiments and Evidence

4.1 Execution and Reference Forecasts

All analyses used Python 3.10.12, NumPy 2.2.6, pandas 2.3.3, and scikit-learn 1.7.2. Scripts were rerun against the same archived inputs to check computational reproducibility. Source records, report hashes, panel construction, fixed configurations, row-level predictions, and diagnostic outputs are retained with the study materials. The literal dictionary and three main model specifications were fixed before model fitting. Persistence and calendar-year specifications serve as descriptive diagnostic baselines.

Table 3 reports the nine-row chronological holdout. The persistence and year-index rows are descriptive diagnostic baselines. Disclosure lowers MAE relative to financial controls in all three fixed model families. Yet persistence has MAE 0.0626: Ridge with disclosure is only slightly lower (0.0615), while random forest

and gradient boosting are higher. The result therefore supports disclosure density as an auxiliary feature to test, rather than a replacement for a simple current-ROA rule.

Table 3: Following-year ROA chronological holdout results for nine transitions. Lower MAE is better. Ensemble MAE is calculated after averaging the five fixed-seed predictions per observation. Changes are calculated from unrounded MAEs and are descriptive, not population tests

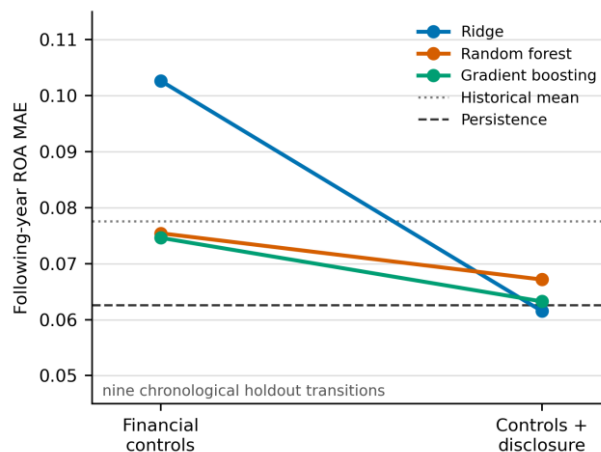
Approach	MAE	Change vs. controls	Change vs. persistence
Historical mean	0.0776	–	+0.0150
Persistence: current ROA	0.0626	–	0.0000
Ridge, controls	0.1027	0.0000	+0.0401
Ridge, plus disclosure	0.0615	-0.0412	-0.0010
Ridge, plus year index	0.0830	-0.0196	+0.0205
Random forest, controls	0.0754	0.0000	+0.0129
Random forest, plus disclosure	0.0672	-0.0083	+0.0046
Random forest, plus year index	0.0757	+0.0003	+0.0131
Gradient boosting, controls	0.0746	0.0000	+0.0121
Gradient boosting, plus disclosure	0.0633	-0.0114	+0.0007
Gradient boosting, plus year index	0.0746	0.0000	+0.0121

Secondary results agree descriptively: controls/disclosure ROA RMSE was 0.1209/0.0772 (Ridge), 0.0936/0.0804 (random forest), and 0.0939/0.0788 (gradient boosting). ROE MAE was 0.2807/0.1482, 0.2064/0.1883, and 0.2106/0.1838, respectively. Five-seed disclosure MAE ranged from 0.0653 to 0.0684 (random forest) and 0.0601 to 0.0657 (gradient boosting). These values describe the same nine rows, not separate replications.

4.2 Feature Ablation and Row-Level Changes

Figure 2 displays the three model-level MAEs without an interval. Across the 27 model–observation comparisons, disclosure had lower absolute error than controls in 16 cases. The median disclosure-minus-controls absolute-error change was -0.0005 for gradient boosting, +0.0010 for random forest, and -0.0239 for Ridge. At company level, disclosure lowers MAE in all three models for Xi'an Catering and Quanjude, but it raises MAE for Ridge and random forest in Tongqinglou. These patterns are useful diagnostics, not independent replications.

Figure 2: Primary following-year ROA MAE. Lines connect the controls-only and disclosure versions within each fixed model family; dashed horizontal lines show the two reference forecasts. Model families are labelled comparisons, not independent statistical observations

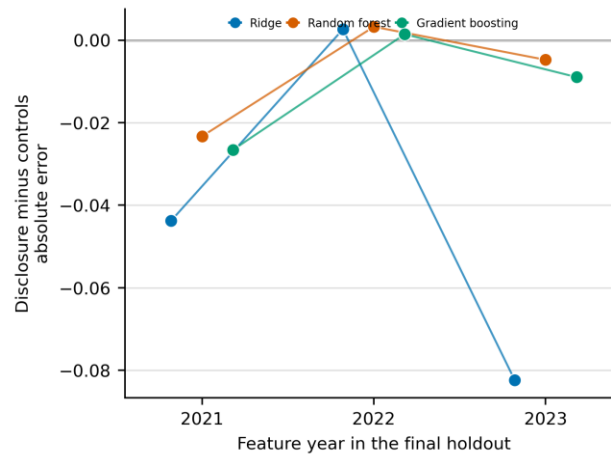


4.3 Calendar-Time Diagnostic and Temporal Sensitivity

Disclosure density is correlated with feature year: Pearson $r = 0.435$ and Spearman $\rho = 0.567$ in all 20 transitions; in the 11 training transitions, the corresponding values are 0.718 and 0.787. This makes calendar time a serious competing explanation. The year-index comparator does not match the disclosure version for Ridge or random forest, but neither comparison identifies independent disclosure information. Figure 3 reports

all three model changes by test feature year. Disclosure lowers error for every model in 2021 and 2023, but raises it for every model in 2022; the direction therefore varies across test years.

Figure 3: Year-specific disclosure-minus-controls absolute-error changes. Each point is the model-specific MAE aggregated across the three firms in that feature year; values below zero indicate lower error with disclosure. The three points per year are shown individually because they are not independent replicates



Across 14 expanding-window forecasts for 2019–2023, pooled controls/disclosure MAE was 0.1378/0.1077 (Ridge), 0.0693/0.0680 (random forest), and 0.0695/0.0661 (gradient boosting). Of 15 model–year comparisons, nine decreased and six increased, consistent with the small-sample feasibility interpretation.

5. Discussion

The frozen disclosure-density feature lowers final-holdout MAE relative to financial controls in all three fixed model comparisons. The baseline check is more mixed: only Ridge is slightly better than persistence, whereas the two tree learners are not. In addition, the feature rises with calendar time and the 2022 error changes move in the opposite direction to 2021 and 2023. Taken together, the results identify disclosure density as an auxiliary signal worth testing in longer panels, rather than a stable independent predictor within this small panel.

The dictionary provides an auditable measurement rule, not a measure of realised digital transformation. The retained literal contexts allow future inspection of each match, but the density cannot distinguish investment, implementation, staff adoption, or reporting style. This makes it a reasonable starting signal for a feasibility study. A more flexible text classifier might refine that distinction, but its added flexibility would be difficult to validate with 20 transitions [15].

The main limitations are interdependent observations, small training size, calendar-time confounding, the fixed 2016 archive boundary, and lack of an external restaurant-only replication panel. Report dates also mean that the outcome year has begun when the feature report is available. A stronger study would use a longer industry panel, hold out more firms and periods, and reassess the same frozen dictionary against persistence before considering more complex models.

6. Conclusion

This study combines financial ratios and annual-report digital-transformation disclosure for following-year ROA estimation in three Chinese listed restaurant firms. Disclosure lowered chronological-holdout MAE versus financial-only specifications in all three model families, while only Ridge slightly outperformed persistence. Test-year variation supports treating disclosure density as an auxiliary signal for validation in longer and broader restaurant-firm panels.

References

- [1] Loughran, T., and McDonald, B., “When is a liability not a liability? Textual analysis, dictionaries, and 10-ks,” *The Journal of Finance*, 66, (1), 35–65, 2011, <https://doi.org/10.1111/j.1540-6261.2010.01625.x>.
- [2] Vial, G., “Understanding digital transformation: A review and a research agenda,” *The Journal of Strategic Information Systems*, 28, (2), 118–144, 2019, <https://doi.org/10.1016/j.jsis.2019.01.003>.
- [3] Verhoef, P. C., Broekhuizen, T., Bart, Y., Bhattacharya, A., Dong, J. Q., Fabian, N., and Haenlein, M., “Digital transformation: A multidisciplinary reflection and research agenda,” *Journal of Business Research*, 122, 889–901, 2021, <https://doi.org/10.1016/j.jbusres.2019.09.022>.
- [4] Liu, H., Zhu, J., and Cheng, H., “Enterprise digital transformation’s impact on stock liquidity: A corporate governance perspective,” *PLOS ONE*, 19, (3), e0293818, 2024, <https://doi.org/10.1371/journal.pone.0293818>.
- [5] Han, G., and Tang, S., *Research on the impact of digital transformation on financial performance of chinese listed companies, Proceedings of the 2023 international conference on digital economy and management science*, 2024, <https://doi.org/10.4108/eai.27-10-2023.2341990>.
- [6] Dam, T. T. H., Dam, T. T., Dang, T. M. N., and Nguyen, T. T. H., “The impact of digital transformation on the financial performance of listed food and beverage companies in vietnam,” *International Journal of Innovative Research and Scientific Studies*, 8, (2), 2025, <https://doi.org/10.53894/ijirss.v8i2.5092>.
- [7] Tashman, L. J., “Out-of-sample tests of forecasting accuracy: An analysis and review,” *International Journal of Forecasting*, 16, (4), 437–450, 2000, [https://doi.org/10.1016/S0169-2070\(00\)00065-0](https://doi.org/10.1016/S0169-2070(00)00065-0).
- [8] Bergmeir, C., Hyndman, R. J., and Koo, B., “A note on the validity of cross-validation for evaluating autoregressive time series prediction,” *Computational Statistics and Data Analysis*, 120, 70–83, 2018, <https://doi.org/10.1016/j.csda.2017.11.003>.
- [9] Tetlock, P. C., “Giving content to investor sentiment: The role of media in the stock market,” *The Journal of Finance*, 62, (3), 1139–1168, 2007, <https://doi.org/10.1111/j.1540-6261.2007.01232.x>.
- [10] Hoerl, A. E., and Kennard, R. W., “Ridge regression: Biased estimation for nonorthogonal problems,” *Technometrics*, 12, (1), 55–67, 1970, <https://doi.org/10.1080/00401706.1970.10488634>.
- [11] Breiman, L., “Random forests,” *Machine Learning*, 45, (1), 5–32, 2001, <https://doi.org/10.1023/A:1010933404324>.
- [12] Friedman, J. H., “Greedy function approximation: A gradient boosting machine,” *The Annals of Statistics*, 29, (5), 1189–1232, 2001, <https://doi.org/10.1214/aos/1013203451>.
- [13] CNINFO, *China listed company information disclosure database*, Online public disclosure database, 2026, <https://www.cninfo.com.cn/>.
- [14] Pedregosa, F., Varoquaux, G., Gramfort, A., and others, “Scikit-learn: Machine learning in python,” *Journal of Machine Learning Research*, 12, 2825–2830, 2011.
- [15] Loughran, T., and McDonald, B., “Textual analysis in accounting and finance: A survey,” *Journal of Accounting Research*, 54, (4), 1187–1230, 2016, <https://doi.org/10.1111/1475-679X.12123>.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Open Access

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

