RAMuST: A Regime-Aware Multiscale and Mixed-Frequency Transformer for Industry-Level Corporate Income Tax Forecasting

Taewan You
Fiscal & Economic Policy Intelligence Research center
ETRI
Daejeon, Republic of Korea
twyou@etri.re.kr

YoungMin Kim
Fiscal & Economic Policy Intelligence Research center
ETRI
Daejeon, Republic of Korea
injesus@etri.re.kr

Yeonhee Lee
Fiscal & Economic Policy Intelligence Research center
ETRI
Daejeon, Republic of Korea
leeyh@etri.re.kr

Abstract

We present the **Regime-Aware Multiscale Transformer (RAMuST)**, a mixed-frequency Transformer-based architecture that couples a multi-scale monthly encoder with a yearly regime/shock head and a Transformer decoder for industry-level corporate income tax forecasting. In mixed-frequency data, monthly macro–economic and financial indicators are patchified at several temporal scales and fused through a soft scale gate integrated with economic cycle, trends, volatility, and seasonality, while yearly regime embeddings and a shock scalar summarize within-year dynamics and transient disturbances. To analyze the model, we designate the 2021–2024 window—covering the post COVID-19 period—as the validation interval and focus on the two most volatile Korean industries, Human health and Real estate activities, for corporate income tax forecasting. As a result of experiments, the model *consistently surpasses* recent strong time-series baselines by at least threefold and up to tenfold.

1 Introduction

Recent advances in time series prediction have produced state-of-the-art models for *short-/long-horizon forecasting*, *classification*, and *anomaly detection*, with Transformer architectures emerging as a unifying backbone for capturing long-range dependencies and complex temporal patterns, and deep neural networks-family baselines (e.g., inferring from [19, 15, 20, 11, 8, 4, 14, 17, 16, 1]). While many benchmark datasets in this line of work are adequately modeled by dependencies in the *first moment* (mean), financial and macroeconomic series are typically governed by the dynamics

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Workshop on Generative AI in Finance.

of both the *first* and the *second moments* (e.g., volatility/conditional variance and co-movement). Because repeated independent observations on the same time span are not available in time series, the theory and models must account for serial dependence, seasonality, heteroskedasticity, structural breaks, and nonlinearities within a single observed trajectory.

Moreover, directly applying off-the-shelf forecasting architectures to economic and financial workloads is challenging, chiefly due to mixed observation frequencies and release calendars. High-frequency market data (e.g., prices, FX) are noisy, and down-sampling (e.g., to daily) often discards intraday volatility information. In contrast, macroeconomic aggregates are published monthly/quarterly/annually with lags (e.g., monetary/interest-rate variables and national accounts), and index-type indicators such as CPI or equity indices follow heterogeneous publication cycles. Consequently, economic and financial pipelines seldom match the homogeneous sampling of physical-sensor benchmarks and require models that respect heterogeneous frequencies and publication constraints.

Classical mixed-frequency regression (e.g., MIDAS) has long been used to connect high-frequency regressors to low-frequency targets [6, 5]; more recent work explores deep and Transformer approaches to jointly exploit short- and long-horizon patterns and nonlinearities [11, 16, 9, 17]. Moreover, economic time series often feature *regimes* (expansion vs. recession) and *regime shifts* [7], as well as *shocks* (e.g., global financial crisis, COVID-19) that induce transient yet systemic effects [3]. For downstream forecasting, early detection of regime changes and shock intensities can be as valuable as the point predictions themselves.

We develop a regime-aware, mixed-frequency model for forecasting industry-level *corporate income tax payments*. The target is annual; inputs combine (i) a large panel of high-frequency economic and financial indicators and (ii) annual firm-level financial statements. We evaluate across 14 industries defined by the Korean standard industrial classification, with particular emphasis on 2021–2024—a period with large volatility and sizable forecast errors—focusing on two stress sectors (Human health and Real estate activities) where fluctuations were most pronounced. The model is evaluated under a strict walk-forward protocol.

We propose the **Regime-Aware Multiscale Transformer (RAMuST)**, which learns short-/mid-/long-horizon patterns from monthly indicators through multiscale patching and shared self-attention applied as the MultiResFormer [4], while detecting *regimes* and *shocks* at the yearly level and fusing them with annual covariates via a TFT-style variable selection module [8]. The architecture is designed to (i) align heterogeneous frequencies without leakage, (ii) isolate interpretable components (scale gate α, regimes R, shocks S), and (iii) operate in a rolling, walk-forward setting suitable for operational forecasting.

2 Methodology

The **Regime-Aware Multiscale Transformer (RAMuST)** is designed to utilize mixed-frequency economic and financial data to forecast the yearly target series—*industry-level corporate income tax payments*. Concretely, RAMuST comprises: (i) a monthly multiscale encoder for high-frequency indicators, (ii) a regime/shock–aware gating module that conditions annual covariates, and (iii) a yearly decoder that produces the final corporate income tax payment forecasting.

Figure 1 illustrates the overall architecture of RAMuST for corporate income forecasting used in mixed-frequency economic and financial data. The model consumes a large panel of monthly sampling indicators and a set of annual covariates from firm financial statements to predict annual corporate income tax payments, while exposing interpretable components for scale importance (α), regimes (R), and shocks (S).

Let batch size B, monthly horizon T, yearly horizon Y, input dimensions d_m, d_y, and model width d. Monthly and yearly inputs are

$$X^{(m)} \in \mathbb{R}^{B\times T\times d_m}, \qquad Z^{(y)} \in \mathbb{R}^{B\times Y\times d_y}. \tag{1}$$

If input widths differ from d, we align them by linear projection:

$$\tilde{X} = X^{(m)}W_x + \mathbf{1}b_x^\top \in \mathbb{R}^{B\times T\times d}, \qquad \tilde{Z} = Z^{(y)}W_z + \mathbf{1}b_z^\top \in \mathbb{R}^{B\times Y\times d}. \tag{2}$$

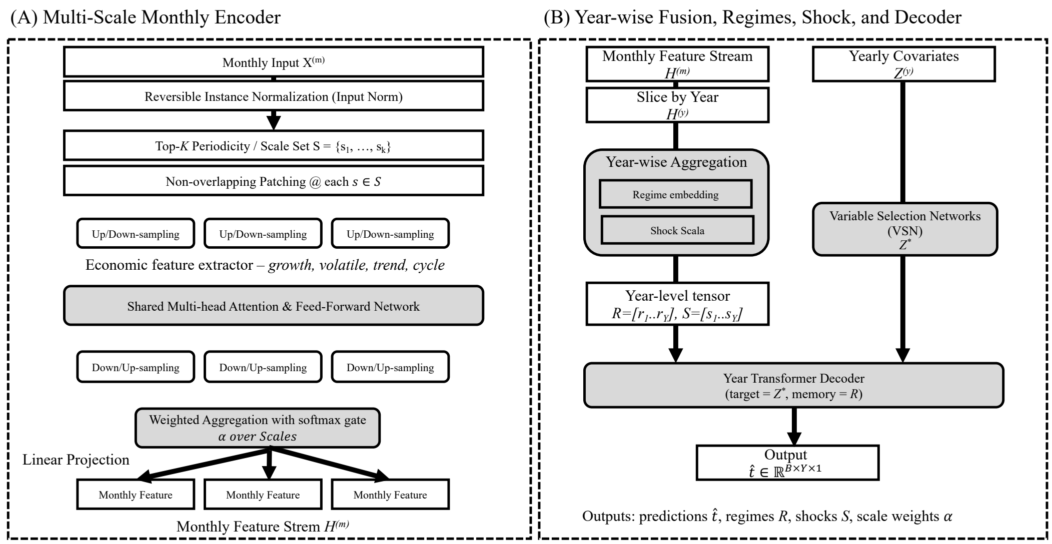


Figure 1: **RAMuST overall architecture.** (**A**) Multi-scale monthly encoder: monthly inputs are patchified at scales $S = \{s_1, \dots, s_K\}$, passed through shared Transformer layers, and fused by a softmax gate over scales to produce the monthly feature stream $H^{(m)}$ and scale weights α. (**B**) Year-wise fusion: monthly features are grouped by year, yielding regime embeddings $R = [r_1, \dots, r_Y]$ and shocks $S = [s_1, \dots, s_Y]$. A TFT-style VSN produces Z^* from annual covariates, and a year Transformer decoder (target $= Z^*$, memory $= R$, optional S) outputs the annual forecasts $\hat{\mathbf{t}}$.

2.1 Step 1: Multi-Scale Monthly Encoder

We analyze monthly data with multiple temporal windows to jointly capture multiple temporal patterns, extracting economically meaningful components (growth, volatility, trend, cycle) and fusing them across scales.

Per-scale processing. Let scales $S = \{s_1, \dots, s_K\}$ (e.g., $6, 12, 24$ months). For scale s, split the horizon into non-overlapping patches $\Pi_s = \{P_1^{(s)}, \dots, P_{n_s}^{(s)}\}$ with $n_s = \lceil T/s \rceil$ and $P_i^{(s)} = \{(i-1)s + 1, \dots, \min(is, T)\}$.

For each month t, apply four linear extractors (growth/volatility/trend/cycle) and fuse:

$$U_t = W_f \,[\, W_g \tilde{X}_t \parallel W_v \tilde{X}_t \parallel W_{\mathrm{tr}} \tilde{X}_t \parallel W_c \tilde{X}_t \,] \in \mathbb{R}^{B \times d}. \tag{3}$$

Then apply a residual self-attention + FFN block to each patch:

$$E_t = \mathrm{LN}\Big(U_t + \mathrm{MHA}(U_t, U_t, U_t)\Big), \qquad \hat{E}_t = \mathrm{LN}\Big(E_t + \mathrm{FFN}(E_t)\Big). \tag{4}$$

Pool within each patch $P_i^{(s)}$ by mean:

$$\bar{E}_i^{(s)} = \frac{1}{|P_i^{(s)}|} \sum_{t \in P_i^{(s)}} \hat{E}_t \in \mathbb{R}^{B \times d}, \qquad E^{(s)} = [\bar{E}_1^{(s)}, \dots, \bar{E}_{n_s}^{(s)}] \in \mathbb{R}^{B \times n_s \times d}. \tag{5}$$

Cross-scale alignment, attention, and fusion. Linearly interpolate each $E^{(s)}$ along time to the maximum length $T_S = \max_s n_s$: $\bar{E}^{(s)} = \mathcal{I}_{T_S}(E^{(s)}) \in \mathbb{R}^{B \times T_S \times d}$. Apply time self-attention per scale (implemented by reshaping to $(B \cdot K) \times T_S$):

$$\tilde{E}^{(s)} = \mathrm{MHA}\big(\bar{E}^{(s)}, \bar{E}^{(s)}, \bar{E}^{(s)}\big). \tag{6}$$

Table 1: Data used for RAMuST benchmarking and baseline inputs.

Block	Frequency	Scope / Size	Source(s)
Monthly macro–financial panel	Monthly	181 indicators	ECOS, KOSIS
Industry tax & financials	Yearly	2 industries; each 10 FS	TASIS, commercial provider

Compute scale importance via MLP and softmax after temporal averaging:

$$m^{(s)} = \frac{1}{T_S}\sum_{t=1}^{T_S} \tilde{E}_t^{(s)}, \quad u^{(s)} = W_2\,\sigma(W_1 m^{(s)}), \quad \alpha^{(s)} = \frac{\exp(u^{(s)})}{\sum_{s'}\exp(u^{(s')})} \in \mathbb{R}^{B\times 1}. \tag{7}$$

Fuse scales:

$$F = \sum_{s\in S} \alpha^{(s)} \odot \tilde{E}^{(s)} \in \mathbb{R}^{B\times T_S\times d}, \qquad H^{(m)} = \mathrm{LN}\big(\mathrm{MLP}(F)\big) \in \mathbb{R}^{B\times T_S\times d}. \tag{8}$$

2.2 Year-wise Aggregation (Regime & Shock)

We slice monthly features $H^{(m)} \in \mathbb{R}^{B\times T\times d}$ into yearly blocks of $M_y = 12$ months (zero–pad if needed), obtaining $H_y \in \mathbb{R}^{B\times M_y\times d}$ for $y = 1,\dots,Y$.

Regime embedding. Compute the yearly mean $\mu_y = \frac{1}{M_y}\sum_{t=1}^{M_y} H_{y,t}$. A light regime head summarizes slow dynamics while a cycle head summarizes periodic variation.

Shock scalar. Four complementary components diagnose transients within the year; neural, rolling stdev, CUSUM-like, and first/second-half drift. Aggregate with nonnegative weights $w_1,\dots,w_4$ (hyperparameters tuned for the target variable), apply gentle bounding, and separate sign/magnitude. Finally, stack $R = [r_1,\dots,r_Y] \in \mathbb{R}^{B\times Y\times d}$ and $S = [s_1,\dots,s_Y] \in \mathbb{R}^{B\times Y\times 1}$.

2.3 2.3. Year Decoder and Output

Given yearly covariates $\tilde{Z} \in \mathbb{R}^{B\times Y\times d}$, gated targets $Z^\star \in \mathbb{R}^{B\times Y\times d}$. A Transformer decoder conditions on regime memory R (keys/values) and receives $Z^\star$ as queries:

$$D = \mathrm{TransformerDecoder}(\mathrm{tgt} = Z^\star,\ \mathrm{mem} = R) \in \mathbb{R}^{B\times Y\times d}, \qquad \hat{\mathbf{t}} = \mathrm{Linear}(D) \in \mathbb{R}^{B\times Y\times 1}.$$

3 Experiments

3.1 Data.

We prepare two datasets to evaluate RAMuST fairly against recent Transformer-based forecasters referred in Table 1 . First, we collect 181 **monthly** macro economic and financial indicators from the national data portals *ECOS* (Bank of Korea) [2] and *KOSIS* (Statistics Korea) [12] in Table 1. Second, we build a **yearly** industry-level panel: (i) *corporate income tax payments* by industry according to Korea's standard industrial classification [10], and (ii) *firm financial statements* aggregated to the same 14 industries from a commercial provider, e.g., ValueSearch [13].

3.2 Training and Evaluation

Monthly inputs and yearly covariates are fed into the RAMuST's multiscale encoder, and the yearly decoder produce the final forecasts. We benchmark **RAMuST** against two representative Transformer-based time–series forecast models - TimeXer and TimeMixer [14, 17], and recently best performance time-series forecasting models - PatchTST and TimesNet [11, 16]. All baselines receive the same inputs and follow the same walk-forward protocol on the test window. Additionally, we report **MAE**, **RMSE**, and **MAPE** over the evaluation horizon, using the standard definitions.

Table 2 reports validation results over the four-year window (2021–2024) for the Human health and Real estate industries corporate income tax payments. Across all three metrics, RAMuST attains the

Table 2: Corporate income tax forecasting results. Best performance is **bolded**, second-best is underlined.

Industry	RAMuST (Ours)			TimeXer			TimeMixer			PatchTST			TimesNet		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
Human health Industry	**0.36[a]**	**0.42[a]**	**16.38**	1.65[a]	2.04[a]	97.47	1.73[a]	2.07[a]	97.98	3.02[a]	4.89[a]	86.06	1.79[a]	1.95[a]	88.65
Real estate Industry	**3.58[a]**	**3.97[a]**	**10.22**	10.0[a]	11.3[a]	26.06	10.0[a]	11.8[a]	26.49	7.75[a]	9.30[a]	19.85	13.0[a]	15.5[a]	32.76

best accuracy. For *Human health* industry, RAMuST reduces MAE by more than $8\times$ and RMSE, and, for *Real estate* industry, RAMuST improves upon the lowest-accuracy baseline (TimesNet) by over more than $3\times$ in both MAE and RMSE. Additionally, Figure 2 visualizes the predictions within the validation window, including the post-COVID-19 period.

4 Acknowledgments

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.2022-0-00857, Development of Financial and Economic Digital Twin Platform based on AI and Data, 50%) and supported by Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korea government(MSIT) (No.2025-25ZB1400, Development of AI and Big Data-Based Policy Intelligence Technology, 50%).

5 Conclusion

We introduced **RAMuST**, a regime-aware, mixed-frequency Transformer designed for operational annual forecasting for corporate income tax payments from heterogeneous monthly/annual data. The model jointly (i) learns short/mid/long patterns via multi-scale patching and shared attention, (ii) aggregates monthly signals into interpretable *regime* embeddings and a *shock* scalar per year, and (iii) fuses gated annual covariates in a decoder that conditions on historical regimes. Empirically, RAMuST achieves state-of-the-art accuracy on Korean industry–level corporate income tax forecasting, particularly in the volatile 2021–2024 window for Human health and Real estate industries. As a result of experiments, across the 2021–2024 validation window, **RAMuST** consistently outperforms recent strong time-series baselines—TimeXer, TimeMixer, PatchTST, and TimesNet—on industry-level corporate income tax forecasting. Averaged over sectors and years, we observe *error reductions* of approximately at least threefold to tenfold across MAE, RMSE, and MAPE, demonstrating clear gains in both point accuracy and robustness. Future work includes expanding model components to enable systematic post-hoc analysis and broader policy insights.

References

[1] Economic growth nowcasting through deep learning: A hybrid model of variational autoencoders and transformers. *ETRI Journal*, n/a(n/a).

[2] Bank of Korea. Ecos: Economic statistics system. `https://ecos.bok.or.kr`, 2025. Data portal and API for Korean economic statistics.

[3] Robert J. Barro, Jose F. Ursua, and Joanna Weng. The coronavirus and the great influenza pandemic: Lessons from the "spanish flu" for the coronavirus's potential effects on mortality and economic activity. *NBER Working Paper 26866*, 2020.

[4] Linfeng Du, Ji Xin, Alex Labach, Saba Zuberi, Maksims Volkovs, and Rahul G. Krishnan. Multiresformer: Transformer with adaptive multi-resolution modeling for general time series forecasting. *arXiv preprint arXiv:2311.18780*, 2023.

[5] Eric Ghysels. *MIDAS Regressions: Theory and Applications*. Princeton University Press, 2016.

[1]MAE and RMSE values are multiplied by 10^6; MAPE values are in percentage points (without the %).

[2a] All MAE and RMSE values are scaled by 10^6.

[6] Eric Ghysels, Pedro Santa-Clara, and Rossen Valkanov. Midas regressions: Further results and new directions. *Econometric Reviews*, 2007.

[7] James D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 1989.

[8] Bryan Lim, Sercan O. Arik, et al. Temporal fusion transformers for interpretable multi-horizon time series forecasting. In *NeurIPS*, 2021.

[9] Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. In *The Twelfth International Conference on Learning Representations*, 2024.

[10] National Tax Service (NTS) of Korea. Tasis: Tax statistics information system. `https://tasis.nts.go.kr`, 2025. Corporate income tax and other tax statistics portal.

[11] Yutong Nie et al. A time series is worth 64 words: Long-term forecasting with transformers. In *ICLR*, 2023.

[12] Statistics Korea (KOSTAT). Kosis: Korean statistical information service. `https://kosis.kr`, 2025. National statistical data portal of Korea.

[13] ValueSearch Co., Ltd. Valuesearch. `https://www.valuesearch.co.kr/`, 2025. Financial statements and industry aggregation data provider.

[14] Yuxuan Wang, Haixu Wu, Jiaxiang Dong, Guo Qin, Haoran Zhang, Yong Liu, Yunzhong Qiu, Jianmin Wang, and Mingsheng Long. Timexer: Empowering transformers for time series forecasting with exogenous variables. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

[15] Haixu Wu et al. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In *NeurIPS*, 2021.

[16] Haixu Wu et al. Timesnet: Temporal 2d-variation modeling for general time series analysis. In *NeurIPS*, 2023.

[17] Kai Zhang et al. Timemixer: Decomposable multiscale mixing for long-term series forecasting. *arXiv:2405.14616*, 2024.

[18] Y. Zhang, X. Li, J. Chen, M. Wang, and M. Long. Mtst: Multi-resolution time-series transformer. *arXiv preprint*, 2024. Preprint; add venue/identifier when available.

[19] Haoyi Zhou et al. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *AAAI*, 2021.

[20] Tian Zhou et al. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *ICML*, 2022.

A Related work

Classical statistical approaches to time–series forecasting (e.g., ARIMA/ETS families) are effective for short horizons but often struggle with long-horizon propagation and time–varying volatility; neural sequence models mitigate some of these issues but face a path-length vs. dependency-capture trade-off. A large body of recent work therefore adopts *Transformer*-based architectures for long-range temporal reasoning. However, naively applying Transformers to multivariate economic and financial data can ignore the distinct dependence structures across variables and frequencies, which is problematic for mixed-frequency, heteroskedastic series.

A first line of work builds *multi-scale* encoders that combine signals observed at different temporal resolutions via parallel branches, patching, and hierarchical attention. *MultiResFormer* [4], a multi-branch Transformer that processes time series in parallel with *patch lengths* chosen to capture short-, mid-, and long-term patterns without naive downsampling. We adopt this principle to structure mixed-frequency economic and financial inputs into scale-aware streams. Closely related, MTST [18] introduce the *Multi-resolution Time-Series Transformer*, which jointly trains branches with different patch sizes so that localized high-frequency patterns and long-term seasonal trends are learned simultaneously; positional encodings are used to emphasize scale-specific periodicities. Such designs enable a single model to encode daily (short-patch) and monthly/quarterly (long-patch) structures.

Patch-based tokenization has proven effective for capturing local temporal correlations. *PatchTST* [11] pioneer, which treats adjacent temporal segments as tokens; this improves the modeling of short-term dynamics and motivates the patch-style encoders used in our architecture. Our architecture is inspired by PatchTST, but we adopt a multiscale design to capture short-, mid-, and long-term patterns.

Macro-economic and financial series often exhibit regime shifts (expansion or recession) and episodic shocks (e.g., crises, pandemics). To incorporate such factors, TFT [8] design the *Temporal Fusion Transformer*, which integrates static and known future covariates via a gating and variable-selection mechanism; we similarly gate economy-derived features (regime and shock descriptors) as exogenous inputs. Timexer [14] further extend Transformers for exogenous-variable modeling: each exogenous variable is embedded as a sequence-level token and interacts with endogenous temporal tokens via cross-attention. Our decoder follows this spirit by letting annual covariate tokens attend to regime embeddings while optionally conditioning on estimated shock signals.

Our model sits at the intersection of these threads. Firstly, it uses multi-resolution, mixed-frequency data, patch-based encoders to discover scale-specific economic structure. Secondly exposes interpretable *regime* and *shock* signals. And finally fuses annual covariates through a gated, exogenous-aware Transformer decoder suitable for mixed-frequency economic forecasting.

B Dataset Details

In this paper, we use monthly macro-economic and financial data and annual corporate income tax payments data. The monthly macro-financial data is from the Korea Statistical Information Service (*KOSIS*) and (*ECOS*). The annual corporate income tax payments data is from the Korea Taxation Information Service (*TASIS*).

B.1 Monthly Data

Economic and financial data encompass both *high-frequency* market/activity series—such as credit-card transaction volumes, equity prices and trading volumes, foreign-exchange rates, and monetary aggregates—and a large body of official statistical series. At the national scale, "big data" resources range from the *national accounts* that compose GDP to detailed *government revenue and expenditure* statistics. For industry-level corporate income tax forecasting on a standardized industry taxonomy, we first construct a dataset of **181** indicators, as summarized in **Table 3**.

The panel comprises macroeconomic drivers including *national accounts*, *monetary aggregates*, *prices & inflation indices*, *external sector* measures, *composite and business-cycle indices*, and *public finance* (government expenditures and related items), spanning **2000–2025**. For the proposed architecture, all indicators are harmonized to a **monthly** sampling frequency to enable multiscale encoding and downstream regime/shock extraction.

B.2 Annual Data

Industry-level corporate income tax payments are publicly released for **14 industries** from **2007–2024** via the National Tax Service portal (*TASIS*). The published amount for year y reflects taxes assessed on firms' operations in year $y-1$, implying a **one–year lag** between economic activity and the recorded payment. To align with this timing, we shift the yearly covariate panel (financial statements and controls) by one year: prior–year financials are aligned to predict the current–year tax outflow.

For forecasting the current year's corporate income tax by industry, we therefore use prior–year firm financial statements together with tax–account variables and sector–specific macro indicators. The concrete set of tax–related fields and additional industry covariates used in our model is summarized in **Table 4**.

Table 3: Compact catalog of the 181 monthly macro/financial indicators used. For each category, we show representative items and indicate that additional series are included ("+ a few more"). Abbrev.: SA = seasonally adjusted; 12MA = 12-month moving average; KTB = Korea Treasury Bond; MSB = Monetary Stabilization Bond.

Category	**Representative items**	**Others**
National Accounts	Real GDP (SA); Real GDP, QoQ (SA)	+ a few
Monetary Aggregates	M1 (avg., SA); M2 (avg., SA); Liquidity aggregate Lf (avg., SA)	+ a few
Banking (Deposits/Loans)	Depository banks: total deposits (end bal.); loans (end bal.)	+ a few
Money Market	Unsecured overnight call rate (overall)	+ a few
Rates & Spreads	KTB 1Y/3Y/5Y yields; MSB 1Y/2Y; 3Y AA– corporate; KTB 3Y–1Y, 5Y–1Y; MSB 2Y–1Y; Corp–KTB 3Y spread	+ a few
Equities & Housing	Equity index (Korea); Housing sale price index; Housing lease-deposit price index	+ a few
Prices & Inflation	PPI (all items); CPI (all items); CPI ex. agri & petroleum; CPI ex. food & energy; Import Price Index	+ a few
Commodities	Crude oil (WTI/Dubai/Brent); Gold; Nickel; Zinc; Aluminum; Soybean; Corn; Wheat; Cotton	+ a few
External Sector & Trade	Current account (SA); Exports/Imports (customs-based); Trade balance; Export/Import volume indices	+ a few
Foreign Exchange	KRW per USD/JPY/EUR/GBP (averages)	+ a few
Sentiment (BSI/CSI)	BSI manufacturing/non-manufacturing (current, SA); BSI manufacturing/non-manufacturing (outlook, SA); Consumer Sentiment Index	+ a few
Investment	Facility investment index (SA)	+ a few
Composite Indices	Composite Leading Index (level & cyclical component); Composite Coincident Index (level & cyclical component)	+ a few
Production & Services	Industrial Production Index: total/mining/manufacturing (SA); excl. chemicals & pharma; electrical equipment; motor vehicles; Service Production (SA)	+ a few
Consumption	Retail sales index (total, SA)	+ a few
Manufacturing (by goods)	Manufacturing production: capital/intermediate/consumer goods (SA)	+ a few
Capacity Utilization	Capacity utilization: manufacturing; chemicals; electrical equipment; motor vehicles (SA)	+ a few
Construction	Orders: total/public/private; Put-in-place: total/public/private (SA)	+ a few
Shipments	Shipments index: total/mining/manufacturing (SA); excl. chemicals & pharma; electrical equipment; motor vehicles	+ a few
Shipments (by goods)	Shipments: capital/intermediate/consumer goods (SA)	+ a few
Domestic Shipments	Domestic shipments: capital/intermediate/consumer goods (SA)	+ a few
Inventories	Inventory index: total; manufacturing; capital/intermediate/consumer goods (SA)	+ a few
Electricity Usage	Electricity consumption: total/residential/services/manufacturing; 12MA counterparts	+ a few
Labor Market	Employed; Unemployed; Unemployment rate; Regular/temporary/daily workers (all SA)	+ a few
Public Finance	Consolidated fiscal: total/current/capital revenue; total/current/capital expenditure; net lending; balance	+ a few

C Full Experimental Results

This section provides additional experimental results and analysis on volatile industries—**Human health** and **Real estate**—for annual corporate income tax forecasting. Our model (**RAMuST**) consumes *yearly* firm financial covariates *and* an additional panel of **181 monthly** macro–financial indicators through a multiscale encoder. For a fair comparison, recent strong time–series baselines (TimeXer, TimeMixer, PatchTST, TimesNet) are trained on the *same yearly covariates*; RAMuST differs only by leveraging the monthly panel to infer regimes and shocks. *expanding* or *rolling-W* training window.

We report MAE, RMSE, and MAPE over the evaluation horizon,**2021–2024** window.

Table 4: Yearly covariates used in the industry panel. Categories group related accounting, credit, and macro variables; representative items are shown with examples of additional fields.

Category	**Representative item**	**Others**
Operating profit buckets	Reported operating profit (filing-based)	+ a few
Investment effects	Investment effect (impact of capital investment)	–
Labor market	Economically active population	–
Corporate tax accounts (prior year)	Corporate income tax *expense* (prior year)	+ a few
Balance sheet (prior year)	Total assets (prior year)	+ a few
Income statement dynamics (prior year)	Sales/revenue growth (prior year)	+ a few
Industry credit (prior year)	Loans by industry (prior year)	–
Macro controls (prior year)	GDP and GNI by economic activity (prior year)	–

$\mathcal{Y}_{\text{test}}$:

$$\text{MAE} = \frac{1}{N}\sum_{y\in\mathcal{Y}_{\text{test}}}\sum_{i\in\mathcal{I}}\big|\hat{t}_{i,y} - t_{i,y}\big|, \quad \text{RMSE} = \sqrt{\frac{1}{N}\sum_{y,i}\big(\hat{t}_{i,y} - t_{i,y}\big)^2},$$

$$\text{MAPE} = \frac{100}{N}\sum_{y\in\mathcal{Y}_{\text{test}}}\sum_{i\in\mathcal{I}}\left|\frac{\hat{t}_{i,y} - t_{i,y}}{\max(\tau,\, t_{i,y})}\right|,$$

where $\tau > 0$ is a small constant to avoid division by zero. We report scores aggregated over the **2021–2024** window and, for stress testing, provide year-specific comparisons for **2021** and **2024**.

C.1 RAMuST full results

We summarize how the main diagnostic variables in the RAMuST are *computed*, what they *mean*, and where they are *used* in the pipeline. Throughout, B is batch size, T monthly length, Y yearly length, and d model width.

Monthly multi–scale mixture and scale weights $\boldsymbol{\alpha}$. The monthly encoder produces a content–adaptive mixture of several temporal scales (e.g., 6/12/24 months). For each scale branch $s \in \mathcal{S}$, an encoded sequence $H^{(m)}_{(s)} \in \mathbb{R}^{B\times T\times d}$ is upsampled to the original monthly grid and combined by soft weights

$$\alpha^{(s)} = \text{softmax}\big(g(H^{(m)}_{(s)})_{s\in\mathcal{S}}\big) \in \mathbb{R}^{B\times|\mathcal{S}|}, \qquad \sum_s \alpha^{(s)} = 1.$$

In Figure 3, these appear `scale_weights`. For the default scales $\{3, 6, 12, 24\}$, we often report

$$\alpha_{\text{s3}},\ \alpha_{\text{s6}},\ \alpha_{\text{s12}},\ \alpha_{\text{s24}},$$

which quantify the relative importance of *half-year*, *annual*, and *bi-annual* periodicities learned from the data. Higher α_{s6} indicates that short-cycle signals dominated the monthly mixing at a given time/batch.

Shock signal and variables `shock_prev`, `shock_values`. It refines the shock into an *intensity* by blending several components:

$$\underbrace{s_y^{\text{neural}}}_{\text{NN head}} \oplus \underbrace{\text{volatility clustering}}_{\text{rolling std}} \oplus \underbrace{\text{structural breaks}}_{\text{CUSUM-like}} \oplus \underbrace{\text{momentum}}_{\text{first/second-half drift}},$$

followed by a bounded normalization and sign extraction. The resulting `shock_values` tensor has shape $B \times Y \times 1$. The auxiliary dictionary `shock_characteristics` stores each component for interpretability (volatility clustering, structural-break score, momentum, raw and normalized magnitudes).

Shock impact. Although the model returns `shock_values` and their components, downstream "impact" is intended for analysis/attribution (e.g., sensitivity of the decoder output to counterfactual changes in the shock) referred in Figure 4. In practice we examine correlations between `shock_values` and the residuals, or ablate the shock input in the legacy decoder where `shock_prev` explicitly enters the fusion block.

Interpretation. *(i)* α *weights* indicate which periodicities the encoder relied on (e.g., large α_{s12} suggests strong annual seasonality). *(ii)* r_y captures the slow-varying within-year regime state that the decoder attends to when predicting $\hat{t}_y$. *(iii)* s_y quantifies transient disturbances; large, positive shock intensity typically co-occurs with jumpy monthly dynamics or detected breaks.

C.2 Human health Industry Forecasting Results

As shown in Figure 5, RAMuST achieves the best accuracy among all models. Human health industry was one of the sectors most affected by COVID–19; incorporating regime signals and shock information enables RAMuST to reduce point errors and stabilize residuals relative to baselines. Across MAE/RMSE/MAPE, RAMuST consistently ranks first, with the smallest cumulative error and a tighter error distribution.

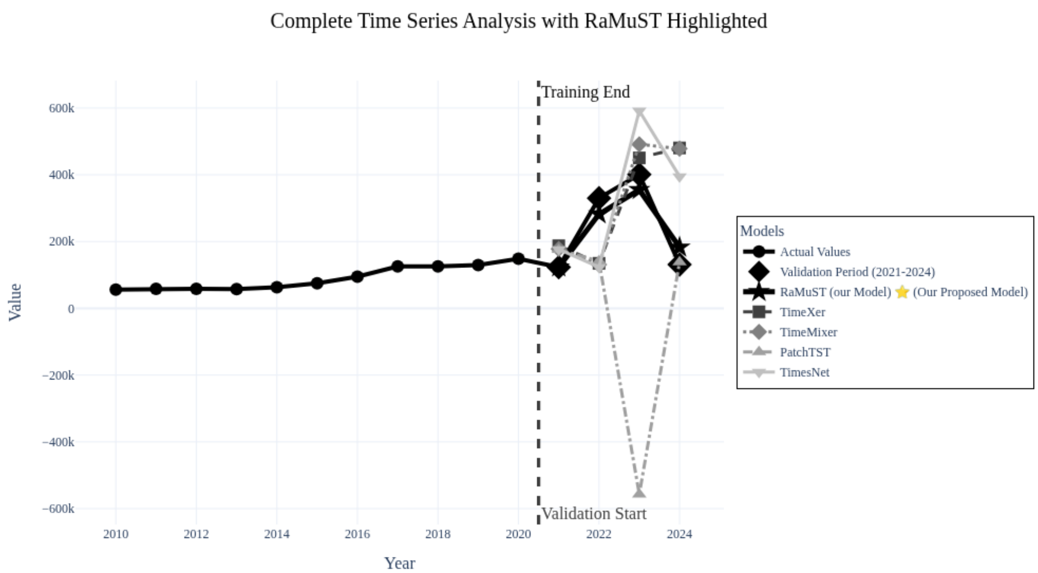
C.3 Real estate Industry Forecasting Results

Figure 6 indicates that recent baselines perform competitively on Real estate, yet RAMuST still achieves the lowest error. Both the cumulative error and the error–distribution panels confirm a clear advantage for RAMuST, yielding the top rank across the three metrics while maintaining stable residual dynamics.

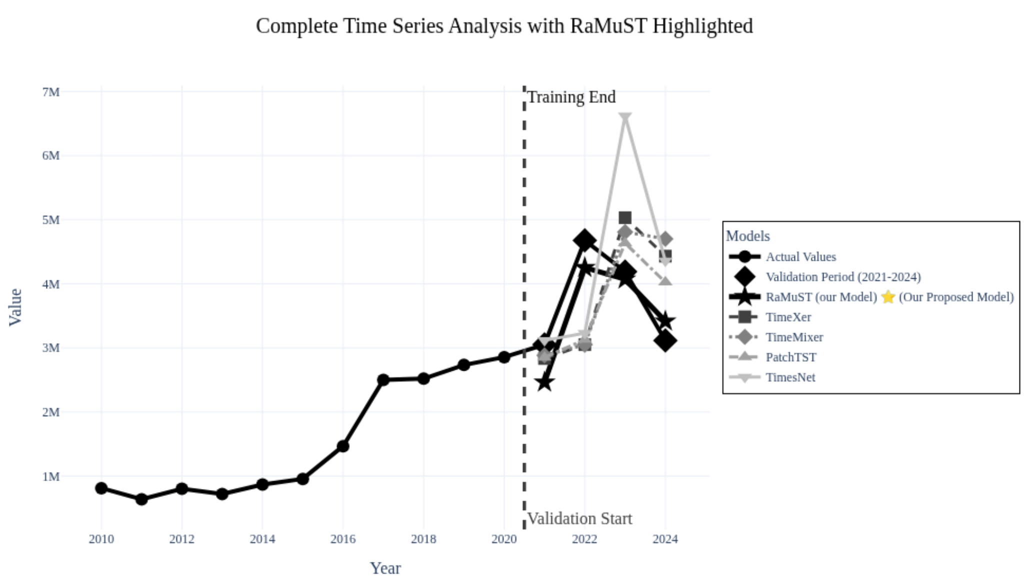
C.4 Future Works for Generative AI in Finance

Recent efforts in industry and academia aim to build *domain–specific LLMs* for financial and economic decision making, as well as *agentic* systems that chain commercial/open LLMs with tools. Our target task—industry–level corporate income tax forecasting—could in principle be tackled by such systems. However, our preliminary tests show that simply feeding large amounts of heterogeneous data (monthly macro indicators, firm financial statements, news) into general–purpose LLMs does *not* yield reliable forecasts. Two obstacles dominate: (i) **lack of grounded numerical modeling**—accurate regression and distributional inference are beyond pattern matching; and (ii) **opacity**—LLMs provide limited visibility into how inputs are used, impeding validation and risk management in finance.

Thus, we believe that RAMuST offers a structured, auditable interface between raw financial "big data" and generative models. Its multiscale encoder produces interpretable signals (*scale-gate weights*, *regimes*, *shocks*) and its decoder yields calibrated forecasts under a strict walk–forward protocol. In short, RAMuST provides the numerically grounded, interpretable signals toward generative Finance AI.

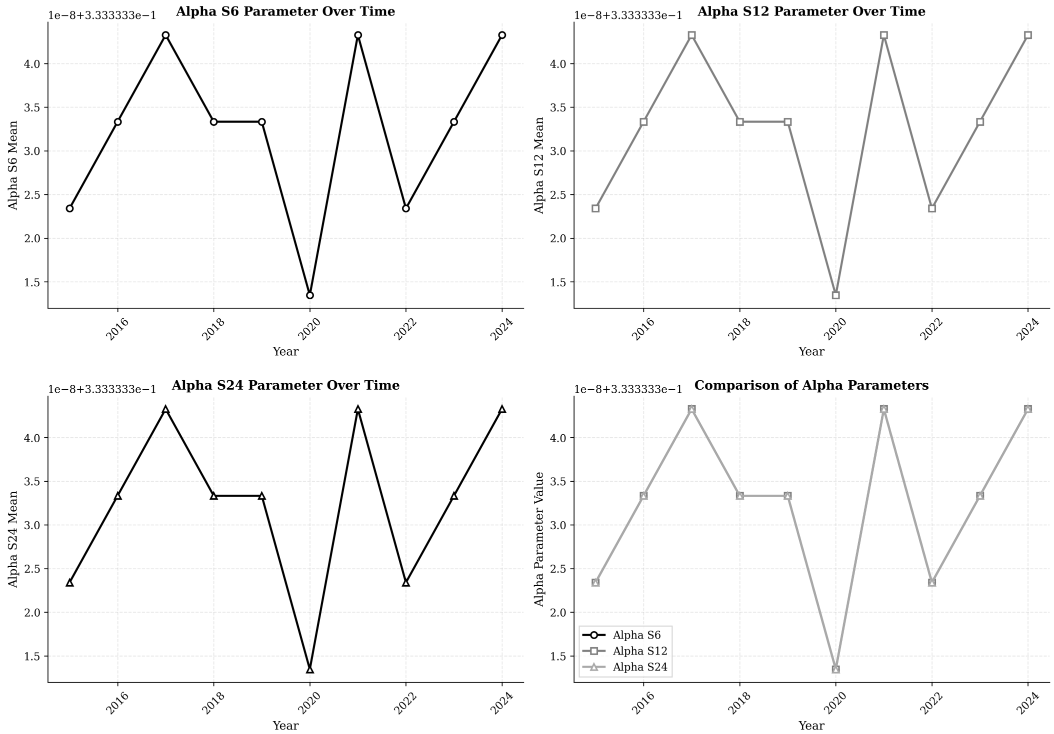


(a) Model-wise forecasts of annual corporate income tax payments for the **Human health** industry (validation window: 2021–2024). RAMuST (ours) is highlighted.

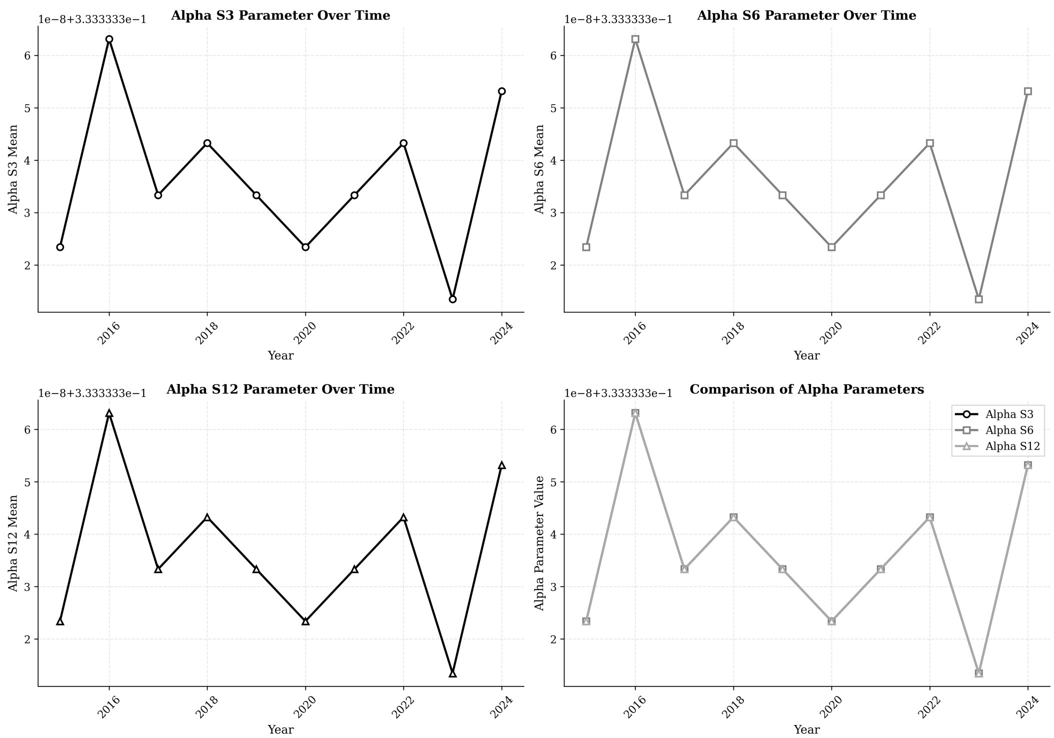


(b) Model-wise forecasts of annual corporate income tax payments for the **Real estate** industry (validation window: 2021–2024). RaMuST (ours) is highlighted.

Figure 2: RAMuST (ours) compared with baseline models on two industries in the validation window (2021–2024).

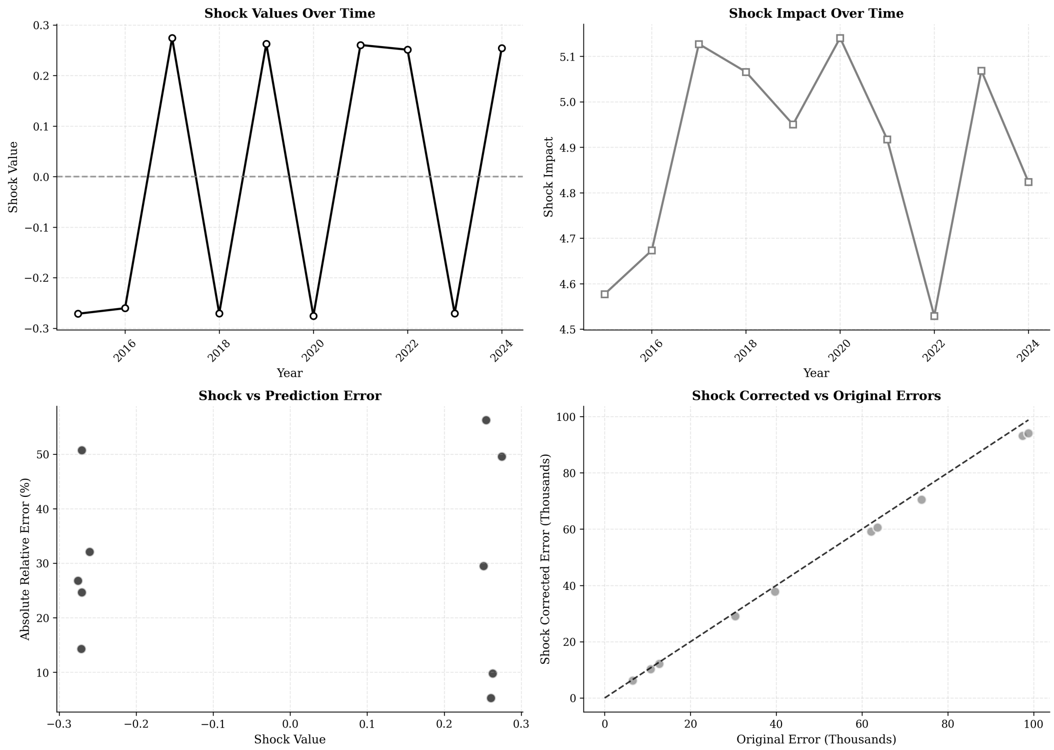


(a) Results of α parameters on Human health industry.

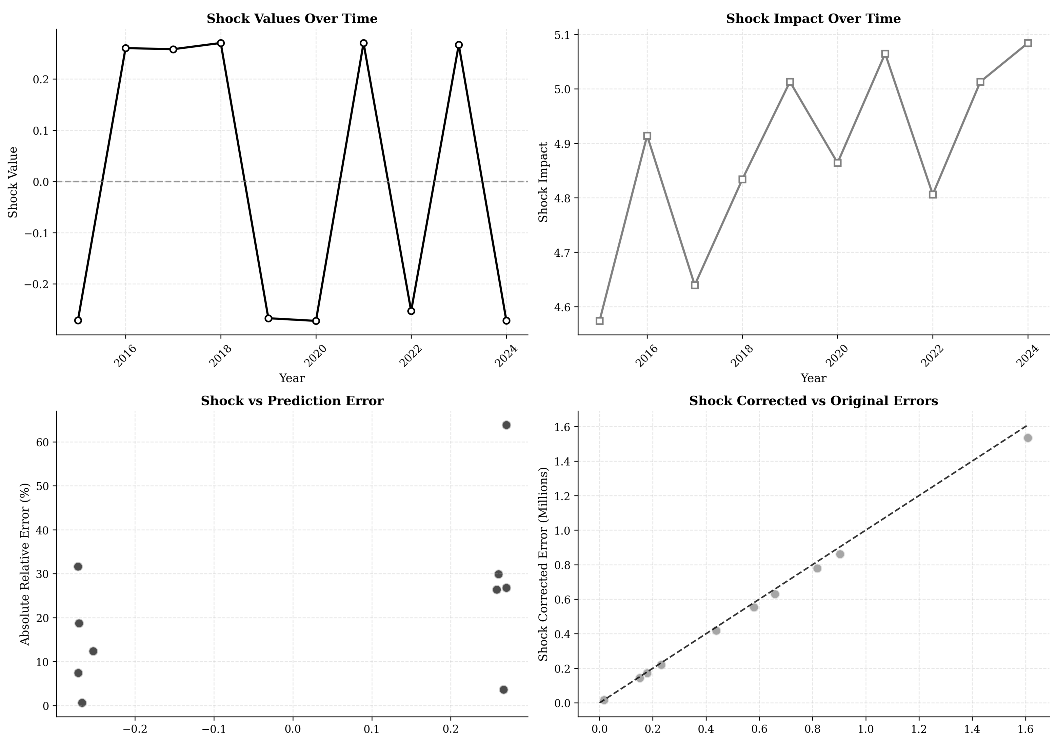


(b) Results of α parameters on Real estate industry.

Figure 3: Results of α parameters on Human health and Real estate industries.

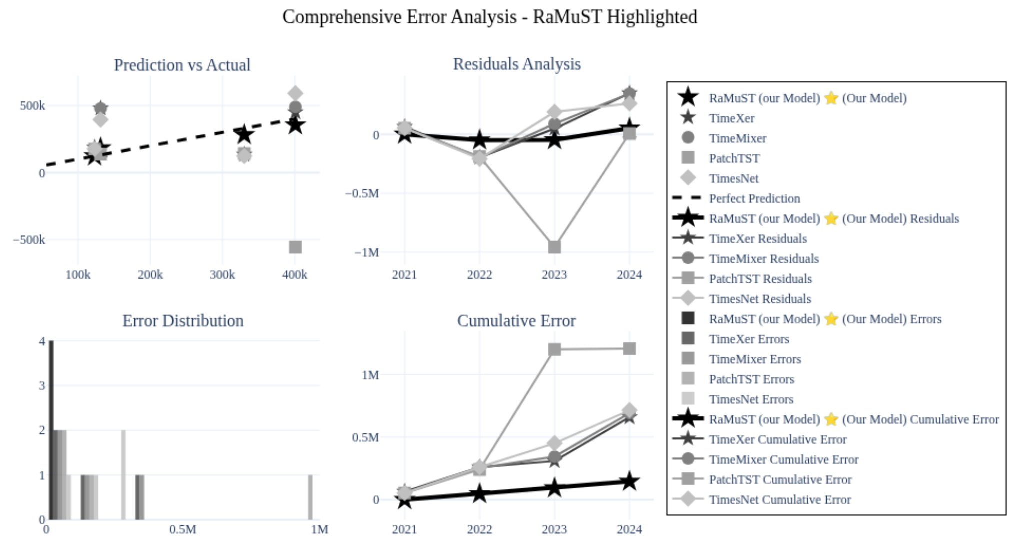


(a) results of shock impact on Human health industry: The temporal shock signal exhibits a sawtooth pattern, whereas its *impact* is concentrated in **2023–2024**. The shock impact is computed as the correlation between the shock values and the residuals.

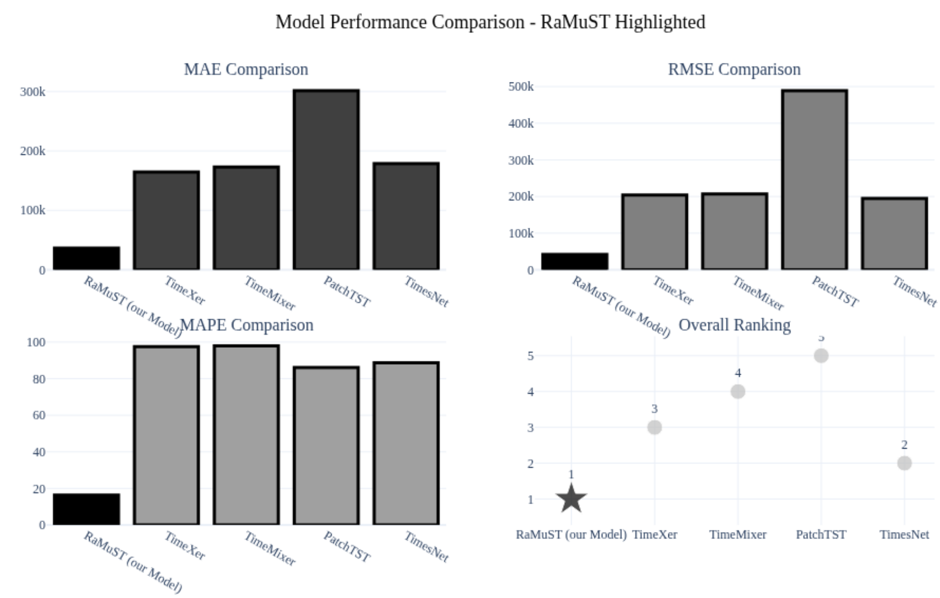


(b) results of shock impact on Real estate industry: Since 2020 the temporal shock follows a similar profile, but its *impact* intensifies during **2022–2024**. The shock impact is computed as the correlation between the shock values and the residuals.

Figure 4: results of shock impact analysis on Human health and Real estate industries.

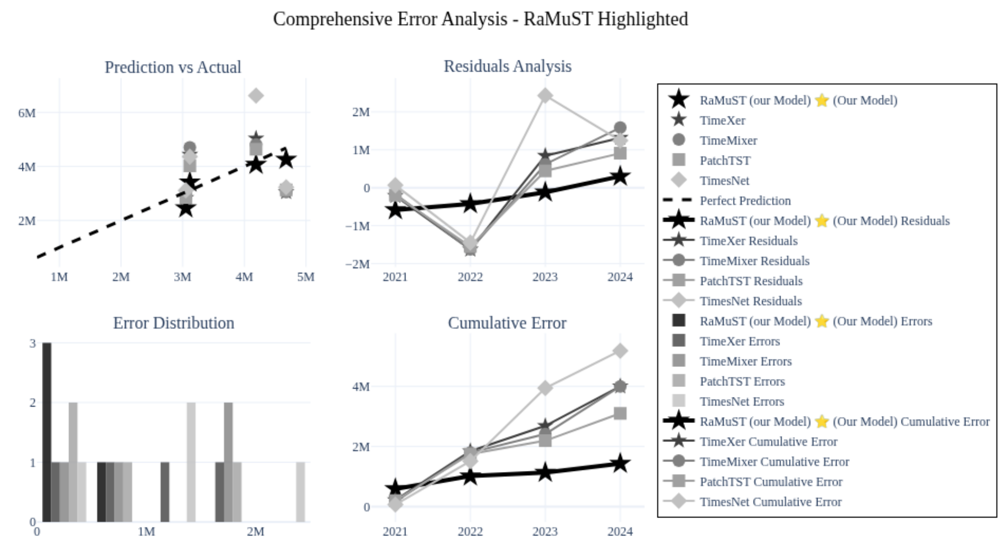


(a) Prediction vs. actual, residuals, error distribution, and cumulative error to compare model behaviors.

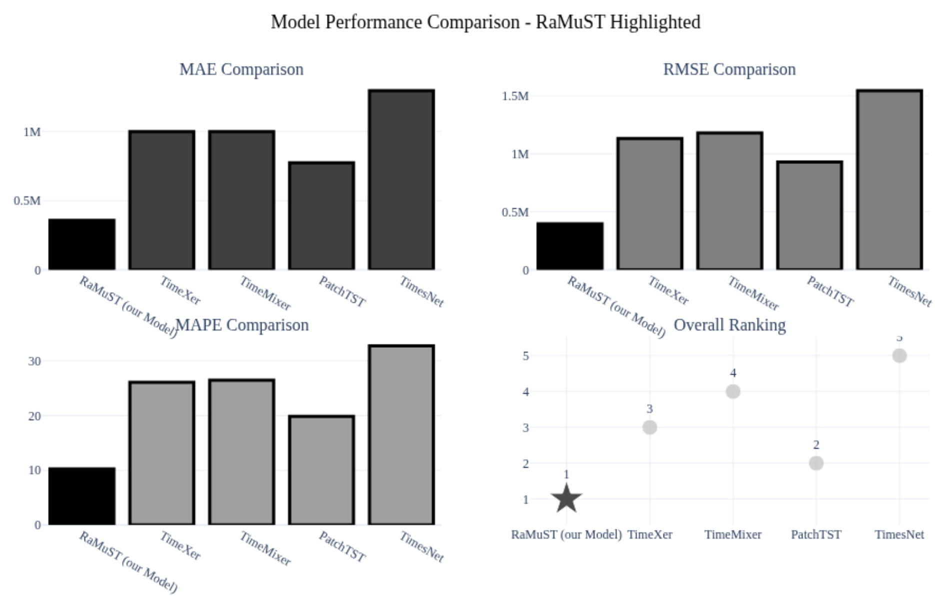


(b) Metric results (MAE, RMSE, MAPE) and overall model ranking.

Figure 5: Side-by-side diagnostic views for short-term corporate income tax forecasting on Human health industry.



(a) Prediction vs. actual, residuals, error distribution, and cumulative error to compare model behaviors.



(b) Metric results (MAE, RMSE, MAPE) and overall model ranking.

Figure 6: Side-by-side diagnostic views for short-term corporate income tax forecasting on Real estate industry.