UNCOVERING AND RESOLVING TEXT COLLAPSE IN MULTIMODAL TIME SERIES FORECASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Multimodal time-series forecasting, which pairs numerical sequences with domain-relevant texts, promises to inject world knowledge into forecasting pipelines. However, we uncover a critical failure mode in existing frameworks that we term **text collapse**: the text branch of the model converges to a content-independent transformation, contributing negligible discriminative signal regardless of the input content. We argue that text collapse arises from two related problems in multimodal time-series forecasting, namely *gradient imbalance* and *insufficient text supervision*. To address these issues, we propose **REST-TS** (**R**esidual-**E**xclusive **S**upervision for **T**ext in **T**ime **S**eries). The numerical backbone first produces an independent forecast, and the text branch is then supervised on the residual between the ground-truth future and this numerical forecast. The resulting auxiliary objective provides explicit supervision and a stronger gradient signal to the text branch, while encouraging it to model predictive information not captured by the numerical backbone. Evaluated across diverse real-world domains and backbone architectures, REST-TS achieves state-of-the-art performance and consistently demonstrates greater text-branch utilization than existing frameworks, providing strong empirical evidence that supervising the text branch on the residual encourages it to extract meaningful information from the input.

1 INTRODUCTION

Forecasting future values of a time series is rarely a purely numerical problem. A central banker predicting inflation, an epidemiologist modeling disease spread, or an energy analyst anticipating grid demand all rely heavily on textual context such as policy statements, outbreak bulletins, and weather advisories that numerical history alone cannot convey (Jiang et al., 2025; Wang et al., 2024b). Recent benchmarks formalize this intuition at scale by pairing historical time series with expert-curated textual content across diverse real-world domains (Liu et al., 2024a), and a growing family of multimodal forecasting frameworks reports consistent improvements over numerical-only baselines when text is incorporated (Liu et al., 2024a; Li et al., 2026; Nguyen et al., 2026), suggesting that text carries a predictive signal. However, these improvements do not necessarily imply that the models consistently exploit text information. Our analysis shows that, in several settings, their forecasts exhibit only limited sensitivity to the accompanying text.

Do multimodal forecasting models actually use the text content? We characterize meaningful text utilization through three linked properties (Section 3). First, different text contents should produce distinguishable representations; we measure this using *effective rank* (Roy & Vetterli, 2007), where a value close to one indicates that the representations collapse to nearly the same direction. Second, differences in text representations should encode information relevant to the forecasting target; we measure this using *Centered Kernel Alignment (CKA)* (Kornblith et al., 2019). Finally, the textual information should influence the forecast; we measure this using *prediction divergence*, which quantifies how much the forecast changes when the report content is replaced by an uninformative input. Figure 1 summarizes these diagnostics. Across four backbones, MM-TSFlib (MMTSF; (Liu et al., 2024a)) and TaTS (Li et al., 2026) exhibit effective ranks close to one, low CKA, and small prediction divergence. These results indicate that the text branch contains little content-dependent, forecasting-relevant information and has limited influence on the final prediction. We refer to this behavior as **text collapse**, and observe it consistently across all nine domains in Time-MMD (Liu et al., 2024a).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

We hypothesize that text collapse is driven primarily by *insufficient text supervision*, and is further exacerbated by *gradient imbalance*. The shared forecasting objective provides little supervision that specifically encourages the text branch to capture information beyond what is already explained by the numerical history. This problem is amplified because the numerical history is strongly predictive of the future, giving the numerical branch a stronger and more consistent learning signal. It can therefore reduce the forecasting loss largely through temporal autocorrelation alone, creating a *numerical shortcut* in which the model increasingly relies on the numerical branch while the text branch remains weakly optimized. Together, insufficient text-specific supervision and the resulting optimization imbalance encourage the model to underutilize the accompanying text.

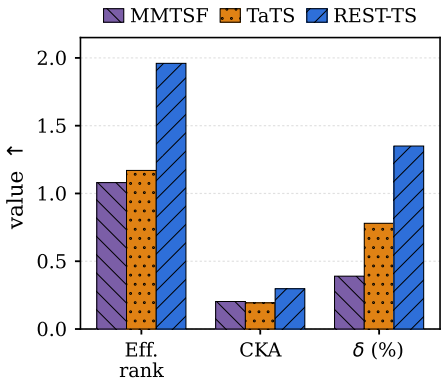


Figure 1: **Text collapse diagnostics** using effective rank, CKA, and prediction divergence, averaged over nine Time-MMD domains, four backbones, and all forecast horizons. Higher values indicate greater text utilization.

To resolve that problem, we propose **REST-TS** (**R**esidual-**E**xclusive **S**upervision for **T**ext in **T**ime **S**eries). Rather than competing over the same forecast target, REST-TS assigns text and numbers complementary roles. Specifically, the numerical backbone produces its own independent numerical forecast while the text branch is supervised to predict the trend and event components of the residual, which is the difference between the future target and the numerical forecast. A residual decomposition provides a heuristic for separating structured residual patterns from high-frequency variation, allowing the text branch to focus on the structured components. A slowly updated copy of the numerical backbone further ensures that the resulting targets remain stable and non-trivial throughout training. Because the auxiliary losses are optimized only through the text, improving them requires the text representation to encode information predictive of the corresponding residual components. As shown in Figure 1, REST-TS substantially improves all three text-collapse metrics, indicating that residual supervision encourages the text branch to learn more content-dependent representations. In summary, our contributions are:

- **Diagnosis of text collapse.** We identify and formally define text collapse, a failure mode in multimodal time-series forecasting, and demonstrate its prevalence across MMTSF and TaTS through effective rank, CKA, and prediction divergence under text removal.
- **REST-TS.** We propose a framework designed to mitigate text collapse by assigning the text branch supervised responsibility for structured components of the backbone residual, using an EMA target network and a residual decomposition module.
- **State-of-the-art results.** Experiments across diverse real-world domains and backbone architectures demonstrate that REST-TS consistently outperforms existing multimodal frameworks across backbones.

2 Related Work

2.1 Unimodal time-series Forecasting

Deep learning for time-series forecasting has progressed through several paradigm shifts. Early Transformer-based models such as Informer (Zhou et al., 2021), Autoformer (Wu et al., 2021), and FEDformer (Zhou et al., 2022) tailored self-attention to long-horizon dependencies. Zeng et al. (2023) then showed that a simple linear model with moving-average decomposition (DLinear) could match or outperform these Transformers, highlighting decomposition as a key predictive ingredient. Subsequent work rehabilitated Transformers through better tokenization and attention design, including PatchTST (Nie et al., 2023), iTransformer (Liu et al., 2024b), TimeMixer (Wang et al., 2024a), and TimeBridge (Liu et al., 2025). Although these methods perform strongly, they are unable to leverage external information from other modalities, such as text, which provides rich semantic context for time-series analysis.

2.2 Multimodal time-series Forecasting

Time-MMD (Liu et al., 2024a) and its companion library MMTSF provide the first systematic framework for text-augmented time-series forecasting at scale, combining numerical backbones with a frozen LLM text encoder through learned prediction-level fusion. TaTS (Li et al., 2026) maps paired text into temporal representations reflecting the series periodicity and injects them at the embedding level, while SpecTF (Nguyen et al., 2026) projects both modalities into the frequency domain and uses cross-attention to reweight frequency bands by textual relevance. GPT4MTS (Jia et al., 2024) injects BERT-encoded summaries as trainable soft prompts into a frozen GPT-2 temporal backbone, attaching text at the prompt level rather than through a fused encoder branch. In contrast, token-native models such as ChatTime (Wang et al., 2025) consume text directly as language-model tokens and therefore fall outside our scope. Despite these different fusion strategies, prior work does not examine whether the text branch learns content-dependent signals from paired text. We study this directly through text collapse and show that existing fusion designs lack an explicit mechanism encouraging the text encoder to preserve input-specific information. REST-TS addresses this by giving the text branch exclusive supervision on residual information not captured by the numerical branch.

2.3 Modality Collapse in Multimodal Learning

The failure of one modality to contribute meaningfully in a jointly trained multimodal model has been extensively studied in vision-language and audio-visual settings under terms including modality imbalance, greedy learning, and modality under-utilization. Early work documented that multimodal classifiers often underperform unimodal counterparts and proposed gradient rebalancing as a remedy (Wang et al., 2020; Peng et al., 2022). Ma et al. (2022); Huang et al. (2022) provide theoretical characterizations of multimodal learning, showing that late-fusion networks under SGD can suppress weaker modalities. More recent methods, including PMR (Fan et al., 2023), MMPareto (Wei & Hu, 2024), MLA (Zhang et al., 2024), and CGGM (Guo et al., 2024), address this problem through gradient manipulation, Pareto optimization, alternating unimodal adaptation, or classifier-guided modulation. These techniques are not inherently restricted to vision or audio and could potentially be adapted to multimodal time-series forecasting. However, they mainly regulate the optimization balance between modalities while retaining the shared prediction objective. In time-series forecasting, this does not directly address the fact that much of the future can already be predicted from the numerical history, leaving little supervision that specifically encourages the text branch to learn complementary information. REST-TS addresses this forecasting-specific issue by giving the text branch separate supervision on the residual information not captured by the numerical forecast.

3 Text Collapse in Multimodal Time-series Forecasting

We use the term *text collapse* when a model gives similar text representations to different texts and its forecasts depend little on the text. We first diagnose text collapse in existing multimodal time-series forecasting models in Section 3.1, and then discuss its likely causes in Section 3.2.

3.1 Diagnosing Text Collapse

We use two measures of the text representation: effective rank and CKA. First, effective rank (Roy & Vetterli, 2007) measures the representation rank, i.e., the variation within the representations. It is the exponential entropy of the normalized singular values of the representation matrix. A value of 1 means that all representations lie in one direction. Second, CKA measures the *target alignment*, i.e., how similar the geometry of the text representations is to that of the forecasting targets (Kornblith et al., 2019). Effective rank can be high even when the representations are not useful for forecasting, while CKA can be high for a low-rank representation.

We additionally measure prediction divergence under text removal, denoted by δ. Let $\mathbf{y}$ be the future, $\hat{\mathbf{y}}_{\text{real}}$ be the forecast from the true text and $\hat{\mathbf{y}}_{\text{no_info}}$ the forecast obtained when every text is replaced by a no-information sentence, so that the text branch still runs and only its input content changes. We define

$$\delta = \frac{\text{MSE}(\hat{\mathbf{y}}_{\text{real}},\, \hat{\mathbf{y}}_{\text{no_info}})}{\text{MSE}(\hat{\mathbf{y}}_{\text{real}},\, \mathbf{y})}, \tag{1}$$

reported as a percentage. These measures capture complementary aspects of text utilization, including representational diversity, alignment with forecasting targets, and sensitivity of the forecast to the text input. Definitions and estimation details are in Appendix 7.3.

Table 21 in Appendix 8.4 reports the three diagnostics for MMTSF (Liu et al., 2024a), TaTS (Li et al., 2026), and REST-TS. The effective ranks of MMTSF and TaTS range from 1.05 to 1.28, close to the minimum value of 1. Their text representations therefore vary mainly along one direction. Their CKA and forecast divergence are also low. REST-TS has a higher effective rank and CKA for every backbone in the table, although the absolute values remain modest. It also has the largest mean divergence. Per-domain results in Appendix 8.4 show that the low ranks occur across most domains rather than in only one dataset.

3.2 WHY TEXT COLLAPSES

We hypothesize that text collapse is driven primarily by insufficient text supervision, and is further exacerbated by gradient imbalance.

Gradient imbalance. Because the historical series $\mathbf{X}$ is strongly predictive of its future $\mathbf{Y}$, the numerical branch can rapidly reduce the forecasting loss using temporal autocorrelation alone. Its gradients are therefore stronger and more consistent than those from the text branch, whose information is often relevant only to less frequent or externally driven changes. Similar optimization imbalance has been observed in multimodal learning, where a dominant modality can suppress the learning of weaker modalities (Peng et al., 2022). During joint optimization, this imbalance encourages the model to rely increasingly on the numerical branch, leaving the text encoder weakly updated. We measure this gradient asymmetry directly in Section 5.3.

Insufficient text supervision. Although the text branch is trained through the shared forecasting objective, it receives little supervision that specifically encourages it to capture information beyond what is already available in $\mathbf{X}$. Related work in multimodal learning has similarly shown that a uniform objective shared across modalities may under-exploit modality-specific information, motivating modality-specific or isolated learning objectives (Lu et al., 2026). In multimodal time-series forecasting, both modalities are optimized to predict the same target $\mathbf{Y}$, even though much of $\mathbf{Y}$ is already predictable from the historical series. Once the numerical branch explains this component, the forecasting objective provides little incentive for the text branch to learn complementary information. Increasing the magnitude of the text-branch gradient can change how strongly the branch is optimized, but it does not change what the branch is asked to explain.

4 METHOD

Given a historical multivariate time series $\mathbf{X} \in \mathbb{R}^{T \times C}$ and an accompanying text S, our goal is to forecast the subsequent H time steps $\mathbf{Y} \in \mathbb{R}^{H \times C}$, where T and C denote the input length and number of variables, respectively. REST-TS addresses *insufficient text supervision* and *gradient imbalance* by training the text branch on residuals from a time-series-only forecaster. This focuses the text branch on complementary information while providing it with an exclusive learning signal.

As illustrated in Figure 2, the time-series branch produces the numerical representation for the joint forecast and maintains an EMA target network to obtain a time-series-only forecast (Section 4.1). The difference between this forecast and the ground truth defines the residual used to supervise the text branch. The text branch encodes the text and contributes to the joint forecast, while also predicting selected residual components during training (Section 4.2). Section 4.3 describes how the residual is decomposed into supervision targets, and Section 4.4 presents the overall training objective. The residual supervision is used only during training.

4.1 TIME-SERIES BRANCH

Joint forecast. A trainable time-series backbone $\mathcal{F}_\theta$ encodes the input as $\mathbf{Z}_{\text{ts}} = \mathcal{F}_\theta(\mathbf{X}) \in \mathbb{R}^{D \times C}$; D and C are the dimension and the number of channels of the time series representation, respectively. The backbone can be a standard forecasting model, such as PatchTST, iTransformer (Nie et al., 2023; Liu et al., 2024b). We combine $\mathbf{Z}_{\text{ts}}$ with the projected text representation $\mathbf{Z}_S$. Our default fusion is element-wise addition. A trainable projection then produces the forecast:

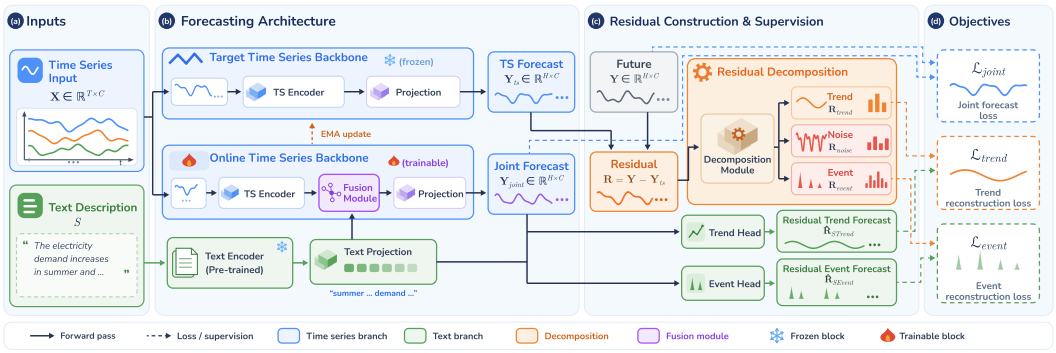


Figure 2: **Overview of REST-TS.** An online network combines the time series $\mathbf{X}$ with the text S and predicts $\mathbf{Y}_{\text{joint}}$. A target network, updated by an exponential moving average, predicts $\mathbf{Y}_{\text{ts}}$ from the time series alone. The residual $\mathbf{R} = \mathbf{Y} - \mathbf{Y}_{\text{ts}}$ is split into trend, noise, and event components. Two text heads learn to predict the trend and event components.

$$\mathbf{Y}_{\text{joint}} = \text{Proj}_{\phi}(\text{Fusion}(\mathbf{Z}_{\text{ts}}, \mathbf{Z}_S)) \in \mathbb{R}^{H \times C}. \tag{2}$$

Time-series-only forecast. We maintain a target copy of the backbone and projection, with parameters (θ', ϕ'), receiving no gradients.

$$\mathbf{Y}_{\text{ts}} = \text{Proj}_{\phi'}(\mathcal{F}_{\theta'}(\mathbf{X})) \in \mathbb{R}^{H \times C}. \tag{3}$$

After each training step, these parameters are updated by EMA: $(\theta', \phi') \leftarrow m(\theta', \phi') + (1-m)(\theta, \phi)$, providing a slowly evolving time-series-only forecast and a more stable target for the text branch.

4.2 Text Branch

Text representation. We tokenize the text S and obtain token embeddings from a pretrained language model. By default, we use the GPT-2 embedding layer with dimension $D_{\text{enc}} = 768$. We mean-pool the token embeddings to obtain $\mathbf{Z}_S^{\text{enc}} = \text{TextEnc}(S)$. The embedding weights remain frozen. A trainable linear layer projects and reshapes this vector to

$$\mathbf{Z}_S = \text{TextProj}(\mathbf{Z}_S^{\text{enc}}) \in \mathbb{R}^{D \times C}. \tag{4}$$

Residual prediction heads. A core observation motivating REST-TS is that texts encode two qualitatively different types of information: slow-moving *trend signals* (e.g., seasonal demand patterns, long-run growth) and abrupt *event signals* (e.g., policy shocks, extreme weather events). We instantiate two lightweight MLP heads operating on $\mathbf{Z}_S$ to predict

$$\hat{\mathbf{R}}_{S,\text{trend}} = \text{TrendHead}(\mathbf{Z}_S) \in \mathbb{R}^{H \times C}, \tag{5}$$

$$\hat{\mathbf{R}}_{S,\text{event}} = \text{EventHead}(\mathbf{Z}_S) \in \mathbb{R}^{H \times C}. \tag{6}$$

The two heads predict the trend and event residual components defined in Section 4.3. Their auxiliary losses provide additional supervision to the text branch, with gradients applied only to the text projection and prediction heads, while the target network remains unaffected.

4.3 Residual Decomposition

The residual $\mathbf{R} = \mathbf{Y} - \mathbf{Y}_{\text{ts}}$ contains all errors made by the target network. Some errors may be related to information in the text, while others may be unpredictable noise. We use a simple signal decomposition to reduce the amount of high-frequency noise in the text targets.

- **Trend** ($\mathbf{R}_{\text{trend}}$): smooth, slow-moving deviations driven by gradual contextual shifts described in S, such as seasonal demand patterns or long-run growth trends;
- **Noise** ($\mathbf{R}_{\text{noise}}$): high-frequency stochastic oscillations that arise from random fluctuations and carry no textual signal;
- **Event** ($\mathbf{R}_{\text{event}}$): sharp, non-periodic spikes caused by exogenous shocks explicitly described in S, such as sudden policy changes, market disruptions, or extreme weather events.

Supervising on $\mathbf{R}_{\text{noise}}$ may encourage the text branch to fit high-frequency variation that is less likely to be explained by the text, injecting gradient noise and weakening content-dependent learning. To extract these components, we decompose the residual. Let MAvg be a moving-average filter and $\mathrm{TopK}_{\mathrm{FFT}}(\cdot, K)$ reconstruct a signal from the K highest-frequency Fourier components (Galib et al., 2024). We compute

$$\mathbf{R}_{\text{trend}} = \mathrm{MAvg}(\mathbf{R}), \tag{7}$$

$$\mathbf{R}_{\text{noise}} = \mathrm{TopK}_{\mathrm{FFT}}(\mathbf{R} - \mathbf{R}_{\text{trend}}, K)\,, \tag{8}$$

$$\mathbf{R}_{\text{event}} = \mathbf{R} - \mathbf{R}_{\text{trend}} - \mathbf{R}_{\text{noise}}. \tag{9}$$

The moving average captures smooth variation. We treat the selected Fourier components as the noise term, although they can also contain periodic variation. The remainder is called the event component. These names describe the role of each component in our method. They do not imply that every extracted event matches an event mentioned in the text. The value of K controls how much oscillatory variation is removed from the event target. Visualizations of the decomposition across domains, together with the variance captured by each component, are provided in Appendix 9.5.

4.4 Training Objective

REST-TS is jointly trained with a composite objective consisting of three terms. The *joint forecast loss* penalizes errors in the primary multimodal prediction:

$$\mathcal{L}_{\text{joint}} = \mathrm{MSE}(\mathbf{Y}_{\text{joint}},\, \mathbf{Y}). \tag{10}$$

For residual supervision, we use only the trend and event components, since the noise component is treated as high-frequency variation that is less likely to be explained by the text. Accordingly, the *trend* and *event supervision losses* are defined as

$$\mathcal{L}_{\text{trend}} = \mathrm{MSE}\left(\hat{\mathbf{R}}_{S,\text{trend}}, \mathrm{sg}\left(\mathbf{R}_{\text{trend}}\right)\right), \quad \mathcal{L}_{\text{event}} = \mathrm{MSE}\left(\hat{\mathbf{R}}_{S,\text{event}}, \mathrm{sg}\left(\mathbf{R}_{\text{event}}\right)\right). \tag{11}$$

Here, $\mathrm{sg}(\cdot)$ denotes the stop-gradient operator (Chen & He, 2021), which treats the decomposition targets as constants during backpropagation. Therefore, gradients from $\mathcal{L}_{\text{trend}}$ and $\mathcal{L}_{\text{event}}$ flow only through the text branch, including the corresponding prediction heads. This prevents the numerical branch from absorbing the residual supervision signal. Because $\mathbf{Y}_{\text{ts}}$ is produced by the slowly updated target network, the residual target also remains informative as the online backbone improves. The total training objective is:

$$\mathcal{L} = \mathcal{L}_{\text{joint}} + \lambda_1\, \mathcal{L}_{\text{trend}} + \lambda_2\, \mathcal{L}_{\text{event}}, \tag{12}$$

where $\lambda_1, \lambda_2 \geq 0$ are scalar hyperparameters controlling the relative weight of each supervision signal. At inference time, only $\mathbf{Y}_{\text{joint}}$ is used as the final forecast; the residual heads and the decomposition module are discarded, incurring no additional inference overhead.

5 Experiments

5.1 Experimental Setup

Datasets. We use three benchmarks. **Time-MMD** pairs multivariate time series with curated reports from nine domains (Liu et al., 2024a). **FNF** and **FNSPID** pair financial series with news articles (Wang et al., 2024b; Dong et al., 2024). Dataset statistics are in Appendix 7.2.

Table 1: Mean squared error on Time-MMD, averaged over all forecast horizons. **Bold** marks the best result. *Promotion* is the percentage change from the strongest baseline TaTS to REST-TS. Results for all backbones and horizons are in Appendix 8.1.

	Method	Agri.	Clim.	Eco.	Ener.	Envi.	Heal.	Sec.	Soc.	Traf.
TimeBridge	unimodal	0.117	1.555	0.0182	0.305	0.350	1.968	130.27	1.291	0.174
	MMTSF	0.121	1.262	0.1014	0.286	0.320	1.504	140.32	1.314	0.213
	TaTS	0.120	1.198	0.0105	0.297	0.274	1.364	126.64	1.124	**0.160**
	REST-TS (ours)	**0.099**	**0.955**	**0.0083**	**0.261**	**0.270**	**1.252**	**108.82**	**0.941**	0.175
	Promotion	17.5%	20.3%	21.0%	12.1%	1.5%	8.2%	14.1%	16.3%	−9.4%
iTransformer	unimodal	0.122	1.183	0.0140	0.269	0.441	1.587	115.94	1.212	0.213
	MMTSF	0.119	1.049	0.1051	0.256	0.307	1.647	140.43	1.219	0.248
	TaTS	0.114	1.035	**0.0082**	0.276	0.271	1.290	119.71	1.039	0.198
	REST-TS (ours)	**0.113**	**1.022**	0.0084	**0.251**	**0.259**	**1.275**	**109.73**	**0.981**	**0.182**
	Promotion	0.9%	1.3%	−2.4%	9.1%	4.4%	1.2%	8.3%	5.6%	8.1%
PatchTST	unimodal	0.120	1.220	0.0172	0.269	0.552	1.652	112.85	1.097	0.188
	MMTSF	0.115	1.016	0.1143	0.263	0.345	1.422	133.15	1.208	0.216
	TaTS	**0.114**	1.019	**0.0087**	**0.258**	0.287	1.357	120.52	0.990	0.180
	REST-TS (ours)	0.115	**0.995**	0.0090	**0.258**	**0.272**	**1.272**	**109.21**	**0.974**	**0.178**
	Promotion	−0.9%	2.4%	−3.4%	0.0%	5.2%	6.3%	9.4%	1.6%	1.1%
DLinear	unimodal	0.223	1.190	0.0580	0.291	0.558	1.737	109.11	1.151	0.230
	MMTSF	0.221	1.040	0.1771	0.313	0.296	1.635	112.63	1.297	0.412
	TaTS	0.229	0.939	0.0241	0.324	0.296	1.466	112.53	1.302	0.196
	REST-TS (ours)	**0.138**	**0.932**	**0.0131**	**0.260**	**0.271**	**1.319**	**107.02**	**0.867**	**0.186**
	Promotion	39.7%	0.7%	45.4%	19.8%	8.4%	10.0%	4.9%	33.4%	5.1%

Table 2: Mean squared error of SpecTF and REST-TS on Time-MMD, averaged over all forecast horizons. **Bold** marks the better result in each domain. Results for each horizon are in Appendix 8.3.

Method	Agri.	Clim.	Eco.	Ener.	Envi.	Heal.	Sec.	Soc.	Traf.
SpecTF	0.103	0.939	**0.0085**	0.246	**0.260**	**1.276**	108.41	0.962	0.171
REST-TS (ours)	**0.100**	**0.932**	0.0092	**0.218**	0.260	1.295	**107.76**	**0.913**	**0.170**
Promotion	+2.3%	+0.7%	-8.2%	+11.1%	0.0%	-1.5%	+0.6%	+5.1%	+0.7%

Baselines and backbones. We compare REST-TS against two multimodal baselines, MMTSF (Liu et al., 2024a) and TaTS (Li et al., 2026), as well as a unimodal baseline for each backbone. Specifically, we evaluate all methods across eight time-series forecasting backbones: TimeBridge (Liu et al., 2025), iTransformer (Liu et al., 2024b), PatchTST (Nie et al., 2023), Crossformer (Zhang & Yan, 2023), DLinear (Zeng et al., 2023), FEDformer (Zhou et al., 2022), Autoformer (Wu et al., 2021), and Informer (Zhou et al., 2021). For each backbone, the unimodal model uses only the time-series input, while MMTSF and TaTS augment the same backbone with text through their respective multimodal frameworks. We additionally evaluate REST-TS in two complementary settings: integration with SpecTF (Nguyen et al., 2026), a standalone frequency-domain multimodal architecture, to test generalization beyond standard forecasting backbones; and comparison with OGM-GE (Peng et al., 2022), a modality-balancing method, to distinguish residual-exclusive supervision from generic multimodal optimization. For all methods, we tune the main optimization and model-specific hyperparameters on the validation set. The main metric is mean squared error (MSE), averaged over all forecast horizons. Training settings and hyperparameter-selection details are in Appendix 7.1.

5.2 Main Results

Table 1 reports the Time-MMD results for four representative backbones. Across all eight backbones, REST-TS improves on the strongest of the three baselines in 57 of 72 domain–backbone settings (Appendix 8.1). Each entry is averaged over all forecast horizons.

Comparison with the unimodal backbone. REST-TS outperforms the unimodal baseline in 35 of 36 backbone-domain configurations in Table 1, with the only exception being a marginal 0.001 MSE difference for TimeBridge on Traffic. Gains are largest in Economy and Climate, where REST-TS

Table 3: Mean squared error on FNF and FNSPID, averaged over all forecast horizons. **Bold** marks the best result and underlining marks the second-best result in each backbone column. Results for each horizon are in Appendix 8.2.

	Dataset	MMTSF		TaTS		**REST-TS (ours)**	
		PatchTST	iTransformer	PatchTST	iTransformer	PatchTST	iTransformer
FNF	Bitcoin	0.013	0.011	0.006	**0.006**	**0.005**	**0.006**
	Electricity	0.195	0.183	0.185	0.179	**0.177**	**0.170**
	Web Traffic	0.704	0.651	**0.675**	0.644	0.681	**0.633**
FNSPID	Delta (DAL)	0.028	0.027	**0.018**	**0.019**	**0.018**	**0.019**
	IBM	0.283	0.336	0.163	**0.165**	**0.161**	0.170
	JPMorgan (JPM)	0.185	0.205	0.104	0.106	**0.097**	**0.103**
	NVIDIA (NVDA)	0.140	0.099	0.061	0.063	**0.059**	**0.061**
	Pfizer (PFE)	0.105	0.101	0.047	**0.045**	**0.045**	**0.045**
	Tesla (TSLA)	1.414	1.686	0.750	**0.825**	**0.726**	0.866

with TimeBridge reduces MSE by 53.9% and 38.6% respectively. In domains with weaker textual signal such as Energy and Traffic, improvements are more modest but remain consistent across all backbones.

Comparison with multimodal methods. REST-TS has the lowest MSE among the three multimodal methods in 32 of 36 settings. Its largest improvement over TaTS is 45.4% for DLinear on Economy. Its largest degradation is 9.4% for TimeBridge on Traffic. These differences vary across domains, so the results do not show that residual supervision helps every backbone equally. Results for all backbones and horizons are in Appendix 8.1.

Financial benchmarks. Table 3 gives results on FNF and FNSPID. With PatchTST, REST-TS reduces MSE over MMTSF by 48% on average across the six FNSPID companies. On FNF, it matches or improves on TaTS in five of the six dataset and backbone settings. These results show that the method also works with news-based financial data, although its advantage over TaTS is not uniform. Results for each forecast horizon are in Appendix 8.2.

Adaptation to standalone architectures. To validate that REST-TS is backbone-agnostic beyond standard Transformer-based backbones, we integrate it with SpecTF (Nguyen et al., 2026), a standalone frequency-domain architecture that fuses text and time series via spectral cross-attention. Replacing SpecTF's original text fusion objective with our exclusive residual supervision improves over the original SpecTF in six of the nine Time-MMD domains (Table 2), with the largest gains in Energy (+11.1%) and SocialGood (+5.1%); per-horizon results are in Appendix 8.3.

Comparison with modality-balancing baselines. We further compare REST-TS against OGM-GE (Peng et al., 2022) on PatchTST and iTransformer. REST-TS outperforms OGM-GE on every domain (Table 23), by a median of 19.2% in MSE across the 18 backbone–domain settings. It demonstrates that residual-exclusive supervision provides advantages beyond modality-balancing approaches alone. Full results and text collapse analysis are in Appendix 8.5.

5.3 Analysis

Text gradient dominance. A direct signature of text collapse is that the text gradient becomes negligible relative to the numerical gradient during training. Figure 3a plots the ratio $\|\nabla_{\text{text}}\|/\|\nabla_{\text{num}}\|$ over training steps for the Environment domain with iTransformer. MMTSF (purple) decays to 10^{-2} within the first 500 steps and never recovers, confirming text collapse. REST-TS (blue) stabilizes near parity throughout, a direct consequence of the exclusive residual supervision. Extended analysis across all nine domains and both PatchTST and iTransformer is in Appendix 9.1.

Robustness to text-branch learning rate. Figure 3b evaluates effective rank, CKA (formal definitions in Section 3.1), and MSE across three text-branch learning rates on the Health domain with iTransformer. For MMTSF, a higher lr_2 reduces MSE, yet effective rank and CKA remain low, indicating that the forecasting improvement is sensitive to the optimization scale of the text branch. REST-TS maintains stable effective rank, CKA, and MSE across all three rates, because its text

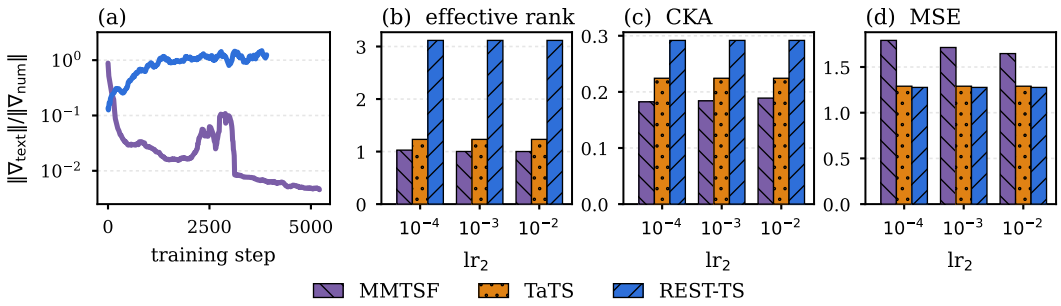


Figure 3: Gradient and learning-rate analysis. Panel (a) reports the text-to-numerical gradient-norm ratio during training. Panels (b) to (d) report effective rank, CKA, and MSE at three text branch learning rates. REST-TS is blue, MMTSF is purple, and TaTS is orange.

Table 4: Ablation study on three domains. The two components are residual decomposition and the target network. **Bold** marks the best result in each column. Variance is reported in Appendix 10.2.

Backbone	Components		Climate		Health		Environment		Avg.	
	Residual Decomp.	Target Network	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
PatchTST	✗	✗	1.003	0.791	1.281	0.749	0.278	0.377	0.854	0.639
	✓	✗	0.996	**0.788**	1.276	0.745	0.273	0.373	0.848	**0.635**
	✗	✓	0.998	0.789	1.277	0.748	0.275	0.374	0.850	0.637
	✓	✓	**0.995**	**0.788**	**1.272**	**0.744**	**0.272**	**0.372**	**0.846**	**0.635**
iTransformer	✗	✗	1.020	0.800	1.282	0.744	0.263	0.369	0.854	0.638
	✓	✗	1.008	0.799	1.276	**0.740**	**0.259**	0.369	0.852	0.635
	✗	✓	1.015	**0.796**	1.279	0.744	0.262	0.370	0.852	0.635
	✓	✓	**0.995**	0.797	**1.275**	**0.740**	**0.259**	**0.361**	**0.842**	**0.633**

branch is anchored to residual targets that are decoupled from the numerical loss scale. Extended results across all nine domains are in Appendix 9.2.

5.4 Ablation Study

Table 4 ablates Residual Decomposition and the Target Network across three domains with PatchTST and iTransformer. Each component improves over the baseline, while combining both performs best for both backbones, suggesting complementary roles: residual decomposition provides structured targets, while the Target Network keeps them informative during training. These results indicate that the Target Network is particularly important when combined with residual decomposition, reducing the average MSE of iTransformer from 0.852 to 0.842. We additionally report CKA and effective rank for the ablation variants in Appendix 10.1. These results suggest that the target network has a clearer effect on representation diversity than on forecasting accuracy.

6 Conclusion

We studied text collapse in fusion-based multimodal time-series forecasting, where the text branch learns weakly varying representations and has limited influence on predictions. Across Time-MMD, MMTSF and TaTS show low effective rank, low CKA, and small prediction divergence. We introduced REST-TS, which provides the text branch with residual-based supervision from a numerical forecaster. REST-TS achieves the lowest MSE in 32 of 36 main settings and improves text utilization across all reported backbones. These results show that multimodal forecasting should be evaluated not only by prediction accuracy, but also by whether the text modality contributes meaningful forecasting information. Future work could extend REST-TS beyond a single static text input by incorporating richer text encoders and multiple or streaming text sources.

AI USE STATEMENT

Generative AI tools were used to improve the writing, grammar, and clarity of this paper. The authors reviewed all generated suggestions and take full responsibility for the final content.

REPRODUCIBILITY STATEMENT

The code is available at the following anonymous repository: https://anonymous.4open.science/r/REST-TS-8E73. Appendix 7 provides the implementation details, experimental setup, dataset information, diagnostic metrics, and training algorithm. Appendix 8 reports the full results, while Appendices 9 and 10 provide additional analyses and robustness checks.

REFERENCES

Xinlei Chen and Kaiming He. Exploring simple siamese representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15750–15758, June 2021.

Zihan Dong, Xinyu Fan, and Zhiyuan Peng. Fnspid: A comprehensive financial news dataset in time series. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '24, pp. 4918–4927, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400704901. doi: 10.1145/3637528.3671629. URL https://doi.org/10.1145/3637528.3671629.

Yunfeng Fan, Wenchao Xu, Haozhao Wang, Junxiao Wang, and Song Guo. Pmr: Prototypical modal rebalance for multimodal learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20029–20038, June 2023.

Asadullah Hill Galib, Pang-Ning Tan, and Lifeng Luo. FIDE: Frequency-inflated conditional diffusion model for extreme-aware time series generation. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=5HQhYiGnYb.

Zirun Guo, Tao Jin, Jingyuan Chen, and Zhou Zhao. Classifier-guided gradient modulation for enhanced multimodal learning. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=oe5ZEqTOaz.

Yu Huang, Junyang Lin, Chang Zhou, Hongxia Yang, and Longbo Huang. Modality competition: What makes joint training of multi-modal network fail in deep learning? (Provably). In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato (eds.), *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 9226–9259. PMLR, 17–23 Jul 2022. URL https://proceedings.mlr.press/v162/huang22e.html.

Furong Jia, Kevin Wang, Yixiang Zheng, Defu Cao, and Yan Liu. Gpt4mts: Prompt-based large language model for multimodal time-series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 23343–23351, 2024.

Yushan Jiang, Kanghui Ning, Zijie Pan, Xuyang Shen, Jingchao Ni, Wenchao Yu, Anderson Schneider, Haifeng Chen, Yuriy Nevmyvaka, and Dongjin Song. Multi-modal time series analysis: A tutorial and survey. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2*, KDD '25, pp. 6043–6053, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400714542. doi: 10.1145/3711896.3736567. URL https://doi.org/10.1145/3711896.3736567.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015. URL https://arxiv.org/abs/1412.6980.

Simon Kornblith, Mohammad Norouzi, Honglak Lee, and Geoffrey Hinton. Similarity of neural network representations revisited. In Kamalika Chaudhuri and Ruslan Salakhutdinov (eds.),

Proceedings of the 36th International Conference on Machine Learning, volume 97 of *Proceedings of Machine Learning Research*, pp. 3519–3529. PMLR, 09–15 Jun 2019. URL https://proceedings.mlr.press/v97/kornblith19a.html.

Zihao Li, Xiao Lin, Zhining Liu, Jiaru Zou, Ziwei Wu, Lecheng Zheng, Dongqi Fu, Yada Zhu, Hendrik Hamann, Hanghang Tong, and Jingrui He. Language in the flow of time: Time-series-paired texts weaved into a unified temporal narrative. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=a1zBg9cBvt.

Haoxin Liu, Shangqing Xu, Zhiyuan Zhao, Lingkai Kong, Harshavardhan Kamarthi, Aditya B. Sasanur, Megha Sharma, Jiaming Cui, Qingsong Wen, Chao Zhang, and B. Aditya Prakash. Time-MMD: Multi-domain multimodal dataset for time series analysis. In *The Thirty-eight Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2024a. URL https://openreview.net/forum?id=fuD0h4R1IL.

Peiyuan Liu, Beiliang Wu, Yifan Hu, Naiqi Li, Tao Dai, Jigang Bao, and Shu-Tao Xia. Timebridge: Non-stationarity matters for long-term time series forecasting. *International Conference on Machine Learning*, 2025.

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*, 2024b. URL https://openreview.net/forum?id=JePfAI8fah.

Nannan Lu, Zhen Tan, Zhiyuan Han, and Song Zhu. Modality-mix learning: Promoting multimodal learning through multilabel objective. *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 04 2026. doi: 10.1109/TNNLS.2026.3677806.

Mengmeng Ma, Jian Ren, Long Zhao, Davide Testuggine, and Xi Peng. Are multimodal transformers robust to missing modality? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 18177–18186, June 2022.

Huu Hiep Nguyen, Minh Hoang Nguyen, Dung Nguyen, and Hung Le. Spectral text fusion: A frequency-aware approach to multimodal time-series forecasting. In *The 29th International Conference on Artificial Intelligence and Statistics*, 2026. URL https://openreview.net/forum?id=w4yjJc06zt.

Yuqi Nie, Nam H Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=Jbdc0vTOcol.

Xiaokang Peng, Yake Wei, Andong Deng, Dong Wang, and Di Hu. Balanced multimodal learning via on-the-fly gradient modulation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. 2019.

Olivier Roy and Martin Vetterli. The effective rank: A measure of effective dimensionality. In *2007 15th European Signal Processing Conference*, pp. 606–610, 2007.

Chengsen Wang, Qi Qi, Jingyu Wang, Haifeng Sun, Zirui Zhuang, Jinming Wu, Lei Zhang, and Jianxin Liao. Chattime: A unified multimodal time series foundation model bridging numerical and textual data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025. arXiv:2412.11376.

Shiyu Wang, Haixu Wu, Xiaoming Shi, Tengge Hu, Huakun Luo, Lintao Ma, James Y Zhang, and JUN ZHOU. Timemixer: Decomposable multiscale mixing for time series forecasting. In *International Conference on Learning Representations (ICLR)*, 2024a.

Weiyao Wang, Du Tran, and Matt Feiszli. What makes training multi-modal classification networks hard? In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12692–12702, 2020. doi: 10.1109/CVPR42600.2020.01271.

Xinlei Wang, Maike Feng, Jing Qiu, Jinjin Gu, and Junhua Zhao. From news to forecast: Integrating event analysis in LLM-based time series forecasting with reflection. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024b. URL https://openreview.net/forum?id=tj8nsfxi5r.

Yake Wei and Di Hu. Mmpareto: boosting multimodal learning with innocent unimodal assistance. In *International Conference on Machine Learning*, 2024.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with Auto-Correlation for long-term series forecasting. In *Advances in Neural Information Processing Systems*, 2021.

Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(9):11121–11128, Jun. 2023. doi: 10.1609/aaai.v37i9.26317.

Xiaohui Zhang, Jaehong Yoon, Mohit Bansal, and Huaxiu Yao. Multimodal representation learning by alternating unimodal adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 27456–27466, June 2024.

Yunhao Zhang and Junchi Yan. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=vSVLM2j9eie.

Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *The Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Virtual Conference*, volume 35, pp. 11106–11115. AAAI Press, 2021.

Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *Proc. 39th International Conference on Machine Learning (ICML 2022)*, 2022.

Appendix for
Uncovering and Resolving Text Collapse in Multimodal Time Series Forecasting

TABLE OF CONTENTS

7 Implementation Details **14**

7.1 Experimental Setup 14

7.2 Dataset Details 14

7.3 Diagnostic Metrics 15

7.4 Algorithm 16

8 Full Results **16**

8.1 Full Forecasting Results on Time-MMD 16

8.2 Full Forecasting Results on Financial Benchmarks 16

8.3 Full SpecTF Adaptation Results 17

8.4 Per-domain Diagnostic Results 29

8.5 Comparison with a Modality-Balancing Baseline 31

9 Full Analysis **32**

9.1 Gradient Analysis Across Domains 32

9.2 Learning Rate Sensitivity Across Domains 34

9.3 Prediction Divergence Analysis 37

9.4 Hyperparameter Sensitivity 39

9.5 Residual Decomposition on Time-MMD 40

9.6 Prediction Visualisation 42

10 Statistical Robustness and Sensitivity Analyses **44**

10.1 Multi-seed Ablation Diagnostics 44

10.2 Ablation with Variance 44

10.3 Seed-level Significance 45

10.4 Text-encoder Sensitivity 45

11 Limitations **46**

12 Broader Impacts **46**

7 Implementation Details

7.1 Experimental Setup

Backbones. We evaluate REST-TS with eight forecasting backbones. They are PatchTST (Nie et al., 2023), iTransformer (Liu et al., 2024b), TimeBridge (Liu et al., 2025), Autoformer (Wu et al., 2021), Informer (Zhou et al., 2021), Crossformer (Zhang & Yan, 2023), FEDformer (Zhou et al., 2022), and DLinear (Zeng et al., 2023). Each backbone is used in the online and target networks without changing its internal layers. The Transformer-based models reuse their forecast heads as Proj_ϕ. DLinear maps directly from $\mathbb{R}^T$ to $\mathbb{R}^H$, so we add a linear projection after its output. The numerical-only model, MMTSF (Liu et al., 2024a), and TaTS (Li et al., 2026) use the same backbone settings.

Text representation. Following MM-TSFlib, we use the frozen token-embedding layer of GPT-2 (Radford et al., 2019). We mean-pool the token embeddings and pass the result through a trainable projection. The same frozen embeddings are used for every dataset, without domain-specific tuning.

Training. We optimize the trainable parameters with Adam (Kingma & Ba, 2015) and a learning rate of 10^{-4}. The batch size is 32. Training runs for at most 20 epochs and stops if validation MSE does not improve for 5 epochs. We use one NVIDIA A100 GPU with 80GB of memory. Experiments that report uncertainty use three random seeds.

Hyperparameters. We set the EMA coefficient to $m = 0.99$. The moving-average width in Eq. 7 is 4, and the Fourier component count in Eq. 8 is 2. We set $\lambda_1 = \lambda_2 = 0.1$ to provide sufficient residual supervision to the text branch without allowing the auxiliary objectives to dominate the primary forecasting loss.

7.2 Dataset Details

We evaluate on Time-MMD, FNSPID, and FNF. Their numerical data are summarised in Table 5, Table 6, and Table 7.

Time-MMD. Time-MMD pairs multivariate time series with curated texts in nine domains (Liu et al., 2024a). The domains are Agriculture, Climate, Economy, Energy, Environment, Health, Security, SocialGood, and Traffic. Their frequencies range from daily to monthly. We use the data splits and evaluation protocol provided with the benchmark.

FNSPID. FNSPID pairs daily stock prices with news articles (Dong et al., 2024). We use data for Delta Airlines, IBM, JPMorgan Chase, NVIDIA, Pfizer, and Tesla. These texts are news articles rather than the curated texts used in Time-MMD.

FNF. FNF pairs Bitcoin price, Web Traffic, and Electricity Demand with news at each time step (Wang et al., 2024b). We use FNF and FNSPID to test the method outside the curated-text setting of Time-MMD.

Table 5: Overview of the numerical data in the Time-MMD datasets (Liu et al., 2024a).

Dataset Name/Domain	Prediction Length	Dimension	Frequency	Number of Samples	Timespan
Agriculture	{6, 8, 10, 12}	1	Monthly	496	1983 to Present
Climate	{6, 8, 10, 12}	5	Monthly	496	1983 to Present
Economy	{6, 8, 10, 12}	3	Monthly	423	1989 to Present
Energy	{12, 24, 36, 48}	9	Weekly	1,479	1996 to Present
Environment	{48, 96, 192, 336}	4	Daily	11,102	1982 to 2023
Health	{12, 24, 36, 48}	11	Monthly	1,889	1997 to Present
Security	{6, 8, 10, 12}	1	Monthly	297	1999 to Present
Social Good	{6, 8, 10, 12}	1	Monthly	900	1950 to Present
Traffic	{6, 8, 10, 12}	1	Monthly	531	1980 to Present

Table 6: Overview of the numerical data in the FNSPID datasets (Dong et al., 2024).

Company/Stock Name	Prediction Length	Dimension	Frequency	Number of Samples	Timespan
Delta Airlines (DAL)	{6, 8, 10, 12}	1	Daily	3,020	2009 to 2020
IBM (IBM)	{6, 8, 10, 12}	1	Daily	1,258	2016 to 2020
JPMorgan Chase (JPM)	{6, 8, 10, 12}	1	Daily	755	2018 to 2020
NVIDIA (NVDA)	{6, 8, 10, 12}	1	Daily	2,516	2011 to 2020
Pfizer (PFE)	{6, 8, 10, 12}	1	Daily	1,258	2016 to 2020
Tesla (TSLA)	{6, 8, 10, 12}	1	Daily	504	2019 to 2020

Table 7: Overview of the numerical data in the FNF datasets (Wang et al., 2024b).

Dataset Name/Domain	Prediction Length	Dimension	Frequency	Number of Samples	Timespan
Bitcoin Price	{6, 8, 10, 12}	1	Daily	4,498	2009 to 2021
Web Traffic	{6, 8, 10, 12}	1	Daily	727	2015 to 2016
Electricity Demand	{6, 8, 10, 12}	1	Daily	1,461	2019 to 2022

7.3 Diagnostic Metrics

We use three complementary metrics to diagnose text utilization: effective rank, CKA, and prediction divergence. The first two characterize the text representations, while prediction divergence measures the sensitivity of the final forecast to the text input.

Effective rank. Let $\mathbf{H} \in \mathbb{R}^{N \times d}$ contain one text representation in each row. Let $\sigma_1, \ldots, \sigma_d$ be its singular values. Effective rank (Roy & Vetterli, 2007) is

$$\mathrm{erank}(\mathbf{H}) = \exp\left(-\sum_{i=1}^{d} p_i \log p_i \right), \qquad p_i = \frac{\sigma_i}{\sum_j \sigma_j}. \tag{13}$$

The normalized singular values $\{p_i\}$ form a distribution over directions of variation. Effective rank measures how evenly this variation is spread. It ranges from 1 to d. A value of 1 means that all representations lie along one direction, while a value of d means that all singular values are equal. A higher effective rank shows greater representation diversity, but does not by itself show that this diversity comes from useful text content.

Effective rank does not require labels or a reference distribution. It can therefore be applied to text representations from different model architectures.

Centred Kernel Alignment (CKA). Effective rank does not show whether representation differences are related to the forecasting target. We therefore also use centred kernel alignment (Kornblith et al., 2019).

For centred text representations $\mathbf{H}_S$ and target representations $\mathbf{H}_Y$ from the same N samples, linear CKA is

$$\mathrm{CKA}(\mathbf{H}_S, \mathbf{H}_Y) = \frac{\|\mathbf{H}_Y^\top \mathbf{H}_S\|_F^2}{\|\mathbf{H}_S^\top \mathbf{H}_S\|_F \cdot \|\mathbf{H}_Y^\top \mathbf{H}_Y\|_F}. \tag{14}$$

CKA is 0 when the two representation matrices are unrelated and 1 when they are identical up to an orthogonal transformation. A high value means that samples have similar pairwise structure in the text and target spaces. This is evidence of target alignment, but it does not prove that the text causes a change in the forecast. CKA is invariant to orthogonal transformations and isotropic scaling.

Together, the metrics distinguish a low-diversity representation from one that is diverse but weakly aligned with the target. We report both because either metric alone can be misleading.

Prediction divergence. Effective rank and CKA characterize the text representations, but do not directly measure how much the text affects the final forecast. We therefore measure prediction divergence by replacing every report with the same no-information sentence while keeping the

text branch active. Let $\hat{\mathbf{y}}_{\text{real}}$ and $\hat{\mathbf{y}}_{\text{no_info}}$ denote the forecasts with the original and replaced text, respectively. We compute

$$\delta = \frac{\text{MSE}(\hat{\mathbf{y}}_{\text{real}}, \hat{\mathbf{y}}_{\text{no_info}})}{\text{MSE}(\hat{\mathbf{y}}_{\text{real}}, \mathbf{y})}, \tag{15}$$

and report it as a percentage. The numerator measures the change in the forecast when the report content is replaced, while the denominator normalizes this change by the forecasting error. A larger δ indicates greater sensitivity of the forecast to the text content, whereas a value near zero indicates that the replacement has little effect.

Estimation protocol. For each framework and backbone, we collect text representations from the full Time-MMD test split of each domain. We compute effective rank without centring. CKA uses feature-centred matrices. For CKA, the target matrix $\mathbf{H}_Y$ is formed by projecting the ground-truth future values $\mathbf{Y}$ to the same embedding dimension. We compute both metrics separately for each domain and backbone. Table 21 averages domains for each backbone, while Appendix 8.4 reports individual settings.

7.4 Algorithm

The full training algorithm is in Algorithm. 1.

Algorithm 1 REST-TS training procedure.

Input. time-series dataset $\mathcal{D} = \{\mathbf{X}, S, \mathbf{Y}\}$; prediction length H; lookback window L; EMA momentum m; loss weights λ_1, λ_2.
Output. Trainable online, text projection, and residual head parameters Θ.

Prepare training samples $\{\mathbf{X}^{(i)}, S^{(i)}, \mathbf{Y}^{(i)}\}_{i=1}^{n}$ of length L and horizon H.
Initialise $\mathcal{F}_\theta$, Proj_ϕ, frozen TextEnc, TextProj, TrendHead, and EventHead; set target parameters $(\theta', \phi') \leftarrow (\theta, \phi)$.
while *not converged* **do**
  for *each training sample* $\{\mathbf{X}^{(i)}, S^{(i)}, \mathbf{Y}^{(i)}\}$ **do**
    $\mathbf{Z}_S^{\text{enc}} \leftarrow \text{TextEnc}(S^{(i)})$; $\mathbf{Z}_S \leftarrow \text{TextProj}(\mathbf{Z}_S^{\text{enc}})$
    $\mathbf{Z}_{\text{ts}} \leftarrow \mathcal{F}_\theta(\mathbf{X}^{(i)})$; $\mathbf{Z}_{\text{joint}} \leftarrow \text{Fusion}(\mathbf{Z}_{\text{ts}}, \mathbf{Z}_S)$; $\mathbf{Y}_{\text{joint}} \leftarrow \text{Proj}_\phi(\mathbf{Z}_{\text{joint}})$
    $\mathbf{Y}_{\text{ts}} \leftarrow \text{Proj}_{\phi'}(\mathcal{F}_{\theta'}(\mathbf{X}^{(i)}))$ ▷ no gradient; $\mathbf{R} \leftarrow \mathbf{Y}^{(i)} - \mathbf{Y}_{\text{ts}}$
    $\mathbf{R}_{\text{trend}} \leftarrow \text{MAvg}(\mathbf{R})$; $\mathbf{R}_{\text{noise}} \leftarrow \text{TopK}_{\text{FFT}}(\mathbf{R} - \mathbf{R}_{\text{trend}}, K)$; $\mathbf{R}_{\text{event}} \leftarrow \mathbf{R} - \mathbf{R}_{\text{trend}} - \mathbf{R}_{\text{noise}}$
    $\hat{\mathbf{R}}_{S,\text{trend}} \leftarrow \text{TrendHead}_{\phi_{\text{trend}}}(\mathbf{Z}_S)$; $\hat{\mathbf{R}}_{S,\text{event}} \leftarrow \text{EventHead}_{\phi_{\text{event}}}(\mathbf{Z}_S)$
    $\mathcal{L} \leftarrow \text{MSE}(\mathbf{Y}_{\text{joint}}, \mathbf{Y}^{(i)}) + \lambda_1 \|\hat{\mathbf{R}}_{S,\text{trend}} - \text{sg}(\mathbf{R}_{\text{trend}})\|_F^2 + \lambda_2 \|\hat{\mathbf{R}}_{S,\text{event}} - \text{sg}(\mathbf{R}_{\text{event}})\|_F^2$
    $\Theta \leftarrow \Theta - \nabla_\Theta \mathcal{L}$; $(\theta', \phi') \leftarrow m(\theta', \phi') + (1-m)(\theta, \phi)$
  end for
end while
return Θ

8 Full Results

8.1 Full Forecasting Results on Time-MMD

The complete Time-MMD results are grouped by backbone. Table 8 and Table 9 report TimeBridge and iTransformer. Table 10 and Table 11 report PatchTST and Crossformer. Table 12 and Table 13 report DLinear and FEDformer. Table 14 and Table 15 report Autoformer and Informer. REST-TS has lower MSE than TaTS in 62 of the 72 backbone and domain settings. The largest reductions occur on Economy and Agriculture. Most cases where REST-TS has higher MSE occur on SocialGood and Traffic.

8.2 Full Forecasting Results on Financial Benchmarks

Table 16 reports MSE on FNF and FNSPID, while Table 17 reports MAE. Both tables include all four forecast horizons for PatchTST and iTransformer. With PatchTST, the largest MSE reductions

Table 8: Full MSE on Time-MMD using TimeBridge and iTransformer as time-series models.

Models		TimeBridge					iTransformer				
Method		Uni-modal	MMTSF	TaTS	REST-TS	*Promotion*	Uni-modal	MMTSF	TaTS	REST-TS	*Promotion*
Agriculture	6	0.071	0.083	0.083	0.058	18.3%	0.077	0.067	0.066	0.065	2.1%
	8	0.098	0.101	0.099	0.087	11.2%	0.104	0.101	0.100	0.098	2.4%
	10	0.124	0.127	0.123	0.105	14.7%	0.142	0.131	0.124	0.130	−5.2%
	12	0.174	0.174	0.174	0.147	15.4%	0.167	0.178	0.168	0.158	5.4%
	Avg	0.117	0.121	0.120	0.099	15.4%	0.122	0.119	0.114	0.113	1.3%
Climate	6	1.593	1.253	1.237	0.927	25.1%	1.127	1.060	1.048	0.995	5.0%
	8	1.650	1.252	1.266	0.952	24.0%	1.191	1.020	1.015	1.028	−1.2%
	10	1.459	1.208	1.174	0.970	17.4%	1.215	1.061	1.044	1.029	1.4%
	12	1.518	1.335	1.114	0.973	12.6%	1.199	1.054	1.034	1.035	−0.1%
	Avg	1.555	1.262	1.198	0.955	20.3%	1.183	1.049	1.035	1.022	1.3%
Economy	6	0.017	0.083	0.0104	0.0081	22.0%	0.015	0.112	0.0078	0.0084	−7.9%
	8	0.016	0.107	0.0104	0.0084	19.2%	0.0140	0.097	0.0083	0.0083	−0.4%
	10	0.018	0.107	0.0104	0.0083	20.1%	0.0140	0.097	0.0085	0.0084	0.8%
	12	0.021	0.108	0.0107	0.0084	21.8%	0.0130	0.113	0.0084	0.0084	0.2%
	Avg	0.018	0.101	0.0105	0.0083	20.8%	0.0140	0.105	0.0082	0.0084	−2.0%
Energy	12	0.127	0.120	0.119	0.103	13.7%	0.112	0.107	0.101	0.092	9.2%
	24	0.223	0.223	0.219	0.216	1.3%	0.222	0.200	0.221	0.197	1.4%
	36	0.409	0.329	0.396	0.304	7.5%	0.306	0.292	0.342	0.301	−3.1%
	48	0.462	0.473	0.454	0.419	7.8%	0.435	0.424	0.441	0.413	2.6%
	Avg	0.305	0.286	0.297	0.261	8.7%	0.269	0.256	0.276	0.251	1.9%
Environment	48	0.322	0.271	0.272	0.269	0.8%	0.415	0.310	0.271	0.263	2.9%
	96	0.364	0.340	0.281	0.277	1.3%	0.439	0.315	0.274	0.264	3.5%
	192	0.357	0.338	0.272	0.269	1.2%	0.454	0.303	0.271	0.255	6.0%
	336	0.358	0.332	0.271	0.266	2.0%	0.455	0.301	0.268	0.253	5.7%
	Avg	0.350	0.320	0.274	0.270	1.5%	0.441	0.307	0.271	0.259	4.4%
Health	12	1.350	1.064	1.021	0.983	3.7%	1.171	1.178	0.854	0.964	−12.9%
	24	2.233	1.365	1.401	1.256	8.0%	1.594	1.627	1.314	1.252	4.7%
	36	2.121	1.667	1.432	1.347	5.9%	1.742	1.880	1.419	1.400	1.3%
	48	2.167	1.921	1.604	1.420	11.5%	1.840	1.902	1.574	1.484	5.7%
	Avg	1.968	1.504	1.364	1.252	8.2%	1.587	1.647	1.290	1.275	1.2%
Security	6	130.05	137.89	127.89	105.65	17.4%	113.57	139.86	120.00	107.70	5.2%
	8	129.49	143.96	125.31	108.43	13.5%	115.88	140.27	117.57	109.91	5.2%
	10	125.40	135.73	126.34	110.25	12.1%	116.95	140.83	120.84	110.18	5.8%
	12	136.14	143.69	127.01	110.92	12.7%	117.40	140.77	120.43	111.11	5.4%
	Avg	130.27	140.32	126.64	108.82	14.1%	115.94	140.43	119.71	109.73	5.4%
Social Good	6	1.289	1.206	0.996	0.854	14.2%	1.129	1.103	0.948	0.905	4.6%
	8	1.223	1.248	1.100	0.926	15.8%	1.133	1.160	1.014	0.968	4.5%
	10	1.310	1.408	1.224	0.976	20.3%	1.275	1.138	1.074	1.021	4.9%
	12	1.341	1.395	1.175	1.010	14.0%	1.313	1.475	1.120	1.031	8.0%
	Avg	1.291	1.314	1.124	0.941	16.2%	1.212	1.219	1.039	0.981	5.6%
Traffic	6	0.150	0.210	0.141	0.164	−16.5%	0.203	0.249	0.175	0.161	7.9%
	8	0.167	0.216	0.148	0.163	−10.5%	0.209	0.252	0.195	0.171	12.4%
	10	0.188	0.214	0.158	0.163	−3.0%	0.211	0.260	0.204	0.170	16.5%
	12	0.191	0.212	0.192	0.210	−9.9%	0.231	0.233	0.217	0.225	−3.8%
	Avg	0.174	0.213	0.160	0.175	−9.6%	0.213	0.248	0.198	0.182	7.9%

over MMTSF occur on NVIDIA, from 0.140 to 0.059, and Pfizer, from 0.105 to 0.045. The results extend the evaluation to financial series paired with news, but the size of the gain varies by dataset.

8.3 Full SpecTF Adaptation Results

Table 2 gives the domain averages in the main paper. Table 18, Table 19, and Table 20 compare REST-TS with SpecTF on all nine Time-MMD domains. REST-TS has lower average MSE in six domains. The largest reductions occur on Energy and SocialGood. SpecTF has lower average MSE on Economy and Health, while the Environment results are equal at the reported precision.

Table 9: Full MAE on Time-MMD using TimeBridge and iTransformer as time-series models.

Models		TimeBridge					iTransformer				
Method		Uni-modal	MMTSF	TaTS	REST-TS	*Promotion*	Uni-modal	MMTSF	TaTS	REST-TS	*Promotion*
Agriculture	6	0.187	0.202	0.201	0.162	13.4%	0.200	0.180	0.175	0.173	1.2%
	8	0.218	0.219	0.217	0.210	3.2%	0.228	0.223	0.224	0.217	2.7%
	10	0.251	0.259	0.251	0.231	7.9%	0.273	0.257	0.249	0.250	−0.5%
	12	0.298	0.306	0.302	0.278	6.7%	0.301	0.318	0.309	0.280	7.0%
	Avg	0.238	0.246	0.243	0.220	7.6%	0.251	0.244	0.239	0.230	3.8%
Climate	6	1.011	0.886	0.885	0.768	13.2%	0.843	0.820	0.805	0.784	2.6%
	8	1.032	0.894	0.899	0.777	13.1%	0.876	0.801	0.797	0.801	−0.6%
	10	0.947	0.860	0.844	0.779	7.7%	0.885	0.820	0.808	0.800	0.9%
	12	0.983	0.918	0.832	0.779	6.4%	0.879	0.823	0.812	0.805	0.8%
	Avg	0.993	0.889	0.865	0.775	10.4%	0.871	0.816	0.805	0.798	0.9%
Economy	6	0.104	0.212	0.085	0.076	10.2%	0.099	0.235	0.075	0.078	−4.2%
	8	0.099	0.242	0.086	0.077	10.2%	0.098	0.262	0.077	0.077	0.3%
	10	0.110	0.284	0.085	0.075	11.6%	0.094	0.260	0.078	0.077	1.6%
	12	0.117	0.239	0.087	0.077	11.5%	0.091	0.276	0.076	0.077	−0.9%
	Avg	0.108	0.244	0.086	0.076	11.1%	0.096	0.258	0.077	0.078	−1.7%
Energy	12	0.246	0.248	0.242	0.238	1.8%	0.231	0.233	0.230	0.216	6.3%
	24	0.351	0.348	0.346	0.340	1.9%	0.352	0.333	0.352	0.328	1.6%
	36	0.471	0.433	0.462	0.415	4.1%	0.409	0.408	0.444	0.403	1.3%
	48	0.528	0.537	0.522	0.504	3.5%	0.509	0.506	0.518	0.495	2.1%
	Avg	0.399	0.392	0.393	0.374	4.5%	0.375	0.370	0.386	0.361	2.4%
Environment	48	0.403	0.378	0.374	0.374	0.0%	0.473	0.387	0.371	0.367	1.1%
	96	0.430	0.418	0.371	0.374	−0.7%	0.493	0.402	0.373	0.368	1.2%
	192	0.418	0.418	0.371	0.369	0.6%	0.506	0.406	0.365	0.369	−1.1%
	336	0.430	0.419	0.370	0.369	0.3%	0.505	0.406	0.366	0.370	−1.1%
	Avg	0.420	0.408	0.372	0.372	−0.1%	0.494	0.400	0.369	0.369	−0.1%
Health	12	0.772	0.727	0.695	0.683	1.8%	0.674	0.717	0.637	0.651	−2.2%
	24	0.947	0.807	0.800	0.748	6.5%	0.807	0.826	0.761	0.724	4.9%
	36	0.949	0.861	0.806	0.775	3.9%	0.862	0.890	0.783	0.774	1.2%
	48	0.997	0.943	0.869	0.815	6.2%	0.923	0.932	0.828	0.810	2.2%
	Avg	0.916	0.834	0.793	0.755	4.7%	0.817	0.842	0.752	0.740	1.6%
Security	6	6.725	6.963	6.359	4.699	26.1%	5.698	7.099	5.976	4.883	14.3%
	8	6.563	7.010	6.259	4.853	22.5%	5.743	6.863	5.697	5.024	11.8%
	10	6.300	6.852	6.176	5.018	18.7%	5.615	7.179	5.859	5.060	9.9%
	12	6.782	7.060	6.268	5.024	19.8%	5.585	7.099	5.790	5.110	8.5%
	Avg	6.593	6.971	6.266	4.899	21.8%	5.660	7.060	5.830	5.019	11.3%
Social Good	6	0.489	0.479	0.416	0.399	4.1%	0.438	0.477	0.385	0.386	−0.3%
	8	0.495	0.528	0.448	0.442	1.4%	0.466	0.492	0.431	0.446	−3.5%
	10	0.539	0.577	0.508	0.470	7.4%	0.496	0.506	0.455	0.471	−3.5%
	12	0.573	0.627	0.520	0.528	−1.5%	0.534	0.615	0.500	0.469	6.1%
	Avg	0.524	0.553	0.473	0.460	2.7%	0.483	0.522	0.443	0.443	0.0%
Traffic	6	0.212	0.273	0.201	0.210	−4.3%	0.228	0.285	0.206	0.213	−3.2%
	8	0.225	0.275	0.208	0.222	−7.0%	0.236	0.292	0.227	0.219	3.4%
	10	0.251	0.274	0.222	0.222	−0.2%	0.243	0.296	0.234	0.218	6.7%
	12	0.220	0.242	0.211	0.221	−4.8%	0.246	0.269	0.217	0.237	−9.0%
	Avg	0.227	0.266	0.210	0.219	−4.1%	0.238	0.285	0.221	0.222	−0.4%

Table 10: Full MSE on Time-MMD using PatchTST and Crossformer as time-series models

Models		PatchTST					Crossformer				
Method		Uni-modal	MMTSF	TaTS	REST-TS	*Promotion*	Uni-modal	MMTSF	TaTS	REST-TS	*Promotion*
Agriculture	6	0.074	0.073	0.072	0.068	5.5%	0.222	0.173	0.330	0.073	57.9%
	8	0.104	0.090	0.092	0.096	−6.5%	0.304	0.242	0.250	0.099	59.1%
	10	0.136	0.131	0.128	0.128	−0.3%	0.357	0.349	0.336	0.130	61.3%
	12	0.168	0.167	0.165	0.167	−0.9%	0.409	0.442	0.341	0.160	53.1%
	Avg	0.120	0.115	0.114	0.115	−0.5%	0.323	0.302	0.314	0.116	61.5%
Climate	6	1.259	1.010	1.014	0.969	4.1%	1.159	0.996	0.949	0.911	4.0%
	8	1.208	0.993	0.986	0.998	−1.2%	1.104	0.961	0.983	0.932	3.0%
	10	1.218	1.035	1.050	1.008	2.6%	1.127	1.016	0.977	0.941	3.7%
	12	1.197	1.025	1.027	1.006	1.8%	1.105	0.989	1.009	0.943	4.6%
	Avg	1.220	1.016	1.019	0.995	2.0%	1.124	0.991	0.980	0.932	4.9%
Economy	6	0.017	0.123	0.0092	0.0095	−3.4%	0.659	0.141	0.170	0.0097	93.1%
	8	0.016	0.093	0.0086	0.0086	0.0%	0.661	0.119	0.114	0.0095	91.7%
	10	0.017	0.123	0.0084	0.0088	−4.5%	0.836	0.244	0.249	0.0096	96.1%
	12	0.018	0.118	0.0085	0.0091	−7.1%	0.875	0.185	0.201	0.0095	94.9%
	Avg	0.017	0.114	0.0087	0.0090	−3.7%	0.758	0.172	0.183	0.0096	94.4%
Energy	12	0.105	0.111	0.107	0.101	3.8%	0.138	0.153	0.162	0.105	23.9%
	24	0.241	0.212	0.213	0.209	1.3%	0.281	0.352	0.240	0.211	12.2%
	36	0.304	0.310	0.293	0.303	−3.4%	0.331	0.354	0.362	0.300	9.4%
	48	0.427	0.421	0.419	0.420	−0.2%	0.422	0.431	0.439	0.415	1.7%
	Avg	0.269	0.263	0.258	0.258	0.0%	0.293	0.323	0.301	0.258	11.9%
Environment	48	0.492	0.345	0.270	0.270	0.2%	0.495	0.281	0.302	0.269	4.2%
	96	0.541	0.353	0.280	0.279	0.3%	0.562	0.294	0.340	0.279	5.1%
	192	0.581	0.339	0.272	0.271	0.3%	0.567	0.310	0.348	0.270	12.8%
	336	0.593	0.341	0.326	0.268	17.7%	0.582	0.312	0.402	0.267	14.4%
	Avg	0.552	0.345	0.287	0.272	5.2%	0.551	0.299	0.348	0.272	9.1%
Health	12	1.267	1.067	1.011	0.979	3.2%	1.453	1.014	0.956	0.997	−4.3%
	24	1.681	1.361	1.331	1.271	4.5%	1.537	1.224	1.453	1.240	−1.3%
	36	1.819	1.627	1.495	1.375	8.0%	1.565	1.238	1.279	1.342	−8.4%
	48	1.842	1.631	1.593	1.465	8.0%	1.586	1.440	1.352	1.446	−6.9%
	Avg	1.652	1.422	1.357	1.272	6.3%	1.535	1.229	1.260	1.256	−2.2%
Security	6	108.80	122.78	117.26	105.91	2.7%	124.97	126.09	125.41	103.75	17.0%
	8	113.53	135.06	118.63	108.83	4.1%	126.51	125.17	125.53	105.45	15.8%
	10	114.82	133.81	120.60	110.84	3.5%	127.70	127.92	127.69	107.33	15.9%
	12	114.25	140.95	125.60	111.24	2.6%	128.69	128.51	131.09	108.19	15.8%
	Avg	112.85	133.15	120.52	109.21	3.2%	126.96	126.92	127.43	106.18	16.3%
Social Good	6	1.047	1.098	0.940	0.864	8.1%	0.791	0.746	0.706	0.830	−17.5%
	8	1.102	1.183	0.969	0.939	3.1%	0.832	0.787	0.742	0.905	−22.0%
	10	1.089	1.299	1.068	1.036	3.0%	0.916	0.808	0.781	0.938	−20.1%
	12	1.148	1.252	0.983	1.057	−7.5%	0.921	0.814	0.815	0.998	−22.7%
	Avg	1.097	1.208	0.990	0.974	1.6%	0.865	0.789	0.761	0.918	−20.6%
Traffic	6	0.182	0.208	0.160	0.165	−3.0%	0.227	0.162	0.168	0.172	−6.4%
	8	0.167	0.210	0.173	0.164	1.8%	0.216	0.162	0.161	0.173	−7.6%
	10	0.178	0.222	0.183	0.170	4.5%	0.202	0.169	0.166	0.173	−4.1%
	12	0.226	0.226	0.203	0.215	−5.9%	0.212	0.181	0.183	0.217	−19.8%
	Avg	0.188	0.216	0.180	0.178	1.0%	0.214	0.169	0.169	0.184	−9.1%

Table 11: Full MAE on Time-MMD using PatchTST and Crossformer as time-series models

Models / Method		PatchTST					Crossformer				
		Uni-modal	MMTSF	TaTS	REST-TS	*Promotion*	Uni-modal	MMTSF	TaTS	REST-TS	*Promotion*
Agriculture	6	0.197	0.188	0.185	0.183	1.1%	0.331	0.321	0.443	0.182	43.3%
	8	0.234	0.211	0.207	0.211	−1.9%	0.406	0.340	0.347	0.226	33.6%
	10	0.262	0.263	0.253	0.252	0.4%	0.435	0.415	0.415	0.273	34.1%
	12	0.294	0.296	0.291	0.289	0.8%	0.451	0.479	0.385	0.295	23.4%
	Avg	0.247	0.239	0.234	0.234	0.0%	0.406	0.389	0.398	0.244	37.2%
Climate	6	0.915	0.789	0.787	0.775	1.5%	0.852	0.760	0.751	0.752	−0.1%
	8	0.878	0.793	0.793	0.782	1.4%	0.829	0.762	0.762	0.755	0.9%
	10	0.894	0.822	0.825	0.787	4.3%	0.829	0.776	0.772	0.764	1.0%
	12	0.891	0.806	0.804	0.799	0.6%	0.837	0.773	0.787	0.763	1.3%
	Avg	0.895	0.803	0.802	0.786	2.0%	0.837	0.768	0.768	0.758	1.3%
Economy	6	0.104	0.289	0.081	0.082	−1.4%	0.749	0.337	0.378	0.081	75.9%
	8	0.104	0.234	0.076	0.077	−1.2%	0.767	0.307	0.296	0.081	72.6%
	10	0.104	0.233	0.077	0.078	−1.9%	0.884	0.464	0.476	0.081	82.6%
	12	0.109	0.251	0.077	0.079	−2.4%	0.912	0.396	0.379	0.081	78.7%
	Avg	0.105	0.252	0.078	0.079	−1.7%	0.828	0.376	0.382	0.081	78.5%
Energy	12	0.229	0.237	0.231	0.223	2.6%	0.262	0.279	0.281	0.231	11.8%
	24	0.363	0.342	0.341	0.336	1.4%	0.399	0.452	0.361	0.338	6.5%
	36	0.408	0.414	0.403	0.406	−0.7%	0.443	0.454	0.467	0.407	8.1%
	48	0.502	0.505	0.501	0.500	0.2%	0.521	0.513	0.524	0.497	3.1%
	Avg	0.376	0.375	0.369	0.366	0.8%	0.406	0.425	0.408	0.368	9.4%
Environment	48	0.500	0.407	0.374	0.376	−0.6%	0.509	0.379	0.409	0.374	1.4%
	96	0.538	0.424	0.374	0.375	−0.2%	0.585	0.405	0.455	0.373	7.8%
	192	0.556	0.431	0.368	0.369	−0.4%	0.607	0.436	0.463	0.368	15.7%
	336	0.553	0.424	0.406	0.369	9.2%	0.621	0.446	0.530	0.366	18.0%
	Avg	0.537	0.421	0.381	0.372	2.2%	0.581	0.417	0.465	0.370	11.2%
Health	12	0.735	0.707	0.667	0.657	1.5%	0.809	0.679	0.633	0.664	−5.0%
	24	0.846	0.788	0.763	0.736	3.6%	0.825	0.732	0.816	0.736	−0.6%
	36	0.918	0.868	0.813	0.772	5.0%	0.831	0.742	0.745	0.781	−5.3%
	48	0.921	0.877	0.858	0.810	5.6%	0.841	0.801	0.778	0.820	−5.3%
	Avg	0.855	0.810	0.775	0.744	4.0%	0.827	0.738	0.743	0.750	−1.6%
Security	6	5.119	6.195	5.626	4.719	7.8%	6.077	6.054	6.003	4.437	26.1%
	8	5.524	6.641	5.579	4.902	11.3%	6.237	6.039	6.078	4.597	23.9%
	10	5.484	6.760	5.802	5.036	8.2%	6.346	6.311	6.279	4.798	23.6%
	12	5.357	7.039	6.130	4.982	7.0%	6.449	6.311	6.498	4.798	24.0%
	Avg	5.371	6.659	5.784	4.910	8.6%	6.277	6.179	6.215	4.658	24.6%
Social Good	6	0.443	0.509	0.421	0.401	4.8%	0.431	0.434	0.351	0.390	−11.2%
	8	0.458	0.539	0.430	0.438	−1.8%	0.414	0.378	0.378	0.428	−13.3%
	10	0.469	0.574	0.462	0.471	−1.9%	0.463	0.430	0.440	0.449	−4.3%
	12	0.610	0.559	0.487	0.490	−0.6%	0.559	0.443	0.439	0.529	−20.5%
	Avg	0.495	0.545	0.450	0.450	0.0%	0.467	0.421	0.402	0.449	−11.8%
Traffic	6	0.252	0.259	0.212	0.219	−3.2%	0.394	0.285	0.301	0.236	17.3%
	8	0.226	0.260	0.210	0.220	−4.9%	0.382	0.285	0.282	0.236	16.2%
	10	0.242	0.275	0.224	0.223	0.5%	0.363	0.305	0.297	0.236	20.5%
	12	0.250	0.235	0.209	0.239	−14.5%	0.365	0.280	0.285	0.261	6.8%
	Avg	0.242	0.257	0.214	0.225	−5.3%	0.376	0.289	0.291	0.242	16.3%

Table 12: Full MSE on Time-MMD using DLinear and FEDformer as time-series models

Models		DLinear					FEDformer				
Method		Uni-modal	MMTSF	TaTS	REST-TS	*Promotion*	Uni-modal	MMTSF	TaTS	REST-TS	*Promotion*
Agriculture	6	0.170	0.164	0.175	0.099	39.6%	0.091	0.093	0.110	0.089	2.2%
	8	0.195	0.214	0.212	0.127	34.9%	0.127	0.188	0.149	0.116	8.7%
	10	0.218	0.231	0.233	0.149	31.7%	0.154	0.134	0.169	0.141	−5.4%
	12	0.308	0.273	0.296	0.175	35.9%	0.180	0.225	0.225	0.176	2.2%
	Avg	0.223	0.221	0.229	0.138	37.4%	0.138	0.160	0.163	0.131	5.1%
Climate	6	1.158	0.909	0.909	0.906	0.3%	1.206	1.031	0.973	0.920	5.4%
	8	1.191	1.142	0.936	0.925	1.2%	1.175	1.025	0.973	0.934	4.1%
	10	1.225	1.148	0.947	0.944	0.3%	1.199	1.039	0.972	0.961	1.1%
	12	1.185	0.961	0.963	0.952	1.0%	1.190	1.036	0.967	0.964	0.3%
	Avg	1.190	1.040	0.939	0.932	0.7%	1.192	1.033	0.971	0.945	2.7%
Economy	6	0.056	0.0199	0.020	0.0127	36.3%	0.042	0.052	0.0188	0.0098	47.9%
	8	0.056	0.255	0.027	0.0132	50.8%	0.039	0.027	0.0181	0.0094	48.0%
	10	0.047	0.230	0.022	0.0126	41.8%	0.036	0.032	0.0172	0.0092	46.5%
	12	0.075	0.202	0.027	0.0138	48.8%	0.053	0.036	0.023	0.0096	58.1%
	Avg	0.058	0.177	0.024	0.0131	45.3%	0.042	0.037	0.0193	0.0095	50.7%
Energy	12	0.136	0.176	0.189	0.107	21.3%	0.095	0.094	0.119	0.090	4.0%
	24	0.261	0.278	0.285	0.214	18.0%	0.170	0.180	0.299	0.191	−12.4%
	36	0.335	0.343	0.349	0.299	10.7%	0.249	0.348	0.355	0.288	−15.7%
	48	0.431	0.455	0.471	0.420	2.6%	0.445	0.449	0.450	0.414	7.0%
	Avg	0.291	0.313	0.324	0.260	10.7%	0.240	0.268	0.306	0.246	−2.5%
Environment	48	0.478	0.271	0.271	0.269	0.7%	0.505	0.274	0.275	0.268	2.3%
	96	0.562	0.299	0.300	0.279	6.8%	0.465	0.287	0.287	0.268	6.5%
	192	0.592	0.308	0.307	0.270	12.1%	0.510	0.289	0.288	0.253	12.2%
	336	0.600	0.306	0.305	0.267	12.6%	0.531	0.286	0.284	0.266	6.2%
	Avg	0.558	0.296	0.296	0.271	8.4%	0.503	0.284	0.283	0.264	6.8%
Health	12	1.595	1.487	1.352	1.059	21.6%	1.051	0.953	0.972	0.938	1.6%
	24	1.778	1.644	1.456	1.323	9.1%	1.493	1.242	1.313	1.286	−3.5%
	36	1.759	1.651	1.462	1.413	3.4%	1.661	1.353	1.443	1.390	−2.8%
	48	1.818	1.756	1.595	1.482	7.1%	1.737	1.420	1.513	1.480	−4.3%
	Avg	1.737	1.635	1.466	1.319	10.0%	1.486	1.242	1.310	1.273	−2.5%
Security	6	107.05	110.71	110.75	105.23	1.7%	112.11	118.40	119.17	107.14	4.4%
	8	107.70	112.24	112.19	106.96	0.7%	113.90	111.83	120.37	107.84	3.6%
	10	109.36	113.70	113.41	107.55	1.7%	117.48	118.90	124.63	110.42	6.0%
	12	112.34	113.87	113.75	108.33	3.6%	114.45	117.04	124.37	110.51	3.4%
	Avg	109.11	112.63	112.53	107.02	1.9%	114.48	116.54	122.13	108.98	4.8%
Social Good	6	1.018	1.347	1.347	0.766	24.8%	0.821	0.827	0.781	0.973	−24.6%
	8	1.137	1.319	1.361	0.842	25.9%	0.929	0.952	0.874	1.047	−19.8%
	10	1.210	1.277	1.268	0.903	25.4%	1.059	1.044	0.970	1.082	−11.6%
	12	1.238	1.244	1.232	0.957	22.3%	1.105	1.079	1.041	1.141	−9.6%
	Avg	1.151	1.297	1.302	0.867	24.7%	0.979	0.976	0.917	1.061	−15.8%
Traffic	6	0.221	0.384	0.186	0.176	5.3%	0.202	0.171	0.164	0.179	−9.3%
	8	0.220	0.448	0.193	0.179	7.3%	0.184	0.177	0.177	0.191	−8.2%
	10	0.217	0.415	0.190	0.180	5.5%	0.198	0.174	0.169	0.181	−6.9%
	12	0.261	0.401	0.215	0.211	1.7%	0.237	0.241	0.224	0.243	−8.4%
	Avg	0.230	0.412	0.196	0.186	5.1%	0.205	0.191	0.184	0.199	−8.4%

Table 13: Full MAE on Time-MMD using DLinear and FEDformer as time-series models

Models		DLinear					FEDformer				
Method		Uni-modal	MMTSF	TaTS	REST-TS	*Promotion*	Uni-modal	MMTSF	TaTS	REST-TS	*Promotion*
Agriculture	6	0.312	0.296	0.308	0.233	21.3%	0.239	0.235	0.263	0.210	10.5%
	8	0.340	0.350	0.344	0.268	21.2%	0.281	0.364	0.321	0.244	13.2%
	10	0.355	0.343	0.347	0.293	14.7%	0.313	0.265	0.332	0.271	−2.1%
	12	0.407	0.378	0.398	0.312	17.5%	0.313	0.394	0.384	0.314	−0.3%
	Avg	0.354	0.342	0.350	0.276	19.3%	0.286	0.315	0.325	0.260	9.1%
Climate	6	0.866	0.751	0.751	0.747	0.5%	0.909	0.804	0.784	0.753	3.9%
	8	0.874	0.845	0.761	0.755	0.8%	0.897	0.818	0.789	0.757	4.0%
	10	0.882	0.855	0.766	0.763	0.4%	0.893	0.818	0.784	0.779	0.6%
	12	0.868	0.772	0.773	0.768	0.5%	0.875	0.810	0.798	0.777	2.6%
	Avg	0.872	0.806	0.763	0.758	0.6%	0.893	0.812	0.789	0.767	2.7%
Economy	6	0.189	0.114	0.115	0.092	19.3%	0.168	0.194	0.114	0.083	27.3%
	8	0.191	0.439	0.132	0.094	28.9%	0.162	0.134	0.111	0.082	26.1%
	10	0.173	0.420	0.118	0.092	22.1%	0.153	0.141	0.108	0.081	25.0%
	12	0.216	0.338	0.131	0.096	26.9%	0.183	0.143	0.124	0.082	33.7%
	Avg	0.192	0.328	0.124	0.093	25.1%	0.166	0.153	0.114	0.082	28.2%
Energy	12	0.264	0.310	0.325	0.235	11.0%	0.212	0.215	0.256	0.205	3.3%
	24	0.385	0.390	0.397	0.341	11.4%	0.303	0.315	0.424	0.316	−4.3%
	36	0.437	0.438	0.444	0.403	7.8%	0.373	0.462	0.470	0.393	−5.4%
	48	0.498	0.512	0.524	0.497	0.2%	0.514	0.539	0.549	0.497	3.3%
	Avg	0.396	0.413	0.422	0.369	6.8%	0.351	0.383	0.425	0.353	−0.6%
Environment	48	0.531	0.383	0.383	0.369	3.8%	0.543	0.372	0.373	0.372	0.1%
	96	0.608	0.425	0.425	0.378	11.1%	0.524	0.388	0.395	0.375	3.3%
	192	0.608	0.446	0.445	0.369	17.2%	0.556	0.409	0.403	0.367	8.9%
	336	0.618	0.447	0.447	0.367	17.8%	0.573	0.400	0.396	0.369	6.9%
	Avg	0.591	0.425	0.425	0.371	12.7%	0.549	0.392	0.392	0.371	5.3%
Health	12	0.808	0.846	0.762	0.687	9.9%	0.756	0.676	0.682	0.639	5.4%
	24	0.835	0.875	0.774	0.767	0.9%	0.933	0.797	0.843	0.739	7.3%
	36	0.850	0.883	0.780	0.791	−1.4%	0.980	0.828	0.890	0.783	5.4%
	48	0.899	0.938	0.871	0.825	5.3%	0.969	0.847	0.909	0.819	3.3%
	Avg	0.848	0.885	0.797	0.768	3.6%	0.909	0.787	0.831	0.745	5.3%
Security	6	4.531	4.700	4.691	4.524	0.2%	4.830	5.452	5.569	4.727	2.1%
	8	4.649	4.826	4.827	4.732	−1.8%	5.200	4.911	5.778	4.741	3.5%
	10	4.759	4.980	4.936	4.821	−1.3%	5.467	5.603	6.163	4.809	12.0%
	12	4.904	4.995	4.968	4.867	0.8%	5.135	5.455	6.117	4.935	3.9%
	Avg	4.711	4.875	4.856	4.736	−0.5%	5.158	5.355	5.907	4.803	6.9%
Social Good	6	0.627	0.816	0.816	0.358	42.9%	0.386	0.420	0.399	0.466	−20.7%
	8	0.702	0.796	0.815	0.406	42.2%	0.451	0.461	0.446	0.507	−13.6%
	10	0.755	0.773	0.770	0.441	41.6%	0.478	0.478	0.499	0.584	−22.3%
	12	0.763	0.769	0.766	0.460	39.7%	0.588	0.560	0.562	0.565	−1.0%
	Avg	0.712	0.789	0.792	0.416	41.6%	0.476	0.479	0.477	0.530	−11.3%
Traffic	6	0.358	0.498	0.310	0.283	8.8%	0.294	0.216	0.214	0.271	−26.5%
	8	0.357	0.515	0.324	0.289	10.7%	0.234	0.220	0.224	0.263	−19.4%
	10	0.347	0.489	0.320	0.237	25.9%	0.251	0.222	0.218	0.240	−10.3%
	12	0.374	0.436	0.316	0.323	−2.2%	0.275	0.272	0.234	0.295	−26.1%
	Avg	0.359	0.485	0.317	0.283	10.9%	0.264	0.232	0.223	0.267	−20.0%

Table 14: Full MSE on Time-MMD using Autoformer and Informer as time-series models

Models		Autoformer					Informer				
Method		Uni-modal	MMTSF	TaTS	REST-TS	*Promotion*	Uni-modal	MMTSF	TaTS	REST-TS	*Promotion*
Agriculture	6	0.109	0.133	0.148	0.082	24.8%	0.451	0.427	0.269	0.068	74.7%
	8	0.135	0.182	0.152	0.111	17.8%	0.569	0.562	0.259	0.095	63.4%
	10	0.173	0.181	0.161	0.144	10.7%	0.618	0.480	0.392	0.118	69.9%
	12	0.213	0.263	0.204	0.175	14.4%	0.756	0.782	0.403	0.150	62.7%
	Avg	0.158	0.190	0.166	0.128	19.0%	0.599	0.562	0.331	0.108	67.3%
Climate	6	1.116	1.076	1.003	0.915	8.8%	1.084	1.160	0.939	0.926	1.4%
	8	1.117	1.070	0.966	0.924	4.4%	1.087	0.949	1.157	0.951	−0.2%
	10	1.170	1.052	0.979	0.954	2.6%	1.117	1.194	1.147	0.964	13.7%
	12	1.123	1.042	0.963	0.982	−2.0%	1.150	1.198	1.259	0.979	14.9%
	Avg	1.131	1.060	0.978	0.944	3.5%	1.110	1.125	1.125	0.955	14.0%
Economy	6	0.083	0.085	0.039	0.0093	75.9%	0.877	0.445	0.185	0.0091	95.1%
	8	0.069	0.167	0.0190	0.0094	50.5%	1.606	0.597	0.299	0.0094	96.9%
	10	0.070	0.128	0.040	0.0091	77.1%	1.409	0.594	0.366	0.0097	97.3%
	12	0.062	0.101	0.061	0.0093	84.7%	1.407	0.562	0.296	0.0095	96.8%
	Avg	0.071	0.120	0.040	0.0093	76.5%	1.325	0.549	0.286	0.0094	96.7%
Energy	12	0.161	0.227	0.164	0.127	21.1%	0.165	0.197	0.193	0.112	32.1%
	24	0.279	0.255	0.300	0.245	3.7%	0.248	0.345	0.243	0.213	12.5%
	36	0.364	0.381	0.385	0.344	5.5%	0.368	0.339	0.349	0.303	10.5%
	48	0.471	0.467	0.457	0.458	−0.2%	0.457	0.407	0.442	0.438	−7.7%
	Avg	0.319	0.333	0.327	0.294	7.8%	0.309	0.322	0.307	0.267	13.0%
Environment	48	0.503	0.301	0.280	0.268	4.3%	0.414	0.276	0.273	0.268	1.7%
	96	0.594	0.312	0.291	0.276	5.0%	0.467	0.276	0.280	0.269	2.6%
	192	0.607	0.348	0.298	0.270	9.5%	0.469	0.314	0.292	0.256	12.4%
	336	0.690	0.297	0.275	0.267	2.9%	0.487	0.287	0.275	0.254	7.8%
	Avg	0.599	0.314	0.286	0.270	5.6%	0.459	0.288	0.280	0.262	6.4%
Health	12	1.389	1.390	1.395	1.088	21.7%	1.173	0.941	0.955	1.067	−13.4%
	24	2.328	1.845	1.487	1.375	7.5%	1.215	1.189	1.214	1.342	−12.8%
	36	1.953	1.591	1.608	1.519	4.5%	1.315	1.168	1.334	1.510	−29.3%
	48	2.179	1.811	1.710	1.535	10.3%	1.408	1.418	1.315	1.582	−20.3%
	Avg	1.962	1.659	1.550	1.379	11.0%	1.278	1.179	1.205	1.375	−16.6%
Security	6	115.54	110.34	122.55	106.55	3.4%	133.17	132.96	140.46	105.58	20.6%
	8	113.36	134.29	125.97	110.07	2.9%	134.38	130.34	141.51	107.36	17.6%
	10	114.41	120.68	122.13	108.93	4.8%	131.38	132.34	132.75	109.87	16.4%
	12	117.80	127.35	118.87	109.91	6.7%	128.20	134.78	133.95	110.01	14.2%
	Avg	115.27	123.17	122.38	108.86	5.6%	131.78	132.61	137.17	108.21	17.9%
Social Good	6	1.234	1.093	1.154	0.923	15.5%	0.773	0.717	0.797	0.893	−24.5%
	8	1.236	1.458	1.338	1.093	11.6%	0.871	0.884	0.775	0.938	−21.0%
	10	1.332	1.190	1.423	1.146	3.7%	0.910	0.839	0.850	1.005	−19.8%
	12	1.312	1.417	1.597	1.087	17.1%	0.926	0.849	0.842	1.011	−20.0%
	Avg	1.278	1.289	1.378	1.062	16.9%	0.870	0.822	0.816	0.962	−17.9%
Traffic	6	0.201	0.214	0.162	0.175	−8.1%	0.197	0.186	0.159	0.161	−1.4%
	8	0.205	0.441	0.173	0.172	0.3%	0.202	0.169	0.164	0.164	−0.3%
	10	0.200	0.234	0.171	0.173	−1.4%	0.196	0.170	0.165	0.153	7.4%
	12	0.241	0.241	0.225	0.210	6.7%	0.213	0.185	0.181	0.190	−5.2%
	Avg	0.212	0.283	0.183	0.182	0.3%	0.202	0.177	0.167	0.167	0.0%

Table 15: Full MAE on Time-MMD using Autoformer and Informer as time-series models

Models		Autoformer					Informer				
Method		Uni-modal	MMTSF	TaTS	REST-TS	*Promotion*	Uni-modal	MMTSF	TaTS	REST-TS	*Promotion*
Agriculture	6	0.255	0.270	0.295	0.197	22.7%	0.557	0.543	0.375	0.180	52.0%
	8	0.278	0.345	0.304	0.239	14.0%	0.633	0.619	0.373	0.219	41.3%
	10	0.311	0.326	0.325	0.282	9.3%	0.633	0.533	0.471	0.232	50.8%
	12	0.345	0.404	0.357	0.317	8.1%	0.698	0.733	0.458	0.290	36.6%
	Avg	0.297	0.336	0.320	0.259	12.8%	0.630	0.607	0.419	0.230	45.2%
Climate	6	0.861	0.826	0.771	0.756	2.0%	0.829	0.865	0.758	0.760	−0.2%
	8	0.856	0.813	0.781	0.767	1.8%	0.829	0.763	0.838	0.773	−1.3%
	10	0.886	0.817	0.788	0.782	0.8%	0.849	0.862	0.831	0.777	6.5%
	12	0.855	0.812	0.794	0.788	0.8%	0.859	0.872	0.875	0.781	9.1%
	Avg	0.865	0.817	0.784	0.773	1.4%	0.841	0.841	0.826	0.773	6.4%
Economy	6	0.222	0.229	0.156	0.081	48.0%	0.896	0.636	0.389	0.080	79.4%
	8	0.210	0.318	0.109	0.081	25.9%	1.238	0.750	0.505	0.080	84.2%
	10	0.209	0.284	0.162	0.078	51.8%	1.153	0.712	0.580	0.083	85.7%
	12	0.186	0.257	0.198	0.080	57.0%	1.153	0.704	0.520	0.081	84.4%
	Avg	0.207	0.272	0.156	0.080	48.8%	1.110	0.700	0.499	0.081	83.8%
Energy	12	0.282	0.353	0.298	0.261	7.4%	0.289	0.317	0.323	0.236	18.3%
	24	0.416	0.384	0.422	0.370	3.6%	0.380	0.449	0.376	0.357	4.9%
	36	0.464	0.491	0.488	0.437	5.8%	0.492	0.462	0.465	0.417	9.7%
	48	0.550	0.554	0.552	0.522	5.1%	0.539	0.506	0.531	0.526	−4.0%
	Avg	0.428	0.446	0.440	0.397	7.2%	0.425	0.433	0.424	0.384	9.4%
Environment	48	0.544	0.385	0.374	0.375	−0.3%	0.481	0.384	0.380	0.377	0.7%
	96	0.605	0.398	0.387	0.373	3.5%	0.520	0.392	0.396	0.377	3.7%
	192	0.601	0.441	0.417	0.368	11.7%	0.519	0.446	0.418	0.372	11.0%
	336	0.643	0.422	0.386	0.365	5.4%	0.529	0.413	0.408	0.374	8.3%
	Avg	0.598	0.412	0.391	0.370	5.3%	0.512	0.409	0.400	0.375	6.4%
Health	12	0.895	0.875	0.872	0.735	15.7%	0.743	0.649	0.667	0.724	−11.5%
	24	1.191	1.008	0.867	0.786	9.4%	0.766	0.765	0.796	0.770	−0.7%
	36	1.005	0.893	0.924	0.834	6.6%	0.773	0.731	0.819	0.839	−14.7%
	48	1.064	0.972	0.941	0.839	10.8%	0.809	0.797	0.808	0.869	−9.0%
	Avg	1.039	0.937	0.901	0.798	11.5%	0.773	0.736	0.772	0.800	−8.7%
Security	6	5.268	4.912	5.753	5.072	−3.2%	6.626	6.643	7.141	4.846	26.9%
	8	4.967	6.743	6.021	4.934	0.7%	6.804	6.549	7.291	4.721	27.9%
	10	4.984	5.620	5.951	4.957	0.5%	6.631	6.772	6.670	5.006	24.5%
	12	5.253	5.966	5.590	4.998	4.9%	6.430	6.929	6.788	4.971	22.7%
	Avg	5.118	5.810	5.829	4.990	2.5%	6.623	6.723	6.973	4.886	26.2%
Social Good	6	0.684	0.616	0.579	0.469	19.0%	0.425	0.356	0.399	0.404	−13.6%
	8	0.664	0.828	0.710	0.595	10.4%	0.449	0.500	0.414	0.458	−10.6%
	10	0.708	0.626	0.791	0.596	4.7%	0.498	0.446	0.465	0.504	−13.0%
	12	0.749	0.769	0.883	0.571	23.8%	0.644	0.503	0.485	0.512	−5.7%
	Avg	0.701	0.710	0.741	0.558	20.4%	0.504	0.451	0.441	0.469	−6.4%
Traffic	6	0.295	0.276	0.212	0.228	−7.7%	0.352	0.345	0.281	0.220	21.8%
	8	0.296	0.469	0.229	0.226	1.3%	0.358	0.302	0.294	0.225	23.4%
	10	0.297	0.298	0.231	0.235	−1.9%	0.351	0.307	0.298	0.216	27.4%
	12	0.303	0.264	0.252	0.266	−5.7%	0.358	0.299	0.271	0.228	16.0%
	Avg	0.298	0.327	0.231	0.239	−3.6%	0.355	0.313	0.286	0.222	22.4%

Table 16: Full MSE results on FNF and FNSPID across all prediction horizons $T \in \{6, 8, 10, 12\}$.

			FNF			*FNSPID*					
Model	Method	T	Bitcoin	Electricity	Web Traffic	Delta	IBM	JPMorgan	NVIDIA	Pfizer	Tesla
PatchTST	MMTSF	6	0.011	0.193	0.676	0.031	0.216	0.165	0.113	0.074	1.011
		8	0.013	0.197	0.730	0.041	0.285	0.177	0.125	0.089	1.297
		10	0.006	0.197	0.669	0.019	0.323	0.184	0.147	0.113	1.572
		12	0.021	0.192	0.743	0.021	0.306	0.213	0.175	0.143	1.775
		Avg	0.013	0.195	0.704	0.028	0.283	0.185	0.140	0.105	1.414
	TaTS	6	0.005	0.188	0.666	0.014	0.134	0.095	0.056	0.037	0.617
		8	0.005	0.182	0.721	0.016	0.147	0.100	0.057	0.044	0.805
		10	0.006	0.185	0.623	0.020	0.176	0.102	0.062	0.051	0.862
		12	0.007	0.187	0.692	0.022	0.195	0.118	0.068	0.057	0.717
		Avg	0.006	0.185	0.675	0.018	0.163	0.104	0.061	0.047	0.750
	REST-TS	6	0.004	0.178	0.645	0.014	0.129	0.083	0.051	0.038	0.647
		8	0.005	0.175	0.666	0.016	0.161	0.093	0.055	0.044	0.724
		10	0.006	0.176	0.733	0.019	0.169	0.101	0.061	0.048	0.750
		12	0.007	0.180	0.680	0.022	0.186	0.112	0.068	0.050	0.784
		Avg	0.005	0.177	0.681	0.018	0.161	0.097	0.059	0.045	0.726
iTransformer	MMTSF	6	0.011	0.178	0.637	0.014	0.258	0.178	0.052	0.081	1.640
		8	0.006	0.182	0.629	0.016	0.315	0.190	0.059	0.091	1.944
		10	0.007	0.186	0.664	0.055	0.365	0.228	0.142	0.107	2.079
		12	0.021	0.187	0.675	0.022	0.407	0.224	0.144	0.123	1.080
		Avg	0.011	0.183	0.651	0.027	0.336	0.205	0.099	0.101	1.686
	TaTS	6	0.005	0.176	0.640	0.015	0.130	0.096	0.053	0.038	0.823
		8	0.005	0.175	0.625	0.016	0.158	0.101	0.058	0.042	0.832
		10	0.006	0.181	0.661	0.019	0.179	0.111	0.068	0.046	0.908
		12	0.007	0.183	0.649	0.024	0.192	0.117	0.072	0.053	0.736
		Avg	0.006	0.179	0.644	0.019	0.165	0.106	0.063	0.045	0.825
	REST-TS	6	0.004	0.168	0.605	0.015	0.142	0.090	0.052	0.037	0.825
		8	0.005	0.169	0.606	0.018	0.162	0.097	0.058	0.042	0.866
		10	0.006	0.172	0.656	0.019	0.174	0.115	0.062	0.047	0.748
		12	0.007	0.173	0.664	0.024	0.204	0.109	0.070	0.053	1.024
		Avg	0.006	0.170	0.633	0.019	0.170	0.103	0.061	0.045	0.866

Table 17: Full MAE results on FNF and FNSPID across all prediction horizons $T \in \{6, 8, 10, 12\}$.

			FNF			*FNSPID*					
Model	Method	T	Bitcoin	Electricity	Web Traffic	Delta	IBM	JPMorgan	NVIDIA	Pfizer	Tesla
PatchTST	MMTSF	6	0.053	0.335	0.596	0.110	0.329	0.295	0.232	0.202	0.737
		8	0.059	0.341	0.636	0.125	0.369	0.305	0.239	0.225	0.874
		10	0.039	0.340	0.602	0.087	0.376	0.305	0.254	0.250	0.977
		12	0.073	0.335	0.639	0.092	0.363	0.307	0.286	0.282	1.007
		Avg	0.056	0.338	0.618	0.103	0.359	0.303	0.253	0.240	0.899
	TaTS	6	0.035	0.327	0.597	0.075	0.253	0.230	0.165	0.141	0.589
		8	0.037	0.323	0.655	0.080	0.265	0.229	0.162	0.155	0.694
		10	0.040	0.327	0.599	0.088	0.281	0.232	0.168	0.169	0.700
		12	0.042	0.328	0.617	0.093	0.312	0.251	0.177	0.179	0.606
		Avg	0.038	0.326	0.617	0.084	0.278	0.235	0.168	0.161	0.647
	REST-TS	6	0.033	0.318	0.616	0.073	0.250	0.207	0.151	0.141	0.594
		8	0.036	0.316	0.629	0.080	0.270	0.219	0.159	0.154	0.629
		10	0.039	0.317	0.668	0.086	0.285	0.231	0.167	0.160	0.649
		12	0.042	0.321	0.640	0.093	0.300	0.243	0.180	0.179	0.659
		Avg	0.037	0.318	0.639	0.083	0.276	0.225	0.165	0.158	0.633
iTransformer	MMTSF	6	0.053	0.320	0.593	0.075	0.350	0.321	0.153	0.210	1.008
		8	0.039	0.324	0.568	0.081	0.401	0.322	0.163	0.224	1.099
		10	0.041	0.327	0.627	0.145	0.411	0.358	0.253	0.247	1.102
		12	0.073	0.330	0.630	0.093	0.438	0.343	0.253	0.272	0.786
		Avg	0.052	0.325	0.604	0.099	0.400	0.336	0.206	0.239	0.999
	TaTS	6	0.035	0.317	0.568	0.076	0.255	0.229	0.154	0.142	0.712
		8	0.037	0.317	0.566	0.081	0.278	0.236	0.160	0.150	0.708
		10	0.039	0.323	0.629	0.087	0.286	0.244	0.173	0.158	0.742
		12	0.043	0.323	0.616	0.098	0.303	0.251	0.181	0.171	0.633
		Avg	0.038	0.320	0.595	0.086	0.281	0.240	0.167	0.155	0.699
	REST-TS	6	0.033	0.310	0.580	0.076	0.264	0.221	0.153	0.140	0.698
		8	0.036	0.310	0.581	0.085	0.281	0.229	0.162	0.151	0.714
		10	0.038	0.312	0.622	0.088	0.289	0.249	0.170	0.161	0.632
		12	0.042	0.313	0.628	0.097	0.313	0.241	0.178	0.171	0.773
		Avg	0.037	0.311	0.603	0.086	0.287	0.235	0.166	0.156	0.704

Table 18: Full MSE and MAE of REST-TS vs. SpecTF on Agriculture, Climate, and Economy. **Bold** is the best.

Dataset	T	SpecTF		REST-TS		*Prom.*
		MSE	MAE	MSE	MAE	
Agriculture	6	**0.064**	**0.170**	0.068	0.177	-5.9%
	8	0.084	**0.192**	**0.083**	0.198	+1.3%
	10	0.110	0.236	**0.106**	**0.223**	+3.8%
	12	0.152	**0.275**	**0.144**	0.293	+5.3%
	Avg	0.103	**0.218**	**0.100**	0.223	+2.3%
Climate	6	**0.904**	**0.748**	0.904	0.752	-0.0%
	8	0.927	**0.756**	**0.926**	0.758	+0.1%
	10	0.965	0.775	**0.947**	**0.764**	+1.8%
	12	0.960	0.770	**0.952**	**0.767**	+0.8%
	Avg	0.939	0.762	**0.932**	**0.760**	+0.7%
Economy	6	**0.0084**	**0.0770**	0.0094	0.0810	-12.2%
	8	**0.0086**	**0.0785**	0.0096	0.0825	-12.0%
	10	**0.0085**	**0.0779**	0.0088	0.0783	-4.1%
	12	**0.0086**	**0.0784**	0.0090	0.0791	-4.5%
	Avg	**0.0085**	**0.0779**	0.0092	0.0802	-8.2%

Table 19: Full MSE and MAE of REST-TS vs. SpecTF on Energy, Environment, and Health. **Bold** is the best.

Dataset	T	SpecTF		REST-TS		*Prom.*
		MSE	MAE	MSE	MAE	
Energy	12	0.096	0.219	**0.087**	**0.206**	+9.6%
	24	0.199	0.326	**0.185**	**0.309**	+7.2%
	36	0.282	0.396	**0.262**	**0.378**	+7.1%
	48	0.406	0.493	**0.340**	**0.448**	+16.3%
	Avg	0.246	0.359	**0.218**	**0.335**	+11.1%
Environment	48	**0.265**	0.370	0.265	**0.370**	-0.1%
	96	**0.265**	**0.368**	0.267	0.373	-0.6%
	192	0.255	**0.365**	**0.255**	0.369	+0.1%
	336	**0.253**	**0.366**	0.255	0.367	-0.7%
	Avg	**0.260**	**0.367**	0.260	0.369	0.0%
Health	12	**0.985**	**0.647**	1.006	0.664	-2.1%
	24	**1.256**	**0.718**	1.292	0.744	-2.9%
	36	**1.395**	**0.769**	1.408	0.790	-0.9%
	48	**1.470**	**0.797**	1.476	0.816	-0.4%
	Avg	**1.276**	**0.733**	1.295	0.753	-1.5%

Table 20: Full MSE and MAE of REST-TS vs. SpecTF on Security, SocialGood, and Traffic. **Bold** is the best.

Dataset	T	SpecTF		REST-TS		*Prom.*
		MSE	MAE	MSE	MAE	
Security	6	106.36	**4.51**	**105.47**	4.51	+0.8%
	8	**106.95**	4.77	107.47	**4.66**	-0.5%
	10	109.77	**4.62**	**108.69**	4.76	+1.0%
	12	110.57	5.09	**109.43**	**4.88**	+1.0%
	Avg	108.41	4.75	**107.76**	**4.70**	+0.6%
SocialGood	6	0.907	**0.404**	**0.852**	0.463	+6.1%
	8	0.939	0.421	**0.858**	**0.406**	+8.6%
	10	0.962	**0.434**	**0.947**	0.465	+1.6%
	12	1.040	**0.514**	**0.994**	0.521	+4.5%
	Avg	0.962	**0.443**	**0.913**	0.464	+5.1%
Traffic	6	0.169	**0.202**	**0.156**	0.228	+7.8%
	8	**0.169**	**0.210**	0.170	0.242	-0.8%
	10	0.173	0.218	**0.163**	**0.217**	+5.9%
	12	**0.175**	**0.210**	0.192	0.261	-9.8%
	Avg	0.171	**0.210**	**0.170**	0.237	+0.7%

Table 21: **Text collapse diagnostics.** Results are averaged over the nine Time-MMD domains and all forecast horizons. Divergence δ is the relative change in the forecast when the report is removed. **Bold** marks the best value for each backbone and metric. Full results are in Appendix 8.4.

	Effective rank ↑			CKA ↑			Divergence δ (%) ↑		
Backbone	MMTSF	TaTS	REST-TS	MMTSF	TaTS	REST-TS	MMTSF	TaTS	REST-TS
Informer	1.11	1.07	**1.71**	0.178	0.159	**0.289**	1.42	1.51	**2.76**
PatchTST	1.12	1.28	**1.86**	0.224	0.245	**0.292**	0.09	0.00	**1.01**
TimeBridge	1.05	1.03	**2.13**	0.207	0.132	**0.298**	0.02	**0.87**	0.68
iTransformer	1.05	1.28	**2.16**	0.203	0.236	**0.309**	0.01	0.75	**0.94**
Mean	1.08	1.17	**1.96**	0.203	0.193	**0.297**	0.39	0.78	**1.35**

8.4 Per-domain Diagnostic Results

Table 21 reports effective rank, CKA and divergence across backbones. Figure 4 reports effective rank by domain and backbone. Figure 5 reports CKA for the same settings. A larger radar area means a higher value, but areas should be compared within each metric only. REST-TS generally has higher effective rank than MMTSF and TaTS. The largest rank differences appear on Energy and Environment. Its CKA is also higher in most settings, although the three methods have similar low values on Environment.

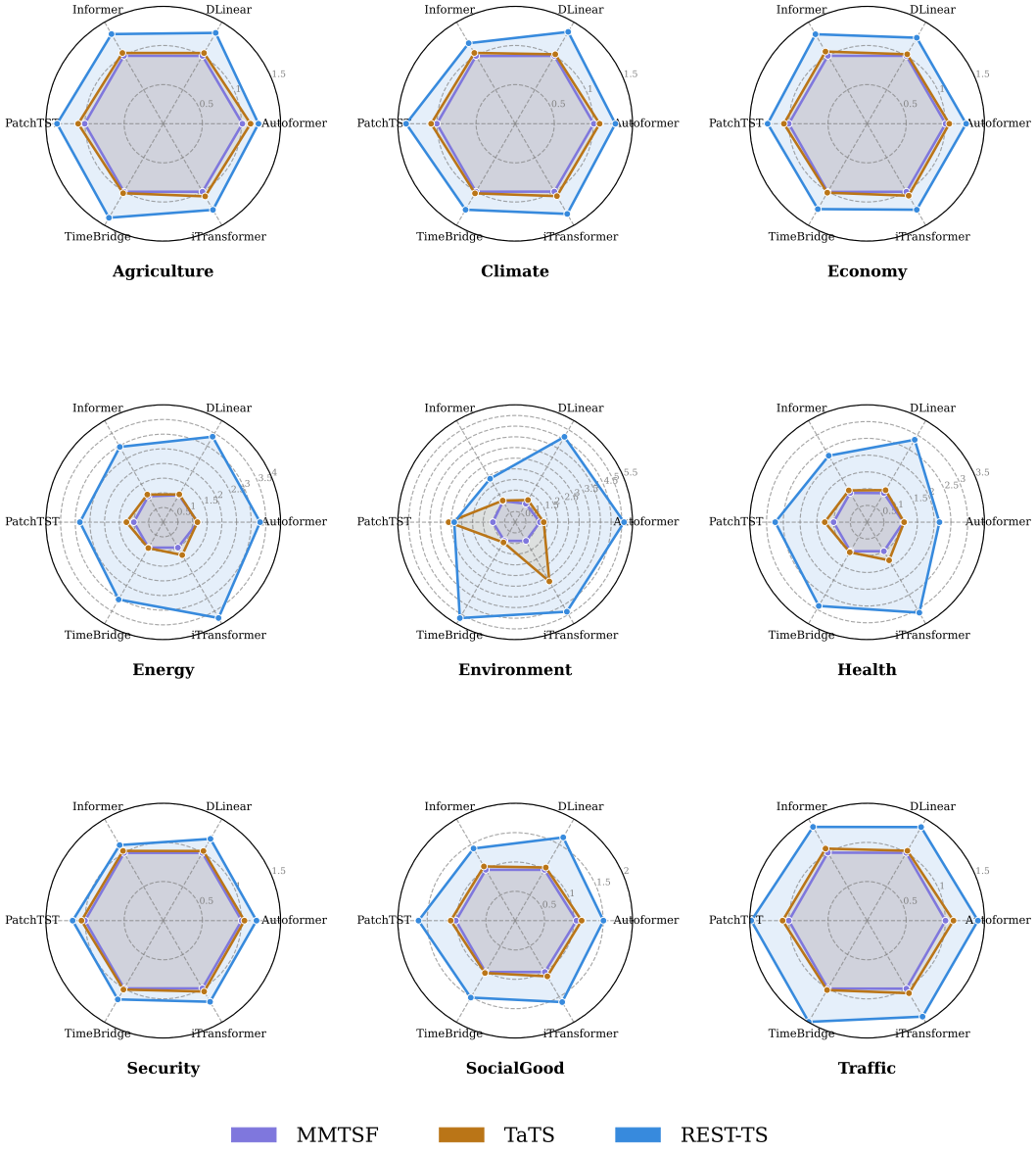


Figure 4: Effective rank of the text representations on nine Time-MMD domains and six backbones. Higher values indicate greater representation diversity.

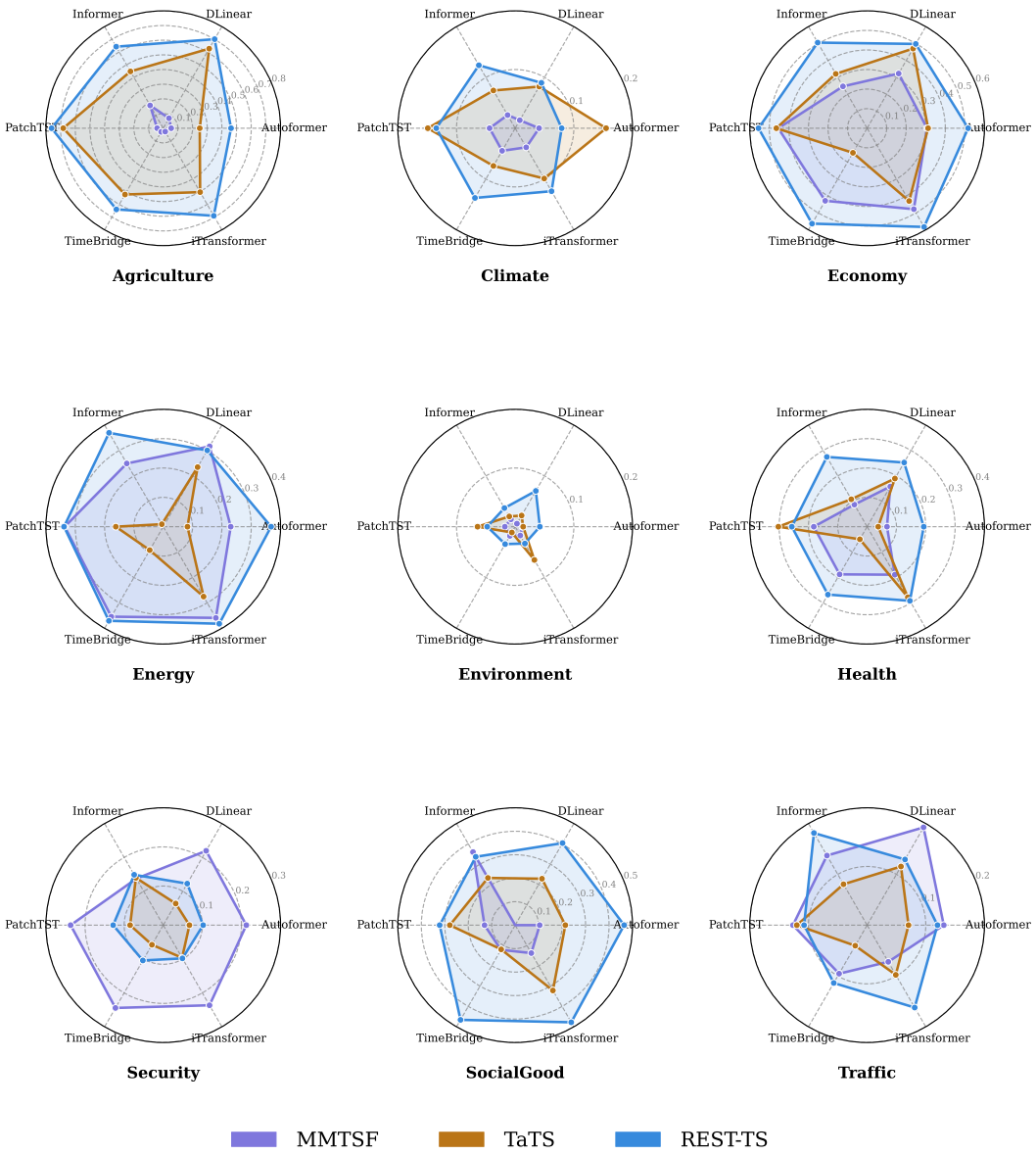


Figure 5: CKA between text representations and forecasting targets on nine Time-MMD domains and six backbones. Higher values indicate stronger alignment.

8.5 Comparison with a Modality-Balancing Baseline

OGM-GE rescales modality gradients during training (Peng et al., 2022). We include it to test whether gradient balancing alone addresses the low-rank text representations. Table 22 reports effective rank and CKA. Table 23 and Table 24 report forecasting error. OGM-GE has mean effective ranks of 1.008 with PatchTST and 1.044 with iTransformer. The corresponding REST-TS values are 1.862 and 2.105. REST-TS also has lower MSE and MAE in every reported domain. This comparison suggests that gradient rescaling is not enough to reproduce the results of residual supervision in these settings.

Table 22: Effective rank and CKA for OGM-GE and the three forecasting methods. **Bold** marks the best result in each row and metric.

Backbone	Dataset	Effective Rank				CKA			
		MMTSF	TaTS	OGM-GE	**REST-TS**	MMTSF	TaTS	OGM-GE	**REST-TS**
PatchTST	Agriculture	1.033	1.082	1.001	**1.358**	0.346	0.677	0.122	**0.761**
	Climate	1.026	1.069	1.003	**1.394**	0.083	**0.138**	0.017	0.134
	Economy	1.027	1.059	1.003	**1.272**	0.459	0.438	0.254	**0.555**
	Energy	1.048	1.212	1.002	**2.842**	**0.340**	0.151	0.313	0.338
	Environment	1.829	2.624	1.042	**2.858**	0.017	**0.052**	0.017	0.024
	Health	1.014	1.220	1.001	**2.741**	0.183	**0.277**	0.235	0.257
	Security	1.010	1.042	1.001	**1.159**	0.233	0.092	**0.269**	0.128
	SocialGood	1.101	1.095	1.017	**1.648**	0.229	0.256	0.309	**0.322**
	Traffic	1.017	1.077	1.001	**1.482**	**0.130**	0.122	0.079	0.107
	Avg	1.123	1.276	1.008	**1.862**	0.224	0.245	0.179	**0.292**
iTransformer	Agriculture	1.026	1.065	1.002	**1.271**	0.231	0.517	0.111	**0.689**
	Climate	1.030	1.064	1.003	**1.331**	0.069	0.102	0.027	**0.124**
	Economy	1.033	1.057	1.110	**1.269**	0.488	0.425	0.498	**0.583**
	Energy	1.031	1.236	1.001	**2.893**	**0.355**	0.261	0.314	0.326
	Environment	1.150	2.669	1.026	**4.845**	0.018	**0.055**	0.018	0.017
	Health	1.010	1.254	1.002	**3.116**	0.185	0.258	0.238	**0.292**
	Security	1.009	1.044	1.001	**1.198**	0.232	0.094	**0.257**	0.098
	SocialGood	1.106	1.093	1.253	**1.603**	0.170	0.308	0.359	**0.478**
	Traffic	1.023	1.068	1.001	**1.417**	0.087	0.102	0.092	**0.162**
	Avg	1.046	1.283	1.044	**2.105**	0.204	0.236	0.213	**0.308**

Table 23: MSE for REST-TS and OGM-GE on Time-MMD, averaged over forecast horizons. **Bold** marks the lower error in each backbone and domain setting.

Backbone	Method	Agri.	Clim.	Eco.	Ener.	Envi.	Heal.	Sec.	Soc.	Traf.
PatchTST	OGM-GE	0.165	1.004	0.120	0.320	0.346	1.406	126.48	1.184	0.291
	REST-TS (ours)	**0.115**	**0.995**	**0.0090**	**0.258**	**0.272**	**1.272**	**109.21**	**0.974**	**0.178**
	Promotion	+30.5%	+0.9%	+92.5%	+19.3%	+21.4%	+9.5%	+13.7%	+17.8%	+38.9%
iTransformer	OGM-GE	0.179	1.028	0.087	0.313	0.308	1.534	129.12	1.210	0.427
	REST-TS (ours)	**0.113**	**1.022**	**0.0084**	**0.251**	**0.259**	**1.275**	**109.73**	**0.981**	**0.182**
	Promotion	+36.8%	+0.6%	+90.3%	+19.8%	+15.8%	+16.9%	+15.0%	+18.9%	+57.4%

Table 24: MAE for REST-TS and OGM-GE on Time-MMD, averaged over forecast horizons. **Bold** marks the lower error. *Promotion* is the percentage reduction from OGM-GE to REST-TS.

Backbone	Method	Agri.	Clim.	Eco.	Ener.	Envi.	Heal.	Sec.	Soc.	Traf.
PatchTST	OGM-GE	0.305	0.799	0.246	0.424	0.426	0.805	6.275	0.561	0.342
	REST-TS	**0.234**	**0.786**	**0.079**	**0.366**	**0.372**	**0.744**	**4.910**	**0.450**	**0.225**
	Promotion	+23.3%	+1.6%	+67.9%	+13.6%	+12.6%	+7.6%	+21.7%	+19.8%	+34.3%
iTransformer	OGM-GE	0.309	0.802	0.235	0.417	0.404	0.818	6.420	0.541	0.483
	REST-TS	**0.230**	**0.798**	**0.078**	**0.361**	**0.369**	**0.740**	**5.019**	**0.443**	**0.222**
	Promotion	+25.5%	+0.5%	+66.9%	+13.5%	+8.7%	+9.6%	+21.8%	+18.0%	+54.0%

9 Full Analysis

9.1 Gradient Analysis Across Domains

Figure 6 shows the gradient ratio for iTransformer, and Figure 7 shows it for PatchTST. The ratio is $\|\nabla_{\text{text}}\|/\|\nabla_{\text{num}}\|$. A value of 1 means that the two gradient norms are equal. For MMTSF, the ratio falls well below 1 early in training on most domains. For REST-TS, it stays closer to or above 1. These results show how the two objectives distribute gradient magnitude, but gradient size alone does not establish semantic text use.

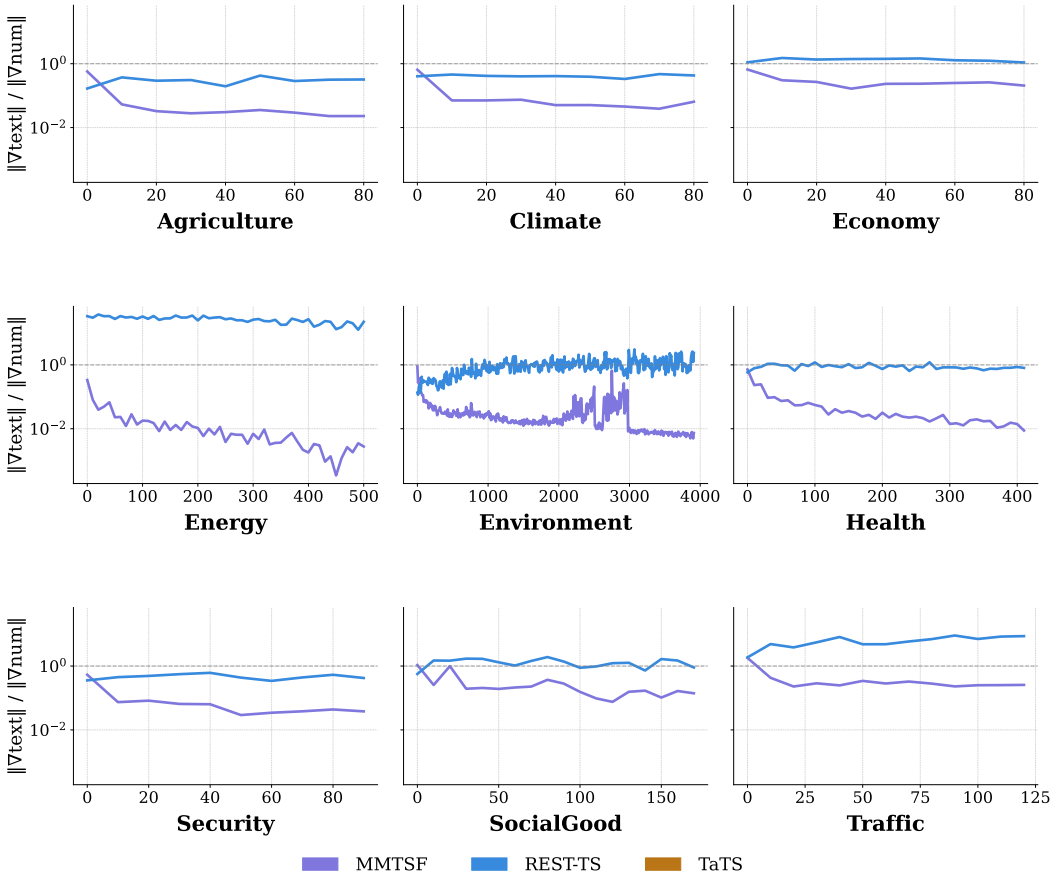


Figure 6: Text-to-numerical gradient-norm ratio during training on nine Time-MMD domains with iTransformer. The dashed line marks equal gradient norms.

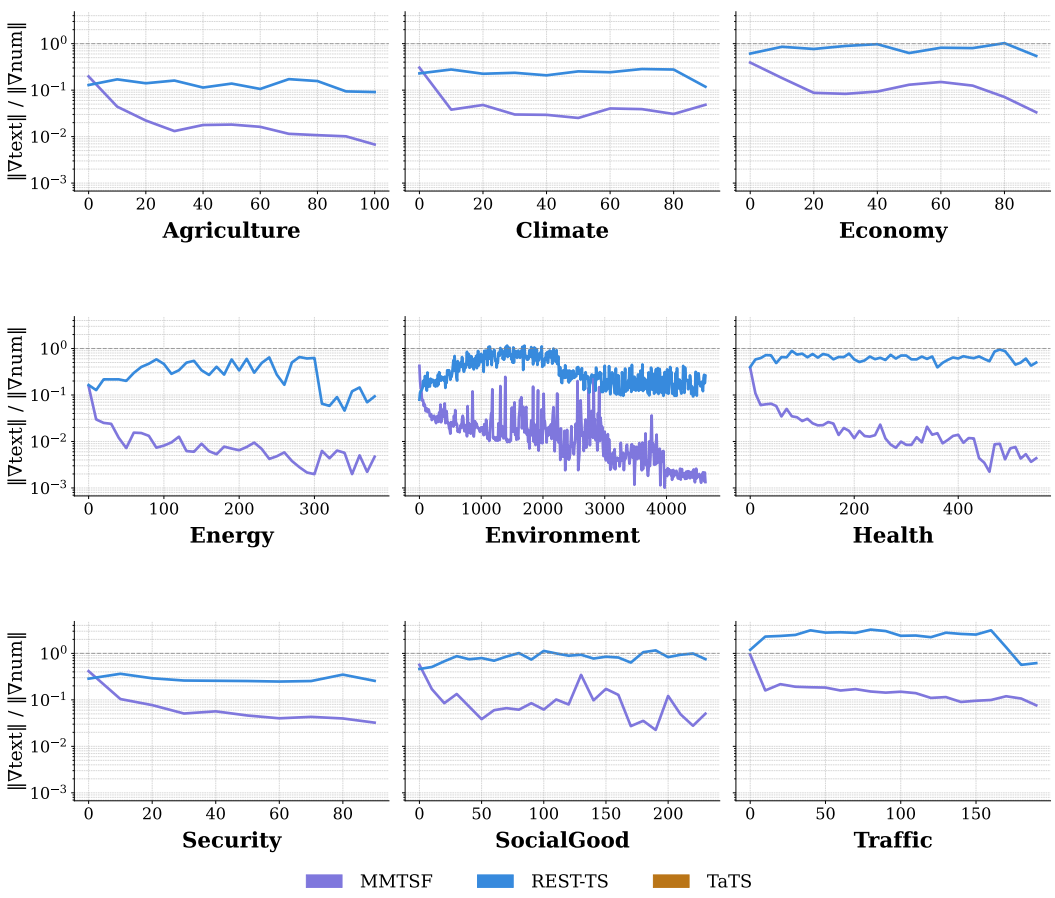


Figure 7: Text to numerical gradient norm ratio during training on nine Time-MMD domains with PatchTST. The dashed line marks equal gradient norms.

9.2 Learning Rate Sensitivity Across Domains

Figure 8, Figure 9, and Figure 10 report effective rank, CKA, and MSE at $\mathrm{lr}_2 \in \{10^{-4}, 10^{-3}, 10^{-2}\}$. Together, the figures cover all nine Time-MMD domains with iTransformer. REST-TS changes less across these three values than MMTSF and TaTS in most panels. This is an empirical sensitivity result within the tested range, rather than evidence that the learning rate is unimportant in general.

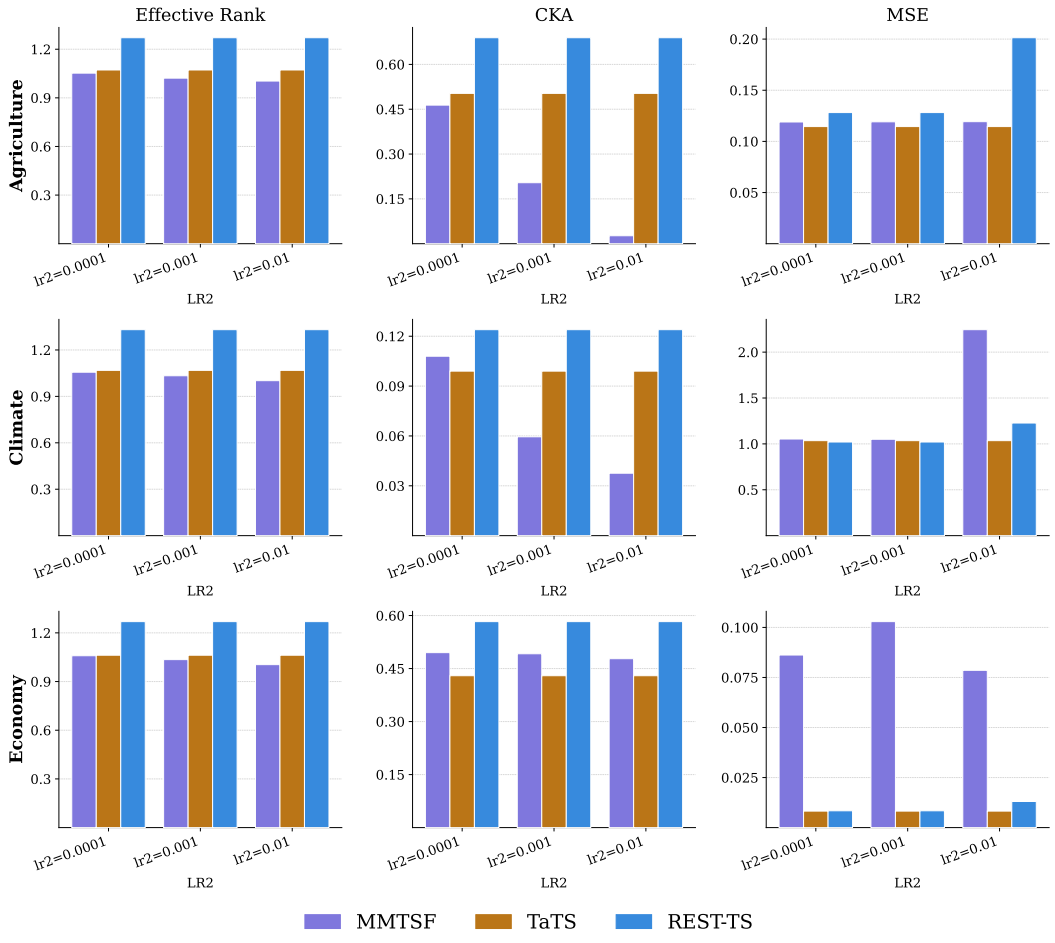


Figure 8: Learning-rate sensitivity on Agriculture, Climate, and Economy with iTransformer. The panels report effective rank, CKA, and MSE.

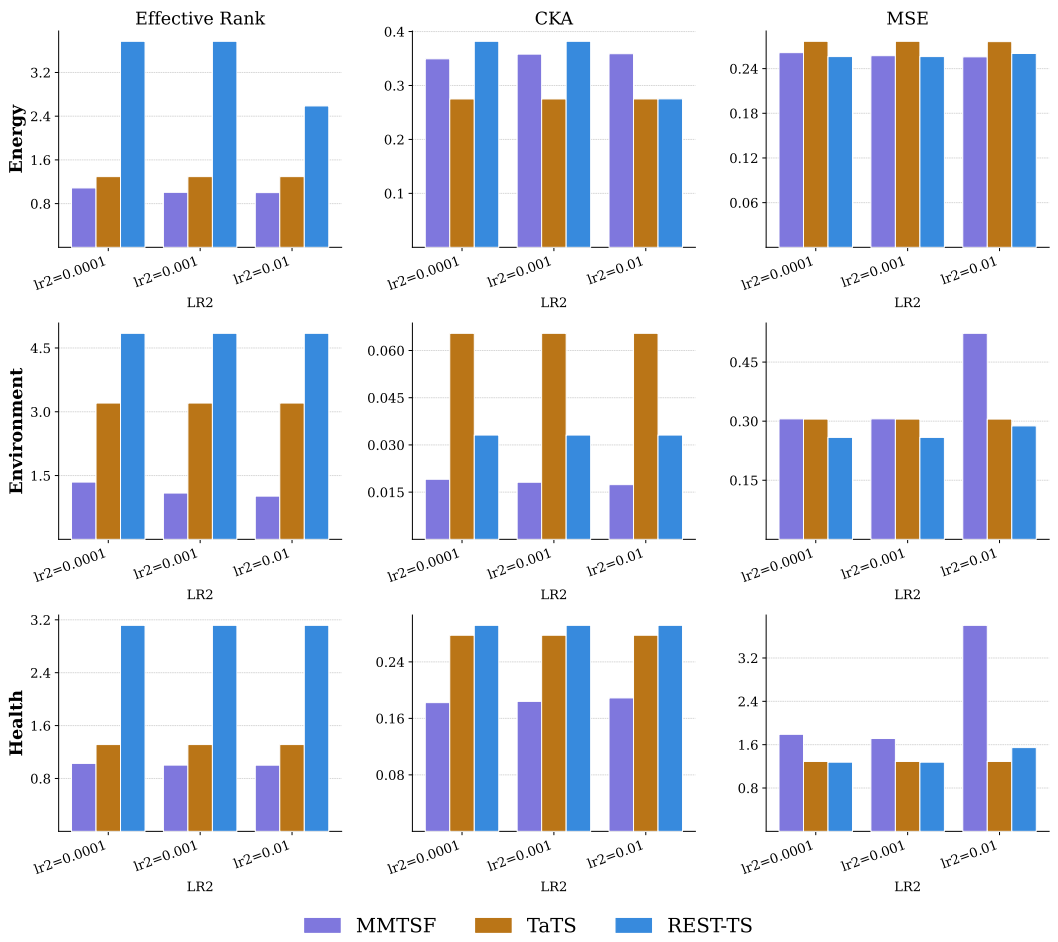


Figure 9: Learning-rate sensitivity on Energy, Environment, and Health with iTransformer. The panels report effective rank, CKA, and MSE.

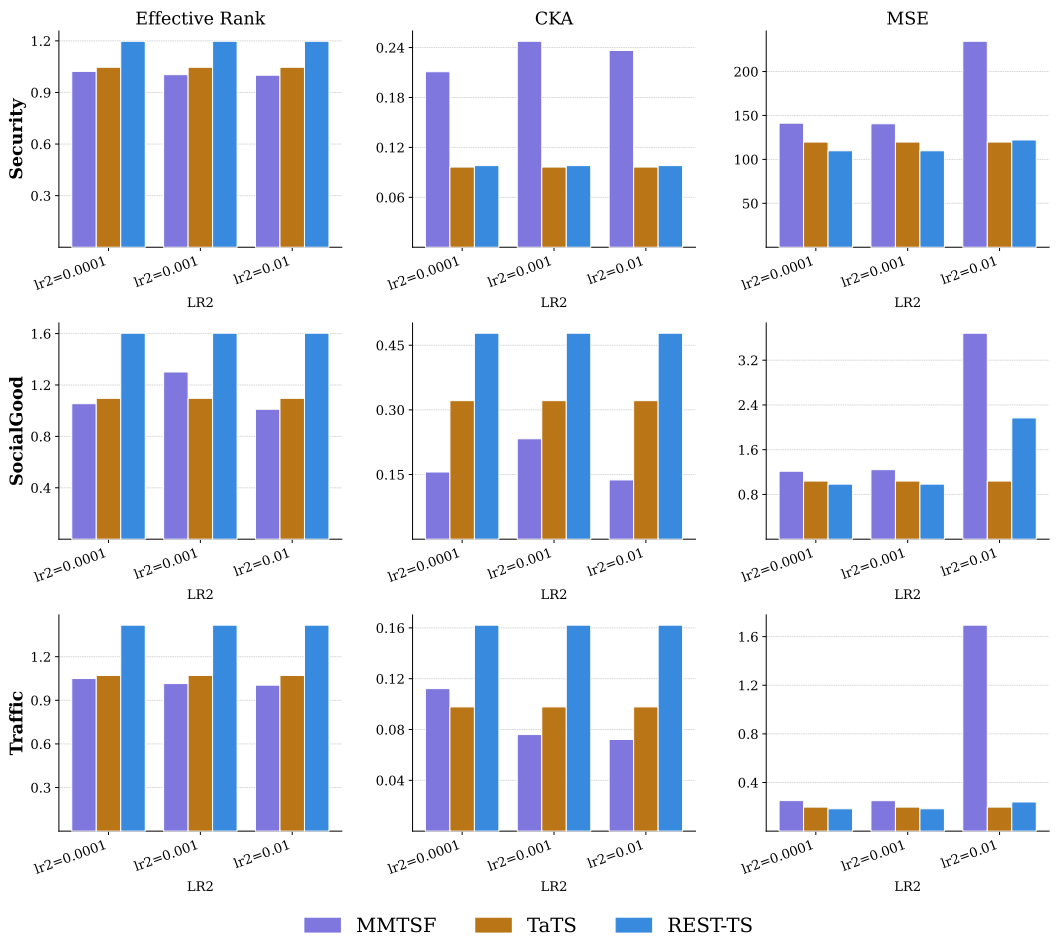


Figure 10: Learning-rate sensitivity on Security, SocialGood, and Traffic with iTransformer. The panels report effective rank, CKA, and MSE.

9.3 Prediction Divergence Analysis

Figure 11 reports prediction divergence for Informer. Figure 12 reports it for PatchTST. We define divergence as

$$\delta = \frac{\mathrm{MSE}(\hat{\mathbf{y}}_{\mathrm{real}}, \hat{\mathbf{y}}_{\mathrm{no_info}})}{\mathrm{MSE}(\hat{\mathbf{y}}_{\mathrm{real}}, \mathbf{y})}, \tag{16}$$

where $\hat{\mathbf{y}}_{\mathrm{real}}$ is the prediction with the report and $\hat{\mathbf{y}}_{\mathrm{no_info}}$ is the prediction without it. The denominator normalizes the change by the model's forecasting error. A value near zero means that this specific removal has little effect on the prediction. It does not show whether the text would matter under other interventions. The figures use a common text branch learning rate, while Table 21 uses the published setting for each method. The two sets of values should therefore not be compared directly. REST-TS has the largest divergence in most, but not all, displayed settings.

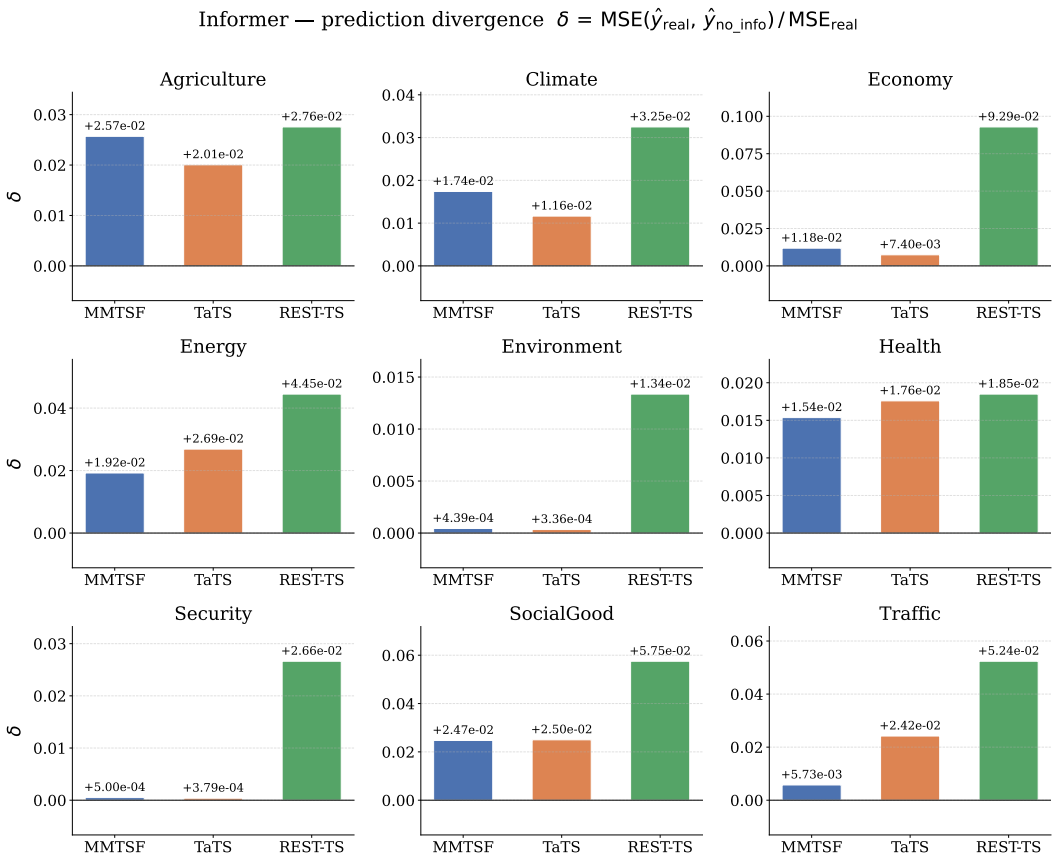


Figure 11: Prediction divergence $\delta = \mathrm{MSE}(\hat{\mathbf{y}}_{\mathrm{real}}, \hat{\mathbf{y}}_{\mathrm{no_info}})/\mathrm{MSE}_{\mathrm{real}}$ on nine Time-MMD domains with Informer. Higher values indicate a larger change when the text is removed.

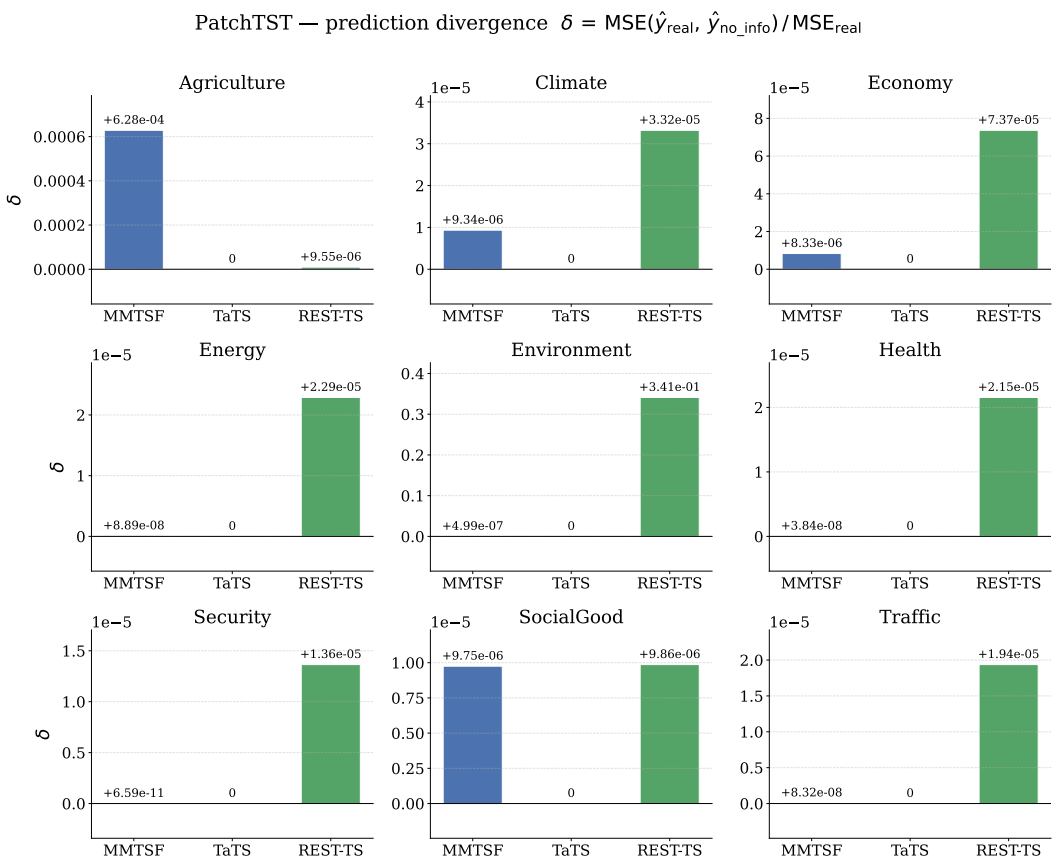


Figure 12: Prediction divergence δ across all nine Time-MMD domains with PatchTST. Higher values indicate a larger change when the text is removed.

9.4 Hyperparameter Sensitivity

We vary the moving-average width w from Eq. (7) and the Fourier component count K from Eq. (8). The experiments cover all nine Time-MMD domains with Informer and TimeBridge.

Moving-average width. Figure 13 reports MAE for $w \in \{2, 4, 8, 12, 16, 24\}$. The changes are small within each backbone, and the best value varies by domain. This supports selecting w with validation data.

Fourier component count. Figure 14 reports MAE for $K \in \{1, 2, 3, 4\}$. TimeBridge changes very little across these values. Informer also changes little on most domains, with larger variation on Energy and SocialGood. No value is best on every domain.

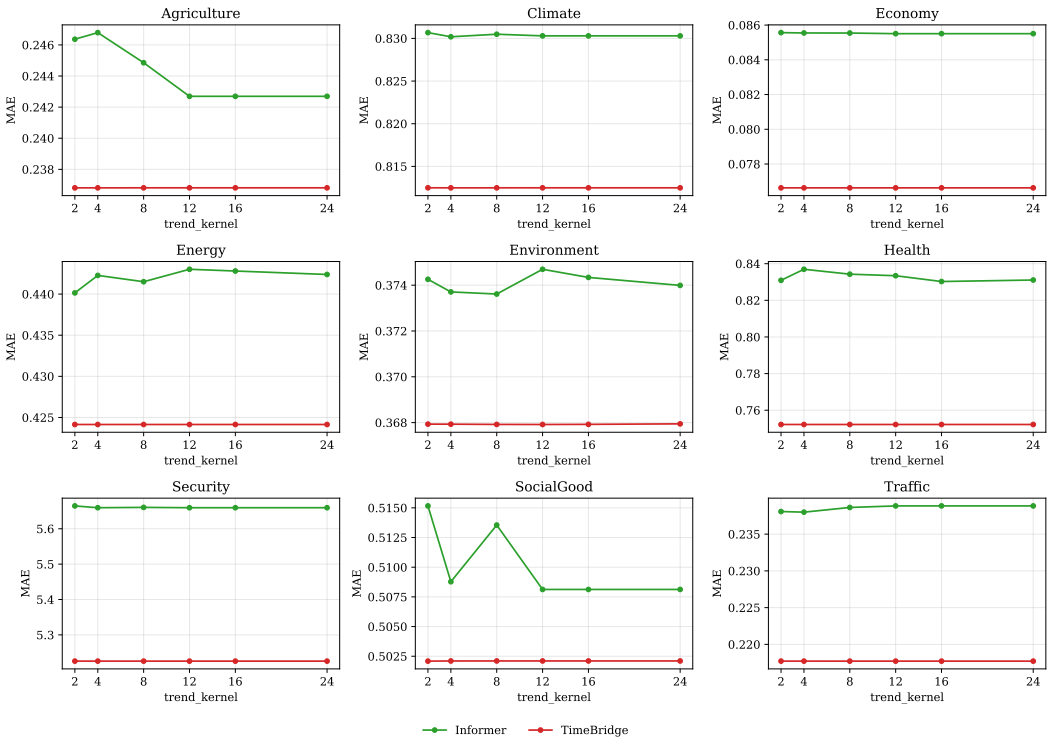


Figure 13: REST-TS MAE across moving-average widths on nine Time-MMD domains. Informer is green and TimeBridge is red.

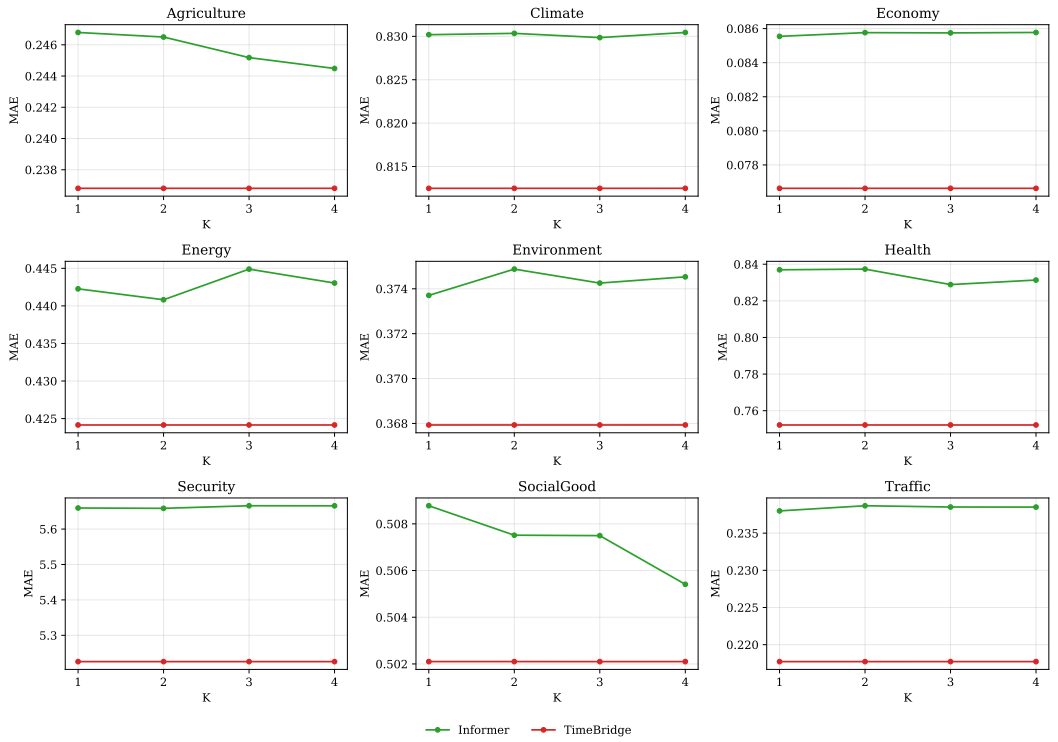


Figure 14: REST-TS MAE across Fourier component counts on nine Time-MMD domains. Informer is green and TimeBridge is red.

9.5 Residual Decomposition on Time-MMD

Figure 15 applies the decomposition of Section 4.3 to PatchTST residuals at the default setting (w=4, K=1) on all nine Time-MMD domains. Each panel shows the test sample with the largest absolute residual in that domain.

Two properties are visible. First, the split between trend and event varies widely across domains, from 97.9% trend on Energy to 70.6% event on Security, so the two text heads carry very different shares of the supervision depending on the domain. Second, the noise component is small everywhere, between 0.00% and 0.13% of residual variance. At K=1 a single low-amplitude bin carries little energy, and this holds at long horizons as well: Environment has 192 steps and 97 frequency bins yet still assigns 0.00% to noise. The noise term therefore acts as a safeguard against supervising the text branch on unstructured variation rather than as a component that removes a substantial part of the residual.

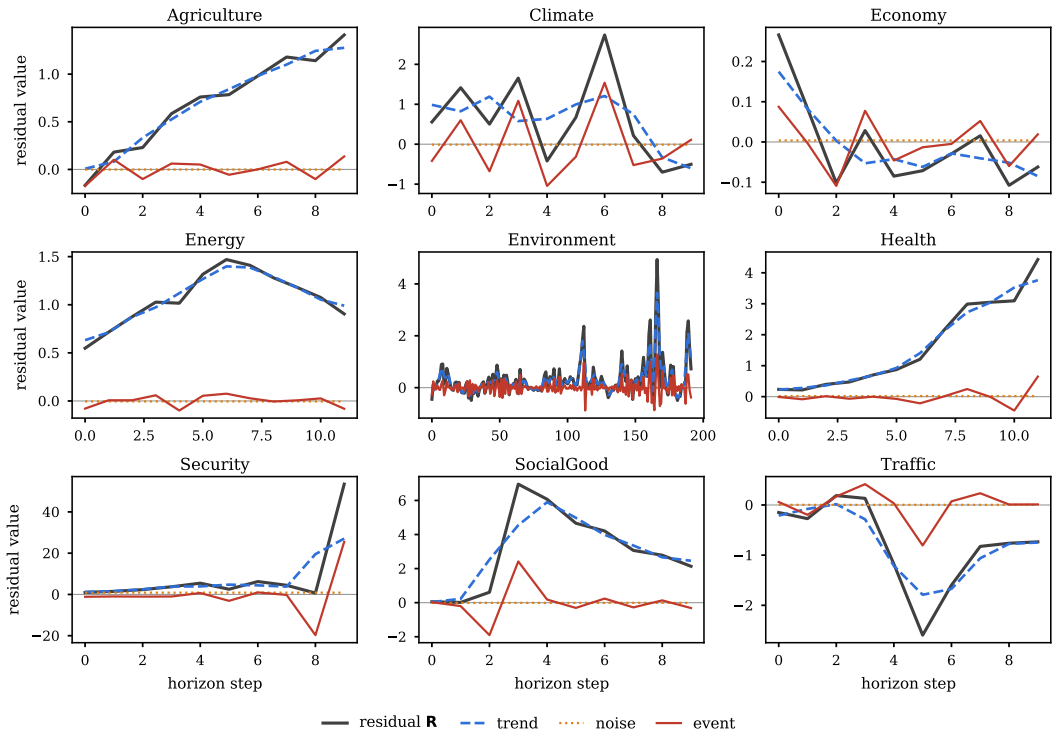


Figure 15: Residual decomposition on all nine Time-MMD domains, PatchTST at w=4, K=1. Each panel shows the residual $\mathbf{R}$ and its trend, noise and event components for the most eventful test sample of that domain. Only trend and event supervise the text branch.

9.6 Prediction Visualisation

Figure 16 shows example predictions from iTransformer, and Figure 17 shows examples from PatchTST. Each panel includes the input history, the ground truth, and the three model predictions. The dashed line marks the start of the forecast window. These examples help show the shape of the errors, but the aggregate results in Table 1 provide the main comparison.

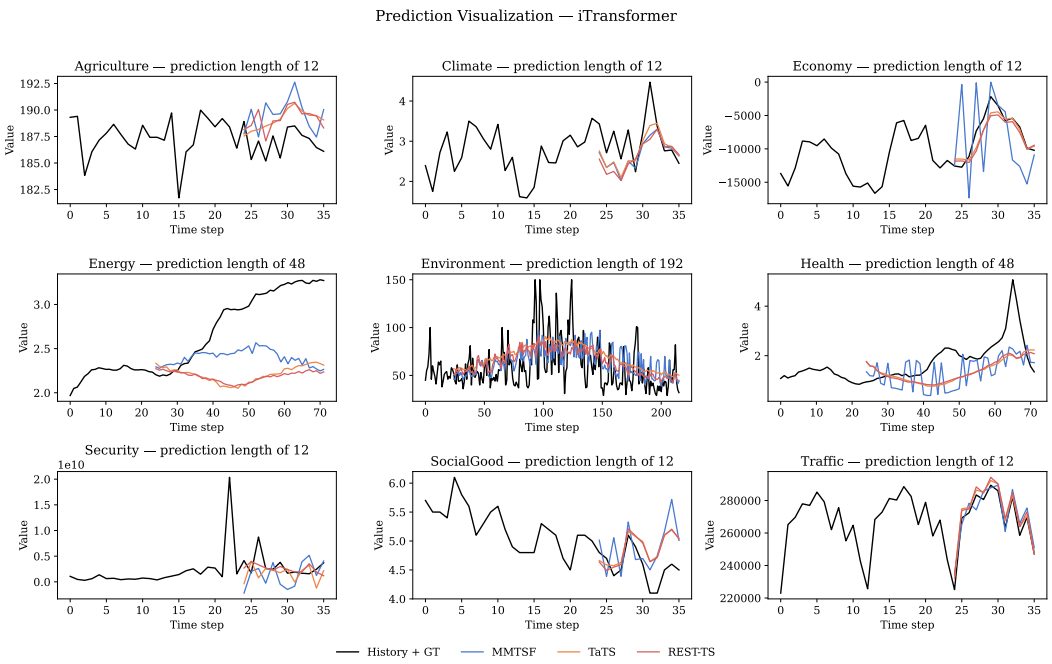


Figure 16: Example predictions on nine Time-MMD domains with iTransformer.

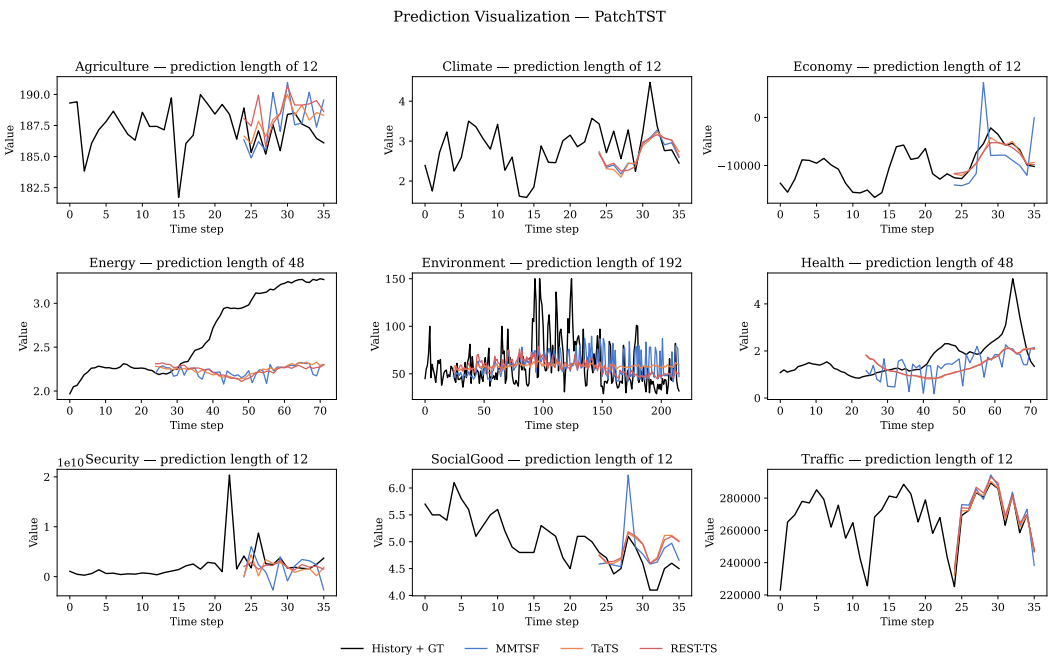


Figure 17: Example predictions on nine Time-MMD domains with PatchTST.

10 Statistical Robustness and Sensitivity Analyses

10.1 Multi-seed Ablation Diagnostics

Table 25 compares the basic model, residual decomposition without the target network, and the full model. The full model has the highest pooled effective rank and the lowest grand mean MSE. The MSE difference is small, so the clearer change is in representation diversity.

Table 25: Ablation results on five domains, two backbones, and all forecast horizons. MSE uses three seeds. Effective rank and CKA use the models from seed 2025.

Variant	Eff. Rank (pooled)	CKA	Eff. Rank, Health / Environment	MSE (grand mean)
no Decomp, no Target Net	1.71	0.332	1.28 / 3.82	0.5354
+ Residual Decomp only	1.72	0.316	1.16 / 4.05	0.5360
Full REST-TS	**3.23**	**0.344**	**2.93 / 9.23**	**0.5344**

10.2 Ablation with Variance

Table 4 in Section 5.4 reports point estimates on three domains. This section repeats the component ablation over all nine Time-MMD domains with three seeds (2025, 1, 2), two backbones and all forecast horizons, so that the differences can be read against the run-to-run variation. Tables 26 and 27 give the standard deviation across seeds of the horizon-averaged MSE and MAE, respectively.

Table 26: Component ablation with three seeds, MSE. Each entry is the standard deviation across seeds of the MSE averaged over the forecast horizons.

Backbone	Components		MSE standard deviation								
	Residual Decomp.	Target Network	Agri.	Clim.	Econ.	Ener.	Envi.	Heal.	Sec.	Soc.	Traf.
PatchTST	✗	✗	0.001	0.005	0.0001	0.001	0.000	0.001	0.3	0.008	0.000
	✓	✗	0.003	0.002	0.0001	0.001	0.000	0.004	0.1	0.015	0.000
	✗	✓	0.002	0.004	0.0001	0.001	0.000	0.005	0.1	0.015	0.000
	✓	✓	0.002	0.004	0.0001	0.001	0.000	0.005	0.1	0.015	0.000
iTransformer	✗	✗	0.003	0.002	0.0001	0.002	0.000	0.004	0.1	0.008	0.003
	✓	✗	0.003	0.007	0.0001	0.003	0.000	0.007	0.3	0.012	0.004
	✗	✓	0.003	0.006	0.0001	0.002	0.000	0.007	0.5	0.012	0.004
	✓	✓	0.003	0.006	0.0001	0.002	0.000	0.007	0.5	0.012	0.004

Table 27: Component ablation with three seeds, MAE. Each entry is the standard deviation across seeds of the MAE averaged over the forecast horizons.

Backbone	Components		MAE standard deviation								
	Residual Decomp.	Target Network	Agri.	Clim.	Econ.	Ener.	Envi.	Heal.	Sec.	Soc.	Traf.
PatchTST	✗	✗	0.003	0.002	0.000	0.001	0.000	0.003	0.015	0.002	0.001
	✓	✗	0.002	0.001	0.001	0.001	0.001	0.002	0.004	0.005	0.000
	✗	✓	0.003	0.002	0.001	0.001	0.000	0.002	0.013	0.006	0.001
	✓	✓	0.003	0.002	0.001	0.001	0.001	0.002	0.013	0.006	0.001
iTransformer	✗	✗	0.008	0.001	0.000	0.002	0.000	0.001	0.012	0.008	0.003
	✓	✗	0.008	0.005	0.001	0.002	0.000	0.002	0.017	0.003	0.002
	✗	✓	0.008	0.005	0.001	0.002	0.000	0.002	0.028	0.004	0.002
	✓	✓	0.008	0.005	0.001	0.002	0.000	0.002	0.028	0.004	0.002

10.3 SEED-LEVEL SIGNIFICANCE

Table 28 reports the three-seed comparison used in Section 5.2. REST-TS has lower mean MSE on Climate and Health. The differences on Economy and Agriculture are smaller and are not significant in these tests. With only three seeds, the estimates of variation remain uncertain.

Table 28: Three-seed MSE comparison. Values are means over forecast horizons with standard deviations. The last column gives a two-sample test of REST-TS and TaTS.

Domain	Backbone	REST-TS	TaTS	MMTSF	p (vs. TaTS)
Climate	PatchTST	0.992±.004	1.022±.012	1.026±.027	.017
Climate	iTransformer	1.021±.007	1.035±.005	1.047±.008	.052
Health	PatchTST	1.281±.008	1.338±.022	1.434±.025	.015
Health	iTransformer	1.283±.007	1.308±.003	1.617±.089	.005
Economy	PatchTST	0.0091±.0002	0.0089±.0003	0.1151±.0028	.264
Economy	iTransformer	0.0085±.0001	0.0084±.0001	0.1016±.0049	.239
Agriculture	PatchTST	0.119±.003	0.116±.002	0.115±.002	.159
Agriculture	iTransformer	0.115±.002	0.116±.005	0.119±.006	.658

10.4 TEXT-ENCODER SENSITIVITY

Table 29 compares frozen GPT-2, BERT-base, and GPT-2-Medium token embeddings on Economy and Health. REST-TS changes by at most 0.7% in these settings. The other methods show larger changes in some settings, particularly MMTSF on Economy with PatchTST.

Table 29: MSE with GPT-2, BERT-base, and GPT-2-Medium token embeddings on Economy and Health. Results use both backbones, all horizons.

Domain	Backbone	REST-TS	TaTS	MMTSF
Economy	iTransformer	0.0085 / 0.0084 / 0.0085	0.0084 / 0.0082 / 0.0084	0.1016 / 0.1047 / 0.1014
Economy	PatchTST	0.0081 / 0.0090 / 0.0091	0.0089 / 0.0092 / 0.0086	0.1151 / 0.0841 / 0.1146
Health	iTransformer	1.283 / 1.277 / 1.280	1.308 / 1.353 / 1.368	1.617 / 1.730 / 1.652
Health	PatchTST	1.281 / 1.274 / 1.273	1.338 / 1.363 / 1.303	1.434 / 1.422 / 1.409

11 LIMITATIONS

REST-TS has several limitations. First, we use frozen mean-pooled text embeddings, which may miss contextual or domain-specific semantics. Second, REST-TS assumes one text input per forecasting window and does not directly address settings with multiple or streaming text sources. Finally, the EMA target network introduces additional training cost, although it is removed at inference.

12 BROADER IMPACTS

Text-aware forecasting may be useful in areas such as public health, energy, and economics. It may also be affected by incorrect, delayed, or adversarial texts. Models used in these settings need human oversight, uncertainty estimates, and checks of their text sources. Effective rank, CKA, and forecast perturbations can support such audits, but they are not substitutes for domain evaluation. Our experiments use public datasets and one NVIDIA A100 GPU. REST-TS requires an additional target network during training, but not during inference.