

TransDyn: Transformer-Driven Modeling of Resolution-Invariant Financial Temporal Dynamics Under Sampling-Induced Distribution Shift

Siyu Zhang^{1,2}, Yujia Wu^{1,*}, Chang Zhou², Lianlei Shan³

¹Sanda University, Shanghai, China

²Viitrix AI Technology, Shanghai, China

³Tsinghua University, Beijing, China

zhangsiyu0066@gmail.com, wuyujia@whu.edu.cn, zhouchang@viitrix.com, shanlianlei18@mails.ucas.edu.cn

Abstract—Cross-frequency financial forecasting is challenging because temporal sampling and aggregation substantially reshape volatility structures and dependency patterns, resulting in sampling-induced temporal distribution shift. Existing methods mainly rely on multi-frequency joint training or implicit multi-scale modeling, while lacking explicit modeling of transferable temporal dynamics across heterogeneous sampling resolutions. To address this issue, we propose TransDyn, a Transformer-driven framework for robust cross-frequency financial forecasting. Trained only on single-asset, single-frequency 5-minute K-bar data, TransDyn directly generalizes to unseen daily-frequency assets without fine-tuning. Specifically, TransDyn collaboratively integrates Multi-Scale Temporal Extraction, Cross-Scale Interaction, Adaptive Temporal Modulation, and Information Bottleneck Regularization to learn resolution-invariant temporal dynamics, while auxiliary objectives further enhance cross-resolution consistency. Extensive experiments on real-world financial markets demonstrate that TransDyn consistently outperforms representative forecasting baselines in zero-shot cross-frequency and cross-asset forecasting. Qualitative analysis further shows that TransDyn effectively alleviates prediction collapse, unstable oscillations, and long-term temporal drift under severe distribution shift.

Index Terms—Financial time-series forecasting, Cross-frequency forecasting, Cross-asset generalization, Transformer, Temporal dynamics modeling

I. INTRODUCTION

Financial time-series forecasting plays a critical role in quantitative trading, risk management, portfolio optimization, and intelligent financial decision-making [1]. With the increasing availability of large-scale market data, learning robust temporal representations from financial observations has become an important research problem in machine learning and data mining [2]. Compared with general time-series scenarios, financial markets exhibit stronger stochasticity, temporal heterogeneity, and complex dependency structures, making robust temporal modeling particularly challenging for real-world quantitative applications [3].

Recent advances in Transformer architectures have substantially improved long-sequence temporal modeling through

self-attention-based global dependency learning [4]. Representative forecasting models, including Informer [5], Autoformer [6], FEDformer [7], PatchTST [8], TimesNet [9], and iTransformer [10], significantly enhance temporal representation learning through sparse attention, decomposition modeling, frequency-domain analysis, and multi-scale temporal modeling. More recently, time-series foundation models such as Chronos [11], TimeGPT [12], TimesFM [13], Timer [14], and Moirai [15] further extend temporal forecasting toward large-scale pretraining and zero-shot generalization. In financial domains, specialized foundation models such as Kronos [16] further demonstrate the potential of large-scale temporal representation learning for modeling complex market dynamics. Collectively, these methods substantially advance financial forecasting under fixed-resolution forecasting settings.

However, despite these advances, most existing methods still face critical limitations in real-world cross-frequency financial forecasting. Existing forecasting frameworks generally assume consistent temporal resolutions and relatively stable observation distributions between training and inference. In financial markets, however, temporal sampling and aggregation operations substantially reshape volatility patterns and temporal dependencies. High-frequency observations usually contain stronger market microstructure noise and rapid local fluctuations, whereas low-frequency observations emphasize smoother long-term trends and macro-level dependencies [17], [18].

Therefore, cross-frequency forecasting is not simply a multi-resolution representation learning problem, but a sampling-induced temporal distribution shift problem. As illustrated in Fig. 1, heterogeneous sampling operators transform a unified latent market process into resolution-specific observations with distinct temporal distributions, making direct transfer across frequencies unstable. The key challenge is thus to learn transferable temporal dynamics that remain stable across heterogeneous temporal resolutions.

Under such sampling-induced temporal distribution shift, existing forecasting methods still exhibit limited generalization capability. Most approaches mainly improve prediction

*Corresponding author: Yujia Wu. E-mail: wuyujia@whu.edu.cn

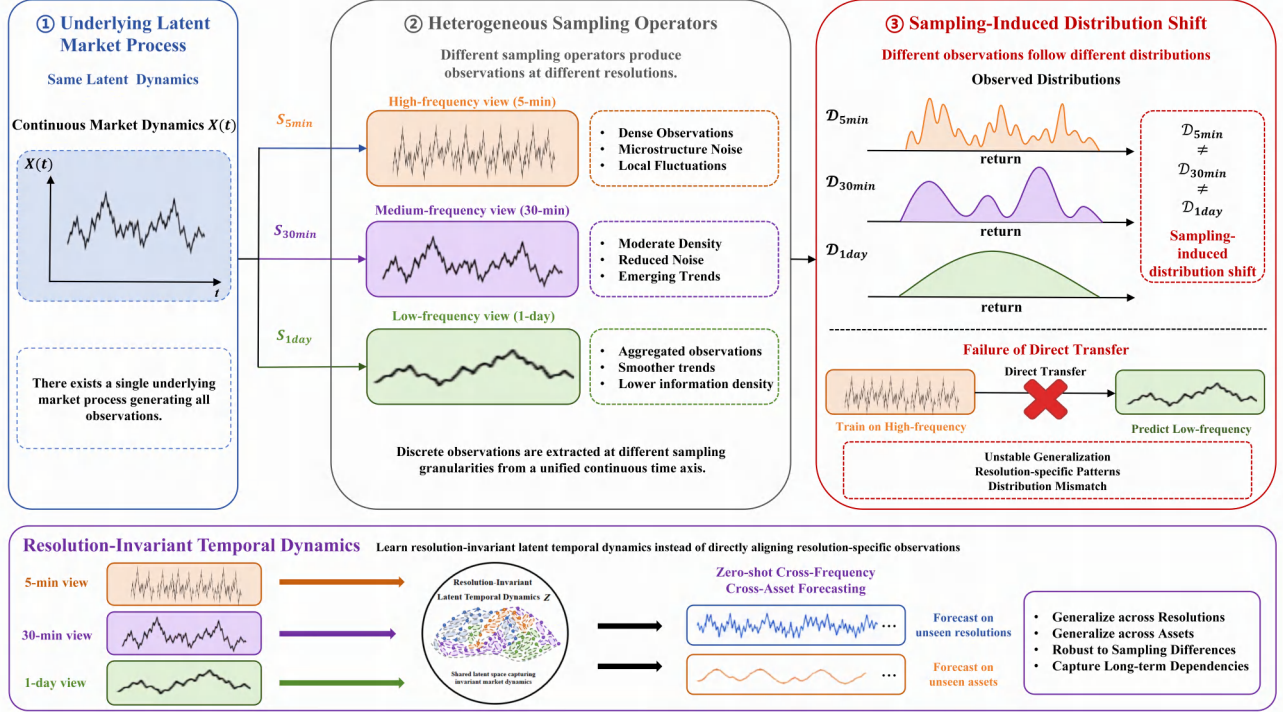


Fig. 1. **Motivation of cross-frequency cross-asset financial forecasting under sampling-induced temporal distribution shift.** Heterogeneous temporal sampling operators transform a unified latent market process into resolution-specific observations with distinct temporal distributions, making direct transfer across frequencies unstable. Instead of directly aligning heterogeneous observations, the proposed framework learns resolution-invariant temporal dynamics for zero-shot cross-frequency and cross-asset forecasting.

performance under fixed temporal resolutions, while current multi-resolution methods rely on joint training over predefined temporal scales and cross-scale interaction within observed resolutions [19]. Consequently, these methods tend to learn resolution-coupled temporal patterns and struggle to generalize toward unseen temporal granularities. Similarly, existing domain generalization methods assume that distribution shifts originate from external environments or semantic domains, which fundamentally differs from the sampling-induced structural temporal shift in cross-frequency financial forecasting [20].

To address these limitations, we propose **TransDyn**, a Transformer-driven framework for robust cross-frequency financial forecasting. Instead of directly aligning observations across resolutions, TransDyn learns transferable temporal dynamics through collaborative Multi-Scale Temporal Extraction, Cross-Scale Interaction, Adaptive Temporal Modulation, and Information Bottleneck Regularization, enabling effective zero-shot generalization across unseen temporal resolutions and unseen financial assets.

The main contributions of this paper are summarized as follows:

- We reformulate cross-frequency financial forecasting as a sampling-induced temporal distribution shift problem from a temporal generalization perspective.
- We propose TransDyn, a Transformer-driven framework that collaboratively models multi-scale, cross-scale,

volatility-aware, and bottleneck-regularized temporal dynamics across sampling resolutions.

- We design collaborative auxiliary objectives to enhance cross-resolution consistency while preserving forecasting-relevant temporal structures.
- Extensive experiments demonstrate that TransDyn achieves strong zero-shot cross-frequency cross-asset forecasting capability across unseen temporal resolutions and unseen financial assets.

II. RELATED WORK

A. Transformer-based Temporal Modeling

Early deep forecasting models mainly rely on recurrent neural networks (RNNs) [21], long short-term memory networks (LSTMs) [22], and temporal convolutional networks (TCNs) [23], which capture temporal dependencies through recurrent state propagation or causal convolution operations. However, such methods are inherently constrained by local temporal inductive biases and often struggle to capture long-range temporal interactions under complex and non-stationary financial dynamics.

With the advancement of Transformer architectures, self-attention-based global dependency modeling has substantially promoted the development of time-series forecasting [24], [25]. Representative models, including Informer [5], Autoformer [6], FEDformer [7], PatchTST [8], TimesNet [9], and iTransformer [10], improve long-sequence forecasting

through sparse attention, decomposition modeling, frequency-domain representation, patch-wise tokenization, and variable-centric interaction. In particular, PatchTST enhances temporal representation learning through patch segmentation, while iTransformer reformulates forecasting from a variable-centric perspective. Despite their effectiveness under fixed-resolution settings, these methods learn representations strongly coupled with predefined sampling granularities, limiting their robustness across heterogeneous temporal resolutions.

Recently, Transformer-based time-series foundation models (TSFMs), including Chronos [11], TimeGPT [12], TimesFM [13], Timer [14], and Moirai [15], have further advanced time-series forecasting toward large-scale pretraining and zero-shot transfer learning. In financial scenarios, domain-specific foundation models such as Kronos [16] further demonstrate the effectiveness of large-scale temporal pretraining for modeling market volatility evolution and long-range dependency structures. Nevertheless, most existing forecasting models and TSFMs still follow fixed-resolution assumptions, where training and inference data are required to maintain consistent temporal granularities or are naively aligned through simple resampling operations [26]. As a result, most existing forecasting frameworks remain strongly dependent on fixed-resolution temporal observations.

B. Multi-Resolution and Cross-Frequency Temporal Modeling

Existing resolution-aware temporal modeling methods can generally be categorized into multi-resolution representation learning and cross-frequency forecasting [19], [27]. Multi-resolution methods mainly focus on jointly modeling temporal patterns at different scales within the same forecasting task, while cross-frequency forecasting concerns whether temporal dynamics learned under one temporal resolution can generalize to another sampling frequency.

Traditional multi-resolution approaches primarily rely on wavelet transformation, frequency decomposition, and multi-scale convolution operations to capture hierarchical temporal structures. More recently, Transformer-based architectures including Pyraformer [28], Crossformer [29], Scaleformer [30], and MICN [31] further improve multi-scale temporal modeling through hierarchical cross-scale interaction and feature aggregation mechanisms. Furthermore, recent cross-frequency forecasting methods, such as CFPT [32], FreqMTA [33], and frequency-domain recurrent attention frameworks [34], improve frequency adaptation through explicit cross-frequency interaction and frequency-aware temporal representation learning.

Although these approaches substantially improve multi-scale temporal modeling, the learned representations remain strongly coupled with observed temporal resolutions and therefore struggle to generalize toward unseen heterogeneous sampling granularities.

C. Transferability and Robust Temporal Dynamics under Distribution Shift

Distribution shift robustness and domain generalization have been extensively studied in machine learning [35]. Existing approaches mainly improve out-of-distribution generalization through feature alignment, adversarial adaptation, invariant representation learning, and distributionally robust optimization. Representative methods such as CORAL [36], DANN [37], IRM [38], and DomainBed [39] demonstrate strong transferability across heterogeneous domains and external environments.

However, cross-frequency financial forecasting fundamentally differs from conventional domain generalization scenarios. Traditional domain shifts mainly originate from external environmental variations, whereas cross-frequency shifts in financial time series are induced by temporal sampling and aggregation operations. Different temporal granularities substantially reshape temporal characteristics and dependency patterns across resolutions. Consequently, conventional domain alignment and invariant representation learning strategies cannot effectively address sampling-induced structural temporal shifts.

Therefore, learning transferable temporal dynamics under sampling-induced temporal distribution shift remains a fundamental challenge in cross-frequency financial forecasting, which directly motivates the proposed TransDyn framework.

III. METHOD

A. Problem Definition

In financial markets, asset prices can generally be regarded as a continuous-time stochastic process, denoted as $X(t)$, where t represents the continuous temporal index. However, in practical applications, market data can only be observed in discrete forms under different temporal resolutions. Different resolutions correspond to different sampling and aggregation mechanisms. Therefore, observations under different temporal resolutions are not merely downsampled versions of the same sequence, but are generated through different temporal partition and aggregation operations.

Formally, given a temporal resolution r , we define a sampling operator $S_r(\cdot)$ that maps the continuous process into a discrete observation sequence:

$$X^{(r)} = S_r(X), \quad X^{(r)} \in \mathbb{R}^{T_r \times d}, \quad (1)$$

where $r \in \mathbb{R}^+$ denotes the sampling interval, T_r denotes the sequence length under resolution r , and d represents the feature dimension.

To further characterize the sampling mechanism, the sampling operator is formulated as a combination of temporal partition and interval aggregation:

$$S_r(X) = \{g(X[(i-1)r : ir])\}_{i=1}^{\lfloor T/r \rfloor}, \quad (2)$$

where $g(\cdot)$ denotes the interval aggregation function. For OHLCV observations, standard candlestick aggregation is

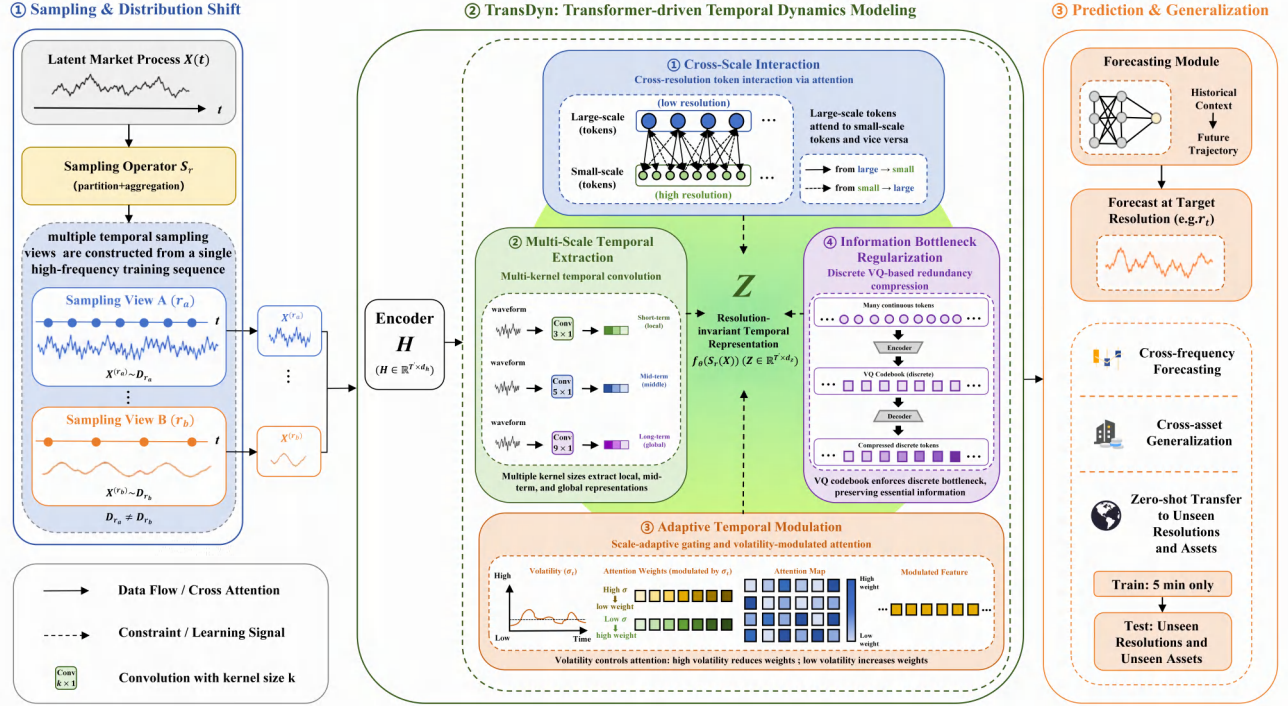


Fig. 2. **Overview of the proposed TransDyn framework.** TransDyn learns resolution-invariant temporal dynamics from heterogeneous sampling resolutions through four collaborative functional components: (1) Multi-Scale Temporal Extraction, (2) Cross-Scale Interaction, (3) Adaptive Temporal Modulation, and (4) Information Bottleneck Regularization. The learned temporal dynamics enable robust zero-shot cross-frequency forecasting and cross-asset generalization under sampling-induced distribution shift.

adopted, where Open, High, Low, Close, and Volume are aggregated using the first, maximum, minimum, last, and summation operations, respectively.

This formulation indicates that temporal resolution changes not only alter the sequence length, but also reshape the statistical structure of the observations through aggregation. Since different sampling intervals modify the noise structure, volatility patterns, and information density, observations under different resolutions generally follow different data distributions. Let D_r denote the data distribution under resolution r . For two different resolutions r_1 and r_2 , we generally have:

$$X^{(r_1)} \sim D_{r_1}, \quad X^{(r_2)} \sim D_{r_2}, \quad D_{r_1} \neq D_{r_2}. \quad (3)$$

This relationship suggests that cross-frequency forecasting is fundamentally a distribution-shift problem induced by sampling transformations rather than a simple sequence mapping task.

Based on this observation, instead of directly aligning observations under different resolutions, we aim to maintain stable temporal dynamics across heterogeneous sampling resolutions. As illustrated in Fig. 2, the proposed TransDyn framework progressively learns resolution-invariant temporal dynamics through collaborative temporal modeling modules together with auxiliary optimization objectives.

Formally, given two observations generated under different sampling operators, the learned temporal dynamics are expected to remain consistent across resolutions:

$$f_\theta(S_{r_1}(X)) \approx f_\theta(S_{r_2}(X)), \quad \forall r_1, r_2, \quad (4)$$

where $f_\theta(\cdot)$ denotes the temporal dynamics modeling function and “ $\approx$ ” denotes cross-resolution feature consistency.

Importantly, resolution invariance does not imply identical observations across sampling frequencies. Instead, we assume that forecasting-relevant latent dynamics, such as persistent trends and medium-to-long-term dependencies, remain partially shared after temporal aggregation, whereas high-frequency noise and local fluctuations are more resolution dependent. Therefore, TransDyn encourages consistency in the learned latent dynamics rather than enforcing observation-level equivalence.

Based on this formulation, cross-frequency forecasting can be viewed as a robust temporal modeling problem under sampling transformations.

The overall optimization objective can thus be formulated as:

$$\min_{\theta} \mathbb{E}_{X, r_1, r_2} [d(f_\theta(S_{r_1}(X)), f_\theta(S_{r_2}(X)))], \quad (5)$$

where $d(\cdot, \cdot)$ denotes a feature-space distance metric. In this work, cosine distance is adopted:

$$d(z_i, z_j) = 1 - \frac{z_i^\top z_j}{\|z_i\| \|z_j\|}. \quad (6)$$

It is important to emphasize that this work focuses on a strict zero-shot cross-frequency and cross-asset generalization setting, where the model is trained only on high-frequency data from a single asset and generalized to unseen assets and unseen resolutions without additional fine-tuning.

B. Resolution-Operator Formulation

According to the formulation above, temporal resolution changes can be interpreted as different sampling operators acting on the same latent market process. The sampling operator not only changes the sequence length, but also implicitly performs temporal partition and interval aggregation. Temporal partition determines the observation granularity, while aggregation compresses local temporal information within each interval.

Since interval aggregation is generally non-invertible, multiple distinct high-frequency trajectories may correspond to the same low-frequency observation. As a result, sampling transformations inevitably introduce irreversible information loss and alter the noise structure, volatility characteristics, and effective information density.

Consequently, observations under different resolutions can be modeled as distinct distributions induced by different sampling operators:

$$S_{r_1} : X \rightarrow D_{r_1}, \quad S_{r_2} : X \rightarrow D_{r_2}. \quad (7)$$

This formulation highlights that cross-frequency forecasting should not directly align raw observations across resolutions, but instead focus on maintaining robust temporal dynamics under varying sampling operators while preserving essential market structures.

Therefore, the proposed TransDyn framework aims to simultaneously satisfy two objectives:

- maintain temporal stability across different resolutions;
- preserve forecasting-relevant market dynamics.

To achieve this goal, TransDyn introduces collaborative temporal modeling modules together with auxiliary optimization objectives to progressively stabilize temporal dynamics under sampling-induced distribution shift.

C. TransDyn Framework

Based on the resolution-operator formulation above, we propose the TransDyn framework, as illustrated in Fig. 2. Specifically, TransDyn consists of several collaborative functional components, including temporal modeling modules and regularization-guided optimization components. Instead of directly learning mappings between sequences of different temporal resolutions, TransDyn progressively learns robust temporal dynamics through collaborative Multi-Scale Temporal Extraction, Cross-Scale Interaction, Adaptive Temporal Modulation, and Information Bottleneck Regularization.

Given an input sequence $X^{(r)}$ under resolution r , the model first extracts temporal features through an encoder:

$$H = \text{Encoder}_r(X^{(r)}), \quad H \in \mathbb{R}^{T' \times d_h}, \quad (8)$$

where T' denotes the number of encoded tokens and d_h denotes the hidden feature dimension.

The encoder adopts a Transformer-based architecture with multi-scale temporal extraction and cross-scale interaction mechanisms to jointly capture local temporal fluctuations and long-term dependencies, which are essential for learning resolution-invariant market dynamics across sampling-induced distribution shifts.

After feature extraction, TransDyn further performs cross-resolution temporal alignment through a shared projection module and contrastive interaction mechanism:

$$Z_{\text{align}} = \text{Align}(H), \quad Z_{\text{align}} \in \mathbb{R}^{T' \times d_z}, \quad (9)$$

where $\text{Align}(\cdot)$ denotes the cross-resolution alignment module and d_z denotes the latent feature dimension.

During training, multiple resolution views are constructed exclusively from the 5-minute training sequences using different sampling operators. Observations corresponding to the same temporal interval across these constructed views are treated as positive pairs, while observations from different intervals are treated as negative samples. This mechanism improves cross-resolution temporal consistency and alleviates sampling-induced statistical discrepancies.

However, financial time series exhibit strong temporal heteroscedasticity. High-volatility regions usually contain substantial short-term noise and stochastic perturbations. To suppress unstable fluctuations, TransDyn further introduces a volatility-aware temporal modulation mechanism:

$$Z_{\text{mod}} = \text{Modulate}(Z_{\text{align}}, \sigma_t), \quad (10)$$

where σ_t denotes the local volatility estimate.

Specifically, the local volatility is computed using the rolling standard deviation of logarithmic returns:

$$\sigma_t = \sqrt{\frac{1}{w} \sum_{i=1}^w (r_{t-i} - \bar{r})^2}, \quad (11)$$

where r_t denotes the logarithmic return at timestamp t , w denotes the rolling window size, and $\bar{r}$ denotes the local mean return.

The volatility-aware modulation adaptively suppresses unstable fluctuations in high-volatility regions, allowing the model to focus on more stable temporal structures shared across resolutions.

After volatility modulation, some resolution-dependent redundant details may still remain in the features. To further compress redundant structures, TransDyn introduces an information bottleneck module:

$$Z = \text{Bottleneck}(Z_{\text{mod}}), \quad Z \in \mathbb{R}^{T' \times d_b}, \quad (12)$$

where d_b denotes the bottleneck feature dimension.

The bottleneck module adopts vector quantization (VQ) to map continuous temporal features into a discrete codebook,

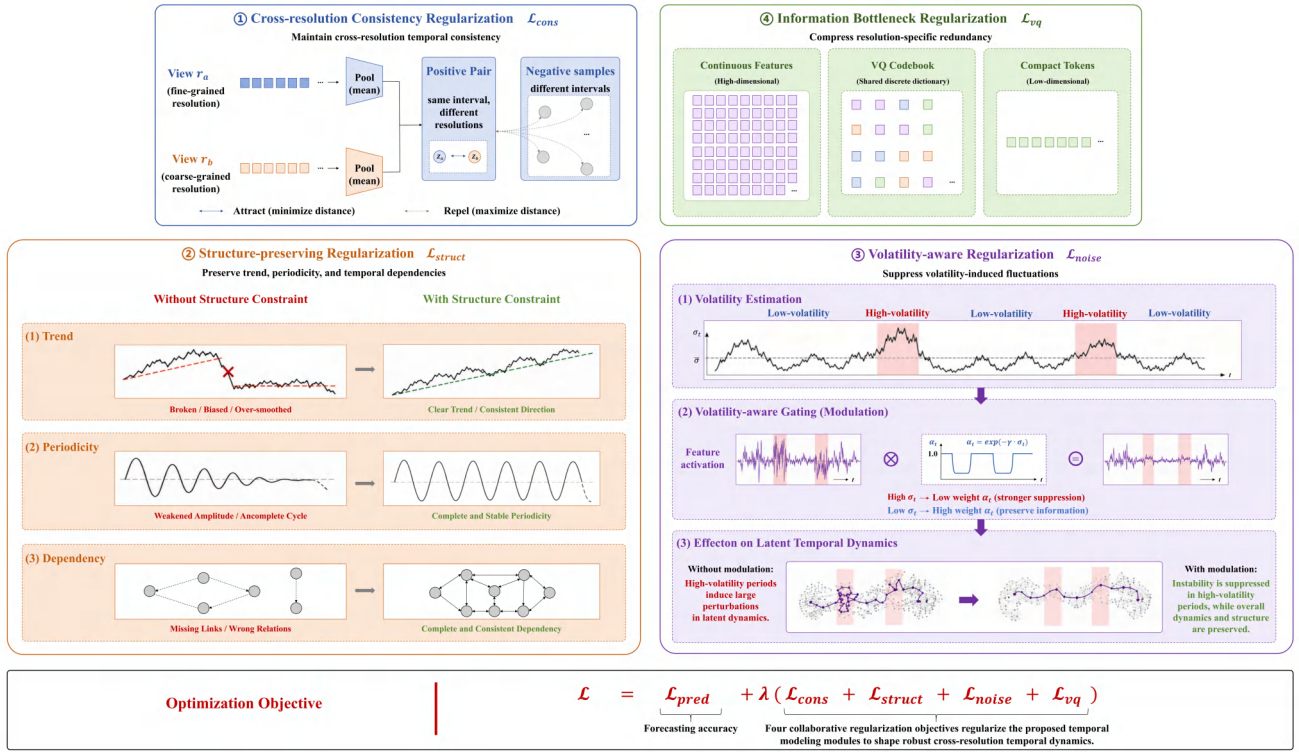


Fig. 3. **Auxiliary optimization objectives in TransDyn.** The overall optimization objective consists of a forecasting loss together with four collaborative regularization objectives for robust temporal modeling under sampling-induced temporal distribution shift, including cross-resolution consistency regularization, structure-preserving regularization, volatility-aware regularization, and information bottleneck regularization.

thereby suppressing resolution-specific redundancy while preserving shared market structures across resolutions and assets.

Overall, the proposed modules collaboratively stabilize temporal dynamics from different perspectives. The multi-scale temporal extraction captures hierarchical temporal patterns, the cross-resolution alignment alleviates sampling discrepancies, the volatility-aware modulation suppresses unstable perturbations, and the bottleneck module further compresses resolution-dependent redundancy.

D. Training Objective

To further stabilize temporal dynamics learning under sampling-induced distribution shift, TransDyn additionally employs several collaborative optimization objectives, as illustrated in Fig. 3.

The overall objective is formulated as:

$$\mathcal{L} = \mathcal{L}_{pred} + \lambda \mathcal{L}_{robust}, \quad (13)$$

where $\mathcal{L}_{pred}$ denotes the forecasting loss implemented using Mean Squared Error (MSE), $\mathcal{L}_{robust}$ denotes the robustness regularization objective, and λ is the balancing coefficient.

To model cross-resolution consistency, sequence-level global features are first constructed from token-level features. Given the temporal features of the i -th sample:

$$Z_i \in \mathbb{R}^{T' \times d_b}, \quad (14)$$

the global feature is obtained through temporal average pooling:

$$z_i = \text{Pool}(Z_i), \quad (15)$$

where $\text{Pool}(\cdot)$ denotes temporal mean pooling.

The cross-resolution consistency objective is defined as:

$$\mathcal{L}_{cons} = -\log \frac{\exp(\text{sim}(z_i, z_i^+)/\tau)}{\sum_{j=1}^N \mathbf{1}_{j \neq i} \exp(\text{sim}(z_i, z_j)/\tau)}, \quad (16)$$

where z_i^+ denotes the positive feature corresponding to the same temporal interval under a different resolution, z_j denotes features from other samples in the mini-batch, τ denotes the temperature coefficient, and $\text{sim}(\cdot, \cdot)$ denotes cosine similarity.

This objective maintains temporal consistency across resolutions by pulling positive pairs together while separating unrelated temporal states.

To further suppress sampling-induced perturbations, TransDyn additionally introduces collaborative regularization objectives for structure preservation, noise suppression, and redundancy compression:

$$\mathcal{L}_{robust} = \mathcal{L}_{cons} + \mathcal{L}_{struct} + \mathcal{L}_{noise} + \mathcal{L}_{vq}. \quad (17)$$

1) Structure-preserving Objective: To avoid destroying important temporal structures during temporal alignment and

feature compression, a structure-preserving objective is introduced:

$$\mathcal{L}_{struct} = \|\phi(H) - \phi(Z)\|_2^2, \quad (18)$$

where $\phi(\cdot)$ denotes a temporal structure extraction function such as autocorrelation estimation or temporal dependency mapping.

This objective preserves trends, periodicity, and multi-scale temporal structures.

2) *Volatility-aware Regularization*: Considering the heteroscedastic nature of financial time series, a volatility-aware regularization objective is introduced:

$$\mathcal{L}_{noise} = \|Z_{mod} - Z_{align} \odot g(\sigma_t)\|_2^2, \quad (19)$$

where $g(\sigma_t)$ denotes the volatility-based scaling function and $\odot$ denotes element-wise multiplication.

This objective suppresses unstable fluctuations in high-volatility regions and improves temporal stability.

3) *Information Bottleneck Objective*: To further remove resolution-dependent redundancy, the vector quantization objective is defined as:

$$\mathcal{L}_{vq} = \|sg[Z_{mod}] - e\|_2^2 + \beta \|Z_{mod} - sg[e]\|_2^2, \quad (20)$$

where $sg[\cdot]$ denotes the stop-gradient operation, e denotes the discrete codebook embedding, and β is the balancing coefficient.

Overall, the proposed objectives collaboratively regularize temporal dynamics from different perspectives. Specifically, the consistency objective maintains cross-resolution stability, the structure-preserving objective preserves forecasting-relevant temporal patterns, the volatility-aware regularization suppresses unstable perturbations, and the bottleneck objective compresses resolution-specific redundancy.

Unlike conventional approaches that directly fit mappings between different temporal resolutions, TransDyn progressively learns resolution-invariant latent temporal dynamics shared across heterogeneous sampling resolutions through collaborative temporal modeling and auxiliary optimization. Consequently, the proposed framework enables robust cross-frequency forecasting and strong zero-shot generalization across unseen temporal resolutions and unseen financial assets under sampling-induced distribution shift.

IV. EXPERIMENTS

In this section, we conduct extensive experiments on real-world financial time-series datasets to evaluate the effectiveness of our proposed TransDyn. We first introduce the experimental setup, followed by comprehensive zero-shot cross-frequency cross-asset performance comparisons with state-of-the-art forecasting methods. Furthermore, we perform detailed ablation studies on our proposed functional components and analyze computational efficiency to demonstrate the robustness and generalization capability of our framework.

A. Experimental Setup

a) Dataset and Cross-Frequency Evaluation Protocol:

Unlike conventional time series forecasting benchmarks that assume consistent temporal resolution between training and inference, we design a *cross-frequency cross-asset* evaluation protocol under sampling-induced distributional shift. The training data are obtained from the publicly released Kronos dataset [16], consisting of 93,912 5-minute OHLCV observations from a single equity (Alibaba 09988.HK, Nov 2019 – Sep 2025), with 65,618 training and 13,967 validation windows. The evaluation data are collected from Tushare, comprising 76,761 daily-frequency OHLCV observations across 11 heterogeneous markets (SH, SZ, HK, IF, DF, ZF, SF, BJ, and INE), from which 50,432 evaluation windows are used. All datasets follow a unified OHLCV representation and preprocessing pipeline. The model is *never exposed* to any evaluation asset or daily-frequency data during training, constituting a zero-shot cross-frequency cross-asset generalization setting.

b) *Baselines*: We compare TransDyn against representative state-of-the-art time series forecasting models spanning diverse architectural paradigms:

- **DLinear** [40] (AAAI 2023), a decomposition-based linear model that serves as a strong simplicity baseline.
- **PatchTST** [8] (ICLR 2023), a patch-based channel-independent Transformer.
- **iTransformer** [10] (ICLR 2024), an inverted Transformer that models cross-variate dependencies.
- **TimesNet** [9] (ICLR 2023), a multi-period temporal 2D-variation model.
- **Kronos** [16] (AAAI 2026), a state-of-the-art financial time-series foundation model built upon discrete tokenization.

Notably, TransDyn is implemented directly upon the original Kronos framework, where all proposed modules are incrementally integrated into the baseline architecture. Therefore, the Vanilla baseline corresponds to the original Kronos model without the proposed temporal modeling and regularization components. This design allows us to directly quantify the effectiveness of the proposed collaborative temporal modeling and regularization mechanisms.

c) *Evaluation Metrics*: We adopt three metrics covering both point-forecast accuracy and financial signal quality:

- **MAE** (Mean Absolute Error): absolute prediction accuracy.
- **sMAPE** (Symmetric Mean Absolute Percentage Error): scale-normalized accuracy bounded in $[0, 200\%]$, replacing the conventional MAPE which diverges when targets approach zero.
- **IC** (Information Coefficient): Pearson correlation between predicted and realized returns across the cross-section at each horizon, averaged over horizons.

d) *Implementation Details*: All experiments are conducted with PyTorch 2.0 on a single NVIDIA RTX 4090 GPU. The input lookback window is $T = 96$ and the default prediction horizon is $H = 24$. TransDyn uses $d_{model} = 128$,

TABLE I
MAIN ZERO-SHOT CROSS-FREQUENCY CROSS-ASSET FORECASTING RESULTS.

Model	MAE↓	sMAPE(%)↓	IC↑
DLinear [40]	76.45	7.15	0.1724
PatchTST [8]	98.41	7.56	0.1393
TimesNet [9]	72.98	6.82	0.1299
iTransformer [10]	80.05	7.25	0.1495
Kronos [16]	72.20	7.15	-0.1219
TransDyn (Ours)	67.45	6.73	0.1810

multi-scale convolution kernels $\{15, 31, 63\}$, hierarchical BSQ with $b_1 = b_2 = 8$ bits, and a 6-layer Transformer predictor with 8 attention heads.

Training proceeds in two stages: the tokenizer is trained for 30 epochs (learning rate 2×10^{-4}), followed by the predictor for 30 epochs (learning rate 4×10^{-5}), both optimized using AdamW with $\beta_1 = 0.9$, $\beta_2 = 0.95$, weight decay 0.1, and OneCycleLR scheduling. All baselines are trained for 30 epochs with the same data pipeline and preprocessing strategy to ensure fair comparison.

During inference, TransDyn generates predictions via Monte Carlo sampling ($K = 10$ samples) with temperature $\tau = 1.0$, top- $k = 50$, and top- $p = 0.9$. All reported results are averaged over three independent runs.

B. Main Results

Table I summarizes the overall zero-shot cross-frequency cross-asset forecasting performance of TransDyn and all baseline models.

TransDyn achieves the best overall forecasting performance with the lowest MAE (67.45), lowest sMAPE (6.73), and highest IC (0.1810) among all compared methods. Compared with the strongest Transformer baseline TimesNet, TransDyn improves MAE from 72.98 to 67.45, while compared with PatchTST, it achieves a substantial 31.4% MAE improvement under severe cross-frequency distribution shift.

Although DLinear obtains relatively competitive IC performance with an extremely lightweight architecture, its significantly larger forecasting error indicates limited temporal representation capability. Meanwhile, Kronos produces negative IC (-0.1219), while PatchTST and iTransformer also exhibit noticeable degradation under cross-frequency forecasting settings. These observations are consistent with the fixed-resolution assumption discussed in the Introduction: existing forecasting models generally assume consistent temporal granularity between training and inference data, resulting in weakened generalization once this assumption is violated.

Overall, TransDyn achieves a favorable balance between forecasting accuracy and financial signal robustness under heterogeneous temporal resolutions, demonstrating strong effectiveness for zero-shot cross-frequency cross-asset forecasting.

C. Ablation Study

We conduct comprehensive ablation experiments to evaluate the contribution of each proposed functional component in TransDyn. Specifically, we remove one component at a time

TABLE II
ABLATION STUDY RESULTS OF TRANSDYN.

Variant	MAE↓	sMAPE(%)↓	IC↑
TransDyn (full)	67.45	6.73	0.1810
w/o Multi-Scale Temporal Extraction	70.69	6.93	0.1053
w/o Cross-Scale Interaction	67.51	7.03	0.1338
w/o Adaptive Temporal Modulation	70.90	7.00	0.1514
w/o Information Bottleneck Regularization	71.62	6.98	0.1559
Vanilla (Kronos)	72.20	7.15	-0.1219

together with its corresponding implementation and optimization mechanism, including Multi-Scale Temporal Extraction, Cross-Scale Interaction, Adaptive Temporal Modulation, and Information Bottleneck Regularization. Quantitative results are reported in Table II.

Overall, the full TransDyn model achieves the best performance across all metrics, while the Vanilla baseline (Kronos) produces negative IC, indicating poor cross-frequency financial signal modeling capability without the proposed components.

Removing Multi-Scale Temporal Extraction causes the largest IC degradation ($0.1810 \rightarrow 0.1053$), suggesting that hierarchical temporal pattern extraction is critical for transferable dynamics learning across resolutions. Removing Adaptive Temporal Modulation substantially increases MAE, demonstrating the importance of stabilizing temporal dynamics under stochastic volatility variations. Although removing Cross-Scale Interaction causes only minor MAE changes, it significantly reduces IC, indicating its primary contribution to cross-frequency signal consistency rather than point-wise regression accuracy. Removing Information Bottleneck Regularization also degrades overall forecasting performance, highlighting the necessity of suppressing resolution-dependent redundancy.

Overall, the proposed components collaboratively improve forecasting robustness and transferable temporal dynamics under severe cross-frequency distribution shift.

D. Qualitative Visualization Analysis

To further evaluate model generalization behavior, Fig. 4 visualizes predictions on the same unseen daily-frequency evaluation asset. Existing baselines exhibit different failure modes under severe cross-frequency distribution shift. PatchTST suffers from prediction collapse after abrupt market fluctuations, while Kronos (Vanilla) produces unstable local oscillations. Although iTransformer and TimesNet preserve partial trend consistency, they still exhibit noticeable temporal drift during the recovery stage.

In contrast, TransDyn maintains more stable temporal evolution despite being trained only on limited single-asset single-frequency data. Although it does not perfectly fit all short-term fluctuations, it better preserves the global recovery trend and avoids catastrophic collapse, excessive oscillations, and long-term drift.

Overall, the visualization suggests that robust cross-frequency cross-asset forecasting requires transferable temporal dynamics beyond local fitting under fixed-resolution forecasting settings.

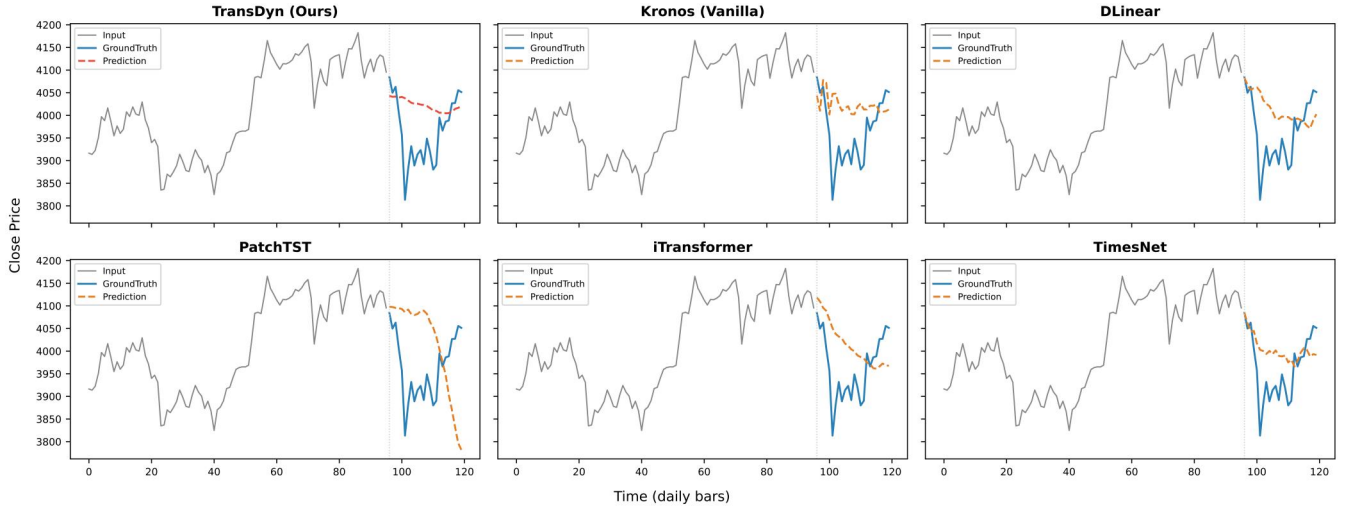


Fig. 4. **Qualitative comparison on the same unseen daily-frequency evaluation asset.** TransDyn is trained from single-asset single-frequency 5-minute K-bar data, while the compared baselines are developed under conventional fixed-resolution or multi-frequency forecasting settings. Under the same zero-shot cross-frequency cross-asset scenario, TransDyn preserves more stable temporal evolution, whereas existing baselines exhibit prediction collapse, local oscillations, or temporal drift.

TABLE III
EFFICIENCY COMPARISON OF DIFFERENT FORECASTING METHODS.

Method	Params	Latency (ms)	Memory (MB)
DLinear [40]	4,656	0.9	21.3
PatchTST [8]	632,216	1.2	23.9
TimesNet [9]	1,658,064	8.3	28.6
iTransformer [10]	610,328	1.1	23.6
Vanilla (Kronos) [16]	3,385,472	7.1	22.2
TransDyn (Ours)	3,834,279	9.2	25.1

E. Efficiency Analysis

We further analyze the computational efficiency of TransDyn in terms of parameter count, inference latency, and GPU memory consumption.

As shown in Table III, although TransDyn introduces additional parameters compared with lightweight baselines, the overall overhead remains moderate. Compared with the vanilla baseline, TransDyn increases parameter count by only 13% (3.39M $\rightarrow$ 3.83M), with merely 2.1 ms additional inference latency and 2.9 MB extra GPU memory consumption.

Despite the increased complexity, TransDyn maintains practical inference efficiency while achieving substantially stronger forecasting performance under cross-frequency distribution shift. In particular, although TimesNet contains fewer parameters, its latency remains close to that of TransDyn, while its forecasting performance is noticeably weaker.

As shown in Table IV, Multi-Scale Temporal Extraction contributes the largest increase in parameter count, while Cross-Scale Interaction and Adaptive Temporal Modulation mainly increase inference latency. Nevertheless, the full TransDyn model still maintains inference latency below 10 ms.

Overall, the proposed functional components moderately increase computational cost while substantially improving forecasting performance, demonstrating a favorable trade-off

between predictive performance and computational efficiency.

V. CONCLUSION

This paper investigates zero-shot cross-frequency cross-asset financial forecasting under sampling-induced temporal distribution shift and proposes TransDyn, a Transformer-driven framework for learning transferable temporal dynamics across heterogeneous sampling resolutions. Extensive experiments on real-world financial markets demonstrate the robustness and effectiveness of TransDyn under severe distribution shift. The current evaluation considers a strict zero-shot setting in which both temporal resolution and asset domain change simultaneously, and thus reflects joint cross-frequency and cross-asset generalization. Future work will explore broader sampling frequencies, training assets, market regimes, and practical trading factors such as transaction costs and risk-adjusted returns.

VI. ACKNOWLEDGMENT

The authors gratefully acknowledge the authors of Kronos for publicly releasing the financial K-line dataset used in this study.

REFERENCES

- [1] M. Praveen, S. Dekka, D. M. Sai, D. P. Chennamsetty, and D. P. Chinta, "Financial time series forecasting: A comprehensive review of signal processing and optimization-driven intelligent models," *Computational Economics*, vol. 67, no. 2, pp. 963–989, 2026.
- [2] Q. Ma, Z. Liu, Z. Zheng, Z. Huang, S. Zhu, Z. Yu, and J. T. Kwok, "A survey on time-series pre-trained models," *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, no. 12, pp. 7536–7555, 2024.
- [3] B. A. Marconi, "Time series foundation models for multivariate financial time series forecasting," *arXiv preprint arXiv:2507.07296*, 2025.
- [4] S. Zhang, L. Shan, R. Qiu, and W. Zhong, "Rah-vla: Resolution-adaptive hierarchical vision–language alignment for multimodal remote sensing understanding," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 64, pp. 5 637 018–5 637 018, 2026.

TABLE IV
EFFICIENCY IMPACT OF DIFFERENT TRANSDYN MODULES.

Variant	Params	Latency (ms)	Memory (MB)
TransDyn (full)	3,834,279	9.2	25.1
w/o Multi-Scale Temporal Extraction	3,389,732	7.5	22.2
w/o Cross-Scale Interaction	3,571,495	7.8	23.1
w/o Adaptive Temporal Modulation	3,784,612	8.4	25.0
w/o Information Bottleneck Regularization	3,830,019	8.6	25.1
Vanilla (Kronos)	3,385,472	7.1	22.2

- [5] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 12, 2021, pp. 11 106–11 115.
- [6] H. Wu, J. Xu, J. Wang, and M. Long, "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting," in *Advances in Neural Information Processing Systems*, vol. 34, 2021, pp. 22 419–22 430.
- [7] T. Zhou, Z. Ma, Q. Wen, X. Wang, L. Sun, and R. Jin, "Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting," in *Proceedings of the 39th International Conference on Machine Learning*, 2022, pp. 27 268–27 286.
- [8] Y. Nie, N. H. Nguyen, P. Sinthong, and J. Kalagnanam, "A time series is worth 64 words: Long-term forecasting with transformers," in *The Eleventh International Conference on Learning Representations*, 2023.
- [9] H. Wu, T. Hu, Y. Liu, H. Zhou, J. Wang, and M. Long, "Timesnet: Temporal 2d-variation modeling for general time series analysis," in *The Eleventh International Conference on Learning Representations*, 2023.
- [10] Y. Liu, T. Hu, H. Zhang, H. Wu, S. Wang, L. Ma, and M. Long, "itransformer: Inverted transformers are effective for time series forecasting," in *The Twelfth International Conference on Learning Representations*, 2024.
- [11] A. F. Ansari, L. Stella, C. Turkmen, X. Zhang, P. Mercado, H. Shen, O. Shchur, S. S. Rangapuram, S. P. Arango, S. Kapoor *et al.*, "Chronos: Learning the language of time series," *Transactions on Machine Learning Research*, 2024.
- [12] A. Garza, C. Challu, and M. Mergenthaler-Canseco, "Timegpt-1," *arXiv preprint arXiv:2310.03589*, 2023.
- [13] A. Das, W. Kong, R. Sen, and Y. Zhou, "A decoder-only foundation model for time-series forecasting," in *Proceedings of the 41st International Conference on Machine Learning*, 2024, pp. 10 148–10 167.
- [14] Y. Liu, H. Zhang, C. Li, X. Huang, J. Wang, and M. Long, "Timer: Generative pre-trained transformers are large time series models," in *Proceedings of the 41st International Conference on Machine Learning*, 2024, pp. 32 369–32 399.
- [15] G. Woo, C. Liu, A. Kumar, C. Xiong, S. Savarese, and D. Sahoo, "Unified training of universal time series forecasting transformers," in *Proceedings of the 41st International Conference on Machine Learning*, 2024.
- [16] Y. Shi, Z. Fu, S. Chen, B. Zhao, W. Xu, C. Zhang, and J. Li, "Kronos: A foundation model for the language of financial markets," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 40, no. 30, 2026, pp. 25 366–25 373.
- [17] Y. Ait-Sahalia, P. A. Mykland, and L. Zhang, "How often to sample a continuous-time process in the presence of market microstructure noise," *The Review of Financial Studies*, vol. 18, no. 2, pp. 351–416, 2005.
- [18] D. Marcek and M. Madera, "Artificial neural networks learning for high-frequency data prediction—big data approach based on genetic and micro-genetic algorithms," *Computational Economics*, 2026.
- [19] Y. Okamoto, "Transformers for time-series forecasting: A comprehensive survey through 2024," *Nonlinear Theory and Its Applications, IEICE*, vol. 17, no. 1, pp. 260–278, 2026.
- [20] Y. Li, X. Wang, R. Zeng, P. Kumar Donta, I. Murturi, M. Huang, and S. Dustdar, "Federated domain generalization: A survey," *Proceedings of the IEEE*, vol. 113, no. 4, pp. 370–410, 2025.
- [21] J. L. Elman, "Finding structure in time," *Cognitive Science*, vol. 14, no. 2, pp. 179–211, 1990.
- [22] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [23] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint arXiv:1803.01271*, 2018.
- [24] Q. Wen, T. Zhou, C. Zhang, W. Chen, Z. Ma, J. Yan, and L. Sun, "Transformers in time series: A survey," in *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, 2023, pp. 6778–6786.
- [25] Y. Wu and J. Wan, "A survey of text classification based on pre-trained language model," *Neurocomputing*, vol. 616, p. 128921, 2025.
- [26] Q. Zhang, Y. Sun, H. Wen, P. Yang, X. Li, M. Li, K.-Y. Lam, S.-M. Yiu, and H. Yin, "Time series analysis in frequency domain: A survey of open challenges, opportunities and benchmarks," *arXiv preprint arXiv:2504.07099*, 2025.
- [27] Y. Zhang, L. Ma, S. Pal, Y. Zhang, and M. Coates, "Multi-resolution time-series transformer for long-term forecasting," in *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*. PMLR, 2024, pp. 4222–4230.
- [28] S. Liu, H. Yu, C. Liao, J. Li, W. Lin, A. X. Liu, and S. Dustdar, "Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting," in *International Conference on Learning Representations*, 2021.
- [29] Y. Zhang and J. Yan, "Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting," in *International Conference on Learning Representations*, 2023.
- [30] M. A. Shabani, A. H. Abdi, L. Meng, and T. Sylvain, "Scaleformer: Iterative multi-scale refining transformers for time series forecasting," in *International Conference on Learning Representations*, 2023.
- [31] H. Wang, J. Peng, F. Huang, J. Wang, J. Chen, and Y. Xiao, "Micn: Multi-scale local and global context modeling for long-term series forecasting," in *International Conference on Learning Representations*, 2023.
- [32] F. Kou, J. Wang, L. Shi, Y. Yao, Y. Li, S. Zhu, Z. Zhang, and J. Du, "Cfpt: Empowering time series forecasting through cross-frequency interaction and periodic-aware timestamp modeling," in *Proceedings of the 42nd International Conference on Machine Learning*, 2025.
- [33] H. Ma, Y. Chen, H. Yu, W. Zhao, W. Zhang, Z. Liu, and G. Yang, "Freqmta: Multi-token attention for stable frequency-domain long-term time series forecasting," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2026, pp. 3611–3615.
- [34] J. Li, H. Deng, J. Xiao, A. Li, S. He, and S. Yu, "Frequency-domain driven recurrent attention with linear complexity for time series forecasting," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2026, pp. 2751–2755.
- [35] J. Wang, C. Lan, C. Liu, Y. Ouyang, T. Qin, W. Lu, Y. Chen, W. Zeng, and P. S. Yu, "Generalizing to unseen domains: A survey on domain generalization," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 8, pp. 8052–8072, 2022.
- [36] B. Sun and K. Saenko, "Deep coral: Correlation alignment for deep domain adaptation," in *European Conference on Computer Vision*. Springer, 2016, pp. 443–450.
- [37] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. March, and V. Lempitsky, "Domain-adversarial training of neural networks," *Journal of Machine Learning Research*, vol. 17, no. 59, pp. 1–35, 2016.
- [38] M. Arjovsky, L. Bottou, I. Gulrajani, and D. Lopez-Paz, "Invariant risk minimization," *arXiv preprint arXiv:1907.02893*, 2019.
- [39] I. Gulrajani and D. Lopez-Paz, "In search of lost domain generalization," in *International Conference on Learning Representations*, 2021.
- [40] A. Zeng, M. Chen, L. Zhang, and Q. Xu, "Are transformers effective for time series forecasting?" in *Proceedings of the AAAI conference on artificial intelligence*, vol. 37, no. 9, 2023, pp. 11 121–11 128.