TF-CoDiT: Conditional Time Series Synthesis with Diffusion Transformers for Treasury Futures

Anonymous ACL submission

Abstract

Diffusion Transformers (DiT) have achieved milestones in synthesizing financial time-series data, such as stock prices and order flows. However, their performance in synthesizing treasury futures data is still underexplored. This work emphasizes the characteristics of treasury futures data, including its low volume, market dependencies, and the grouped correlations among multivariables. To overcome these challenges, we propose TF-CoDiT, the first DiT framework for language-controlled treasury futures synthesis. To facilitate low-data learning, TF-CoDiT adapts the standard DiT by transforming multi-channel 1-D time series into Discrete Wavelet Transform (DWT) coefficient matrices. A U-shape VAE is proposed to encode cross-channel dependencies hierarchically into a latent variable and bridge the latent and DWT spaces through decoding, thereby enabling latent diffusion generation. To derive prompts that cover essential conditions, we introduce the Financial Market Attribute Protocol (FinMAP) - a multi-level description system that standardizes daily/periodical market dynamics by recognizing 17/23 economic indicators from 7/8 perspectives. In our experiments, we gather four types of treasury futures data covering the period from 2015 to 2025, and define data synthesis tasks with durations ranging from one week to four months. Extensive evaluations demonstrate that TF-CoDiT can produce highly authentic data with errors at most 0.433 (MSE) and 0.453 (MAE) to the ground-truth. Further studies evidence the robustness of TF-CoDiT across contracts and temporal horizons.

1 Introduction

Treasury futures (TF) represent a cornerstone of the fixed-income market. Beyond hedging against interest rate volatility, it serves as a critical gateway for general investors and quantitative investment institutions to participate in fixed income investments, earning widespread interest. Recently, emerging foundation models (FMs) (Qi et al., 2025; Jin et al., 2024) with billions of parameters have catalyzed a transformative shift in financial time-series analysis (Zhang et al., 2025; Islam et al., 2024), unlocking a simultaneous opportunity for the deep study of TF markets (Du and Hong, 2025). However, the efficacy of FMs relies heavily on the vast volume of real-world financial data for training and validation. The gap between the drastically growing data needs and the daily renewing markets increases, posing a significant challenge to the research community.

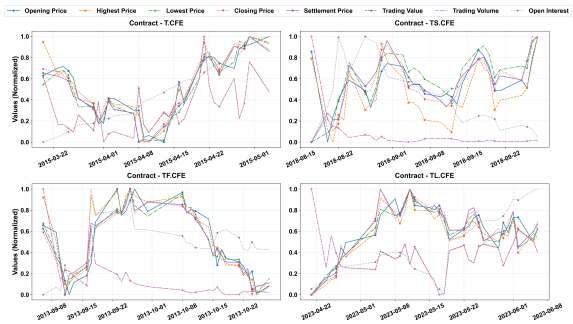


Figure 1: Monthly tendency line of treasury futures in four types of contracts (T, TS, TF, and TL).

Considerable effort has been dedicated to generating synthetic financial time-series data. Generative adversarial networks (GANs) were first explored: TimeGAN (Yoon et al., 2019) and Quant GANs (Wiese et al., 2019) generate high-fidelity time series of stock prices for data augmentation purposes. WGAN (Naritomi and Adachi, 2020) simulates the market by generating synthetic order events. Subsequently, Wang and Ventre (2024); Takahashi and Mizuno (2024) leverage diffusion models to overcome unstable training and mode collapse of GANs, significantly improving stock price generation. Furthermore, time-series FMs, e.g., FinCast (Zhu et al., 2025) and CALF (Liu et al., 2025), generate time series through temporally continuous next token prediction, and their results are also promising. Neverthe-

less, these approaches primarily focus on unconstrained generation, assuming physical prototypes (like GBM (Kim et al., 2025)) to model price trends, and strive to satisfy the *stylized facts* (Kwon and Lee, 2024) observed in fluctuating financial markets. However, they are inadequate for accurately tracing the inscrutable movement of TF markets.

We highlight three key characteristics that set treasury futures apart from financial time series observed in the equity market. First, the available data is intrinsically sparse and limited. Existing TFs (in China) encompass only four types of contracts, covering periods from 2 to 30 years. Additionally, unlike high-frequency trading that leverages vast tick-by-tick data, decision-making in TF markets is primarily driven by inter-day market data, resulting in a low-data regime. Second, the dynamics of financial derivatives are fundamentally exogenous, heavily governed by shifting economic policies and geopolitics. This necessitates a conditional generative approach to ensure synthesized outputs remain economically grounded. Finally, as illustrated in Figure 1, the TF market is a multi-dimensional system depicted by eight interrelated variables. Effectively modeling the intricate, non-linear correlations between these variables - particularly their co-movement during market shocks - remains a formidable and unresolved challenge.

Based on the above foundations, we propose TF-CoDiT, the first controllable DiT framework for time series (TS) synthesis of treasury futures with language as conditions. The main idea of TF-CoDiT is to adapt text-to-image DiT for text-to-TS generation, thereby facilitating low-data learning. To this end, we train TF-CoDiT to generate the 2D spectrogram of the 1D TS data, and the invertible Discrete Wavelet Transform is hired to map the signal between time and frequency domains. Additionally, we propose a U-shape VAE that transforms the spectrogram into a latent vector by hierarchically channel-wise encoding, which not only effectively captures inter-variable correlations but also aligns our framework to latent diffusion (Rombach et al., 2022a). Finally, to cover the most essential market conditions using precise prompt words, we introduce the Financial Market Attribute Protocol (FinMAP). This is a hierarchical description system that traces daily market dynamics using 17 economic indicators from seven perspectives, while reviewing the market during a period using 23 indicators across eight dimensions.

We collect all four types of treasury futures from 2015 to 2025. We also review the daily economic environment during this interval by accessing open websites and our in-house databases, deriving over 30,000 prompt-TS pairs regularized by our FinMAP. Extensive experiments are conducted on this dataset by defining four synthesis tasks across varying horizons: one week and one to four months. Results demonstrate that: 1) Our TF-CoDiT framework can synthesize high-fidelity TF time series with at most 0.433 MSE and 0.453 MAE to the ground-truth. 2) TF-CoDiT outperforms previous approaches, reducing the MSE/MAE of T2S (Ge et al., 2025) by 13.4%/12.8% at month-level generations. 3) The advancement of TF-CoDiT appears consistent, regardless of contacts and durations.

2 Related Works

2.1 Diffusion Transformers

The appearance of Denoising Diffusion Probabilistic Models (DDPMs) (Ho et al., 2020) and Latent Diffusion Models (LDMs) (Rombach et al., 2022a) has fundamentally revolutionized generative modeling. While early diffusion models relied heavily on convolutional U-Net backbones, recent research has shifted towards transformer-based architectures due to their superior scalability and global dependency modeling capabilities. Peebles and Xie (Peebles and Xie, 2023) first introduced the Diffusion Transformer (DiT), demonstrating that replacing the U-Net with a Vision Transformer (ViT) operating on latent patches significantly improves image synthesis fidelity and scales predictably with computational budget. Subsequent works, such as PixArt-α (Chen et al., 2024) and Stable Diffusion (Rombach et al., 2022b), have further refined this by integrating multimodal conditioning via cross-attention. Our TF-CoDiT is inspired by FuseDiT (Tang et al., 2025). Unlike standard DiTs that treat text as a secondary condition, FuseDiT enables a deeper fusion of linguistic semantics and generative denoising by a pre-trained LLM, which we hypothesize is crucial for interpreting complex financial narratives.

2.2 Financial Times-series Synthesis

Generating financial time-series data presents unique challenges due to "stylized facts" such as volatility clustering, heavy tails, and non-stationarity (Safari et al., 2025). Traditional approaches centered on generative adversarial networks, with TimeGAN (Yoon et al., 2019) and

WGAN (Naritomi and Adachi, 2020) serving as benchmarks for capturing temporal correlations. However, GANs often suffer from mode collapse and training instability when applied to the high-dimensional, noisy environment of futures markets. Recent efforts (Wang and Ventre, 2024; Tanaka et al., 2025; Cao et al., 2024) have explored diffusion models for time-series tasks. Takahashi and Mizuno (2024) first join wavelet transforms with DDPMs to generate stock prices, where the wavelet coefficients of log returns, spreads, and trading volumes are viewed as the luminance of RGB colors. Furthermore, time-series models designed for forecasting, such as FinCast (Zhu et al., 2025), Time-LLM (Jin et al., 2024), and TimeHF (Qi et al., 2025), are capable of performing continuous zero-shot predictions within specified time intervals to generate time series data. Despite these advances, there remains a gap in models that can capture cross-channel dependencies (e.g., 8-variable treasury futures) while being directly controllable by natural language. Our TF-CoDiT achieves this by combining the hierarchical channel encoding of U-VAE with the semantic reasoning of LLM-based diffusion.

3 Methodology

The overview of TF-CoDiT is illustrated in Figure 2. It comprises a signal transformation unit, which bridges a time signal and its frequency representation through the invertible Discrete Wavelet Transform (DWT); a U-shape variational autoencoder (VAE), which compresses multi-channel DWT coefficients into a compact latent variable that serves as the generation target; and a backbone LLM ϵ_θ that performs latent diffusion.

3.1 Problem Definition

Given a TF time-series $X = \{x_1, x_2, \ldots, x_T\} \in \mathbb{R}^{C\times T}$, where $C = 8$ in our context, indicating eight variables observed in the TF market cross T time steps, specifically: opening, closing, highest, and lowest prices; settlement price; transaction amount; trading volume; and open interest. Our objective is to learn a conditional distribution $p(X|c)$, where c is a structured natural language description standardized by the Fin-MAP protocol.

3.2 Signal Transformation

Wavelet Transformation. To expose the multi-scale volatilities of treasury futures, we decompose X in the frequency domain using the Discrete

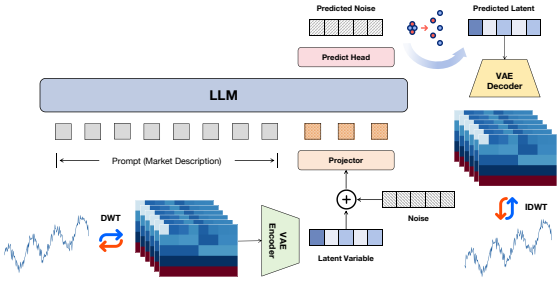


Figure 2: Overview of the TF-CoDiT framework.

Wavelet Transform (DWT) with the Haar wavelet. For each channel $i \in \{1, \ldots, C\}$, the signal $x^{(i)}$ is passed through a series of *high-pass* filters $\phi_k(\cdot)$ and *low-pass* filters $\psi_k(\cdot)$. At the decomposition level J, the signal is represented as:

$$x^{(i)}(t) = \sum_k a^{(i)}_{J,k}\phi_{J,k}(t) + \sum_{j=1}^{J}\sum_k d^{(i)}_{j,k}\psi_{j,k}(t)$$

where $a_{J,k}$ are the **approximation coefficients**, which outlines low-frequency trends; and $d_{j,k}$ are the **detail coefficients**, indicating high-frequency shocks. We concatenate these coefficients into a 3-D tensor $\mathbf{W} \in \mathbb{R}^{C\times\Gamma\times T}$, where $\Gamma = J+1$ represents the total number of frequency scales. Specifically, for each channel i, the coefficient matrix $\mathbf{W}_i \in \mathbb{R}^{\Gamma\times T}$ is constructed by stacking the approximation and detail coefficients along the vertical axis: $\mathbf{W}_i = \left[\mathbf{a}^{(i)}_J, \mathbf{d}^{(i)}_J, \mathbf{d}^{(i)}_{J-1}, \ldots, \mathbf{d}^{(i)}_1\right]^\top$, where $\mathbf{a}^{(i)}_J \in \mathbb{R}^{T/2^J}$, $\mathbf{d}^{(i)}_j \in \mathbb{R}^{T/2^j}$. This transformation enables TF-CoDiT to learn from the informative "energy distribution" of time-series data, instead of information-less point-in-time values.

Inverse Wavelet Transformation. Since the DWT is invertible, given the coefficient matrix generated by TF-CoDiT, the Inverse Discrete Wavelet Transformation (IDWT) is used to recover the time signal. This facilitates our model to treat a 1-D multivariate time series as a multi-channel 2-D image, effectively reducing the synthesis of time series to text-to-image generation.

3.3 Backbone LLM

The core architecture of TF-CoDiT is identical to FuseDiT (Tang et al., 2025), where a pre-trained LLM is used to model textual conditions and perform denoising. As shown in Figure 2, the inputs of this LLM consist of text embeddings followed by latent variables. At the attention layer l, the hidden states from the text $\mathbf{h}^{text}_l \in \mathbb{R}^{N\times D}$ and the latent $\mathbf{h}^{latent}_l \in \mathbb{R}^{M\times D}$ are concatenated together:

$\mathbf{h}_l = [\mathbf{h}_l^{text}; \mathbf{h}_l^{latent}] \in \mathbb{R}^{(N+M)\times D}$, and undergoes a shared self-attention with an asymmetric mask:

$$\mathbf{h}'_l = \text{Attention}(\text{LN}(\mathbf{h}_l), \mathcal{M}) + \mathbf{h}_l$$

where $\text{LN}(\cdot)$ is layer normalization. $\mathcal{M}$ is a causal mask on text tokens, while a full mask on latent tokens:

$$\mathcal{M}_{i,j} = \begin{cases} 0 & \text{if } j \le i \le M \\ 0 & \text{if } i > N \\ -\infty & \text{otherwise} \end{cases}$$

Subsequently, the two modalities diverge in their normalization and modulation:

$$\begin{aligned} \mathbf{h}_{l+1}^{text} &= \text{FFN}_{text}(\text{LN}(\mathbf{h}_l^{text'})) + \mathbf{h}_l^{text'}, \\ \mathbf{h}_{l+1}^{latent} &= \text{FFN}_{latent}(\text{AdaLN}(\mathbf{h}_l^{latent'}, t)) + \mathbf{h}_l^{latent'} \end{aligned}$$

where an adaptive LN layer $\text{AdaLN}(\cdot)$ injects the diffusion timestep t into the forward propagation.

Training. We define a forward diffusion process that adds noise ϵ to the latent variable $\mathbf{z}_0$ over T steps, resulting in $\mathbf{z}_t$. The LLM is trained to predict the added noise via the following loss function:

$$\mathcal{L}_{\text{diff}} = \mathbb{E}_{\mathbf{z}_0, \epsilon\sim\mathcal{N}(0,\mathbf{I}), t}\left[\|\epsilon - \epsilon_\theta(\mathbf{z}_t, t, c)\|^2\right]$$

By freezing the LLM's pre-trained weights (i.e., only train the embedding and output layers), TF-CoDiT can capture the non-linear correlations between macro-economic narratives and the resulting price movements in the TF market.

3.4 U-shape Variational Autoencoder

TF-CoDiT necessitates a VAE model for: 1) bridging the gap between wavelet coefficients (a frequency variable) and the diffusion results (a latent variable), 2) modeling the elusive correlations among multiple time variables.

Before this work, U-Cast (Ni et al., 2025) introduced a *Latent Query Attention* (LQA) mechanism that can effectively encode inter-channel dependencies of time series into the Transformer's hidden states. Inspired by U-Cast, we propose the U-shape VAE (U-VAE). It is a novel VAE architecture that embeds channels of the DWT spectrum into a dense vector. Additionally, U-VAE shares learnable latent queries across the aligned layers of encoder and decoder. These queries vary in size with increasing layer depth, which effectively enables channel-wise encoding and recovery.

Encoder $\mathcal{E}$. Given the wavelet coefficients $\mathbf{W} \in \mathbb{R}^{C\times\Gamma\times T}$, each channel $\mathbf{W}_c$ is a structured time-frequency map, which aligns with ordinary images. To preserve local multi-scale features, we partition each channel into non-overlapping 2D patches of size $P_f \times P_t$. These patches are then projected into a sequence of time-and-frequency-specific vectors:

$$\mathbf{h}_0^{(c)} = \text{Linear}(\text{Patchify}(\mathbf{W}_c)) + \mathbf{E}_{pos} \in \mathbb{R}^{N\times d_c}$$

where $N = \frac{\Gamma}{P_f} \times \frac{T}{P_t} = N_f \times N_t$ is the number of pathces per channel, and $\mathbf{E}_{pos}$ are 2D positional embeddings that encode the time and frequency coordinates (see Appendix A). To capture the complex inter-channel dependencies while performing spatial compression, we follow U-Cast and build the U-VAE encoder by stacking L Transformer layers. LQA is implemented at each layer ℓ:

$$\mathbf{Q} = \mathbf{W}_q^{(\ell)}\mathbf{Q}^{(\ell)}, \mathbf{K} = \mathbf{W}_k^{(\ell)}\mathbf{H}^{(\ell-1)}, \mathbf{V} = \mathbf{W}_v^{(\ell)}\mathbf{H}^{(\ell-1)}$$

$$\mathbf{H}^{(\ell)} = \text{LN}(\text{MHA}(\mathbf{Q}, \mathbf{K}, \mathbf{V})) \in \mathbb{R}^{C_\ell\times d}$$

where $\mathbf{H}^{(0)} = \text{Concat}(\mathbf{h}_0^{(1)}, \ldots, \mathbf{h}_0^{(C)}) \in \mathbb{R}^{C\times d}$ are channel-wise embeddings, and $d = N \times d_c$. $\mathbf{Q}^{(\ell)} \in \mathbb{R}^{C_\ell\times d}$ is a learnable query, $C_\ell = C/r$, and $r \ge 1$ is a reduction ratio. This hierarchical attention mechanism forces the model to distill global dependencies into an increasingly compact hidden space. Consequently, the hidden state at the deepest layer, $\mathbf{H}^{(L)} \in \mathbb{R}^{1\times d}(C_L = 1)$, distills the multi-channel wavelet features into a compact representation. This state is then parameterized into the mean μ and variance σ of the latent distribution, from which the latent variable $\mathbf{z}_0$ is sampled:

$$\mathbf{z}_0 = \mu + \sigma \odot \epsilon, \quad \epsilon \sim \mathcal{N}(0, \mathbf{I})$$

To align with the standard latent diffusion, we recover $\mathbf{z}_0$ into a structured 3D spatial-temporal tensor $\mathcal{Z} \in \mathbb{R}^{N_f\times N_t\times d_c}$. The dimensions N_f and N_t scale with the frequency and temporal axes of the partitioned patches. This enables the diffusion generation on a structured grid that preserves the topological essence of the financial signals[1].

Decoder $\mathcal{D}$. The decoder inverts the compression process, transforming the sampled latent variable $\mathbf{z} \sim q(\mathbf{z}|\mathbf{W})$ back to the wavelet manifold while preserving multi-scale structures. We first project $\mathbf{z}$ to match the encoder's deepest representation:

$$\mathbf{U}^{(L)} = \text{Linear}_{\text{in}}(\mathbf{z}) \in \mathbb{R}^{C_L\times d}$$

To restore the original channel dimension and reconstruct the wavelet coefficients, we employ an

[1] Our diffusion model performs a similar 2D embedding.

Up-sampling Latent Query Attention that mirrors the encoder's compression hierarchy in reverse. At each decoding layer $\ell \in \{L, \ldots, 1\}$:

$$\mathbf{Q} = \mathbf{W}_q^{(\ell)}\mathbf{Q}^{(\ell)}, \tilde{\mathbf{K}} = \mathbf{W}_k^{(\ell)}\mathbf{U}^{(\ell)}, \tilde{\mathbf{V}} = \mathbf{W}_v^{(\ell)}\mathbf{U}^{(\ell)}$$

$$\mathbf{U}^{(\ell-1)} = \mathrm{LN}(\mathrm{MHA}(\tilde{\mathbf{Q}}, \tilde{\mathbf{K}}, \tilde{\mathbf{V}})) \in \mathbb{R}^{C_{\ell-1} \times d}$$

where $\mathbf{Q}^{(\ell)}$ is drawn from the same layer of the encoder. This architecture enables the decoder to selectively retrieve information from the compressed latent space while leveraging scale-aligned representations from the encoder for accurate reconstruction. Finally, the reconstructed wavelet matrices are obtained through channel-wise projection and reshaping:

$$\hat{\mathbf{W}} = \{\mathrm{Unpatchify}_{\Gamma \times T}\left(\mathrm{Linear}_{\mathrm{out}}\left(\mathbf{U}_c^{(0)}\right)\right)\}_{c=1}^{C}$$

Learning Objective. U-VAE is trained by maximizing the evidence lower bound:

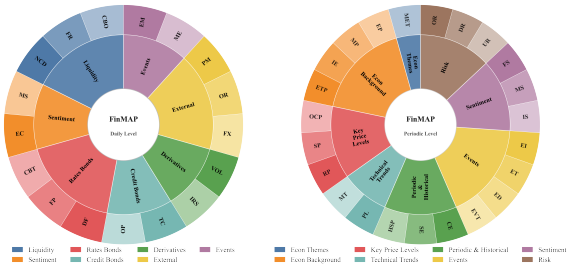
$$\mathcal{L}_{\mathrm{vae}} = \underbrace{\|\mathbf{W} - \hat{\mathbf{W}}\|_1}_{\text{Reconstruction term}} - \beta \cdot \underbrace{\mathrm{KL}\left(q(\mathbf{z}|\mathbf{W}) \| p(\mathbf{z})\right)}_{\text{Regularization term}}$$

Notably, we adopt an L_1 reconstruction loss (rather than the standard MSE) to promote sharpness in the reconstructed wavelet coefficients, ensuring that high-frequency "edges" in the time series are not smoothed out during the compression process, which is vital for high-fidelity synthesis in the TF-CoDiT framework.

3.5 FinMAP

The Financial Market Attribute Protocol (FinMAP) is not only an economic taxonomy that guides the tracing of market dynamics but also a structured language system, which facilitates writing prompt words that cover the essential conditions to promote treasury futures generation.

As illustrated in Figure 3, the FinMAP's taxonomy is organized into a hierarchical architecture encompassing two distinct temporal resolutions: the daily snapshot and the periodic narrative. **The daily snapshot** focuses on high-frequency liquidity and sentiment shifts that drive intra-day and day-to-day fluctuations. It categorizes market attributes into 7 primary categories, encompassing 17 types of economic factors. **The periodical narrative** captures long-term structural trends - spanning weeks, months, or quarters - that define the fundamental regime of the TF market. This level recognizes 23 distinct economic indicators reviewed



(a) Daily-level FinMAP (b) Periodic-level FinMAP

Figure 3: Architecture of FinMAP taxonomy.

from 8 perspectives. Based on this, we implement a temporal aggregation pipeline. Environmental data is first collected chronologically according to daily-level factors; subsequently, these daily snapshots are distilled and reformulated into periodic indicators. This process culminates in the generation of structured prompt anchors, which serve as the high-context conditioning inputs for TF-CoDiT, steering the model's generation toward financially grounded and temporally consistent market analysis. (Appendix A.3 provides more details).

3.6 Inference

The inference process of TF-CoDiT consists of three steps: 1) **Iterative Denoising**. Starting from Gaussian noise $\mathbf{z}_T \sim \mathcal{N}(0, \mathbf{I})$, the LLM ϵ_θ recovers the clean latent $\mathbf{z}_0$ conditioned on a prompt c following Fin-MAP. We employ the DPM-Solver (Lu et al., 2022) to solve the reverse-time stochastic differential equation:

$$\mathbf{z}_{t-1} = \Psi(\mathbf{z}_t, \epsilon_\theta(\mathbf{z}_t, t, c), t, \Delta t)$$

where Ψ is the solver transition. 2) **Latent Decoding**. The LLM-predicted latent $\mathbf{z}_0$ is projected into the wavelet coefficient space via the pre-trained VAE decoder $\mathcal{D}$:

$$\hat{\mathbf{W}} = \mathcal{D}(\mathbf{z}_0), \quad \hat{\mathbf{W}} \in \mathbb{R}^{C \times L}$$

3) **Signal Reconstruction**. We use the Inverse Discrete Wavelet Transform to map the coefficients back to the time domain and produce the 8-variable treasury futures sequence $\hat{X}$:

$$\hat{X} = \mathrm{IDWT}(\hat{\mathbf{W}}), \quad \hat{X} \in \mathbb{R}^{C \times T}$$

4 Experiment

4.1 Datasets

We constructed a comprehensive dataset comprising treasury futures contracts for TS (2-year), TF

Table 1: Evaluation results of time series synthesis for contracts TS, TF, T, and TL. Best results are in **bold**, second best are underlined. Outlier results are in red. The maximum errors of each model are in blue.

Models	Metric	TS			TF			T			TL		
		8	32	64	8	32	64	8	32	64	8	32	64
TimeGAN	MSE	0.408	0.423	0.465	0.455	0.422	0.437	0.413	0.430	0.458	0.450	0.480	0.460
	MAE	0.434	0.434	0.483	0.481	0.412	0.431	0.446	0.430	0.479	0.514	0.527	0.537
WGAN	MSE	0.429	0.472	0.488	0.403	0.457	0.478	0.451	0.480	0.400	0.406	0.448	0.452
	MAE	0.251	0.413	0.484	0.439	0.467	0.433	0.455	0.415	0.491	0.533	0.505	0.549
QuatGAN	MSE	0.525	0.557	0.537	0.570	0.548	0.500	0.425	0.464	0.459	0.487	0.509	0.560
	MAE	0.458	0.452	0.467	0.463	0.466	0.463	0.561	0.552	0.592	0.595	0.514	0.627
GBMDiff	MSE	0.388	0.664	0.614	0.794	1.433	1.295	0.390	0.863	0.920	0.936	1.610	1.376
	MAE	0.435	0.601	0.625	0.903	1.169	1.041	0.443	0.797	1.139	0.931	1.847	1.144
FinDDPM	MSE	0.456	0.422	0.490	0.425	0.491	0.400	0.433	0.442	0.669	0.428	0.439	0.461
	MAE	0.370	0.394	0.397	0.421	0.405	0.538	0.415	0.431	0.479	0.465	0.466	0.475
TimeDiT*	MSE	0.404	0.427	0.464	0.441	0.311	0.329	0.301	0.366	0.277	0.437	0.522	0.505
	MAE	0.277	0.323	0.293	0.335	0.316	0.368	0.296	0.343	0.320	0.353	0.463	0.488
T2S*	MSE	**0.244**	**0.293**	**0.283**	0.289	0.241	0.383	**0.262**	0.297	0.275	0.404	0.553	0.424
	MAE	**0.175**	0.370	0.376	0.369	0.380	0.380	0.380	0.373	0.423	0.392	0.346	0.427
TimeLLM*	MSE	0.429	0.306	0.440	0.310	0.447	0.313	0.430	0.305	0.451	0.535	0.663	0.525
	MAE	0.356	0.322	0.390	**0.225**	0.291	0.300	0.333	0.342	**0.269**	0.328	**0.339**	0.361
CALF*	MSE	0.370	0.394	0.397	0.421	0.605	0.538	0.415	0.431	0.479	0.465	0.466	0.475
	MAE	0.404	0.427	0.464	0.441	0.411	0.429	0.501	0.466	0.677	0.537	0.722	0.605
TF-CoDiT*	MSE	0.277	0.323	0.293	**0.235**	**0.216**	**0.268**	0.296	**0.243**	**0.220**	**0.353**	**0.433**	**0.418**
	MAE	0.234	**0.273**	**0.283**	0.275	**0.241**	**0.283**	**0.242**	**0.279**	0.279	**0.301**	0.384	**0.324**

(5-year), T (10-year), and TL (30-year), covering the period from March 20, 2015, to December 31, 2025. This dataset consists of four distinct subsets with the daily observations: 1,789 for TS, 2,995 for TF, 2,625 for T, and 655 for TL. To facilitate the training and evaluation of our model, we created high-quality prompt-time series pairs by aggregating daily market reviews from the Wind Terminal[2] and our internal proprietary databases. Additionally, an automatic labeling pipeline is used to extract these unstructured narrative reviews into our structured FinMAP protocol. Appendix A.4 details specifications regarding the labeling heuristics and data cleaning procedures.

4.2 Experimental Settings

In experiments, we evaluate the model's capability to generate TF time series across multiple temporal horizons, ranging from one business week to one fiscal quarter. To facilitate the implementation, we run the model to generate sequences of lengths $L \in \{8, 32, 64, 128\}$ and report the averaged **MSE** and **MAE** scores across opening/highest/lowest/closing prices (OHLC). The test set comprises time series data spanning the last 200 days in our dataset. Appendix A.5 provides more details.

The prompt construction follows a hierarchical temporal aggregation strategy: for the 8-day generation task, the prompt absorbs the information from 8 consecutive daily FinMAP attributes. For extended horizons, such as the 32-day or 64-day tasks, the prompts are constructed by recursively aggregating lower-level periodic FinMAP records (e.g., a 16-day prompt aggregates two 8-day periodic situational descriptions). This hierarchical prompt engineering ensures that the model maintains a consistent "macro-narrative" while accounting for long-term structural shifts in the market regime.

Comparison models. We compared TF-CoDiT with three types of models: (1) **Time-series GANs**: TimeGAN (Yoon et al., 2019), WGAN (Naritomi and Adachi, 2020), and QuatGAN (Wiese et al., 2019). (2) **Diffusion Models**: GBMDiff (Wang and Ventre, 2024) and FinDDPM (Takahashi and Mizuno, 2024), TimeDiT (Cao et al., 2024), T2S (Ge et al., 2025). (3) **Time-series Foundation Models**: TimeLLM (Jin et al., 2024), CALF (Liu et al., 2025). Appendix A.6 details our training strategy and implementation of these models.

4.3 Evaluation Results

Table 1 reports the TF time-series synthesis results across temporal horizons ranging from 8 to 64 days. A notable observation is that the specific contract type does not significantly impact model performance, except for TL due to its exclusively limited

[2] https://www.wind.com.cn/

Table 2: Evaluation results of 128-day time series synthesis.

Contract	Metric	TimeGAN	WGAN	QuatGAN	GBMDiff	FinDDPM	TimeDiT*	T2S*	TimeLLM*	CALF*	TF-CoDiT*
TS	MSE	0.475	0.470	0.476	0.699	0.480	0.480	0.480	0.473	0.423	**0.402**
	MAE	0.469	0.496	0.440	0.710	0.447	0.513	0.430	0.505	0.451	**0.405**
TF	MSE	0.627	0.543	0.681	0.750	0.560	0.611	0.494	0.569	**0.425**	0.447
	MAE	0.582	0.618	0.600	0.733	0.694	0.727	0.577	0.615	**0.439**	0.448
T	MSE	0.525	0.434	0.572	0.550	0.400	0.417	0.426	0.476	0.417	**0.391**
	MAE	0.474	0.423	0.568	0.546	0.399	0.426	**0.314**	0.352	0.444	0.372
TL	MSE	0.560	0.509	0.563	0.601	0.463	0.508	0.469	0.503	0.498	**0.418**
	MAE	0.579	0.568	0.578	0.583	0.476	0.475	0.481	0.477	0.502	**0.453**

Table 3: Evaluation results of conditioned (Cond.) generation versus unconditional generation of TF-CoDiT.

Cond.	Metric	TS			TF			T			TL		
		8	32	64	8	32	64	8	32	64	8	32	64
✓	MSE	0.277	0.323	0.293	0.235	0.216	0.268	0.296	0.243	0.220	0.353	0.463	0.418
	MAE	0.234	0.273	0.238	0.275	0.241	0.283	0.242	0.279	0.279	0.301	0.384	0.324
✗	MSE	0.448	0.415	0.482	0.431	0.485	0.407	0.426	0.437	0.462	0.422	0.444	0.458
	MAE	0.376	0.389	0.402	0.418	0.411	0.529	0.422	0.436	0.473	0.459	0.471	0.469

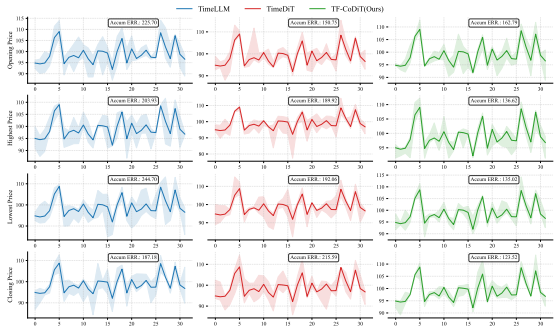
Table 4: Evaluation results of TF-CoDiT using FinMAP, partial FinMAP, and without using FinMAP.

FinMAP		Metric	TS			TF			T			TL		
Daily	Period		8	32	64	8	32	64	8	32	64	8	32	64
✓	✓	MSE	0.277	0.323	0.293	0.235	0.216	0.268	0.296	0.243	0.220	0.353	0.463	0.418
		MAE	0.234	0.273	0.238	0.275	0.241	0.283	0.242	0.279	0.279	0.301	0.384	0.324
✗	✓	MSE	0.391	0.419	0.379	0.331	0.491	0.438	0.392	0.447	0.383	0.353	0.473	0.462
		MAE	0.362	0.392	0.376	0.397	0.398	0.418	0.387	0.439	0.383	0.372	0.388	0.403
✓	✗	MSE	0.402	0.417	0.398	0.347	0.483	0.447	0.421	0.461	0.461	0.411	0.417	0.421
		MAE	0.388	0.420	0.385	0.327	0.476	0.429	0.416	0.418	0.429	0.408	0.415	0.423
✗	✗	MSE	0.448	0.415	0.482	0.431	0.485	0.407	0.426	0.437	0.462	0.422	0.444	0.458
		MAE	0.376	0.389	0.402	0.418	0.411	0.529	0.422	0.436	0.473	0.459	0.471	0.469

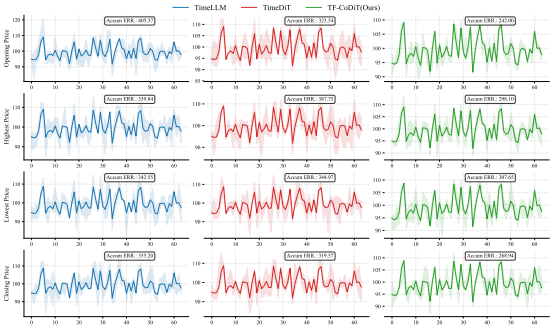
volume (see Table 6). GAN-based models (e.g., TimeGAN) exhibit higher error rates, with MSE values often exceeding 0.45 and MAE around 0.5. In contrast, diffusion models and foundation models perform much better; for instance, the average 32-day MSE of FinDDPM is 0.42. Moreover, conditional generative models (marked with *) exhibit significant superiority across the contract TS, TF, and T. TF-CoDiT outperforms autoregressive models TimeLLM and CALF by addressing the prevalent error accumulation. It also surpasses the previous best model, T2S, with an average improvement of 13.4% in MSE and 12.8% in MAE. Particularly, on the 32-day horizon, TF-CoDiT achieves an average MSE of 0.31 and an MAE of 0.29, whereas T2S achieves 0.35 and 0.37, respectively.

In Table 2, we further compare the models' performance in generating ultra-long time series of 128 days. While most of the compared models experienced significant performance degradation - the average MSE of T2S climbing to 0.47 - TF-CoDiT maintained its structural consistency with a leading MSE of 0.41 and MAE of 0.42.

Ablation Studies are first conducted to investigate the necessity of conditioned generation. As reported in Table 3, removing the semantic prompts degrades the performance of TF-CoDiT to a level comparable to the unconditional FinDDPM, achieving an average MSE of 0.44 and MAE of 0.43 in 32-day generation tasks. These errors increase 0.14 (0.3 → 0.44) and 0.16 (0.27 → 0.43) compared to the standard TF-CoDiT. Given that text conditions are required, we further evaluate the functions of our FinMAP. We drop the daily FinMAP by asking the labeling agent to generate free-form prompts directly from raw market reviews, bypassing our structured taxonomy. To drop the periodic FinMAP, the model generates a macro-narrative that summarizes all daily reviews within the span, bypassing our recursive aggregation logic. Table 4 indicates that the daily-level FinMAP is a primary driver of performance. With the daily-level FinMAP, the average MSE of unconditional generations declines 0.01 (row 3 vs. row 4); it also improves the generations with solely periodic FinMAP by reducing 0.11 average MSE (row 1 vs. row 2).



(a) Error band of 32-day generation.



(b) Error band of 64-day generation.

Figure 4: Error band analyses in the contract-T.

Case Study. In Figure 4, we compare TimeLLM, TimeDiT, and TF-CoDiT by generating 20 trajectories for the T-contract across 32-day and 64-day horizons. We analyze the resulting error bands relative to the ground-truths. TF-CoDiT achieves the lowest cumulative reconstruction error, outperforming TimeDiT by 190.37 in the 32-day setting. Notably, this performance gap more than doubles at the 64-day horizon. This non-linear scaling demonstrates the robustness of TF-CoDiT in suppressing the variance explosion typically observed in long-term financial generation.

5 Discussions

MSE Loss vs. L1 Loss in U-VAE. We evaluate the impact of the reconstruction loss on the U-VAE by comparing the widely adopted MSE loss against the L1 loss. Figure 5 visualizes the element-wise reconstruction errors for the DWT matrix of opening price, with the 32-day T-contract. L1 loss significantly reduces reconstruction errors, achieving an 1.24% (0.0304 $\rightarrow$ 0.0308) reduction in approximation and 10.94% (0.0275 $\rightarrow$ 0.0309) in detail (D_0) coefficients compared to MSE. This improvement stems from the **heavy-tailed** distribution of treasury futures. MSE over-penalizes outliers, resulting in *blurred* reconstructions that average out sharp transitions. In contrast, by maintaining a constant gradient, L1 preserves high-frequency *detail* coefficients essential for capturing intraday volatility and structural breakouts.

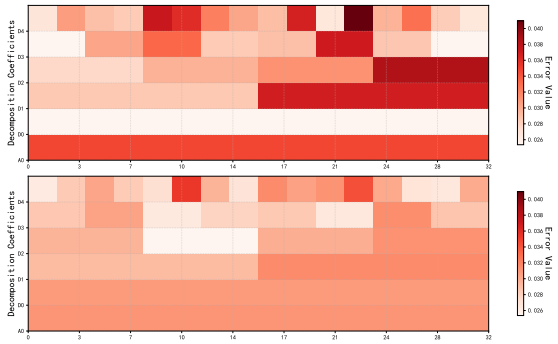


Figure 5: Element-wise reconstruction errors achieved by U-VAE with MSE loss (upper) and L1 loss (lower).

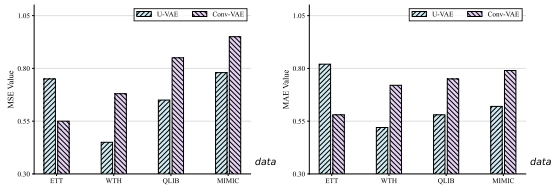


Figure 6: Visualization of VAE reconstruction errors.

Convolutional VAE vs. U-VAE. To test the effectiveness of our proposed channel-wise encoding technology, we compare hierarchical LQA-based U-VAE with the Conv-VAE widely used for image encoding (which consists of four 2D convolutional layers in both the encoder and decoder, with channels $c \in [128, 256, 512, 512]$ across the layers). Their reconstruction losses across four domain-specific multivariate time series datasets, ETT (Zhou et al., 2021), WTH (Wu et al., 2021), QLIB (Yang et al., 2020), and MIMIC (Johnson et al., 2016), are compared in Figure 6. On three of these datasets, our model achieved better results. This comparison empirically proves the importance of modeling cross-channel correlations.

6 Conclusion

This work introduces TF-CoDiT, a novel LLM-DiT framework for conditioned time-series generation (TSG) of treasury futures. We establish a robust link between linguistic narratives of economics and time-continuous prices modeling by: 1) a U-shaped VAE that encodes cross-channel correlations via latent query attention; and 2) the FinMAP protocol, which provides a criterion that standardizes contextual prompts. Experimental results across real contracts demonstrate that TF-CoDiT significantly outperforms existing state-of-the-art TSG models, achieving improved fidelity of generation results.

Limitations While TF-CoDiT exhibits superiority over existing time series generation models, it possesses several limitations that warrant further investigation:

- **Domain Knowledge Integration**. As the core architecture of TF-CoDiT is adapted from a text-to-image model, the pre-trained parameters lack intrinsic alignment with complex financial and economic theories. This gap necessitates extensive fine-tuning and can hinder the model's ability to generalize to "black-swan"-like events or rare market regimes not heavily observed in the training distribution.

- **Unimodal Conditioning Constraints**. TF-CoDiT relies exclusively on textual narratives to steer market dynamics. While this interface is human-friendly, textual prompts are inherently discrete and information-sparse, compared to continuous cross-asset signals. Incorporating multi-modal conditioners—such as synchronous time series of commodity prices (e.g., gold and oil) or foreign exchange rates remains a promising but unexplored avenue for further enhancing synthesis fidelity.

- **Evaluation of FinMAP**. While the efficacy of the FinMAP protocol is empirically validated through our ablation studies, its evaluation remains largely intrinsic. Currently, there is a lack of an external, standardized benchmark to quantitatively assess the semantic quality and coverage of such financial taxonomies. Establishing a rigorous, objective benchmark for market-description protocols is a vital task left for our future work.

References

Defu Cao, Wen Ye, Yizhou Zhang, and Yan Liu. 2024. Timedit: General-purpose diffusion transformers for time series foundation model. *CoRR*, abs/2409.02322.

Junsong Chen, Jincheng Yu, Chongjian Ge, Lewei Yao, Enze Xie, Zhongdao Wang, James T. Kwok, Ping Luo, Huchuan Lu, and Zhenguo Li. 2024. Pixart-α: Fast training of diffusion transformer for photorealistic text-to-image synthesis. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net.

Jiawei Du and Yi Hong. 2025. Market-consistent economic scenario generation system for treasury bond yields with the smith-wilson method. *Expert Syst. Appl.*, 286:128002.

Yunfeng Ge, Jiawei Li, Yiji Zhao, Haomin Wen, Zhao Li, Meikang Qiu, Hongyan Li, Ming Jin, and Shirui Pan. 2025. T2S: high-resolution time series generation with text-to-series diffusion models. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2025, Montreal, Canada, August 16-22, 2025*, pages 5208–5216. ijcai.org.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*.

Md. Khairul Islam, Ayush Karmacharya, Timothy Joowon Sue, and Judy Fox. 2024. Large language models for financial aid in financial time-series forecasting. In *IEEE International Conference on Big Data, BigData 2024, Washington, DC, USA, December 15-18, 2024*, pages 4892–4895. IEEE.

Ming Jin, Shiyu Wang, Lintao Ma, Zhixuan Chu, James Y. Zhang, Xiaoming Shi, Pin-Yu Chen, Yuxuan Liang, Yuan-Fang Li, Shirui Pan, and Qingsong Wen. 2024. Time-llm: Time series forecasting by reprogramming large language models. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net.

Alistair EW Johnson, Tom J Pollard, Lu Shen, Li-wei H Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. 2016. Mimic-iii, a freely accessible critical care database. *Scientific data*, 3(1):1–9.

Gihun Kim, Sun-Yong Choi, and Yeoneung Kim. 2025. A diffusion-based generative model for financial time series via geometric brownian motion. *CoRR*, abs/2507.19003.

Sohyeon Kwon and Yongjae Lee. 2024. Can gans learn the stylized facts of financial time series? In *Proceedings of the 5th ACM International Conference on AI in Finance, ICAIF 2024, Brooklyn, NY, USA, November 14-17, 2024*, pages 126–133. ACM.

Peiyuan Liu, Hang Guo, Tao Dai, Naiqi Li, Jigang Bao, Xudong Ren, Yong Jiang, and Shu-Tao Xia. 2025. CALF: aligning llms for time series forecasting via cross-modal fine-tuning. In *AAAI-25, Sponsored by the Association for the Advancement of Artificial Intelligence, February 25 - March 4, 2025, Philadelphia, PA, USA*, pages 18915–18923. AAAI Press.

Ilya Loshchilov and Frank Hutter. 2019. Decoupled weight decay regularization. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. 2022. Dpm-solver: A fast ODE solver for diffusion probabilistic model sampling in around 10 steps. In *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*.

Yusuke Naritomi and Takanori Adachi. 2020. Data augmentation of high frequency financial data using generative adversarial network. In *IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology, WI/IAT 2020, Melbourne, Australia, December 14-17, 2020*, pages 641–648. IEEE.

Juntong Ni, Shiyu Wang, Zewen Liu, Xiaoming Shi, Xinyue Zhong, Zhou Ye, and Wei Jin. 2025. U-cast: Learning hierarchical structures for high-dimensional time series forecasting. *Preprint*, arXiv:2507.15119.

William Peebles and Saining Xie. 2023. Scalable diffusion models with transformers. In *IEEE/CVF International Conference on Computer Vision, ICCV 2023, Paris, France, October 1-6, 2023*, pages 4172–4182. IEEE.

Yongzhi Qi, Hao Hu, Dazhou Lei, Jianshen Zhang, Zhengxin Shi, Yulin Huang, Zhengyu Chen, Xiaoming Lin, and Zuo-Jun Max Shen. 2025. Timehf: Billion-scale time series models guided by human feedback. *CoRR*, abs/2501.15942.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022a. High-resolution image synthesis with latent diffusion models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022*, pages 10674–10685. IEEE.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022b. High-resolution image synthesis with latent diffusion models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022*, pages 10674–10685. IEEE.

Sara A. Safari, Maximilian Janisch, and Thomas Lehéricy. 2025. International financial markets through 150 years: Evaluating stylized facts. *Preprint*, arXiv:2504.08611.

Tomonori Takahashi and Takayuki Mizuno. 2024. Generation of synthetic financial time series by diffusion models. *Preprint*, arXiv:2410.18897.

Yuki Tanaka, Ryuji Hashimoto, Takehiro Takayanagi, Zhe Piao, Yuri Murayama, and Kiyoshi Izumi. 2025. Cofindiff: Controllable financial diffusion model for time series generation. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2025, Montreal, Canada, August 16-22, 2025*, pages 9357–9365. ijcai.org.

Bingda Tang, Boyang Zheng, Sayak Paul, and Saining Xie. 2025. Exploring the deep fusion of large language models and diffusion transformers for text-to-image synthesis. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2025, Nashville, TN, USA, June 11-15, 2025*, pages 28586–28595. Computer Vision Foundation / IEEE.

Zhuohan Wang and Carmine Ventre. 2024. A financial time series denoiser based on diffusion models. In *Proceedings of the 5th ACM International Conference on AI in Finance, ICAIF 2024, Brooklyn, NY, USA, November 14-17, 2024*, pages 72–80. ACM.

Magnus Wiese, Robert Knobloch, Ralf Korn, and Peter Kretschmer. 2019. Quant gans: Deep generation of financial time series. *CoRR*, abs/1907.06673.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. 2021. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pages 22419–22430.

Xiao Yang, Weiqing Liu, Dong Zhou, Jiang Bian, and Tie-Yan Liu. 2020. Qlib: An ai-oriented quantitative investment platform. *CoRR*, abs/2009.11189.

Jinsung Yoon, Daniel Jarrett, and Mihaela van der Schaar. 2019. Time-series generative adversarial networks. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 5509–5519.

Xu Zhang, Zhengang Huang, Yunzhi Wu, Xun Lu, Erpeng Qi, Yunkai Chen, Zhongya Xue, Qitong Wang, Peng Wang, and Wei Wang. 2025. Multi-period learning for financial time series forecasting. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining, V.1, KDD 2025, Toronto, ON, Canada, August 3-7, 2025*, pages 2848–2859. ACM.

Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. 2021. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*, pages 11106–11115. AAAI Press.

Zhuohang Zhu, Haodong Chen, Qiang Qu, and Vera Chung. 2025. Fincast: A foundation model for financial time-series forecasting. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management, CIKM 2025, Seoul, Republic of Korea, November 10-14, 2025*, pages 4539–4549. ACM.

A Appendix

A.1 Treasury Futures

Treasury futures (TF) are standardized exchange-traded contracts to buy or sell a government bond (the underlying asset) at a predetermined price on a specified future date. They are a primary type of interest rate future, originating in the 1970s to help investors manage the risk of fluctuating interest rates. As a pivotal instrument for interest rate risk management, their pricing dynamics are fundamentally derived from the present value of the underlying bond's cash flows, making them highly sensitive to yield curve fluctuations. Contracts are categorized by the maturity of the notional deliverable bond. Table 5 outlines the key structures in China and the United States - the two largest markets in the world.

This work focuses on treasury futures traded on the China Financial Futures Exchange (CFFEX). Additionally, the dynamics of the TF market are traced by eight key daily variables:

- **Opening price (O)**. The price at which the first transaction of the trading day occurs, indicating the initial market sentiment.
- **Highest price (H)**. The maximum transaction price recorded during the trading session, reflecting the upper boundary of intraday price volatility.
- **Lowest price (L)**. The minimum transaction price recorded during the trading session, reflecting the lower boundary of intraday price volatility.
- **Close price (C)**. The price of the last transaction before the market closes, widely used as a daily benchmark.
- **Settlement price**. The official price calculated by the exchange at the close, typically as a volume-weighted average, which serves as the daily mark-to-market benchmark for margin calculations.
- **Trading value**. The total monetary value of all contracts traded during the day (measured in millions of CNY in the sample), representing the overall market activity scale.
- **Trading volume**. The total number of contracts traded during the day (labeled as "stock" in the sample but representing the standardized contract unit), reflecting market liquidity and participation intensity.
- **Open interest**. The total number of outstanding contracts held by market participants at the end of the trading day, indicating the market's depth and the persistence of capital commitments.

Together, these variables provide a comprehensive view of price action, liquidity, and market structure for the CFFEX treasury futures market. Particularly, the **opening price**, **highest price**, **lowest price**, and **close price** (OHLC) are pivotal for valuation, while the remaining variables collectively gauge trading activity and market sentiment.

A.2 Time-Frequency Position Embedding

The structured map M, resulting from our Multi-scale Temporal Alignment (MTA), presents a grid-like topology. We partition this map into non-overlapping patches of size $P_l \times P_t$. Each patch acts as a single discrete token in the Transformer, where (I, J) denotes its grid coordinates in the frequency-temporal domain: $I \in \{0, 1, \ldots, \frac{J+1}{P_f} - 1\}$ and $J \in \{0, 1, \ldots, \frac{R}{P_t} - 1\}$ ($R = \frac{1}{T}$ in our implementation is the scaled time steps after DWT). To ground these tokens in the global time-frequency space, we implement a composite embedding strategy. First, the content of each patch is aggregated via a linear projection layer. This layer effectively captures the local relative positioning within the patch. Then, we assign a 2D axial embedding based on the patch's grid indices:

$$\begin{aligned}\mathbf{E}_{I,J} =& \text{Linear}(\text{flatten}(\text{Patch}_{I,J})) + \mathbf{PE}_{freq}(I) \\ &+ \mathbf{PE}_{time}(J)\end{aligned}$$

where $\mathbf{PE}_{freq}(I) \in \mathbb{R}^D$ is the Frequency-axis Embedding. It assigns a unique signature to each row of the grid. This allows the self-attention mechanism to recognize that tokens in the same row share the same spectral scale (e.g., long-term trends vs. short-term noise), facilitating intra-scale (horizontal) dependency modeling. $\mathbf{PE}_{time}(J) \in \mathbb{R}^D$ is the Time-axis Embedding. It assigns a unique signature to each column. This enables the model to identify tokens across different frequency bands that occur at the same logical timestamp, thereby enhancing inter-scale (vertical) synchronization.

By assigning embeddings at the patch level rather than the element level, we allow the model to

Table 5: Overview of treasury futures contracts.

Contract	Maturity	Features
China (Traded on China Financial Futures Exchange)		
TS	2-Year	Tracks short-term rates
TF	5-Year	Tracks medium-term rates
T	10-Year	Tracks benchmark long-term rates
TL	30-Year	Tracks bltra-long term rates
United States (Traded on CME Group)		
T-Note Futures	2-year, 5-year, 10-year	Tracks medium to long-term yields
T-Bond Futures	30-year	Tracks long-term interest rate expectations

capture high-level structural correlations while significantly reducing the computational complexity of the self-attention mechanism. This coordinate system ensures that the model can effectively learn complex cross-scale correlations - such as how a macro-economic shift in the low-frequency trend layer triggers specific volatility patterns in high-frequency detail layers - while maintaining strict temporal alignment. This spatial awareness is crucial for synthesizing 8-variable treasury futures that are both statistically plausible and economically consistent.

A.3 Financial Market Attribute Protocol

FinMAP is a description system that facilitates the development of prompt words for TF data generation. It comprises a **taxonomy** of market factors that impact the fluctuation of treasury futures and a **structured language** that standardizes the prompt writing. As a taxonomy, FinMAP features a hierarchical architecture that captures market signals across two temporal resolutions: the Daily Snapshot and the Periodic Situation.

A.3.1 Hierarchical Taxonomy

To derive a daily snapshot of the market, we focus on seven types of factors:

- **Liquidity**: Quantifies central bank interventions (OMO/MLF) and monitors interbank borrowing costs (FR007/DR001) to assess systemic funding stress.
- **Sentiment**: Evaluates the qualitative "trading tone," including risk-on/off shifts and safe-haven demand, often reflected in equity market correlations.
- **Rates Bonds**: Tracks the yield curve movements of sovereign benchmarks and the technical performance (volume, open interest) of lead futures contracts.
- **Credit Bonds**: Evaluates credit spread fluctuations and secondary market liquidity across various rating grades and bond types (e.g., Tier-2 Capital Bonds).
- **Derivatives**: Monitors Interest Rate Swaps (IRS) and volatility metrics (implied vs. realized) to capture long-term benchmark expectations.
- **External Linkages**: Records global spillovers from foreign exchange markets (USDCNH), U.S. Treasury yields, and commodity price actions.
- **Event-Driven Factors**: Documents discrete policy shifts, macroeconomic data releases, and the market's internal pricing of upcoming scheduled events.

On the other hand, to review the market situation over extended horizons (e.g., a week, month, or quarter), FinMAP underscores eight macroscopic categories:

- **Economic Themes**: Identifies the dominant economic narrative (e.g., "Cooling Inflation") that dictates investor behavior throughout the period.
- **Economic Environment**: Assesses the structural economic cycle, monetary policy stances, and the international geopolitical environment.
- **Key Prices**: Defines the boundary constraints of the market, including contract amplitude and critical technical/psychological support and resistance levels.
- **Technical Trends**: Describes the geometric evolution of price action (e.g., W-bottom pat-

terns) and monthly-scale momentum indicators like MACD.

- **Cyclical Factors**: Accounts for historical analogies and seasonal anomalies, such as quarter-end liquidity tightening or "Spring Fever" effects.

- **Events and Timeline**: Maps the chronological chain of catalysts, from the initial release of data to its eventual absorption into market prices.

- **Market Sentiment**: Tracks the emotional trajectory of the market (Initial, Mid-term, and Late-phase) as influenced by evolving supply-demand pressures.

- **Risk Analysis**: Profiles potential tail risks, identifying asymmetric upside/downside triggers and systemic shocks that could disrupt the prevailing trend.

A.3.2 Structured Language

Based on the factors that trace the market dynamics, it is not hard to develop structured prompts for conditional generation. Example prompts are demonstrated in Table 12, Table 13, and Table 14, with a horizon of one day, one week, and one month, respectively. It is noteworthy that, when preparing the dataset, we utilize daily snapshots of market services as a foundation for summarizing periodical market dynamics. This process is detailed in Appendix A.4.

A.4 Data Collection

We aggregated treasury futures market data from official open-access platforms[3] and synthesized conditioning inputs by extracting latent factors from daily market reviews sourced from the Wind Terminal and proprietary in-house databases[4].

To manage the high-throughput requirement of processing thousands of unstructured financial reports, we developed an automated labeling pipeline powered by a multi-agent collaborative framework. As illustrated in Figure 7, the framework utilizes a Labeler-Reviewer architecture to ensure the fidelity of the extracted attributes. The process mirrors the hierarchical structure of the FinMAP protocol, executing in two distinct stages: 1) daily snapshot extraction, and 2) recursive periodic synthesis. At each stage, the agents engage in an iterative feedback loop, capped at three cycles, to achieve high-quality semantic grounding.

A.4.1 Daily-Level Snapshot Extraction

The daily labeling process follows a structured three-step protocol:

- **Pre-processing**: For each target date, raw market reviews are retrieved. Non-machine-readable formats (e.g., PDF reports) are pre-processed using Optical Character Recognition (OCR) to ensure text integrity before being passed to the Labeler agent.

- **Structured Extraction**: The Labeler agent is tasked with parsing the review passages to identify and extract crucial economic factors defined by the FinMAP daily schema, outputting the result in a structured format.

- **Iterative Quality Audit**: The Reviewer agent performs a quantitative assessment of the Labeler's output, assigning a score (1–5) based on its alignment with the source text. If the output receives a score below 4, the Reviewer generates specific revision comments. The Labeler then re-processes the prompt, incorporating this feedback to refine the extraction.

This cycle terminates when the Reviewer assigns a score of ≥ 4 or the iteration limit of three loops is reached.

A.4.2 Periodic Narrative Synthesis

The period-level labeling leverages a hierarchical aggregation logic to capture long-term market regimes:

- **Recursive Data Retrieval**: Given a specified start date and temporal window, the pipeline retrieves previously generated snapshots. For weekly narratives, the system aggregates daily FinMAPs; for monthly narratives, it recursively aggregates weekly FinMAPs to maintain temporal coherence; for quarterly narratives, it recursively aggregates monthly FinMAPs, and so on.

- **Macro-Factor Distillation**: The Labeler agent synthesizes the temporally ordered snapshots into a periodic narrative, focusing on macroscopic indicators such as economic themes and structural risk analysis.

[3] https://www.wind.com.cn/

[4] The daily market review is published the next day.

- **Consistency Review**: The Reviewer agent audits the periodic summary against the underlying snapshots to ensure no significant daily signals were lost in the synthesis. The scoring and feedback mechanism follows the same iterative logic as the daily-level extraction.

We implement the Labeler and Reviewer agents with Tongyi-DeepResearch [5] and Qwen3-235B [6], respectively. Section A.7 reports the prompts developed for these two agents.

A.5 Datasets

Processing. While treasury futures share similarities with equity markets (e.g., stock price), their price dynamics exhibit fundamental differences. Unlike the often non-stationary nature of stock prices, TF prices are bounded by the underlying bond's par value, typically oscillating within a stable range (e.g., 90 to 110) with significant mean-reverting characteristics. Consequently, traditional log-return normalization - while standard for equities - may fail to capture the structural stationarity of TF markets. We apply a stratified normalization strategy to three distinct feature groups. For price-based variables: opening price, highest price, lowest price, closing price, and settlement price, we normalize them based on the previous day's opening price ($Open_{t-1}$):

$$X_t^{norm} = \frac{X_t - Open_{t-1}}{Open_{t-1}} \times 100$$

This approach preserves the magnitude of relative fluctuations while centering the data to facilitate model convergence. For volume and liquidity metrics, trading volume and trading value, we apply a log-transformation to compress the dynamic range and mitigate the impact of outliers:

$$X^{norm} = \log_{10}(X + 1)$$

For the open interest that captures the momentum of market participation and positioning shifts, we normalize it using the first-order growth rate:

$$X_t^{norm} = \frac{X_t - X_{t-1}}{X_{t-1}}$$

Splitting. Our collected dataset comprises 2,625 prompt-time series pairs across four contracts, spanning from March 20, 2015, to December 31, 2025. For each contract, we retain the data from the last 200 days for testing purposes, while the remaining data is utilized for model training. Table 6 presents the data distributions based on the duration of synthesis tasks.

A.6 Implementation Details

All our experiments are conducted on 8 Nvidia H200 GPUs. Our codes are mainly implemented with Pytorch[7] and Huggingface Transformers[8]. We train the U-VAE model using the AdamW optimizer (Loshchilov and Hutter, 2019) with a learning rate of 0.001 and a cosine learning rate scheduler. Similarly, the TF-CoDiT model is trained with the same optimizer and scheduler, but with a learning rate of 0.0005 and a warming-up rate of 0.05. The hyperparameters in terms of model architectures are reported in Table 7 and Table 8. More details are as follows.

A.6.1 Model Training

Because the U-VAE is a separate component of our TF-CoDiT, we trained U-VAE and TF-CoDiT (the LLM that performs conditional diffusion generation) using distinct data recipes and strategies.

U-VAE. We train U-VAE to encode DWT coefficients of time series, especially their inter-channel correlation, into a latent variable. For this purpose, we curate the training set based on our in-hand TF data. We apply a sliding window approach with a size $S \in \{4, 8, 16, 32, 64, 96, 128\}$ and a stride of 1 on our four-contract TF time-series data, where each time series has a length of 2,625. This process yields 72,448 samples for the model training.

TF-CoDiT. We train TF-CoDiT to generate the latent variable derived from the DWT coefficients that decomposes the time series. Specifically, TF-CoDiT starts from FuseDiT, which is based on a Gemma-2B backbone. We freeze this LLM while adding a novel predicting head and replacing the visual embedding layer with a novel one initialized randomly. Additionally, the 2D RoPE position embedding at each attention layer, which originally locates the pixels by latitude and longitude, is redirected to the channels by time and frequency. We train TF-CoDiT using the training sets listed in Table 6, by mixing varying contracts and durations, with only the predicting head and visual embedding layer updated.

[5] https://www.modelscope.cn/models/iic/Tongyi-DeepResearch-30B-A3B

[6] https://www.modelscope.cn/models/Qwen/Qwen3-235B-A22B-Thinking-2507

[7] https://pytorch.org/

[8] https://huggingface.co

Table 6: Data distribution of our dataset. *Cont.*: contract categories.

Cont.	# Training Samples				# Test Samples			
	8	32	64	128	8	32	64	128
TS	1,582	1,558	1,526	1,462	193	169	137	73
TF	2,788	2,764	2,732	2,668	193	169	137	73
T	2,418	2,394	2,362	2,298	193	169	137	73
TL	448	424	392	328	193	169	137	73
Total	7,236	7,140	7,012	6,756	772	676	548	292

Table 7: Hyperparameter specifications for U-VAE.

Item	Encoder	Decoder
Hierarchical Layers (L)	3	3
Reduction Ratio (r)	2	2
Hidden Dimension (d)	64	64
Attention Heads (h)	$\{16, 8, 4\}$	$\{4, 8, 16\}$

Table 8: Architectural specifications of Gemma2.

Item	Value
Hidden Layers	26
Hidden Dimensions	2,304
Attention Heads	8
Projection Dimensions	9,216
Vocabulary Size	256,000
Context Size (Tokens)	8,192
Parameters (Billion)	~ 2.61

A.6.2 Comparison Models

We compared TF-CoDiT in experiments with the following models:

WGAN: a generative model formulated for the optimal transport problem that uses the Wasserstein distance to provide a more stable training process and generate high-fidelity synthetic data.

QuantGAN: a GAN framework that utilizes temporal convolutional networks to capture long-range dependencies and volatility clusters in financial time series.

TimeGAN: a novel framework that preserves temporal dynamics by combining unsupervised adversarial training with a stepwise supervised loss and a learned embedding space.

Time-LLM: an autoregressive framework that adapts frozen large language models (LLMs) for time-series forecasting by reprogramming input data into text prototypes and using "Prompt-as-Prefix" to provide natural language context and task instructions.

CALF: a training framework for LLM, which aligns the distribution discrepancies between textual and temporal modalities through a multi-level cross-modal fine-tuning approach that includes input matching, feature regularization, and output consistency.

FinDDPM: utilizes Denoising Diffusion Probabilistic Models (DDPMs) to generate synthetic financial data by converting multiple time series into images via wavelet transformation and then reconstructing them through inverse transformation.

GBMDiff: a diffusion framework that explicitly incorporates Geometric Brownian Motion into the forward noising process of a score-based model to reflect the price-dependent heteroskedasticity inherent in financial markets.

TIMEDIT: a "proto-foundation" diffusion transformer that combines temporal dependency learning with a unified masking mechanism and a fine-tuning-free editing strategy to handle diverse tasks like forecasting, imputation, and generation.

T2S: a domain-agnostic diffusion framework that bridges natural language and time series by using a length-adaptive VAE and DiT-based Flow Matching to generate high-resolution sequences of arbitrary lengths.

In our experiments, we utilized the *repro-gan* library to implement WGAN and QuantGAN, while TimeGAN was deployed using its official implementation. FinDDPM, GBMDiff, TIMEDIT, and T2S are reproduced based on the Huggingface Diffusers library. A significant architectural adjustment was made to FinDDPM: while the original framework is constrained to 3-channel inputs to align with standard RGB image formats, we extended the architecture to 4 channels to generate OHLC prices. For controllable models (all of them controlled using natural language), we truncate our FinMAP prompt to fit their context size. Conversely, for uncontrollable models, the generation process is initiated from a standard Gaussian noise distribution to synthesize either raw time series sequences or wavelet coefficient matrices.

A.6.3 Evaluation

In our experiments, we evaluate the models' performance in reconstructing the standard Open, High,

Table 9: Comprehensive comparison of baselines. *GAN*: Generative Adversarial Network. *AR*: Autoregressive model. *DM*: Diffusion Model.

Model	Architecture	Controlled	Multimodal	Pretrained
WGAN	GAN	✗	✗	✗
QuantGAN	GAN	✗	✗	✗
TimeGAN	GAN	✗	✗	✗
Time-LLM	AR	✓	✓	✓
CALF	AR	✓	✓	✓
GBMDiff	DM	✗	✗	✗
FinDDPM	DM	✗	✗	✗
TimeDiT	LLM-DiT	✓	✓	✓
T2S	LLM-DiT	✓	✓	✓
TF-CoDiT	LLM-DiT	✓	✓	✗

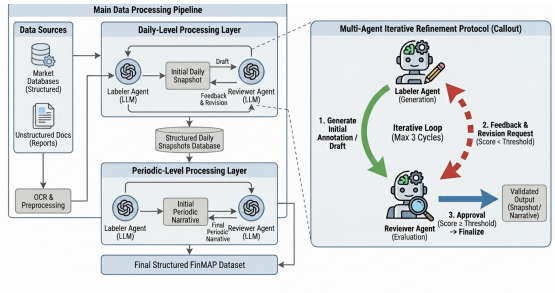


Figure 7: A multi-agent framework for prompt labeling of treasury futures data.

Low, and Close (OHLC) price quartet for a given time span. We select these four variables as the primary targets because they constitute the fundamental price discovery signals utilized in technical analysis and risk management strategies. Accuracy in these benchmarks is critical for capturing intraday volatility and the structural boundaries of the market regime. Quantitatively, we measure the divergence between the generated synthetic trajectories and the ground-truth observations using:

- **Mean Squared Error (MSE)**: MSE penalizes significant outliers in price generation.
- **Mean Absolute Error (MAE)**: MAE provides a robust measure of the average linear error magnitude.

A.7 Prompts

Category	Items	Abbr.	Indicators	Details (Market Logic & Function)
Liquidity	Central Bank Ops	CBO	Scale, interest rates, and net injection/drain of OMO and MLF.	Monitors price and volume signals of short-to-medium term liquidity tools; serves as the primary window for policy fine-tuning.
	Funding Rates	FR	Key rates including DR007, FR007, and R001.	Benchmarks for interbank borrowing costs. DR007 is the central policy target; its deviation reflects instantaneous liquidity stress.
	NCD Market	NCD	1Y State-owned/Joint-stock NCD rates, issuance spreads, and net financing.	Core metric for commercial banks' medium-term liability costs; reflects bank-side liquidity demand and future interest rate expectations.
Sentiment	Market Sentiment	MS	Bond leverage levels (repo volume), futures basis/calendar spreads, and primary auction results.	Quantitative synthesis of risk appetite. Leverage reflects trading heat, while auction results indicate the strength of real demand.
	Equity Correlation	EC	Correlation between CSI 300 and 10Y T-bond yields; turnover ratios; "Stock-Bond Seesaw" intensity.	Monitors cross-asset capital flows. Marginal shifts in relative value drive capital reallocation and impact bond demand.
Rates Bonds	Cash Bond Trends	CBT	10Y T-bond active yields, yield curve slope (10Y-1Y spread), and key level breakouts.	The pricing benchmark and trend core of the market. Curve morphology reflects economic outlooks and trend reversals.
	Futures Perf.	FP	Price and open interest (OI) of lead contracts, volume-to-OI ratio, and net basis (CTD).	Price discovery and sentiment amplifier. OI changes indicate capital direction, while net basis reflects arbitrage and delivery intent.
	Driving Factors	DF	Liquidity conditions, net bond supply, macro data, and policy expectations.	A systematic synthesis of multi-dimensional drivers used for attribution analysis and forward-looking judgment.
Credit Bonds	Overall Perf.	OP	Credit spreads of AAA/AA+ notes; spread divergence between industrial and LGFV bonds.	Systematic risk pricing dashboard. Spread movements reflect credit environment tightening/easing; divergence shows risk appetite structure.
	Trading Char.	TC	High-yield trading ratio, valuation deviations, and primary buyer types (Wealth Mgmt, Funds).	Characterizes market microstructure. Valuation deviations reveal institutional positioning and the sustainability of trends.
Derivatives	Int. Rate Swaps	IRS	1Y/5Y Repo IRS rates and their spreads against Treasury yields.	Reflects expectations for future short-term funding rate averages; isolates pure interest rate expectations from liquidity premiums.
	Volatility	VOL	Implied volatility from bond options and historical yield volatility.	Quantifies market uncertainty. Implied volatility represents future pricing of risk, while historical volatility validates past realized fluctuations.
External	FX Market	FX	USDCNH spot/midpoint deviation; 1Y NDF implied depreciation expectations.	Constraint indicator for cross-border capital flows and policy room. Exchange rate strength influences foreign demand for domestic bonds.
	Overseas Rates	OR	10Y U.S. Treasury (UST) yields; 10Y China-U.S. yield spread.	The global risk-free rate anchor. UST yields exert external pressure on domestic rates via capital flows and exchange rate expectations.
	Precious Metals	PM	International gold prices (USD) and real interest rate (TIPS yield) fluctuations.	Mirror for global risk aversion and real rates. Significant gold price moves often corroborate global recession or haven expectations.
Events	Major Events	ME	Domestic macro data releases, PBOC meetings, and global financial events.	A calendar of deterministic schedules and idiosyncratic risk sources likely to trigger market re-pricing or volatility.
	Expectation Mgmt	EM	Rate hike/cut probabilities implied by derivatives; market reaction before/after events.	Analyzes "priced-in" expectations versus actual outcomes to identify trading opportunities stemming from expectation gaps.

Table 10: FinMAP Daily Market Attribute Taxonomy (L_{day}) with Key Indicators and Financial Logic.

Categories	Items	Abbr.	Details (Description & Indicators)
Economic Theme	Macro-Economic Theme	MET	A high-level synthesis and naming of the core macroeconomic drivers influencing the market during the observation phase.
Economic textbfEnvironment	Economic Period	EP	Identifies the current economic stage (e.g., Recession, Recovery) and its impact on asset allocation and capital flows.
	Monetary Policy	MP	Specific operations, adjustments, and policy rhetoric from the central bank regarding liquidity management, interest rates, and RRR.
	International Env.	IE	Market performance of major overseas economies (e.g., U.S. Treasury yields) and the transmission of global sentiment to the domestic market.
	External Policies	ETP	Monetary policy dynamics of major global central banks (e.g., Fed, ECB) and the resulting shifts in global liquidity.
Key Prices	Open/Close Prices	OCP	Critical price points of the lead contract within the phase, used to characterize the price benchmark and overall amplitude.
	Support Price	SP	Significant price floors derived from technical indicators, historical lows, or psychological thresholds.
	Resistance Price	RP	Significant price ceilings derived from historical highs, upper technical boundaries, or psychological thresholds.
Technical textbfTrends	Moving Tendency	MT	Predictions of short-term price direction and targets based on chart patterns (e.g., W-bottoms, triangles).
	Periodic Lines	PL	Assessment of medium-to-long term trends based on weekly or monthly technical indicators (e.g., MACD, moving averages).
Cyclical textbfFactors	Hist. Same Period	HSP	Statistical average performance or rising probability for the same calendar period over previous years as a historical reference.
	Seasonal Effect	SE	Recurrent market fluctuation patterns caused by seasonal or periodic factors (e.g., quarter-ends, major holidays).
	Calendar Effect	CE	Periodic spikes in market volatility triggered by fixed schedules (e.g., data release dates, central bank meetings).
Events & textbfTimeline	Crucial Events	EVT	Identification and scheduling of critical events occurring within the phase that are likely to impact the market.
	Events Development	ED	Actual progress, results, or specific data released regarding the identified crucial events.
	Expected Trend	ET	Prevailing market expectations or mainstream views regarding ongoing or upcoming key events.
	Event Impact	EI	Specific, quantifiable short-term effects of an event's actual outcome on market prices.
Market textbfSentiment	Initial Sentiment	IS	The dominant market sentiment and its primary drivers at the start of the observation phase.
	Mid-term Sentiment	MS	Shifts in market sentiment and their corresponding triggers during the middle of the phase.
	End-term Sentiment	ES	The final state of market sentiment and its drivers at the conclusion of the phase.
Risk Analysis	Upside Risks	UR	Potential bullish factors or scenarios that could trigger a structural upward breakout in market prices.
	Downside Risks	DR	Potential bearish factors or scenarios that could trigger a structural downward breakdown in market prices.
	Other Risks	OR	Potential risk events that may trigger severe market volatility beyond direct bullish or bearish directions.

Table 11: The Proposed FinMAP Periodic Situation Taxonomy (L_{phase}) for Macro-Level Market Analysis.

Instructions for the Labeler to generate prompts that follow the daily-level FinMAP.

[Documents (public market daily review)]

You are a helpful financial assistant, excellent at information retrieval and summarization. There is a table that illustrates crucial factors that affect the movement of treasury futures:

[Daily-level Financial Market Attribute Protocol]

Please follow this table to identify crucial information from the given documents, and then write a novel market review like this example:

[An example of market review]

Start your writing:

Instructions for the Labeler to generate prompts that follow the period-level FinMAP.

[Documents (daily-level FinMAP prompts)]

You are a helpful financial assistant, excellent at information retrieval and summarization. There is a table that illustrates crucial factors that trace the movement of treasury futures during a period:

[Period-level Financial Market Attribute Protocol]

Please follow this table to collect the relevant information exhibited in the given documents, and then write a [weekly/monthly/quarterly] market review like this example:

[An example of market review]

Start your writing:

Instructions for the Reviewer to score the Labeler-generated prompts.

You are an expert financial analyst specializing in treasury futures. Please evaluate the quality of an AI-generated market review by comparing it against raw documents and the provided FinMAP Taxonomy.

Raw Documents: [Public market reviews]

FinMAP Taxonomy: [Daily/Period-level FinMAP, Table format]

AI-Generated Market Review: [Labeler-generated text]

Assess how effectively the generated prompt extracts and structures information according to the FinMAP schema. Provide a score from 1 to 5 for each of the following criteria:

C1: Taxonomy Alignment: Does the prompt accurately map information to the specific FinMAP categories (e.g., Liquidity, Technical Trends, Risk Analysis)? (*5: Perfectly categorized; 1: Completely ignores the schema.*)

C2: Information Density: Does the prompt capture all key indicators mentioned in the review (e.g., specific OMO volumes, yield basis points, or specific event development)? (*5: No critical data lost; 1: Highly superficial or missing key values.*)

C3: Financial Logicality: Does the prompt maintain the causal relationships described in the review (e.g., "Yields rose *because* of net liquidity drain")? (*5: Causal logic preserved; 1: Causal links are broken or hallucinated.*)

Output your evaluations:

- Taxonomy Alignment Score: [1-5]
- Information Density Score: [1-5]
- Financial Logicality Score: [1-5]
- Revision Comments: [less than 5 sentences]

Table 12: A sample of market review regularized by our daily-level FinMAP.

Market Review of Chinese Treasury Futures (On December 26, 2025)

Liquidity

- **Central Bank Operations:** On December 25, the PBOC conducted a 177.1 billion CNY 7-day reverse repo (rate: 1.40%) and a 400 billion CNY 1-year MLF operation (300 billion CNY MLF matured), achieving a net MLF injection of 100 billion CNY. This marked the 10th consecutive month of MLF net injections. The 2025 MLF net injection totaled 1161 billion CNY, compared to -1986 billion CNY in 2024.
- **Funding Rates:** Overnight rate fell 0.55bps to 1.2189% (lowest since August 2025); 7-day rate rose 10.42bps to 1.4442% (1-month high). DR007 weighted average remained below 1.26%.
- **CD Market:** 1-month NCD rate increased 2.81bps to 1.85% (1-month high), while 1-year NCD rates remained stable.

Market Sentiment

- **Risk Appetite:** Year-end trading activity remained muted. Risk sentiment was generally neutral, with a brief selloff in the afternoon due to market rumors ("small essay") followed by a rebound after clarification.
- **Equity Correlation:** Shanghai Composite fell 0.44%, with over 4000 stocks declining across both exchanges. The Wind Real Estate Bond Index dropped 1.2%.

Interest Rate Bonds

- **Bond Yields:** 10-year yield showed divergence, with long-term bonds (10Y+) weakening (10Y rose 0.5bp, 30Y rose 0-0.5bp).
- **Futures Performance:** Treasury bond futures (TL) closed down 0.09%, with the 10-year main contract trading at 101.24%.
- **Key Drivers:** Stable liquidity failed to drive yields lower, while year-end supply pressure weakened short-term bonds.

Credit Bonds

- **Overall Performance:** High-grade bonds remained stable, while lower-grade spreads widened. Vanke bonds underperformed ("22 Vanke 14", "22 Vanke 02", etc.). The Wind High-Yield Municipal Bond Index fell 0.01%.
- **Trading Characteristics:** High-grade bonds traded at par, while lower-grade bonds traded 0.5-1bp below valuation.

External Linkage

- **FX Market:** Onshore CNY closed at 7.0050 (+5bps), while offshore CNY surged past 7.00 (15-month high). USD Index stabilized above 99.
- **Overseas Rates:** 2Y US Treasury rose 1.47bps, 10Y US Treasury fell 0.39bps.
- **Precious Metals:** Gold reversed gains to return to opening levels; silver traded sideways.

Events

- **Key Events:**
  - MOF and 8 other ministries issued climate disclosure guidelines for enterprises (voluntary implementation).
  - CSRC reported 2 trillion CNY in sci-tech innovation bonds issued since 2021.
- **Expectation Management:** Market focused on PBOC liquidity operations and fiscal policy intensity. Expectations for improved liquidity post-holiday.

Key Takeaways

- Bond market remained range-bound, with stable liquidity but limited yield declines.
- Real estate bonds underperformed (index fell 1.2%), while offshore CNY broke above 7.00.
- Attention shifted to PBOC's post-holiday liquidity operations and US CPI data.

Table 13: A sample of a weekly market review regularized by our period-level FinMAP.

Market Review of Chinese Treasury Futures (December 2–6, 2024)

Economic Theme

- **Year-end liquidity gaming and policy expectation building.** Weak PMI below 50 reinforced weak recovery consensus. CPI stayed low, deflationary pressure persisted. Market focus shifted from policy magnitude to sustainability of cross-cycle monetary adjustment. RRR cut expectations diverged; MLF oversubscription widely expected.

Economic Background

- **Economic Cycle:** Policy-bottom to economic-bottom transition. Credit expansion below seasonal. Corporate medium-to-long term loan growth stable but not rebounding. Institutional allocation demand dominant.
- **Monetary Policy:** PBOC maintained minimal daily reverse repos (RMB 20–40bn), net weekly withdrawal of RMB 12bn. Q4 MPC meeting minutes (Dec 7) shifted wording from "strengthening counter-cyclical adjustment" to "precise and effective implementation," interpreted as lower RRR probability but stronger structural tools.
- **International Environment:** US 10Y yield ranged 4.2%–4.3%. China-US spread (10Y) deeply inverted at -130 to -140bp. Foreign outflow pressure marginally eased.
- **External Policy:** Eurozone inflation cooled, markets pricing ECB cut in Q1 2025. USD Index 103–104. USD/CNY near 7.25, neutral impact.

Key Levels

- **Start/End:** T2503 opened 102.85, closed 102.92 (+0.07, range 0.25%). Active 10Y cash yield fell 1.5bp to 2.63%.
- **Support:** Futures: 102.70 (20DMA), 102.55 (Nov consolidation low). Cash: 2.65% psychological.
- **Resistance:** Futures: 103.00 handle, 103.15 (Nov high). Cash: 2.60% strong resistance.

Technical Trend

- **Pattern:** Late-stage wedge consolidation on daily chart. Volume shrank from 38k to 21k lots. Open interest -3k to 185k lots, reflecting pre-policy window deleveraging.
- **Monthly:** Small bullish candle. MACD red bars narrowing, fast line nearing bearish crossover. Price remained above 10M and 20M MA, medium-term trend intact.

Cyclical Factors

- **Historical Seasonality:** Past 5 years (2019–2023) first week of December average +0.12%, 60% win rate, driven by year-end liquidity easing.
- **Seasonal Effects:** December faces triple funding pressure: bank year-end accounting, wealth management repatriation, government bond settlement. Money market rates typically rise 20–30bp, constraining long-end; short-end (1–3Y) outperforms on allocation demand.
- **Calendar Effects:** Early December pre-Central Economic Work Conference quiet period, highly sensitive to policy rumors on deficit ratio or special bonds.

Timeline and Key Events

- **Dec 2:** Caixin Services PMI 49.8 (prev 50.2), weaker than expected. Futures opened +0.05.
- **Dec 4:** Trade data: Nov exports -3.5% YoY (USD). MLF: RMB 400bn matured, PBOC injected RMB 450bn (net +50bn) at 2.50% unchanged. Market interpreted as "quantity-up, price-flat" neutral stance. Futures printed long upper shadow, long profit-taking evident.
- **Event Expectations:** Focus shifted to Dec 9 CPI/PPI release and mid-December Central Economic Work Conference. Consensus expects "stability with progress" tone, deficit ratio 3.2%–3.5%, below aggressive forecasts of 3.8%.
- **Event Impact:** PBOC net withdrawal offset by MLF injection. DR007 rose from 1.75% to 1.82% but stayed near policy rate 1.80%, limited bond impact.

Market Sentiment

- **Early (Dec 2–3):** Cautiously optimistic. Weak PMI reinforced "weak reality" narrative, but concerns over concentrated local government special bond issuance in December (RMB 300bn quota) subdued position-taking.
- **Mid (Dec 4–5):** Wait-and-see. MLF outcome cooled RRR cut expectations, profit-taking emerged. Declining open interest reflected active long liquidation, but prop desks and rural commercial banks provided strong support above 2.65% in cash bonds.
- **Late (Dec 6):** Marginally improved. Friday afternoon rumor of "PBOC symposium on RRR cuts over weekend" plus A-share decline drove safe-haven flows. T2503 closed near session high. Wind Bond Sentiment Index rebounded from 48 to 52.

Risk Analysis

- **Upside:** If Central Economic Work Conference sets 2025 deficit ratio above 3.5% with ultra-long special bonds, plus 25bp RRR cut, 10Y yield breaks below 2.60%, futures target 103.30 (cash 2.55%).
- **Downside:** If CPI YoY >0.5% or Fed dot plot shows only one 2025 cut, China-US spread widens beyond -150bp, foreign selling triggers 10Y yield rebound to 2.75%, futures retreat to 102.40.
- **Other:** Year-end credit bond default risk (property sector) could spark liquidity crunch and indiscriminate rates selling. Surge in NCD issuance (1Y NCD rate 2.52% to 2.58%) may divert bond allocation funds.

Table 14: A sample of a monthly market review regularized by our period-level FinMAP.

Market Review of Chinese Treasury Futures (December 2–31, 2024)

Economic Themes

Escalating Supply Pressure. The Ministry of Finance announced an intensified issuance schedule for ultra-long special treasury bonds in early December, with a focus on the 30-year tenor in Q4. Markets widely anticipated a significant increase in net supply. Primary market demand weakened, with bid-to-cover ratios falling to 2.1x, triggering a self-reinforcing feedback loop: "harder to sell → lower prices → greater investor reluctance."

Pronounced Policy Divergence. On December 10, the U.S. Federal Reserve cut rates by 25 bps to 3.50%–3.75%, yet internal dissent was evident. Meanwhile, on December 15, the PBOC maintained the MLF rate unchanged while injecting only CNY 100 billion net liquidity—dashing market expectations for easing. This policy misalignment led to oscillating sentiment between caution and optimism.

Credit Risk Flare-up. On December 15, Vanke's proposed extension for two medium-term notes totaling CNY 5.7 billion was rejected by bondholders. Although granted a 30-day grace period, the event intensified concerns over real estate sector solvency.

Economic Background

Economic Cycle: Late Recession The economy remains in the late stage of recession, with deteriorating corporate earnings. Some institutional investors began reducing long-duration holdings to manage NAV volatility, creating nonlinear selling pressure ("selling more as prices fall").

Monetary Policy: Moderately Accommodative but Constrained In December, the PBOC conducted a net MLF injection of CNY 100 billion and purchased CNY 50 billion in government bonds—marking the third consecutive month of such operations. However, the unchanged MLF rate on December 15 fueled fears of an end to the "asset shortage" regime. On December 18, the PBOC Monetary Policy Committee signaled flexibility, stating it would "efficiently deploy a variety of tools, including RRR cuts and rate reductions".

International Environment: Weaker USD, Volatile UST Yields Following the Fed's rate cut, the U.S. dollar index weakened, accelerating foreign inflows into Chinese assets. However, the U.S. 10-year Treasury yield remained elevated at 4.127%, sustaining a deep inversion in the China-U.S. yield spread of approximately –150 bps, limiting the appeal of RMB-denominated bonds.

External Policy: Fed Dissent Amplifies Volatility The Fed's December meeting minutes revealed significant disagreement among officials regarding economic risks and the policy path forward. While the median FOMC projection for 2026 implied only one rate cut, market pricing anticipated 50 bps of easing—widening the expectation gap and amplifying market swings.

Key Price Levels

Opening/Closing Levels

- Early December (Dec 1): T2603 opened at 108.33 and closed at 108.21; TL2603 opened near 113.89.
- Mid-December (Dec 12): T2603 closed at 107.985 (–0.21% from Dec 1); TL2603 closed at 112.470 (–1.25%).

Support Levels

- Technical: T2603 found short-term support near 107.60 (Dec 4 low); TL2603 broke below its key 112.80 support on Dec 4.
- Psychological: T2603 at 107.00; TL2603 at 112.00.

Resistance Levels

- Technical: T2603 faces resistance at 108.30 (prior high); TL2603 at 113.00.
- Psychological: T2603 at 108.50; TL2603 at 113.50.

Technical Trends

Trend Evolution

- T2603 formed a "V-shaped" rebound after the Dec 4 low of 107.61, recovering to 108.08 by Dec 18—but failed to breach the 108.30 resistance.
- TL2603 broke below 112.80 on Dec 4 and continued declining to 112.45, confirming bearish momentum on the long end.

Monthly Chart Perspective

- T2603 traded below the Bollinger Band midline in early December but moved back above it by Dec 12—yet remained capped by the upper band.
- TL2603 exhibited narrowing Bollinger Band width, reflecting sharply reduced risk appetite for ultra-long duration.

Cyclical Factors and Historical Patterns

Historical Performance (First Half of December)

- 2020: T2103 +0.7%, 2021: T2203 +0.5%, 2023: T2403 +0.2%, 2024: T2412 +0.8%
- **2025: T2603 –0.47%**, significantly underperforming the historical average gain of 0.5%–0.8% and contrasting with the typical 70%–80% upside probability.

Seasonal Effects

- Historically, early December faces tax payments and year-end funding withdrawals, pushing DR007 higher.
- In 2025, DR007 stayed below 1.4%, indicating markedly eased liquidity pressure due to sustained PBOC net injections.

Calendar Effects

- Key recurring dates—MLF operations around Dec 15 and PBOC committee meetings around Dec 18—aligned with 2025's timeline.
- The broad selloff on Dec 4 resembled historical patterns during "data-dense windows" with insufficient policy communication.

Timeline of Key Events

Key Events

- Dec 4: TL2603 dropped to 113.670, a >1-year low, down >2.3% from mid-November highs.
- Dec 10: Fed cut rates by 25 bps amid internal dissent.
- Dec 15: PBOC kept MLF rate unchanged (net CNY 100B injection); Vanke's debt extension failed.
- Dec 18: PBOC signaled potential RRR/rate cuts.
- Dec 25: PBOC conducted an additional CNY 400B MLF operation (interpreted as year-end liquidity management).

Event Developments & Impact

- Dec 4: TL2603's plunge reflected concerns over long-term growth, fiscal sustainability, or rate trajectory.
- Dec 10: Fed cut weakened USD, boosting foreign inflow expectations—but domestic long-end yields remained pressured by supply.
- Dec 15: T2603 hit 107.61; TL2603 broke 112.80—driven by policy disappointment and credit shock.
- Dec 18: T2603 rebounded to 108.08 on PBOC's dovish signal; TL2603 remained weak.
- Dec 25: T2603 showed limited recovery; TL2603 fell further to 112.51—highlighting the gap between policy expectations and reality.

Market Expectations vs. Reality

- Early Dec: Expected PBOC easing → bullish sentiment.
- Mid-Dec: Anticipated unchanged MLF rate + rising supply → bearish long end.
- Late Dec: Hoped for extra liquidity → range-bound outcome.

Market Sentiment

- **Early (Dec 1–3): Cautiously Optimistic** — Supported by PBOC liquidity injections and lack of negative data. Drivers: PBOC operations, data vacuum.
- **Mid (Dec 4–15): Turning Cautious** — Long-end weakness and "end of asset shortage" fears dominated. Credit concerns emerged but did not directly spill into rates. Drivers: supply, policy divergence, credit risk.
- **Late (Dec 18–25): Sentiment Recovery** — Short-end stabilized on PBOC signals; long-end remained depressed. Drivers: policy rhetoric, liquidity support.

Risk Analysis

Upside Risks

- Persistent economic underperformance could push T2603 above 108.50.
- Surprise PBOC rate/RRR cut could ignite bullish momentum.

Downside Risks

- Inflation rebound could drive T2603 below 107.50.
- Fed policy reversal strengthening USD could trigger foreign outflows.

Other Risks

- Credit contagion (e.g., Vanke default) causing liquidity shocks.
- Geopolitical escalation increasing rate volatility via safe-haven flows.