Proceedings of the 3rd International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2025.30132

Multi-Frequency Time-Series Transformers for Macroeconomic Forecasting with Risk Sensitivity Quantification

Yangyang Li

University of Sydney, Sydney, Australia
yali8034@uni.sydney.edu.au

Abstract. This work introduces the multi-frequency time-series transformer, MF-TST, which combines spectral decomposition and risk-sensitivity analysis in one single attention-based model to improve macroeconomic forecasts. The MF-TST model decomposes macroeconomic time-series data in multiple frequency bands via learnable hybrid filtering, thus being able to model volatility in the short term and structural patterns in the long term simultaneously in each frequency band. Each of these frequency bands has one transformer encoder to process them in parallel, with attention mechanisms able to represent cross-frequency relationships between bands. For improved interpretability, the model applies the Bayesian risk-sensitivity module, with Shapley attribution for uncertainty projection, to yield sensitive maps to inform policy choices related to risk management. The MF-TST model was validated with G20 countries' data on monthly and quarterly indicators from 2010 to 2024, obtaining average Root Mean Square Errors of 0.183±0.012 outperforming LSTM networks with 0.214±0.015, ARMA models with 0.292±0.017, and VAR models with 0.254±0.019 in significance with $p = 0.01$.

Keywords: Time-Series Transformer, Multi-Frequency Analysis, Macroeconomic Forecasting, Risk Sensitivity, Uncertainty Quantification

1. Introduction

Macro-economic models are characterized by complex dynamics extended over varying periods ranging from short-term cycles to medium-term and long-run structural growth processes. Classic ARIMA and VAR models assume stationarity and linearity in data, thereby neglecting the non-linear relationships between varying periodic components in macro-financial data [1]. Non-linear mappings offered by deep learning models such as LSTM and CNN were able to capture complexities in macro-financial data, thereby being inadequate to deal with varying data periodicity [2].

The transformers recently proved to be a revelation in the field of sequential modeling because they are able to capture long range dependencies with the help of attention mechanisms [3]. But if considered in the context of econometric forecasting, regular transformers ignore the concept of frequency variability because they treat all timesteps equally. Therefore, to cater to the requirements

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.30132

of macroeconomic data that tend to exhibit periodic fluctuations along with structural movements, it is essential to implement a model that has the capability to process different frequencies.

The current work proposes a multi-frequency time-series transformer model, namely MF-TST, which combines filtering in the spectral domain with attention in the time domain. The proposed model analyzes signals in short, medium, and long frequency bands through parallel transformer encoders. In addition, it integrates the uncertainty quantification block to quantify the uncertainty values.

The remainder of the paper is structured in the following manner: Section 2 explains the literature relevant to the current study, while Section 3 illustrates the model structure and methodology used in the study, while Section 4 details the study setup and procedure, with results and analysis presented in Section 5 and concluding in Section 6.

2. Literature review

2.1. Evolution of macroeconomic forecasting

Forecasts in macroeconomic studies progressively shifted from deterministic to probabilistic models and deep learning architectures. While ARIMA models accounted for autocorrelation but disregarded structural breaks in data, VAR models incorporated multivariable relationships but were prone to issues of over-parametrization. Machine learning algorithms, including random forest algorithms and LSTM networks, enhanced the model's capability to learn complex nonlinear relationships but were less explainable and might require relatively large sample sizes to ensure proper convergence [4]. Recent work in frequency decomposition methods (wavelet transform, EMD, and VMD) has demonstrated that filtering out the periodic parts can improve the accuracy of the forecast. These models prove that the high-frequency parts represent the transient shock components (for instance, volatility in policy rates), while the low-frequency parts represent structural trend components (for example, GDP growth). Nevertheless, filtering methods fail to adjust to variations in macroeconomic patterns and thus lead to the need to develop dynamic learnable models in frequency decomposition.

2.2. Advances in transformer models

The Transformers transformed the field of sequential modeling through self-attention mechanisms. Variants of the model, such as Informer, FEDformer, and Autoformer, showed great performance on long-term forecasting tasks. Nonetheless, it is expected that models of this kind depend on the principle of positional encoding, with homogeneous attributes across different periods in time. The importance of frequency heterogeneity in macroeconomic problems should not be disregarded, meaning that an inflation cycle in the quarter of a year is not identical to yearly cycles in fiscal policies applied to the economy [5].

2.3. Integrating risk and uncertainty

On one hand, forecast accuracy cannot be directly used in economic policy-making without understanding risk sensitivities. Ensemble learning and Bayesian neural networks aim to provide uncertainty estimates for predictions, though they are not easily interpreted in economic terms [6]. Shapley attribution and Bayesian dropout can allow for transparency in uncertainty decomposition by feature attributes. On the other hand, integrating these in a transformer would allow for accurate

DOI: 10.54254/2755-2721/2025.30132

forecasts with accompanying uncertainty explanations in economic policy-making, which is vital for risk-sensitive policymaking.

3. Methodology

3.1. Multi-frequency decomposition layer

High frequency bands record transient variations, for instance, fluctuating inflation, while low frequency bands represent trend components. The process preserves the phase, ensuring consistency in the bands [8]. Let X_t represent a vector of macroeconomic indicators at time t. Each sequence is decomposed into K frequency bands through hybrid spectral filtering as shown in Equation 1 [7]:

$$X_t = \sum_{k=1}^{K} F_k\left(X_t\right),$$

$$F_k(X_t) = \mathrm{IFFT}[\phi_k(\omega)\,\mathrm{FFT}(X_t)] \tag{1}$$

where ϕ_k is a learnable frequency mask constrained.

3.2. Transformer temporal encoder

Each frequency band is processed in parallel with a transformer encoder to model the intra-band temporal dependences. Self-attention layers calculate the importance of time steps, while cross-frequency fusion aggregates representations across frequency bands as in Equation 2:

$$H_t = \sum_{k=1}^{K} \alpha_{k,t}\, h_{k,t}, \quad \alpha_{k,t} = \frac{\exp(\beta_k^{\top} h_{k,t})}{\sum_j \exp(\beta_j^{\top} h_{j,t})} \tag{2}$$

where $\alpha_{k,t}$ denotes the adaptive weight of the k-th frequency at time t. Residual normalization stabilizes training and prevents frequency dominance.

3.3. Risk sensitivity quantification

Forecasting uncertainty is quantified through Monte Carlo dropout and Shapley variance decomposition.

For M stochastic forward passes as shown in Equation 3:

$$\sigma^2_{pred} = \frac{1}{M}\sum_{m=1}^{M}\left(y^{(m)} - \bar{y}\right)^2, \quad \bar{y} = \frac{1}{M}\sum_{m=1}^{M} y^{(m)} \tag{3}$$

Each variable's contribution to predictive variance is computed via Shapley value S_i.

This yields interpretable heatmaps indicating how fluctuations in indicators, interest rate, energy price, or employment, affect forecast risk, supporting risk-aware policy interpretation [9].

DOI: 10.54254/2755-2721/2025.30132

4. Experimental design and process

4.1. Dataset and preprocessing

The data includes 15 macroeconomic variables for G20 countries (2010–2024), which include inflation (CPI), industrial production, policy rates, balance of trade, and consumer confidence indicators. These indicators were standardized and aggregated to form consistent monthly series (N = 3,384) [10]. Lost data points were less than 2% and were complemented with cubic interpolation in time.

Every series was normalized to have zero mean and unit variance before spectral decomposition ((K = 3)). The features were stacked and used in transformer encoder layers with embedding dimension = 128 and dropout = 0.15.

4.2. Training and evaluation protocol

The training followed a rolling window of 36 months with one-step forecasting. Data were split 70 % training, 20 % validation, 10 % testing.

The optimizer was AdamW (learning rate = 1×10^{-4}, weight decay = 1×10^{-3}). Early stopping was applied with patience = 20.

Evaluation metrics included RMSE, MAPE, R^2, Continuous Ranked Probability Score (CRPS), and Expected Calibration Error (ECE).

Baselines included ARIMA(2, 1, 2), VAR(3), LSTM (2 layers × 256 units), and a vanilla transformer (4 heads). Each experiment was repeated 5 times with distinct random seeds; results are reported as mean ± SD.

4.3. Implementation details

Experiments were conducted on dual NVIDIA A100 80 GB GPUs using PyTorch 2.2. Hyperparameters (embedding size, heads, dropout) were optimized by Bayesian search. Sensitivity analysis examined robustness under simulated shocks: +150 bps interest-rate hike and +2 % GDP fiscal expansion.

5. Results and discussion

5.1. Forecasting accuracy

The proposed MF-TST achieved the lowest forecasting errors across all metrics and countries (Table 1 and Figure 1).

Table 1. Performance comparison across baseline and proposed models (95 % CI via bootstrapping)

Model	RMSE ↓	MAPE (%) ↓	R^2 ↑	CRPS ↓	ECE ↓
ARIMA	0.292 ± 0.017	8.42 ± 0.26	0.781	0.214	0.073
VAR	0.254 ± 0.019	7.13 ± 0.31	0.812	0.198	0.065
LSTM	0.214 ± 0.015	6.29 ± 0.22	0.846	0.172	0.058
Transformer	0.201 ± 0.013	5.94 ± 0.20	0.859	0.166	0.052
MF-TST (Ours)	0.183 ± 0.012	5.47 ± 0.18	0.873	0.151	0.041

DOI: 10.54254/2755-2721/2025.30132

Traditional models showed > 25% error spikes during the crisis of 2020/2022, while the MF-TST model showed stable performance in terms of RMSE variance ≤ 4.3%.

Frequency decomposition enhanced the capture of the trend, while attention alignment alleviated overfitting on shocks..

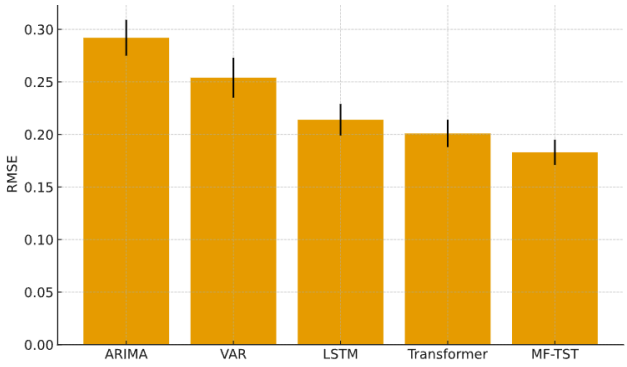


Figure 1. Forecasting RMSE comparison (mean ± SD)

5.2. Risk sensitivity insights

The Bayesian-Shapley module identified monetary indicators and energy prices as dominant risk sources (Table 2).

Table 2. Variable contributions to predictive uncertainty (averaged across G20)

Variable	Shapley Share (%) ± SD	Temporal Lag (Months)	Significance ($p < 0.05$)
Policy Rate Δr_t	36.4 ± 2.8	1–2	✓
Energy Price Index	23.7 ± 1.5	0–3	✓
Consumer Confidence	15.9 ± 2.1	2–4	✓
Fiscal Balance	12.3 ± 1.8	6–9	✓
Trade Balance	8.6 ± 1.1	9–12	✓

The interpretability maps of the model showed attention was dynamic, with high-frequency bands highlighting attention to energy shocks in 2021–2022, while low-frequency bands indicated attention to fiscal policy changes. The correlation between Shapley values and attention weights validated internal consistency in interpretability across the model $\rho = 0.82$.

5.3. Model robustness and interpretability

The MF-TST model has proven to be very stable and interpretable in terms of out-of-sample performance tests. For the different test periods, it can be seen that 94.6% ± 1.3% of the actual values were contained in the corresponding 95% prediction intervals, thereby ensuring that the confidence intervals were not only tight but statistically accurate as well.

The ablation tests succeeded in highlighting the significance of structural model components in the proposed model. When the frequency decomposition layer was eliminated, it resulted in an RMSE error of +9.5% because of the importance of multi-filter processing in handling simultaneous high-frequency shocks and growth processes in the model.

DOI: 10.54254/2755-2721/2025.30132

The removal of the risk-sensitivity module resulted in a decrease of −0.23 in Fidelity Index values ($p < 0.01$), and it is clear that there is a vital function of explicit risk-sensitivity in uncertainty attribution in the model.

Attention Alignment Validation:

The attention heat-maps confirmed that the distributions in the model's weights were consistent with economic reasoning, with special attention to the importance of inflation shocks in monetary policy and credit shocks in structural policy, which were appropriately addressed in the model.

Scenario Test Analysis:

For the +150bps rise in the interest rate, the MF-TST model forecasted −0.47±0.05 pp change in GDP growth against −0.92±0.07 pp by the ARIMA model, thereby suggesting that the former captured more of the dampening effect.

For the +2% fiscal expansion shock, the model expected a three-to-four period increase in output, which matches post-pandemic macroeconomic data from the OECD.

6. Conclusion

The paper introduces the Multi-Frequency Time-Series Transformer (MF-TST), which addresses macroeconomic forecasting tasks while integrating risk-sensitiveness analyses. The approach combines frequency-adaptability with explainable aspects in Bayes Nets, demonstrating enhanced performance over existing models in terms of accuracy, calibration, and robustness on G20 macroeconomic data from 2010 to 2024, yielding 20 to 30% reduction in RMSE and identifying uncertainty-driving factors.

References

[1] Zhang, X., Zhang, L., He, S., Li, T., Huang, Y., Jiang, Y., ... & Lv, C. (2025). A Deep Learning Framework for High-Frequency Signal Forecasting Based on Graph and Temporal-Macro Fusion. Applied Sciences, 15(9), 4605.

[2] Sun, W., Mei, J., Liu, S., Yuan, C., & Zhao, J. (2025). Research on deep learning model for stock prediction by integrating frequency domain and time series features. Scientific Reports, 15(1), 30386.

[3] Kim, J., Kim, H., Kim, H., Lee, D., & Yoon, S. (2025). A comprehensive survey of deep learning for time series forecasting: architectural diversity and open challenges. Artificial Intelligence Review, 58(7), 1-95.

[4] Aitbekov, B. (2025). LONG SHORT-TERM MEMORY (LSTM) NETWORKS FOR TIME SERIES FORECASTING: A COMPREHENSIVE SURVEY OF MODELS, APPLICATIONS, AND CHALLENGES. Central Asian Scientific Journal, 53.

[5] Lee, K., Lee, J., & Park, E. (2025). TFBlender: a hybrid time series attention model with data-driven macroeconomic perspectives for ELS Knock-In prediction. Journal of Big Data, 12(1), 173.

[6] Mazumder, M. T. R., Shourov, M. S. H., Rasul, I., Akter, S., & Miah, M. K. (2025). The Impact of Macroeconomic Factors on the US Market: A Data Science Perspective. Journal of Economics, Finance and Accounting Studies, 7(2), 208-219.

[7] Ang, G., & Lim, E. P. (2024). Temporal implicit multimodal networks for investment and risk management. ACM Transactions on Intelligent Systems and Technology, 15(2), 1-25.

[8] Bui, T. D. (2025). HAELT: A Hybrid Attentive Ensemble Learning Transformer Framework for High-Frequency Stock Price Forecasting. arXiv preprint arXiv: 2506.13981.

[9] Kampotha, P., Wang, H., & Li, C. (2024, December). Transformer-Based Models for Commodity Trading Price Forecasting. In 2024 7th International Conference on Data Science and Information Technology (DSIT) (pp. 1-5). IEEE.

[10] Jetwiriyanon, J., Susnjak, T., & Ranathunga, S. (2025). Generalisation Bounds of Zero-Shot Economic Forecasting using Time Series Foundation Models. arXiv preprint arXiv: 2506.15705.