APPLIED ECONOMICS LETTERS
https://doi.org/10.1080/13504851.2025.2497429

Routledge
Taylor & Francis Group

Neural Titans in market prediction: MLP, transformer, & hybrid models across G-7 and China

Arab Dahir Hassan and Mahat Maalim Ibrahim

School of Business, Department of Economics, Ibn Haldun University, Istanbul, Türkiye

ABSTRACT

This study aims to conduct a comparative evaluation of eight state-of-the-art forecasting models – TimeMixer, PatchTST, iTransformer, NHITS, NBEATS, SOFTS, RMoK, and BiTCN – representing diverse deep learning architectures, neural basis expansion techniques, and hybrid approaches, for predicting stock market prices. We assess their performance across seven major global indices: the Shanghai Stock Exchange, S&P/TSX Composite, FTSE 100, DAX, CAC 40, S&P 500, and Nikkei 225, using rigorous metrics (MAE, SMAPE and RMSE). Our findings indicate that neural basis expansion models (NHITS, NBEATS) achieve superior overall accuracy (aggregated MAE: 0.013–0.014), particularly in North American and Asian markets. In contrast, transformer-based architectures exhibit market-specific strengths, with iTransformer delivering exceptional performance on Canada's S&P/TSX (MAE: 0.003). Notably, European indices (DAX, CAC 40) present significant challenges, where BiTCN and RMoK underperform (MAE: 0.032–0.038), suggesting limitations in modelling abrupt volatility shifts characteristic of these markets. These results highlight critical regional performance variations and provide insights into architectural efficacy under diverse market conditions.

KEYWORDS
Financial market forecasting; TSMixer; iTransformer; N-BEATS; N-HiTS

JEL CLASSIFICATION
G17; G15; G00

I. Introduction

Stock price prediction has emerged as a critical domain that captivates a diverse array of professionals, from financial analysts and economists to institutional and retail investors (Harel and Harpaz 2024). This widespread interest is fundamentally rooted in the potential for substantial returns, coupled with the inherent market inefficiencies that create exploitable opportunities (Fama 1970). The past decade's financial landscape has faced unprecedented shocks, testing market resilience and investor confidence. Endogenous crises like the 2008 financial meltdown and exogenous events such as COVID-19 and geopolitical tensions have fuelled uncertainty, often triggering herding behaviour among investors (F. Wu et al. 2024), underscoring the crucial role of market sentiment in investment decisions, a phenomenon extensively documented in the literature (John and Li 2021; Zheng, Osmer, and Zhang 2018).

The evolution of stock market forecasting methodologies reflects the growing sophistication of analytical techniques. Traditional approaches, such as the univariate ARIMA models (Box et al. 1970), while foundational, have demonstrated limitations in capturing complex cross-variate relationships. Generative Additive Models (GAM) gained prominence as an alternative framework (Parray et al. 2020), followed by the widespread adoption of traditional Machine Learning (ML) algorithms (Vijh et al. 2020). Deep learning architectures, particularly Long Short-Term Memory (LSTM) networks, have shown promising results in stock price prediction (Sharma, Gupta, and Gupta 2021). However, these approaches often exhibit diminishing accuracy in long-horizon forecasting scenarios (Harel and Harpaz 2024).

The limitations of conventional models have catalysed a shift towards Transformer-based architectures, which have demonstrated remarkable success across various domains, notably in natural language processing (Kalyan, Rajasekharan, and Sangeetha 2021). In the time series domain, innovations such as Autoformer (H. Wu et al. 2022), Informer (H. Zhou et al. 2021), and FEDformer (T. Zhou et al. 2022) have established new performance

CONTACT Arab Dahir Hassan arabhassan@ihu.edu.tr School of Business, Department of Economics, Ibn Haldun University, Istanbul, Türkiye
Supplemental data for this article can be accessed online at https://doi.org/10.1080/13504851.2025.2497429

benchmarks. However, recent empirical evidence has challenged the presumed superiority of multivariate Transformer architectures in leveraging cross-variate information (Zeng et al. 2022). Notably, multilayer perceptron (MLP) models, particularly N-BEATS, have demonstrated superior performance across various forecasting tasks. N-BEATS uses stacked fully connected layers with backward and forward residual connections, excelling in capturing trends and seasonality without complex feature engineering (Oreshkin et al. 2020). Its evolutionary successor, N-HiTS, further enhances efficiency and accuracy by introducing hierarchical interpolation and multi-rate sampling, making it particularly effective for long-horizon forecasts, demonstrating superior performance across various forecasting tasks (Challu et al. 2022).

In parallel, novel architectures like TimeMixer have emerged, introducing a multiscale approach to time series forecasting. TimeMixer employs Past-Decomposable-Mixing (PDM) blocks to capture historical patterns and Future-Multipredictor-Mixing (FMM) blocks to predict future trends, enabling effective information extraction and accurate forecasting (Chen et al. 2023). Similarly, iTransformers have adapted the Transformer architecture for time series forecasting by applying attention and feed-forward networks to inverted dimensions, enhancing performance while preserving the original structure (Y. Liu et al. 2024). PatchTST further advances this domain by dividing input data into smaller patches, treating each as an independent time series to capture local patterns and dependencies, improving accuracy and computational efficiency (Nie et al. 2023).

Further advancing multivariate forecasting, SOFTS introduces the STAR module, which aggregates series into a global representation via centralized processing. This design enables linear-complexity channel interactions, robustness to distribution shifts, and achieves state-of-the-art results (Z. Liu et al. 2024). Moreover, Decision-makers often prefer probabilistic forecasts for their uncertainty quantification, but these often require separate models for each time series, increasing complexity. BiTCN by Sprangers, Schelter, and de Rijke (2023) addresses the challenge of probabilistic forecasting for multiple time series by combining bidirectional processing with temporal convolutions, capturing past and future temporal dependencies to improve accuracy and efficiency.

Finally, the RMoK architecture, based on Kolmogorov-Arnold Networks (KAN), introduces a novel approach by combining reversible neural networks with mixture-of-experts principles. This design reduces memory usage while enhancing model flexibility and scalability, making it suitable for complex forecasting tasks (L. Han et al. 2024).

This study makes two key contributions. First, it implements an ensemble of advanced time series forecasting techniques to analyse stock market trajectories across G-7 nations and China. While previous studies have explored various models for stock market analysis using traditional ARIMA models (Ho, Darman, and Musa 2021), machine learning (Khan et al. 2022), and deep learning techniques (Gupta 2024). This study integrates recently developed methodologies that remain underutilized in the literature, offering a comprehensive assessment of their effectiveness in capturing complex market dynamics. Second, it compares these advanced techniques, evaluating their predictive accuracy and computational efficiency across varying market conditions. To the best of our knowledge, this is among the first studies to apply these models to multi-market forecasting, an area largely unexplored using these emerging methodologies introduced within the past 3 years.

II. Dataset and description

This study utilizes daily closing price data (in natural logarithms) from the G-7 countries and China, obtained from Yahoo Finance from 2 July 1997, to 6 March 2025. Following data cleaning, each series contains 5,923 observations as demonstrated in Table 1. The dataset was partitioned into a training set and a holdout sample comprising the most recent 24 observations for validation and testing purposes.

III. Methodology

This investigation employs a sophisticated ensemble of contemporary forecasting architectures consisting of MLP-based (N-BEATS (Oreshkin et al. 2020), N-HiTS (Challu et al. 2022), TSMixer (Chen

Table 1. Descriptive statistics.

Stock	Obs	Mean	Std. Dev.	Min	Max
GDAXI	5923	8690.5	4206.804	2202.96	23419.5
GSPTSE	5923	13003.3	4518.454	5336.2	25698.5
N225	5923	17356.7	7503.688	7054.98	42224
FTSE	5923	6137.53	1088.074	3287	8871.3
FCHI	5923	4797.54	1269.798	2403.04	8239.99
SS	5923	2498.39	901.973	1011.499	6092.06
GSPC	5923	2047.96	1250.128	676.53	6144.15

et al. 2023)), Transformer-based (PatchTST (Nie et al. 2023), BiTCN (Sprangers, Schelter, and de Rijke 2023), iTransformer (Z. Liu et al. 2024)), and hybrid models (SOFTS (X. Han et al. 2024), and RMoK (X. Han et al. 2024)) to analyse G-7 and China's stock market price dynamics. These models were selected for their distinctive methodological contributions to time series forecasting.

N-BEATS enables interpretable forecasting via its neural backbone, facilitating cross-temporal transfer learning. N-HiTS extends this with hierarchical decomposition, improving multi-scale dependency modelling. PatchTST employs subseries patching and channel independence for long-sequence efficiency. BiTCN combines bidirectional TCNs with attention to capture forward/backward dependencies. iTransformer swaps temporal/feature axes, boosting long-sequence efficiency and feature interactions. TSMixer blends temporal/cross-variate data through dimension mixing, yielding a lightweight solution.

SOFTS introduces a state-space framework optimized for multivariate time series forecasting with Series-Core Fusion, effective in volatile market conditions. RMoK implements a robust mixture of Kolmogorov-Arnold Network (KAN) Y. Liu et al. (2024) approach, excelling in handling non-stationary time series patterns.

IV. Results and analysis

Model evaluation is essential for assessing accuracy, reliability, and real-world applicability. This

Table 2. Model evaluation.

model	mae	smape (%)	rmse
TimeMixer	0.016	0.084	0.018
PatchTST	0.015	0.077	0.017
iTransformer	0.016	0.085	0.018
NHITS	0.014	0.073	0.016
NBEATS	0.013	0.07	0.016
SOFTS	0.014	0.072	0.016
RMoK	0.017	0.09	0.019
BiTCN	0.02	0.106	0.022

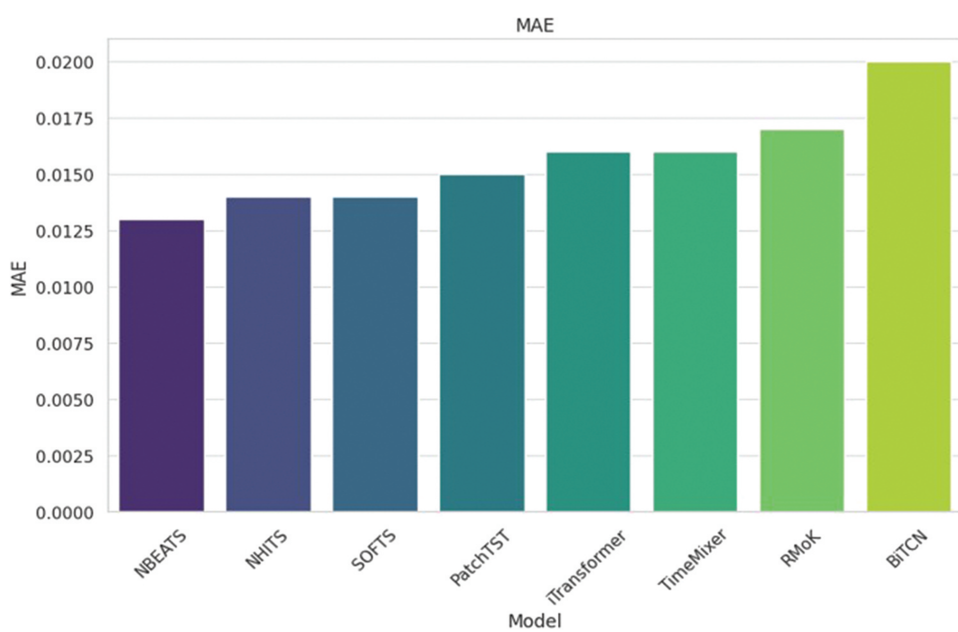

Figure 1. Model comparison based on Mean Absolute Error (MAE).

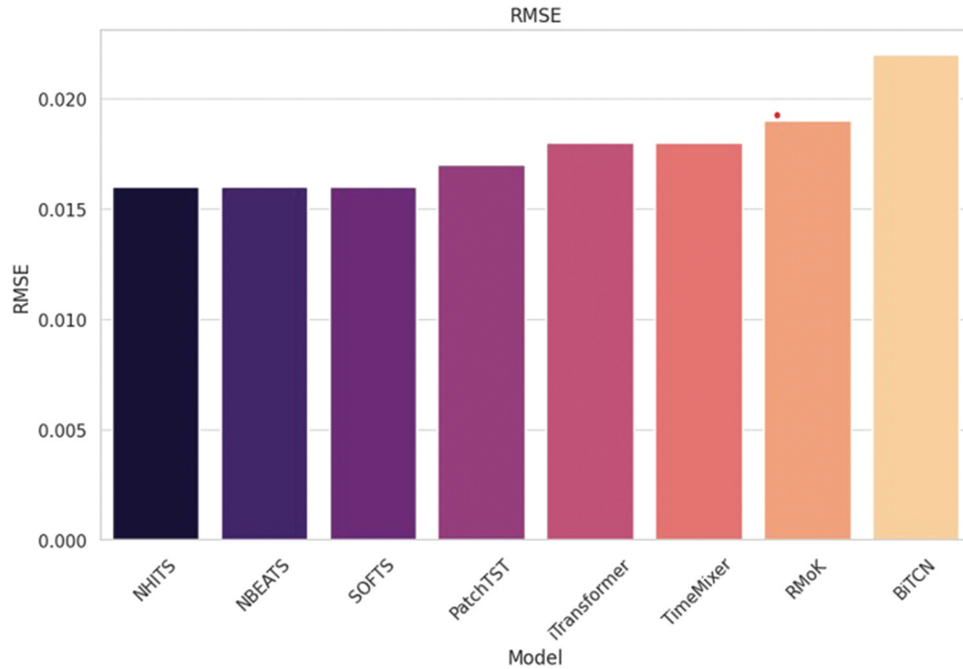

Figure 2. Model comparison based on Root Mean Square Error (RMSE).

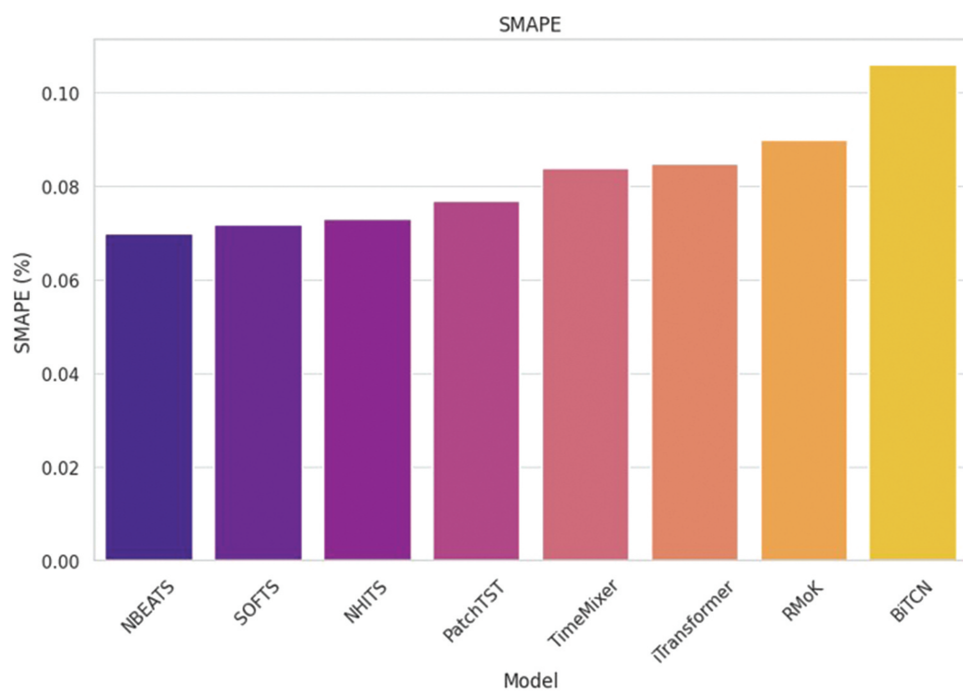

Figure 3. Model comparison based on SMAPE.

study employs MAE, SMAPE, and RMSE to compare forecasting models, and the results are presented in Table 2, Figures 1–3.

NBEATS achieves the lowest MAE (0.013), SMAPE (7.0%), and RMSE (0.016), demonstrating the highest predictive accuracy. NHITS (MAE: 0.014, SMAPE: 7.3%, RMSE: 0.016) and SOFTS (MAE: 0.014, SMAPE: 7.2%, RMSE: 0.016) closely follow, showing strong performance. PatchTST (MAE: 0.015, SMAPE: 7.7%, RMSE: 0.017), TimeMixer (MAE: 0.016, SMAPE: 8.4%, RMSE: 0.018), and iTransformer (MAE: 0.016, SMAPE: 8.5%, RMSE: 0.018) exhibit moderate accuracy, while RMoK (MAE: 0.017, SMAPE: 9.0%, RMSE: 0.019) and BiTCN (MAE: 0.020, SMAPE: 10.6%, RMSE: 0.022) perform the worst, with BiTCN showing the highest errors across all metrics.

Furthermore, Table S1 (Appendix 1) highlights regional performance disparities: iTransformer excelled on Canada's S&P/TSX (MAE: 0.003), while SOFTS led the US S&P 500 (MAE: 0.004). European indices (DAX, CAC 40) had higher model errors, likely due to macroeconomic/geopolitical sensitivity. BiTCN and RMoK struggled most with abrupt market shifts.

V. Conclusion

This study evaluated eight advanced forecasting models – TimeMixer, PatchTST, iTransformer, NHITS, NBEATS, SOFTS, RMoK, and BiTCN – across seven major global stock indices. The results highlight significant performance variations, with NHITS, NBEATS, and SOFTS demonstrating superior predictive accuracy. This aligns with prior research that shows the efficacy of neural hierarchical architectures in financial time-series forecasting (Souto 2023). The success of NHITS and NBEATS is likely due to their multi-scale decomposition mechanisms, which capture both short-term fluctuations and long-term trends in stock market data.

Author contributions

CRediT: **Arab Dahir Hassan:** Conceptualization, Formal analysis, Software, Writing – original draft, Writing – review & editing; **Mahat Maalim Ibrahim:** Formal analysis, Methodology, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported that there is no funding associated with the work featured in this article.

ORCID

Arab Dahir Hassan http://orcid.org/0000-0003-2397-4126
Mahat Maalim Ibrahim http://orcid.org/0000-0001-8450-7551

References

Box, G. E., G. M. Jenkins, G. C. Reinsel, and G. M. Ljung. 1970. *Time Series Analysis: Forecasting and Control*. John Wiley & Sons.

Challu, C., K. G. Olivares, B. N. Oreshkin, F. Garza, M. Mergenthaler-Canseco, and A. Dubrawski. 2022. "NHITS: Neural Hierarchical Interpolation for Time Series Forecasting." *Proceedings of the AAAI Conference on Artificial Intelligence* 37 (6): 6989–6997. https://doi.org/10.1609/aaai.v37i6.25854.

Chen, S.-A., C.-L. Li, N. Yoder, S. O. Arik, and T. Pfister. 2023. "TSMixer: An All-MLP Architecture for Time Series Forecasting (arXiv: 2303.06053)." *arXiv*. https://doi.org/10.48550/arXiv.2303.06053.

Fama, E. F. 1970. "Efficient Capital Markets: A Review of Theory and Empirical Work." *The Journal of Finance* 25 (2): 383–417. https://doi.org/10.2307/2325486.

Gupta, H. 2024. "A Study on Stock Forecasting Using Deep Learning and Statistical Models." *arXiv*. https://doi.org/10.48550/arXiv.2402.06689.

Han, L., X. Y. Chen, H. J. Ye, and D. C. Zhan. 2024. "SOFTS: Efficient Multivariate Time Series Forecasting with Series-Core Fusion." *arXiv Preprint arXiv: 2404.14197*. https://arxiv.org/abs/2404.14197.

Han, X., X. Zhang, Y. Wu, Z. Zhang, and Z. Wu. 2024. "Kan4tsf: Are Kan and Kan-Based Models Effective for Time Series Forecasting?" *arXiv Preprint arXiv: 2408.11306*. https://arxiv.org/abs/2211.14730.

Harel, A., and G. Harpaz. 2024. "Why Stock Analysts May Make Wrong Predictions?" *Economics Letters* 244:111956. https://doi.org/10.1016/j.econlet.2024.111956.

Ho, M. K., H. Darman, and S. Musa. 2021. "Stock Price Prediction Using ARIMA, Neural Network, and LSTM Models." *Journal of Physics: Conference Series* 1988 (1): 012041. https://doi.org/10.1088/1742-6596/1988/1/012041.

John, K., and J. Li. 2021. "COVID-19, Volatility Dynamics, and Sentiment Trading." *Journal of Banking & Finance* 133:106162. https://doi.org/10.1016/j.jbankfin.2021.106162.

Kalyan, K. S., A. Rajasekharan, and S. Sangeetha. 2021. "AMMU: A Survey of Transformer-Based Biomedical Pretrained Language Models." *Journal of Biomedical Informatics* 126:103982. https://doi.org/10.1016/j.jbi.2021.103982.

Khan, W., M. A. Ghazanfar, M. A. Azam, A. Karami, K. H. Alyoubi, and A. S. Alfakeeh. 2022. "Stock Market Prediction Using Machine Learning Classifiers and Social Media, News." *Journal of Ambient Intelligence and Humanized Computing* 13 (7): 3433–3456. https://doi.org/10.1007/s12652-020-01839-w.

Liu, Y., T. Hu, H. Zhang, H. Wu, S. Wang, L. Ma, and M. Long. 2024. "iTransformer: Inverted Transformers are Effective for Time Series Forecasting (arXiv: 2310.06625)." *arXiv*. https://doi.org/10.48550/arXiv.2310.06625.

Liu, Z., Y. Wang, S. Vaidya, F. Ruehle, J. Halverson, M. Soljačić, T. Y. Hou, and M. Tegmark. 2024. "KAN: Kolmogorov-Arnold Networks [Preprint]." *arXiv: 2404.xxxxx [cs.LG]*. https://arxiv.org/abs/2404.19756.

Nie, Y., N. H. Nguyen, P. Sinthong, and J. Kalagnanam. 2023. "A Time Series is Worth 64 Words: Long-Term Forecasting with Transformers (arXiv: 2211.14730)." *arXiv*. https://doi.org/10.48550/arXiv.2211.14730.

Oreshkin, B. N., D. Carpov, N. Chapados, and Y. Bengio. 2020. "N-BEATS: Neural Basis Expansion Analysis for Interpretable Time Series Forecasting (arXiv: 1905.10437)." *arXiv*. https://doi.org/10.48550/arXiv.1905.10437.

Parray, I. R., S. S. Khurana, M. Kumar, and A. A. Altalbe. 2020. "Time Series Data Analysis of Stock Price Movement Using Machine Learning Techniques." *Soft Computing* 24 (21): 16509–16517. https://doi.org/10.1007/s00500-020-04957-x.

Sharma, S., K. Gupta, and D. Gupta. 2021. "Recommender System: A Bibliometric Analysis." *IOP Conference Series: Materials Science & Engineering* 1022 (1): 012057. https://doi.org/10.1088/1757-899X/1022/1/012057.

Souto, H. G. 2023. *Nhits for Forecasting Stock Realized Volatility*. SSRN Electronic Journal.

Sprangers, O., S. Schelter, and M. de Rijke. 2023. "Parameter-Efficient Deep Probabilistic Forecasting." *International Journal of Forecasting* 39 (1): 332–345. https://doi.org/10.1016/j.ijforecast.2021.11.011.

Vijh, M., D. Chandola, V. A. Tikkiwal, and A. Kumar. 2020. "Stock Closing Price Prediction Using Machine Learning Techniques." *Procedia Computer Science* 167:599–606. https://doi.org/10.1016/j.procs.2020.03.326.

Wu, F., Y. Shi, X. Xiong, Y. Li, and Z. Dong. 2024. "Benefiting from the Frightened Herd: Dynamic Asset Allocation Amid Panic Sentiment." *Economics Letters* 245:112052. https://doi.org/10.1016/j.econlet.2024.112052.

Wu, H., J. Xu, J. Wang, and M. Long. 2022. "Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting (arXiv: 2106.13008)." *arXiv*. https://doi.org/10.48550/arXiv.2106.13008.

Zeng, A., M. Chen, L. Zhang, and Q. Xu. 2022. "Are Transformers Effective for Time Series Forecasting? (arXiv: 2205.13504)." *arXiv*. https://doi.org/10.48550/arXiv.2205.13504.

Zheng, Y., E. Osmer, and R. Zhang. 2018. "Sentiment Hedging: How Hedge Funds Adjust Their Exposure to Market Sentiment." *Journal of Banking & Finance* 88:147–160. https://doi.org/10.1016/j.jbankfin.2017.11.016.

Zhou, H., S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang. 2021. "Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting (arXiv: 2012.07436)." *Proceedings of the AAAI Conference on Artificial Intelligence* 35 (12): 11106–11115. https://doi.org/10.1609/aaai.v35i12.17325.

Zhou, T., Z. Ma, Q. Wen, X. Wang, L. Sun, and R. Jin. 2022. "FEDformer: Frequency Enhanced Decomposed Transformer for Long-Term Series Forecasting (arXiv: 2201.12740)." *arXiv*. https://doi.org/10.48550/arXiv.2201.12740.