Decision Making Advances

Journal homepage: www.dma-journal.org
ISSN: 2956-2384

Forecasting Aluminum Futures Prices: A Comparative Evaluation of Classical, Deep Learning and Transformer-based Models

László Vancsura[1], Tibor Tatay[2,*], Tibor Bareith[3,4], Kornélia Mészáros[5]

[1] Department of Marketing and Supply Chain Analysis; Institute of Agricultural and Food Economics, Hungarian University of Agriculture and Life Sciences, Kaposvár, Hungary
[2] Department of Statistics, Finances and Controlling, Széchenyi István University, Győr, Hungary
[3] ELTE, Centre for Economic and Regional Studies, Institute of Economics, 1097 Budapest, Hungary
[4] Economic Geography and Urban Marketing Centre, John von Neumann University, 1117 Budapest, Hungary
[5] Department of Marketing and Supply Chain Analysis, Institute of Agricultural and Food Economics, Hungarian University of Agriculture and Life Sciences, Gödöllő, Hungary

ARTICLE INFO

Article history:
Received 15 May 2026
Received in revised form 28 July 2026
Accepted 3 August 2026
Available online 7 August 2026

Keywords:
Aluminum Futures; Commodity Forecasting; Machine Learning; Deep Learning; Time Series Forecasting; Decision Making; Structural Breaks

ABSTRACT

Forecasting commodity prices remains a challenging task due to market volatility, structural breaks, and changing economic conditions. This study evaluates the forecasting performance of classical econometric, deep learning, convolutional, and Transformer-based models for aluminum futures prices. Daily aluminum futures data are analysed using 11 forecasting approaches. Forecasting experiments are conducted for two distinct evaluation periods representing the years 2022 and 2025 to assess the robustness of model performance under different market environments. The empirical results reveal substantial differences in forecasting performance across model families. Recurrent neural network architectures, particularly GRU and RNN, consistently achieve the lowest forecasting errors across most horizons. The classical ARIMA model remains highly competitive despite its relative simplicity. In contrast, Transformer-based models generally fail to outperform simpler alternatives and frequently produce higher forecast errors. Statistical comparisons based on Diebold–Mariano tests and Model Confidence Set procedures indicate that performance differences among the best-performing models are often limited, suggesting that increased model complexity does not necessarily translate into superior predictive accuracy. Overall, the findings highlight the importance of empirical model evaluation and demonstrate that parsimonious forecasting approaches can remain effective competitors to substantially more complex machine learning architectures in aluminum futures price forecasting.

1. Introduction

The decarbonization of the global economy, the rapid expansion of electromobility, and the advancement of industrial digitalization are fundamentally transforming the demand structure for

* *Corresponding author.*
E-mail address: tatay@sze.hu

https://doi.org/10.31181/dma412026191

strategic raw materials. Aluminum plays a distinguished role in this transition, simultaneously serving the objectives of lightweight vehicle manufacturing, improved energy efficiency, and the circular economy. Due to its low density, favorable strength-to-weight ratio, corrosion resistance, and recyclability, aluminum is a key material for automotive innovation [1-3]. With the proliferation of electric vehicles (EVs), the importance of lightweight solutions continues to grow, as reducing vehicle mass directly contributes to extending driving range and moderating energy consumption. Projections suggest that automotive aluminum consumption could expand significantly by 2030 [4], forecasting a structural growth in demand for the metal. However, the strategic importance of aluminum is not solely defined by the demand side. Production is highly energy-intensive; thus, aluminum prices are closely linked to developments in global energy markets. Energy price shocks, geopolitical conflicts, and regional supply concentration collectively exacerbate market vulnerability. While London Metal Exchange (LME) reference prices function as global benchmarks, the extreme volatility experienced in recent years has highlighted that pricing mechanisms are increasingly driven by external, structural factors [5-9]. The energy crisis following 2022, supply chain disruptions, and geopolitical fragmentation have created an environment of uncertainty where price movements reflect systemic rather than merely cyclical instability. Trends in industrial metal prices have a direct impact on manufacturing cost structures, investment decisions, and national industrial policies. In the case of aluminum, fluctuations in procurement costs influence the competitiveness of automotive and electronics manufacturers, the terms of long-term supplier contracts, and the return on capacity expansion investments. The literature indicates that price movements of industrial metals move in close tandem with industrial cycles and generate significant macroeconomic feedback effects [10]. At the same time, recent results from machine learning and econometric models suggest that the predictability of metal prices is limited, particularly during turbulent periods [11]. In this context, the reliability of price forecasts carries resource-policy significance. Government programs concerning strategic raw materials, critical raw materials lists, subsidy schemes, and industrial resilience strategies are all, implicitly or explicitly, based on assumptions regarding future price trajectories. If these forecasts become unstable during regime shifts, or if model performance deteriorates significantly due to structural breaks, it may result in a distorted information base for public policy decision-making. Consequently, the reliability of model-based expectations is directly linked to the risk profile of fiscal planning, industrial support policies, and supply security strategies [12-14].

This study is predicated on the recognition that the structural instability observed during and after 2022 is not merely a transient episode of volatility, but rather signifies a transformation in market operational mechanisms. In such an environment, ranking forecasting models solely based on aggregated accuracy metrics is insufficient. It is necessary to examine the extent to which model performance is regime-robust, how stable the dominance hierarchy remains across different market conditions, and whether the increase in forecasting errors stems from methodological limitations or the inherent unpredictability of the market. The central problem of this research is thus to determine whether advanced, high-capacity forecasting architectures, including recurrent neural networks and Transformer-based models, can provide an economically meaningful and structurally robust advantage in forecasting aluminum prices, or if their performance is primarily confined to stable market environments. If model dominance is regime-dependent and unstable, it suggests that model selection itself becomes a risk factor in industrial policy and corporate decision-making [15].

The objective of this study is to compare the forecasting performance of a broad set of econometric, recurrent neural network, convolutional, and Transformer-based architectures for aluminum futures prices. Particular attention is given to the stability of model rankings across different forecasting horizons and market environments. Rather than proposing a new forecasting methodology, the study provides a systematic benchmark comparison of widely used forecasting approaches under conditions characterized by structural breaks and changing market dynamics.

2. Literature Review

Global demand for mineral raw materials has increased significantly over recent decades, driven by economic expansion, structural industrial transformation, and the proliferation of material-intensive technologies [16-19]. According to the USGS [20], this growth is not merely a cyclical phenomenon but is structural in nature, closely linked to the expanding demand for conductive, lightweight, and high-performance metals. Population growth, urbanization, and infrastructure development exert further pressure on primary raw material reserves, while the issues of long-term supply security and sustainability become increasingly prominent. In this context, price volatility is not simply a market phenomenon but a macroeconomic and resource-policy risk factor. For developing economies, where a significant portion of export revenues is derived from mineral resources, fluctuations in metal prices have a direct impact on fiscal balance, foreign exchange reserves, and growth trajectories [21-24]. The cyclicity of commodity prices can thus result in structural vulnerabilities that necessitate industrial policy and fiscal adjustment. Consequently, the evolution of global metal prices is a decisive factor not only at the corporate level but also within a national economic dimension. The question of substitutability is a central element of resource economics and critical raw materials discourse. Although many metals can be replaced to some extent, substitution typically involves trade-offs in technology, efficiency, or cost. For instance, aluminum serves as an alternative to copper in several electrical applications due to its lower density and cost advantages [25-27]. However, substitution is not solely a consequence of physical scarcity; sustained price increases can incentivize material innovation, technological adaptation, and the expansion of recycling [28-30]. While these processes may reduce dependence on primary extraction, they can also generate new types of supply chain exposures and secondary price pressures.

Iron, aluminum, and copper, as the most widely used industrial metals globally, play a pivotal role in industrial production and international trade. Their prices and liquidity are significantly influenced by international commodity exchanges and futures markets [31-33], indicating a strong link between financial markets and the real economy. Price movements in industrial metals directly affect the profitability of metal-intensive sectors, investment decisions, and corporate risk management strategies. Market analyses of aluminum have demonstrated that price developments are substantially influenced not only by fundamental supply-demand factors but also by global business cycles, the US dollar exchange rate, and geopolitical uncertainty [34-36]. This complex interplay suggests that metal price forecasting cannot be reduced solely to market fundamentals. Reliable price forecasts are therefore of strategic importance in both corporate and public policy decision-making. More accurate predictions of supply and demand shocks can contribute to the design of more effective regulatory instruments, the mitigation of investment risks, and long-term industrial policy planning [37]. From a resource-policy perspective, predictive capability is a fundamental prerequisite for supply security strategies, stockpiling decisions, and the strengthening of industrial resilience. In the early stages of metal price forecasting, classical statistical and econometric models dominated. Research findings indicate that the classical ARIMA model frequently underperforms compared to neural network approaches, particularly in the case of non-linear time series [38-40]. Machine learning and deep learning methods, especially LSTM and GRU architectures, are playing an increasingly decisive role in commodity market forecasting. Empirical studies suggest that these models are capable of capturing long-term dependencies and complex dynamics, often yielding lower error metrics than linear benchmark models [41,42]. However, performance is heterogeneous and regime-dependent, suggesting that there is no universally dominant model [43]. Hybrid approaches - integrating linear and non-linear components - often exhibit more stable out-of-sample results. Decomposition-based and optimized frameworks, such as CEEMDAN–VMD–LSTM or optimized XGBoost models, can further enhance forecasting performance by isolating the structural components of price movements [12,44,45]. The quantification of uncertainty has also come to the fore: probabilistic forecasts allow for more informative risk assessments [46], while incorporating cross-market information, such as mining company stock prices, can improve aluminum price forecasting across certain horizons [11]. Nevertheless, the literature emphasizes that commodity markets are characterized by structural instability. Storage dynamics and speculative mechanisms can result in non-linear price movements [47], while structural breaks and regime shifts - such as those during crises or geopolitical shocks - can lead to parameter instability [48-50]. The literature suggests that metal price forecasting is a structurally constrained task: model performance is a function of not only methodological factors but also systemic uncertainty [51-55]. This realization is particularly relevant for resource-policy strategies concerning critical raw materials, where decisions are often implicitly built upon future price trajectories. Beyond predictive accuracy, robustness, regime sensitivity, and the explicit treatment of uncertainty thus become crucial evaluative dimensions for strengthening industrial resilience and supply security.

The post-2022 period highlights the structural vulnerability of metal markets with particular intensity. Energy market shocks, geopolitical fragmentation, and disruptions in global supply chains have created an environment in which pricing mechanisms previously considered stable have partially transformed. The structural breaks and regime shifts documented in the literature can be interpreted in the post-2022 period not merely as theoretical possibilities but as empirically relevant phenomena [56-60]. Rising energy costs, especially for aluminum, which is characterized by a highly energy-intensive production structure, have resulted in direct supply-side pressure, while uncertainty in global demand and heightened volatility in financial markets have opened new transmission channels [61-64]. In this environment, the performance of forecasting models cannot be evaluated solely based on static accuracy metrics. If parameters are unstable, the model superiority demonstrated on historical samples may easily lose its validity in a new regime. Furthermore, increasing financial integration and the interconnectedness of commodity markets raise the probability that price dynamics become more sensitive to external shocks, leading to a deterioration in predictive performance [21,65]. These literature-based insights justify the empirical design of this study. The investigation covers several partially distinct market regimes, allowing for an analysis of how stable the relative performance of models remains amidst structural changes. The application of multiple forecasting horizons is not merely a technical choice but an approach aligned with industrial and resource-policy decision-making, as short-term procurement decisions and long-term investment or capacity planning strategies have distinct information requirements [66-68]. Furthermore, the study does not rely exclusively on traditional error metrics but also interprets predictive performance within the dimension of structural predictability. This is consistent with findings in the literature suggesting that the predictability of commodity markets can itself be regime-dependent and structurally constrained during certain periods [69-71]. Thus, the empirical framework does not merely conduct a model comparison; it investigates the extent to which predictive dominance can be considered stable in an environment characterized by heightened uncertainty. This approach is particularly relevant for resource-policy decision-making. If forecasting performance is regime-sensitive, it is insufficient for decision-makers to rely on the historical superiority of a given model; instead, an explicit evaluation of robustness and structural adaptability becomes necessary. The turbulent post-2022 period, therefore, serves not only as an empirical sample but as a stress-test environment in which the economic policy relevance of the models can be meaningfully assessed.

A critical synthesis of the forecasting literature reveals three unresolved issues. First, most comparative studies report average out-of-sample errors for a single test period, which makes it difficult to determine whether the apparent superiority of a model persists when the underlying market regime changes. Second, although structural breaks are widely acknowledged as a defining characteristic of commodity prices, they are often treated as a diagnostic property of the data rather than as a central dimension of model evaluation. Consequently, limited evidence is available on how the relative ranking of classical, recurrent, convolutional, and Transformer-based models changes across periods containing multiple breaks. Third, the literature frequently equates methodological sophistication with predictive progress, while giving less attention to whether any accuracy advantage is statistically distinguishable, economically meaningful, and sufficiently stable to justify the additional computational and implementation burden. These limitations are particularly important for aluminum futures, where energy intensity, global supply-chain exposure, and geopolitical sensitivity can cause rapid changes in the data-generating process.

The research gap addressed by this study therefore concerns the regime robustness of model dominance in aluminum futures forecasting. Existing studies provide valuable evidence on individual algorithms and hybrid frameworks, but they do not offer a sufficiently broad, internally consistent comparison of classical econometric, recurrent neural, convolutional, decomposition-based, and Transformer architectures across multiple forecasting horizons and distinct structurally unstable evaluation periods. Nor do they consistently combine conventional forecast-error measures with formal tests of predictive differences and model-set membership. By benchmarking eleven models on a common univariate information set for 2022 and 2025, and by interpreting the results alongside structural-break evidence, this study separates architecture-related performance from changes attributable to the market regime. In this way, the analysis extends the literature from a static accuracy contest towards a decision-oriented assessment of accuracy, robustness, parsimony, and stability.

Accordingly, the study addresses the following research questions:

i. (RQ1) *How does the relative forecasting performance of classical econometric, recurrent, convolutional, and Transformer-based models differ across two structurally distinct market regimes (2022 and 2025) and across multiple forecasting horizons?*
ii. (RQ2) *Are the observed differences in forecasting accuracy between competing models statistically significant, or do the best-performing models form a set of comparable predictive quality?*
iii. (RQ3) *Does increasing architectural complexity, as embodied by Transformer-based architectures, translate into a measurable and stable forecasting advantage once structural breaks and multiple market regimes are explicitly taken into account, or does model dominance instead prove regime-dependent?*

3. Methodology

3.1 Data

The analysis focuses on two distinct periods of aluminum prices. The first interval covers the period from June 1, 2014, to December 31, 2022, representing the market environment before and during the pandemic. The second interval spans from June 1, 2017, to December 31, 2025, capturing the post-pandemic aluminum market characterized by heightened volatility and structural changes.

These two samples enable a comparison of model performance across different market regimes, with a particular focus on the temporal evolution of exchange rate dynamics.

Daily frequency aluminum futures price data were sourced from the Yahoo Finance database. As the study was conducted within a sequential forecasting framework, the time series were utilized in their natural chronological order, and no data imputation or synthetic oversampling was applied. This approach ensured that the models relied exclusively on actually observed historical information, in alignment with the methodological requirements of time-series forecasting. When interpreting the samples, it is crucial to distinguish between the full, descriptive, and forecasting subsets. Table 1 summarizes the stationarity and structural break tests performed on the full available time series (1 June 2014 to 31 December 2025), while Table 3 presents a comparison of two partially overlapping descriptive intervals (2014–2022 and 2017–2025).

Table 1
Results of unit root, structural break and stationarity tests for aluminum (1 June 2014 to 31 December 2025)

Variable	Values
ADF stat	-1.6992
ADF *p*-value	0.4315
KPSS stat	2.8841
KPSS *p*-value	0.0100 *
PP stat	-1.7355
PP *p*-value	0.4129
DFGLS stat	-1.6847
DFGLS *p*-value	0.0904 *
Structural breaks (Bai-Perron)	08.02.2022; 28.04.2022; 10.06. 31.03.2025; 11.06.2025; 06.10
Structural breaks (PELT)	08.02.2022; 28.04.2022; 10.06. 31.03.2025; 11.06.2025; 06.10

***p<= 0.01; **p <= 0.05; *p <= 0.1

In contrast, the training and test sets for the models were constructed using time-based separation: training was conducted on the periods 01.06.2014–31.12.2021 and 01.06.2017–31.12.2024, while testing was based on the years 2022 and 2025, respectively. Table 1 presents the results of the unit root, structural break, and stationarity tests conducted for each category. Specifically, the Augmented Dickey–Fuller (ADF), Phillips–Perron (PP), Kwiatkowski–Phillips–Schmidt–Shin (KPSS), and Elliott–Rothenberg–Stock Dickey–Fuller Generalized Least Squares (DF–GLS) tests were applied to assess the presence of unit roots and mean reversion, while the Pelt and Bai-Perron tests were employed to detect structural breaks in the series. The results of the stationarity tests consistently indicate non-stationary price series. Neither the ADF nor the PP tests reject the null hypothesis of a unit root, which is clearly confirmed by high p-values (0.41–0.43). In the case of the DF–GLS test, a marginal result is observed for aluminum ($p \approx 0.09$); however, this is insufficient to accept stationarity. Accordingly, the KPSS test rejects the null hypothesis of stationarity at the 1% significance level, further supporting non-stationarity. The presence of structural instability is corroborated by the breakpoints identified by the PELT and Bai-Perron methods, discussed in detail in Section 4.3. For aluminum, a particularly dense distribution of breakpoints is observed, suggesting frequent market regime shifts and strong external shocks. Regarding the two study intervals (Period 1 and Period 2), we examined the breakpoints detected jointly by these two methods. These results justify the rejection of classical stationarity assumptions and the application of models capable of handling flexible, non-stationary processes. Methodologically, the test results further support the relevance of advanced time-series and machine learning models in forecasting industrial metal prices.

3.2 Applied Methods

Although the inclusion of exogenous variables may improve forecasting performance in certain settings, the present study adopts a univariate forecasting framework for methodological reasons. The primary objective is not to identify the economic determinants of aluminum price movements but to compare the predictive performance of alternative forecasting architectures under a common information set. Using only historical aluminum futures prices ensures that all models are evaluated on identical inputs, thereby facilitating a transparent and consistent benchmark comparison across classical econometric methods, recurrent neural networks, convolutional architectures, and Transformer-based models. Furthermore, a univariate framework reduces the influence of model-specific variable selection procedures and data availability constraints, allowing the analysis to focus on differences arising from the forecasting architectures themselves. While previous research has demonstrated the potential predictive value of macroeconomic, financial, and commodity-market variables, incorporating such information would introduce additional sources of heterogeneity across model specifications. Therefore, the present study intentionally concentrates on the information contained in the historical futures price series to provide a comparable assessment of forecasting performance across different forecasting horizons and market environments.

3.2.1 Autoregressive integrated moving average

The ARIMA model represents the classical econometric approach, widely utilized in commodity and metal market analyses due to its robustness and mathematical transparency. The fundamental assumption of the model is that the future values of a time series (y_t) can be described as a linear combination of past observations and past forecasting errors, provided the series can be rendered stationary. The model is composed of three components: the autoregressive (AR) part, which measures the impact of past values; the integrated (I) part, responsible for handling non-stationary trends through differencing; and the moving average (MA) component, which models the decay of short-term shocks. In the context of industrial metals, such as aluminum, ARIMA is highly effective at capturing inertia effects and mean-reversion tendencies resulting from adjustments in inventory levels or market expectations. Mathematically, the ARIMA(p,d,q) process can be formalized using the lag operator (L) as follows:

$$\Phi(L)(1-L)^d y_t = c + \theta(L)\epsilon_t \tag{1}$$

where the autoregressive polynomial is $\Phi(L) = 1 - \sum_{i=1}^{p} \Phi_i L^i$ and the moving average polynomial is $\theta(L) = 1 - \sum_{j=1}^{q} \Phi\theta_j L^j$. In this expression, $(1-L)^d$ denotes the d-th order differencing responsible for removing stochastic trends, while ϵ_t represents the white noise process, where $\epsilon_t \sim WN(0, \sigma^2)$. In the present study, ARIMA serves not only as a predictive tool but also as a benchmark, allowing for the exact measurement of the added value provided by subsequent complex, non-linear deep learning models. Due to its low computational requirements and transparent parameterization, it remains a reference model in production planning and procurement management to this day [72].

3.2.2 Simple Recurrent Neural Network

The Vanilla or simple Recurrent Neural Network (RNN) represents the most fundamental neural architecture capable of modeling sequential dependencies in time-series analysis. The distinguishing feature of RNNs is the hidden state (h_t), which functions as a form of internal memory, updating itself recursively at each time step. This theoretically enables the identification of temporal correlations beyond fixed lag structures [73]. In the context of commodity price forecasting, this mechanism empowers the model to dynamically map momentum effects and short-term market reactions. The mathematical transition of the network can be described as follows:

$$h_t = \sigma(W_{hh}h_{t-1} + W_{xh}x_t + b_h) \tag{2}$$

$$\hat{y}_t = W_{hy}h_t + b_y \tag{3}$$

where x_t is the current input vector, W_{xh} and W_{hh} are the weight matrices, b represents the bias terms, and σ denotes a non-linear activation function. Although RNNs are theoretically capable of processing sequences of arbitrary length, in practice, they are severely affected by the vanishing and exploding gradient problems during backpropagation through time (BPTT). This limitation is particularly relevant for industrial metals, where pricing often depends on long-term cycles, such as production capacity expansions or macroeconomic trends. In this study, the simple RNN serves as a conceptual bridge between linear statistical models and more sophisticated gated architectures. Its straightforward structure allows for a clearer interpretation of performance differentials within the recurrent model family. The primary objective of including the RNN is to evaluate the extent to which a basic neural network structure provides incremental information over ARIMA before introducing the more complex solutions of LSTM and GRU.

3.2.3 Long Short-Term Memory

The Long Short-Term Memory (LSTM) network is an extension of the simple RNN architecture, specifically designed to handle long-term temporal dependencies and to mitigate the vanishing gradient problem [74]. The distinguishing feature of the LSTM is the cell state (C_t), which allows information to be preserved or discarded over long periods through specialized gating mechanisms. In the case of aluminum futures prices, such dependencies arise from multi-year production cycles, strategic investment decisions, and sustained macroeconomic trends. The model regulates the flow of information using three primary gates: the forget gate (f_t) determines which past information should be discarded; the input gate (i_t) selects new relevant data; and the output gate (o_t) controls the amount of information passed to the hidden state. Mathematically, the process can be described by the following system of equations:

$$f_t = \sigma(W_f \times [h_{t-1}, x_t] + b_f) \tag{4}$$

$$i_t = \sigma(W_i \times [h_{t-1}, x_t] + b_i) \tag{5}$$

$$\tilde{C}_t = tanh(W_C \times [h_{t-1}, x_t] + b_C) \tag{6}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \tag{7}$$

$$o_t = \sigma(W_o \times [h_{t-1}, x_t] + b_o) \quad (8)$$

$$h_t = o_t \odot \tanh(C_t) \quad (9)$$

where σ denotes the sigmoid activation function, and $\odot$ represents element-wise multiplication. From a production economics perspective, the LSTM's ability to capture persistent effects enables more reliable medium-term forecasts for procurement and cost management. Although the computational requirements of LSTM are significantly higher than those of linear models, its robustness and adaptability to volatile commodity market environments justify its application. In the present research, the role of the LSTM is to verify whether the introduction of memory cells indeed yields quantifiable improvements in forecasting complex industrial raw material prices compared to standard neural networks [75].

3.2.4 Gated recurrent unit

The Gated Recurrent Unit (GRU) is a streamlined alternative to the LSTM architecture, merging memory and gating mechanisms into a more compact structure. The GRU utilizes two primary gates: the update gate (z_t) and the reset gate (r_t), which collectively regulate the modifications of the hidden state (h_t). By reducing the number of parameters, the GRU often achieves predictive performance comparable to the LSTM, yet with lower computational costs and faster convergence. This is particularly advantageous for large-scale forecasting tasks or real-time industrial decision support systems. In commodity markets, the GRU is capable of modeling non-linear price dynamics while remaining more stable during training. Due to its simpler architecture, it mitigates the risk of overfitting in cases of limited datasets [76]. The mathematical foundation of the model is formed by the following equations:

$$z_t = \sigma(W_z \times [h_{t-1}, x_t] + b_z) \quad (10)$$

$$r_t = \sigma(W_r \times [h_{t-1}, x_t] + b_r) \quad (11)$$

$$\tilde{h}_t = tanh(W \times [r_t \odot h_{t-1}, x_t] + b) \quad (12)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t) \quad (13)$$

From a management perspective, the GRU offers a practical balance between model complexity and accuracy. Its inclusion in this study allows for a direct comparison between various gated recurrent architectures, helping to determine whether the more complex structure of the LSTM is justified within the context of industrial metal prices.

3.2.5 Temporal convolutional network

The Temporal Convolutional Network (TCN) offers a modern alternative to recurrent architectures, modeling sequential data by leveraging causal and dilated convolutions. One of the primary advantages of TCNs over RNN-based models is that they process the entire input window in parallel, resulting in more stable gradients and significantly faster training. The application of dilated convolutions enables the capture of long-term temporal dependencies without requiring excessively deep networks, thereby avoiding the vanishing gradient problem [77]. In the case of aluminum prices, this structure is capable of simultaneously handling short-term volatility and longer-term fundamental cyclical effects. The TCN equation can be defined as follows:

$$y_t = x_t + \sum_{l=1}^{L} \sum_{k=0}^{K-1} w_k^{(l)} x_{t-d_l \times k} \tag{14}$$

where x_t is the value of the input time series, y_t is the TCN output, L represents the number of TCN layers (blocks), l is the layer index, K denotes the length of the convolutional kernel, k is the kernel index, w represents the weight, and d is the dilation factor. The network also employs residual blocks, which facilitate the effective training of deeper structures and prevent model degradation. From an operational perspective, TCNs are excellently suited for the rolling-window forecasting frameworks used in production planning, as they utilize a fixed-length historical context to estimate future values. Their deterministic receptive field and the parallelization of sequential data processing enhance the reproducibility of results and increase robustness against noise in industrial data.

3.2.6 Neural basis expansion analysis

Neural Basis Expansion Analysis (N-BEATS) is a deep learning architecture specifically designed for univariate time-series forecasting, which relies not on recursion or convolution, but on a deep hierarchy of fully connected layers. A key characteristic of the model is the sequential decomposition of the forecast, implemented through a series of stacked residual blocks. This structure allows the model to iteratively remove explained patterns from the data and focus on the remaining residual errors in subsequent blocks [78]. In commodity market time series, this decomposition is implicitly capable of separating trend-like processes from periodic seasonality. The model expresses the forecast as follows:

$$\theta^{(l)} = g^{(l)}(x^{(l)}) \tag{15}$$

$$\hat{x}^{(l)} = B^{(l)}(\theta^{(l)}) \tag{16}$$

$$\hat{y}^{(l)} = F^{(l)}(\theta^{(l)}) \tag{17}$$

$$x^{(l+1)} = x^{(l)} - \hat{x}^{(l)} \tag{18}$$

$$\hat{y} = \sum_{l=1}^{L} \hat{y}^{(l)} \tag{19}$$

where x represents the historical time series, $\hat{x}$ is the explanation of the past (backcast), is $\hat{y}$ the future prediction, L is the number of blocks, ϑ denotes the learned parameter vector, g represents the Multi-Layer Perceptron (MLP), B and F are the linear mappings. From a production economics perspective, this additive structure harmonizes well with the theory that price movements emerge as the resultant of multiple economic forces acting over different time horizons (macroeconomic trends, seasonality, market shocks). The flexibility of N-BEATS and its architecture - free from domain-specific assumptions - allow it to capture the complex pricing dynamics of industrial raw materials with high precision [79].

3.2.7 Informer

The Informer is a Transformer-based architecture specifically developed for the efficient and accurate handling of Long Sequence Time-Series Forecasting (LSTF). The primary limitation of traditional Transformers is the quadratic computational complexity of the attention mechanism, which the Informer addresses by replacing it with the ProbSparse Self-attention mechanism. This solution exploits the sparsity of the relationships between Queries (Q) and Keys (K), computing interactions only between the most informative elements, thereby reducing complexity to O(L log L) [80]. In the aluminum futures market, this empowers the model to incorporate several months of historical data when determining future trends. The attention mechanism is formalized as:

$$Attention(Q, K, V) = Softmax(\frac{\bar{Q}K^T}{\sqrt{d_k}})V \tag{20}$$

where V represents the value and d denotes the embedding dimension. Furthermore, the model applies a "distilling" operation between layers, which further compresses temporal information and reduces redundancy. The Informer's generative-style decoding enables the production of the entire forecasting range in a single step, thereby avoiding cumulative errors stemming from autocorrelation. This attribute becomes particularly critical during the planning of long-term strategic raw material procurement, where accuracy across multiple horizons is indispensable.

3.2.8 Autoformer

The Autoformer radically reimagines the Transformer architecture by embedding time-series decomposition into the deep layers of the network. Instead of seeking point-wise attention relationships, the Autoformer relies on an Auto-Correlation mechanism, which leverages the periodicity and similarity of signals by analyzing time delays [81]. This approach aligns directly with the nature of commodity market prices, where cyclic seasonality and recurring market patterns dominate price movements. The model's internal series-decomposition block continuously separates the trend and the seasonal component within each layer:

$$Auto - Correlation(Q, K, V) = \sum_{\tau \in T} Softmax(R_{Q,K}(\tau))Roll(V, \tau) \tag{21}$$

where R denotes the autocorrelation coefficients, τ represents the time delay, and T signifies the selected lags. The Autoformer is more robust against noise than its counterparts employing point-wise attention, as it relies on aggregated temporal information. From a production economics perspective, this stability and the enhanced interpretability provided by decomposition assist managers in distinguishing sustained price changes from short-term market noise, thereby enabling more informed inventory management decisions.

3.2.9 Patch time series transformer

Patch Time Series Transformer (PatchTST) belongs to the latest generation of Transformer models, treating time series not as point-wise data streams but as fixed-length "patches". Semantically, this shift is significant because individual data points often carry little meaning within commodity market noise, whereas a patch (e.g., a week's price movement) encapsulates local trends and momentum. By utilizing patches, the sequence length is significantly shortened, enabling the model to process a much longer historical context using the same computational capacity [82]. The mathematical foundation of the model is based on channel-independent processing:

$$x_p \in R^{N \times P} \tag{22}$$

where P is the patch length and N is the number of generated patches. This structure implicitly reduces the risk of overfitting and enhances the model's generalization capability in volatile markets. PatchTST is particularly effective in forecasting aluminum prices, where market information is often integrated into prices intermittently, in blocks. From a management perspective, this model represents stability and high-level pattern recognition, which are essential for robust production scheduling.

3.2.10 DLinear

The DLinear model challenges the complexity of Transformer architectures by returning to the power of linear models, augmented with a modern decomposition framework. The model's fundamental premise is that, in many cases, complex attention mechanisms only add noise to the time series, whereas a well-structured linear layer can capture trends more accurately. DLinear decomposes the time series into trend and seasonal components using a moving average, then applies separate weighted linear transformations to each [83]:

$$\hat{y} = W_t x_{trend} + W_s x_{season} + b \tag{23}$$

where W_t is the trend weight matrix and W_s is the seasonal weight matrix. This simplicity results in extraordinary robustness and interpretability, which are often more important in production economics decisions than a fractional percentage advantage in accuracy. Due to its low computational requirements, it is excellently suited for running large-scale scenario analyses and sensitivity tests. The inclusion of DLinear in this research serves as an "Occam's Razor" principle, helping to determine whether nonlinear complexity is truly necessary for reliable results in the case of aluminum prices.

3.2.11 Temporal Fusion Transformer

The Temporal Fusion Transformer (TFT) is the most holistic model in this study, specifically designed to handle heterogeneous data sources (historical prices, known future events, and static metadata) within a single framework. The TFT combines the sequential power of LSTMs with the attention-based capabilities of Transformers, while utilizing Gated Residual Networks (GRN) to select the importance of input variables. This mechanism allows the model to ignore momentarily irrelevant factors and focus only on the most critical market drivers [84]:

$$\hat{y}_t = \sum_{\tau=1}^{T} \beta_{t,\tau}\, GRN(LSTM(\sum_{i=1}^{M} \alpha_{t,i} x_{\tau,i})) \tag{24}$$

where M is the number of input variables, T is the number of time steps considered, $\beta_{t,\tau}$ represents temporal attention weights, and $\alpha_{t,i}$ denotes the variable-selection weight. A distinguishing feature of the TFT is its multi-horizon attention mechanism, which highlights which past time points were most significant for a specific future estimate. This "glass-box" transparency is extremely valuable for decision-makers, as it not only estimates the price but also provides an explanation for why the model predicted that specific value. In modeling the regime shifts and cyclicality of the aluminum markets, the TFT currently appears to be one of the most sophisticated and comprehensive solutions.

3.3 Performance Evaluation

The most commonly used metrics in the literature for evaluating the predictive models and assessing their accuracy are root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) [85].

i. *Root mean squared error* (RMSE) – This performance indicator shows an estimation of the residuals between actual and predicted values.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{25}$$

where $\hat{y}_i$ is the estimated value produced by the model, y_i is the actual value and n is the number of observations.

ii. *Mean Absolute Error* (MAE) – This indicator measures the average magnitude of the error in a set of predictions.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{26}$$

iii. *Mean Absolute Percentage Error* (MAPE) – This indicator measures the average magnitude of the error in a set of predictions and shows the deviations in percentage.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{27}$$

The forecasts are more reliable and accurate if these indicators have lower values. It is important to note that RMSE penalises larger deviations more due to squaring, so this metric can give more extreme values than MAE. The former should be interpreted as a value, while the MAPE should be interpreted as a percentage (deviations expressed as a percentage of the original value). For this reason, the MAPE can be used to compare different instruments because it does not depend on the nominal size of the price. As our study examined indices from around the world and the effects of two different negative economic events, we used the MAPE indicator for comparability in the overall assessment of the models.

3.3 Hyperparameters

Table 2 summarizes the hyperparameter configurations employed for the models. For the investigated periods (2022 and 2025), each model was optimized separately using the full time series of the given regime. During forecasting, a 50-day rolling window procedure was applied, which not only allowed for the continuous integration of the latest historical information but also served as a time-series cross-validation method. This approach contributed to mitigating the risk of overfitting while ensuring the temporal robustness of the models. The training and testing sets were partitioned using time-based separation: the training sets covered the periods from 01.06.2014 to 31.12.2021 and 01.06.2017 to 31.12.2024, while the test periods fell within the intervals of 01.01.2022 to

31.12.2022 and 01.01.2025 to 31.12.2025, respectively. Random sampling was intentionally avoided to maintain the integrity of the temporal structure of the time series. Hyperparameter tuning was conducted via grid search, where each validation split was based exclusively on past observations, thereby completely eliminating the possibility of information leakage between the training and testing phases.

Table 2
Hyperparemeters

Model	Parameters	Value
ARIMA	*p*, *d*, *q*	1, 1, 1
RNN LSTM GRU	Hidden Layers	2
	Hidden layer neuron count	100, 150
	Batch size	16, 32
	Epochs	100
	Activation	ReLU, Tanh, Sigmoid, Linear
	Dropout rate	0.1; 0.01; 0.001
	Optimizer	Adam
TCN	Number of layers	2
	Number of filters	16, 32, 64
	Kernel size	3
	Dilation factors	1, 2, 3, 4
	Dropout rate	0.1; 0.01; 0.001
	Batch size	16, 32
	Activation	ReLU, Tanh, Sigmoid, Linear
	Optimizer	Adam
N-BEATS	Hidden layer size	2
	Hidden units	100, 150
	Number of stacks	2
	Number of blocks per stack	2, 3
	Dropout rate	0.1; 0.01; 0.001
	Batch size	16, 32
	Activation	ReLU, Tanh, Sigmoid, Linear
	Optimizer	Adam
Informer	Embedding dimension (d_model)	32, 64
	Number of encoder/decoder layers	2
	ProbSparse attention factor	1, 3, 5
	Number of attention heads	3, 4
	Dropout rate	0.1; 0.01; 0.001
	Batch size	16, 32
	Activation	ReLU, Tanh, Sigmoid, Linear
	Optimizer	Adam
Autoformer	Embedding dimension (d_model)	32, 64
	Number of encoder/decoder layers	2
	Series decomposition kernel size	25
	Number of attention heads	3, 4
	Dropout rate	0.1; 0.01; 0.001
	Batch size	16, 32
	Activation	ReLU, Tanh, Sigmoid, Linear
	Optimizer	Adam

Table 2 (*continued*)

Model	Parameters	Value
ARIMA	*p*, *d*, *q*	1, 1, 1
PatchTST	Embedding dimension (d_model)	32, 64
	Patch length	5, 10, 20
	Number of transformer layers	2
	Number of attention heads	3, 4
	Dropout rate	0.1; 0.01; 0.001
	Batch size	16, 32
	Activation	ReLU, Tanh, Sigmoid, Linear
	Optimizer	Adam
DLinear	Kernel size (multi-scale temporal kernels)	5, 10, 25, 50
	Dropout rate	0.1; 0.01; 0.001
	Batch size	16, 32
	Activation	Linear
	Optimizer	Adam
TFT	Embedding dimension (d_model)	32, 64
	Number of LSTM layers	2
	Number of LSTM units	100, 150
	Number of attention heads	3, 4
	Dropout rate	0.1; 0.01; 0.001
	Batch size	16, 32
	Activation	ReLU, Tanh, Sigmoid, Linear
	Optimizer	Adam

Wherever applicable, particularly for deep learning models and their integrated variants, early stopping was implemented based on the validation loss (MAE). The training process was terminated when no further improvement was observed, and in all cases, the model parameters from the epoch yielding the best performance were retained. This procedure enhanced the stability and generalization of the forecasts. The models were re-optimized for each investigated period and aluminum-related category to ensure that hyperparameters adequately adapted to varying market and macroeconomic environments.

Prior to modeling, the time series were normalized using Min-Max scaling, transforming the data into a fixed [0, 1] interval. This preserves the shape of the original distribution and the relative differences within the time series, thereby facilitating learning stability and model convergence, especially for deep learning architectures.

Analyses were conducted in a Python 3.13 environment. For model implementation, the TensorFlow and Scikit-learn libraries were utilized, including modules necessary for hyperparameter tuning and time-series validation.

4. Results

Table 3 presents the descriptive statistics for aluminum across two partially overlapping time intervals: 2014–2022 (Period 1) and 2017–2025 (Period 2). The high sample size ($N > 2100$) provides a robust foundation for descriptive comparison; however, the partial overlap necessitates caution in interpreting the results. The mean value increased in the second period (from 2127.70 to 2295.67), which is consistent with the elevated price levels during the investigated interval. The rise of the median to the 2300 level also indicates an upward shift in the central tendency. The moderate difference in standard deviation values suggests a change in the dispersion of price levels, though this does not inherently imply a clear moderation of volatility. The identical minimum and maximum values indicate that the same extreme values occurred across the full samples. The shift in quartiles (Q1 and Q3) is compatible with the general appreciation of price levels, while the decrease in the skewness coefficient in the second period suggests a more symmetrical distribution. Based on the kurtosis values, the distribution remained leptokurtic. Overall, the descriptive statistics show that the post-2017, partially overlapping period is characterized by higher price levels and slightly different distributional characteristics.

Table 3
Descriptive statistics for period 1 (1 June 2014 to 31 December 2022) and period 2 (1 June 2017 to 31 December 2025)

Statistic	Period 1 (2014-2022)	Period 2 (2017-2025)
N	2125	2155
Mean	2127.7048	2295.6713
Median	2189.7500	2300.0000
Standard Deviation	403.6736	364.6779
Min	1452.0000	1452.0000
Max	3873.0000	3873.0000
25th Percentile (Q1)	1786.5000	2155.2500
75th Percentile (Q3)	2328.2500	2494.5000
Skewness	0.8486	0.3730
Kurtosis	1.1217	1.2184

4.1 Forecasting Results for Aluminum in 2022 and 2025

Based on the results of the forecasting experiments for Period 1 (2022), model performance exhibits significant variations across different forecasting horizons (Figure 1).

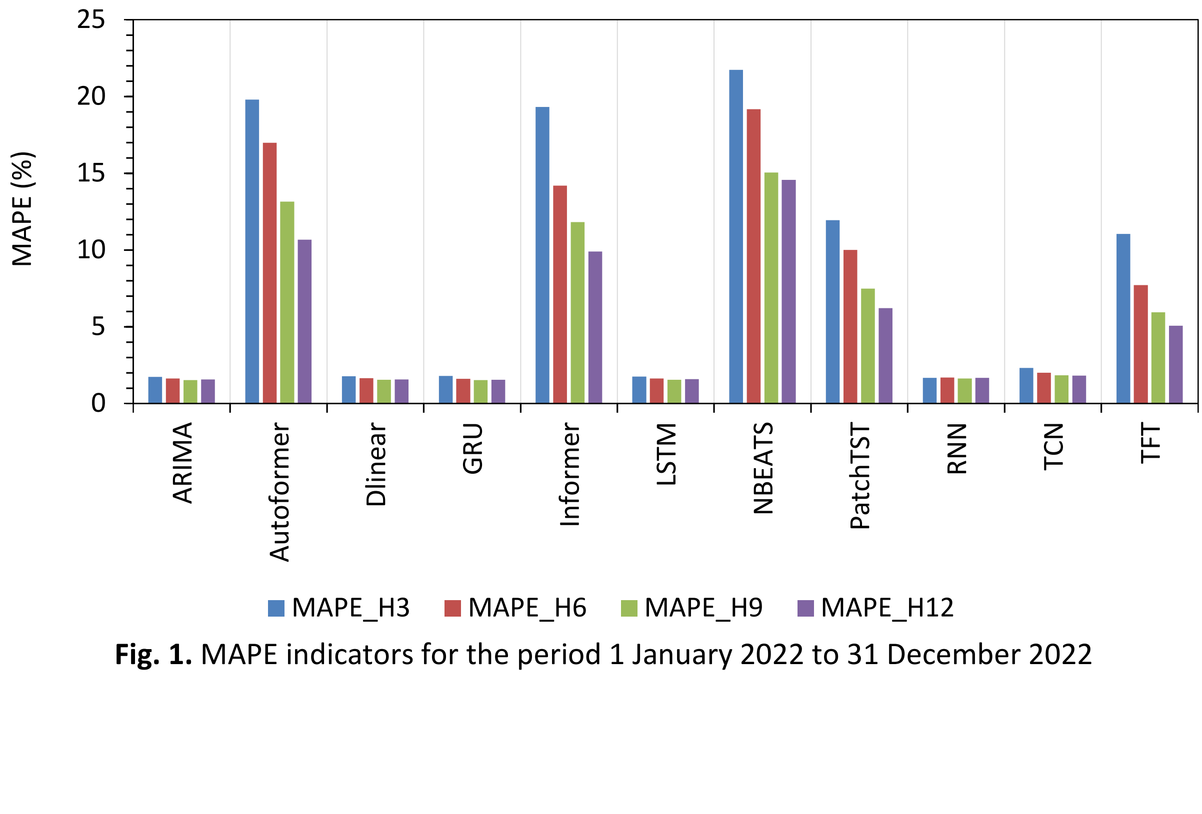

Fig. 1. MAPE indicators for the period 1 January 2022 to 31 December 2022

In interpreting the MAPE metrics, it must be emphasized that lower values indicate superior forecasting accuracy. In the short term (H3), the ARIMA (1.72%), DLinear (1.75%), LSTM (1.74%), and RNN (1.65%) models achieve the lowest errors. In contrast, for the Autoformer and Informer, the MAPE exceeds 19%, while the N-BEATS shows an error of 21.7%. The weaker performance of complex models likely stems from the tension between sample size and the extreme noise inherent in commodity market data, underscoring the importance of "Data Efficiency" in predicting critical raw materials. In the medium term (H6), the GRU (1.60%) and DLinear (1.63%) continue to perform exceptionally well, while the errors for the Autoformer (16.97%) and Informer (14.18%) remain orders of magnitude higher. Over the longer forecasting horizon (H9), the error metrics for ARIMA and GRU hover around 1.52%, indicating sustained predictive stability. A similar pattern is observed for H12, where the GRU (1.53%) and ARIMA (1.55%) maintain low error levels. Although Transformer-based models show improvement as the horizon increases (e.g., TFT: from 11.03% to 5.05%), their performance still lags behind classical and RNN-based approaches. The N-BEATS consistently produces high errors across all timeframes (H12: 14.56%), which may suggest a poorer fit to the time-series structure. Overall, the results indicate that simpler linear or recurrent neural models are more robust in forecasting aluminum prices.

The results of the forecasting experiments conducted on data for Period 2 (2025) indicate that model performance continues to exhibit significant disparities across the various forecasting horizons (Figure 2).

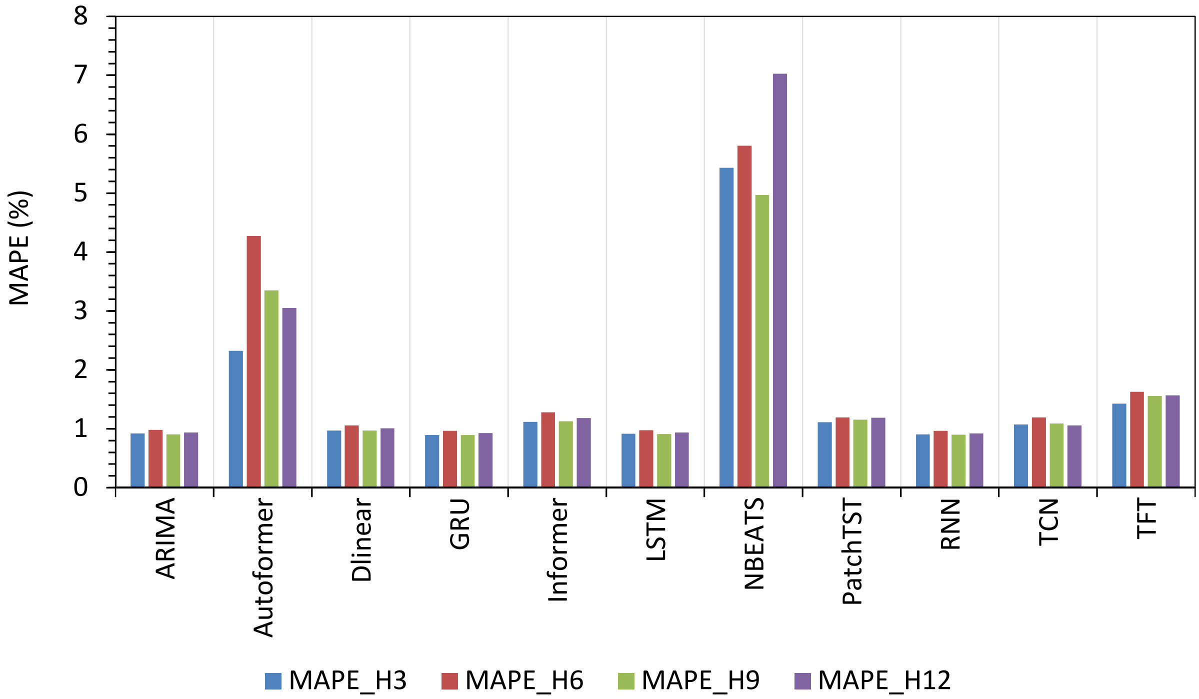

Fig. 2. MAPE indicators for the period 1 January 2025 to 31 December 2025

In the short term (H3), the GRU (0.89%), RNN (0.90%), LSTM (0.92%), and ARIMA (0.92%) achieve the lowest error rates. In contrast, the Autoformer (2.32%) and N-BEATS (5.43%) demonstrate substantially weaker performance, even over short forecasting horizons. A similar pattern is observed in the medium term (H6), where the GRU (0.96%), RNN (0.96%), and ARIMA (0.98%) maintain a consistently low error level. In this case, the MAPE for the Autoformer is 4.27%, while the N-BEATS exceeds 5.8%. Over longer horizons (H9), the GRU and RNN produce MAPE values around 0.89%, signaling exceptionally stable performance. This tendency persists for H12 forecasts, where the GRU

(0.92%), RNN (0.92%), and ARIMA (0.93%) yield the best results. Although the performance of Transformer-based models improves as the horizon extends, they still significantly lag behind the accuracy of recurrent neural models, even at H12. Across all examined timeframes, the N-BEATS consistently shows high MAPE values (H12: 7.02%), indicating weaker empirical performance within the current data environment and configuration. Overall, the results suggest that the dominance of simpler and more robust models in forecasting aluminum prices remains sustained in 2025. This is particularly relevant for decision support systems, where stable, low-error forecasts hold paramount economic significance. The methodological takeaway of the investigation is that increasing model complexity does not inherently result in superior forecasting performance. Consequently, parsimony remains a key design principle in predicting industrial metal prices. Compared to 2022, the 2025 period exhibited more moderate error reactions; however, based on the presence of structural breaks, it cannot be considered a completely tranquil phase.

The comparison of forecasting results between 2022 and 2025 clearly illustrates the temporal stability and robustness of the models' performance. In the short term (H3), the MAPE values of the top-performing models typically ranged between 1.6% and 1.8% in 2022, whereas by 2025, this range decreased to 0.89–0.92%, indicating a significant improvement in accuracy. A similar trend is observable in the medium term (H6), where the error of leading models was still around 1.6% in 2022 but fell between 0.96% and 0.98% in 2025. Over longer horizons (H9–H12), the GRU, RNN, and ARIMA models exhibit stable performance in both years; however, in 2025, their MAPE values consistently hover around 0.9–0.94%, compared to the 1.5–1.6% levels seen in 2022. Transformer-based models underperform in both years, although a moderate reduction in error is also observable for them in 2025. The N-BEATS model produces exceptionally high MAPE values in both periods, suggesting persistently weaker performance under the examined configurations. Overall, the results show that the relative ranking among models remains largely intact, though its stability cannot be considered absolute across all market conditions. This observation is particularly valuable for decision-support applications. The investigation confirms that simpler, parsimonious models remain competitive in the long run for forecasting aluminum prices.

4.2 Statistical Comparison of Model Performances

Based on the results of the Diebold-Mariano (DM) test conducted on the 2022 aluminum dataset (Table 4), model pairs can be clearly categorized into those where predictive differences are statistically verifiable and those where no significant deviation can be detected. During the pairwise comparison of ARIMA, DLinear, GRU, LSTM, and RNN models, the DM statistics are typically low and show no significant difference, suggesting that the error performance of these models is closely aligned across the investigated horizon. In contrast, the N-BEATS, PatchTST, and TFT models exhibit significant differences at the 1% level compared to several of the aforementioned models, indicating that their error structures differ substantially from the more parsimonious approaches. In the case of the TCN model, a significant difference at the 1% level is also observed relative to ARIMA (2.93). It is observable that the DM values between fundamental types of neural networks (RNN, LSTM, GRU) consistently remain below critical thresholds, suggesting similar predictive characteristics despite their methodological diversity. The data suggest that, within the investigated horizon, the complexity of the model architecture does not always lead to significantly different results. According to Table 4, Transformer-based models, such as Informer and Autoformer, show a significant difference even when compared to each other at the 1% level (-5.68), indicating that divergent error dynamics can prevail even within the same model family. The N-BEATS model showed strongly significant deviations in several examined relations, pointing to its unique statistical position within the sample.

Table 4
Results of DM tests for Period 1 (Year 2022, 12-month horizon)

	ARIMA	Autoformer	Dlinear	GRU	Informer	LSTM	NBEATS	PatchTST	RNN	TCN	TFT
ARIMA	-										
Autoformer	9.87***	-									
Dlinear	0.42	-9.85***	-								
GRU	-0.42	-9.86***	-1.40	-							
Informer	9.77***	-5.68***	9.75***	9.77***	-						
LSTM	0.15	-9.86***	-0.30	1.01	-9.77***	-					
NBEATS	12.43***	5.99***	12.43***	12.44***	12.47***	12.44***	-				
PatchTST	9.47***	-9.66***	9.43***	9.49***	-9.33***	9.51***	-13.32***	-			
RNN	1.12	-9.83***	1.57	2.12**	-9.72***	1.17	-12.41***	-9.37***	-		
TCN	2.93***	-9.85***	3.52***	3.93***	-9.75***	3.39***	-12.44***	-9.40***	2.80***	-	
TFT	7.11***	-10.01***	7.08***	7.12***	-11.55***	7.14***	-14.11***	-5.04***	7.03***	7.01***	-

****p*<= 0.01; ***p* <= 0.05; **p* <= 0.1

The investigation concerning the 2025 horizon (Table 5) shows a partial realignment of significant differences, reflecting the divergent dynamics of the later stages of the time series. While ARIMA continues to show no significant deviation compared to the DLinear, GRU, LSTM, and RNN models, previous differences - such as those between TCN and DLinear (1.51), or PatchTST and Informer (-1.36) - become statistically negligible over this timeframe. The N-BEATS model remains an exception, as it exhibits significant differences at the 1% level compared to several examined architectures, pointing to a consistently distinct predictive characteristic. An interesting observation is the DM statistic of 2.96 between the TFT and the Informer, which indicates significance at the 1% level, signaling a detectable difference even between these two Transformer-based solutions. Within the relationship between the DLinear model and recurrent neural networks (RNN, LSTM), significant differences emerge at the 5% and 10% levels, suggesting a more differentiated error structure. The significance pattern among Transformer-based models partially persists, implying that the handling of long-term dependencies differs in a statistically measurable way across various algorithms. Based on the aggregated data, the absence of significant differences becomes dominant for a substantial portion of the models over the 2025 horizon.

Table 5
Results of DM tests for Period 2 (Year 2025, 12-month horizon)

	ARIMA	Autoformer	Dlinear	GRU	Informer	LSTM	NBEATS	PatchTST	RNN	TCN	TFT
ARIMA	-										
Autoformer	7.63***	-									
Dlinear	1.44	-7.62***	-								
GRU	-0.25	-7.68***	-1.91*	-							
Informer	4.14***	-7.55***	3.43***	4.30***	-						
LSTM	-0.12	-7.68***	-2.04**	0.28	-4.33***	-					
NBEATS	10.91***	8.85***	10.91***	10.92***	10.80***	10.92***	-				
PatchTST	4.31***	-7.25***	4.55***	4.37***	-1.36	4.70***	-10.83***	-			
RNN	-0.24	-7.67***	-1.73*	0.00	-4.24***	-0.13	-10.92***	-4.22***	-		
TCN	3.02***	-7.58***	1.51	3.54***	-3.09***	3.80***	-10.85***	-1.94*	3.43***	-	
TFT	7.08***	-6.74***	7.64***	7.42***	2.96***	7.72***	-10.65***	6.14***	7.16***	6.25***	-

****p*<= 0.01; ***p* <= 0.05; **p* <= 0.1

4.2.1 MCS tests in 2022

The summary evaluation of the relative efficiency of forecasting models for the 2022 period is provided by the *p*-values of the Model Confidence Set (MCS) procedure (Table 6). A higher value for these indicators denotes stronger empirical support for a given model to remain within the set of superior models, although it should not be interpreted as a direct probability. Based on the data, the GRU model achieves the highest p-value (p = 1.0), clearly placing it within the superior set for the examined sample. This is followed by simpler neural networks and benchmarks, such as ARIMA and LSTM (p = 0.689), as well as DLinear (p = 0.68), all of which receive relatively strong empirical support in this study. At the other end of the spectrum are the more complex Transformer-based and decomposition models. The extremely low values for N-BEATS (p = 0.001) and PatchTST (p = 0.024) suggest that these architectures receive significantly weaker support for remaining in the superior set. The indicators for Informer, Autoformer (p = 0.025), and TFT (p = 0.032) confirm that these models are less competitive in this empirical environment. The TCN (p = 0.063) and RNN (p = 0.076) show slightly more favorable, yet still moderate support. Overall, the MCS test results highlight that for the examined data, simpler models, particularly GRU, ARIMA, and LSTM, provide more consistent forecasting performance than several of the more complex architectures.

4.2.2 MCS tests in 2025

Based on the MCS procedure for the 2025 period, the GRU model again yields the highest p-value (p = 1.0), indicating that this architecture clearly remains within the set of superior models for the examined sample (Table 6). This is closely followed by the RNN (p = 0.999), as well as the ARIMA and LSTM (p = 0.953) group, which also receive strong empirical support. The value for DLinear (p = 0.131) is perceptibly lower but still indicates a more favorable position than the more complex architectures. At the opposite end of the scale are the complex decomposition and Transformer-based models. The extremely low p-value for N-BEATS (p = 0.003) suggests that this architecture is not among the most strongly supported models in the sample. The Autoformer, TFT, PatchTST, and Informer models consistently produced values of p = 0.028, also indicating weaker empirical support. These results highlight an important methodological lesson: in modeling aluminum exchange rates, recurrent and traditional models (GRU, RNN, ARIMA) performed more stably in this empirical environment than more intricate algorithms built on attention mechanisms. The MCS test confirms the findings of the previous DM tests, showing that the lack of significant differences among simpler models is consistent with their collective inclusion in the superior set.

Table 6
Results of MCS tests for Periods 1 and 2 for 12-month horizon

Models	*p*-value (2022)	*p*-value (2025)
NBEATS	0.001	0.003
PatchTST	0.024	0.028
Informer	0.025	0.028
Autoformer	0.025	0.028
TFT	0.032	0.028
TCN	0.063	0.131
RNN	0.076	0.999
Dlinear	0.68	0.131
LSTM	0.689	0.953
ARIMA	0.689	0.953
GRU	1	1

4.3 Structural Breaks in 2022

Using the PELT and Bai-Perron methods, we examined only those breakpoints identified by both tests to ensure the robustness of the detected regime shifts. We present a detailed analysis of the three identified breakpoints. Based on the results (Figure 3), a positive ΔMAPE is observed for nearly all models following the break on February 8, 2022, suggesting that the initial shock significantly degraded forecasting accuracy. The most substantial error anomalies occurred in the complex, Transformer-based models (Autoformer, Informer, N-BEATS, TFT), whereas ARIMA and the simpler recurrent and convolutional architectures (GRU, RNN, TCN, DLinear) exhibited substantially lower sensitivity. This indicates that higher model complexity is initially more vulnerable to abrupt regime shifts. However, a marked turnaround is observable at the April 28, 2022 break; several deep learning models show a negative ΔMAPE, indicating that forecasting performance in the new post-shock regime surpassed pre-shock levels. Significant improvements are particularly evident for the Autoformer, Informer, and N-BEATS, suggesting an adaptive learning capability. Regarding the June 10, 2022 break, a mixed pattern emerges: some complex models continue to show improvement, while recurrent models exhibit a slight decline in performance. The ARIMA model shows relatively moderate shifts across all three breakpoints, pointing to a more stable but less dynamic adaptation pattern.

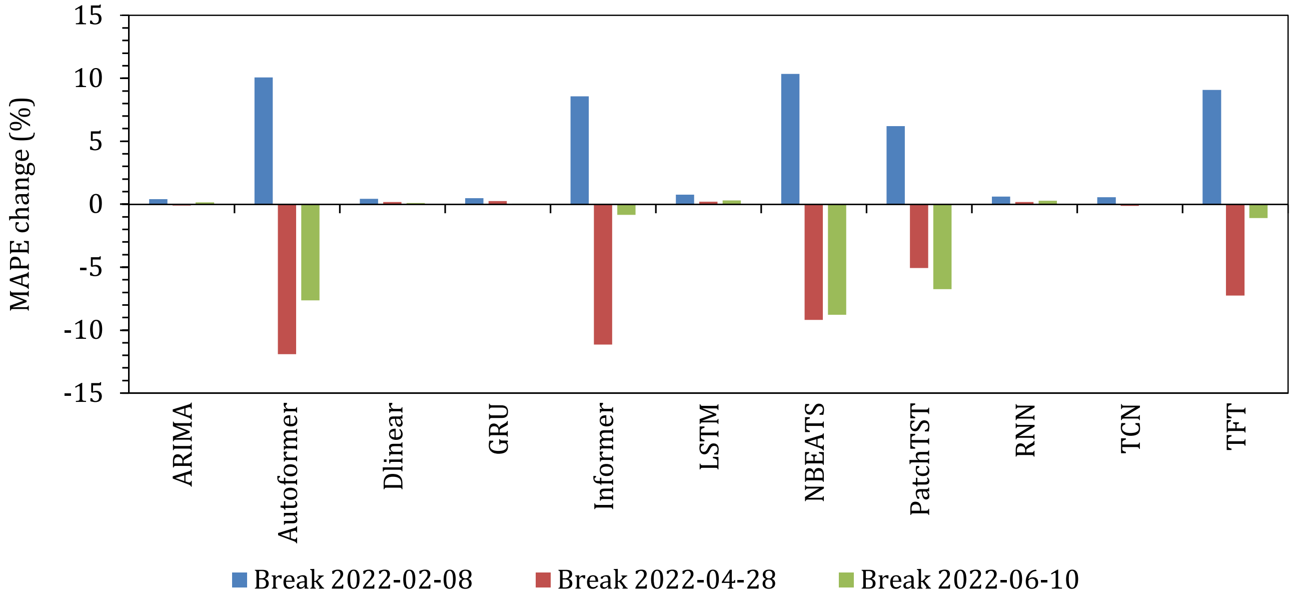

Fig. 3. Shock sensitivity analysis. Delta MAPE for each model around detected structural breaks in Period 1 (2022)

The heatmap (Figure 4) reveals clearly regime-dependent error dynamics, transcending simple differences in average performance. Following the structural break on February 8, 2022, a pronounced error clustering is observed, particularly for high-capacity Transformer-based architectures (Autoformer, Informer, N-BEATS, TFT). The abrupt surge in MAPE suggests that the representation space of these models is heavily calibrated to pre-shock patterns; consequently, the regime shift initially triggers generalization instability. This phenomenon can be interpreted as learning over-specialization within a quasi-stationary environment, which rapidly loses validity upon the impact of an exogenous shock. However, after the second breakpoint (28 April 2022), a visible error convergence occurs across several complex models, indicating adaptive restructuring. The gradual decline in error levels shows that these models are capable of internalizing the new volatility

Decision Making Advances
Volume 4, Issue 1 (2026) 154-190

and trend structures, suggesting a regime-dependent learning capacity. Nevertheless, the speed of adaptation is architecture-dependent. While the Autoformer and Informer exhibit faster error reduction, the high error level for N-BEATS persists longer, pointing to divergent inductive biases. In contrast, classical and lower-complexity models (ARIMA, DLinear, GRU, RNN, TCN) show less extreme error swings, indicating greater short-term robustness, though the magnitude of their error reduction is also more moderate. This pattern supports a dynamic interpretation of the bias–variance trade-off: high-capacity models respond with higher variance during a shock, yet they may achieve lower regime-specific bias after successful adaptation. Following the third breakpoint, the error structure becomes more balanced, suggesting that the market is converging toward a new equilibrium and the models' predictive spaces are stabilizing. The heatmap empirically confirms that structural breaks do not merely cause temporary performance degradation but generate architecture-specific adaptation trajectories. The brightening of colors after the shock is not just a reduction in error, but visual evidence of the architecture's recalibration capability. This supports the recommendation that shifting toward more complex models is advisable during the post-shock consolidation phase. Model complexity, therefore, is not an inherent advantage or disadvantage but a regime-dependent performance factor, interpretable through the interaction between learning dynamics and the market's structural instability.

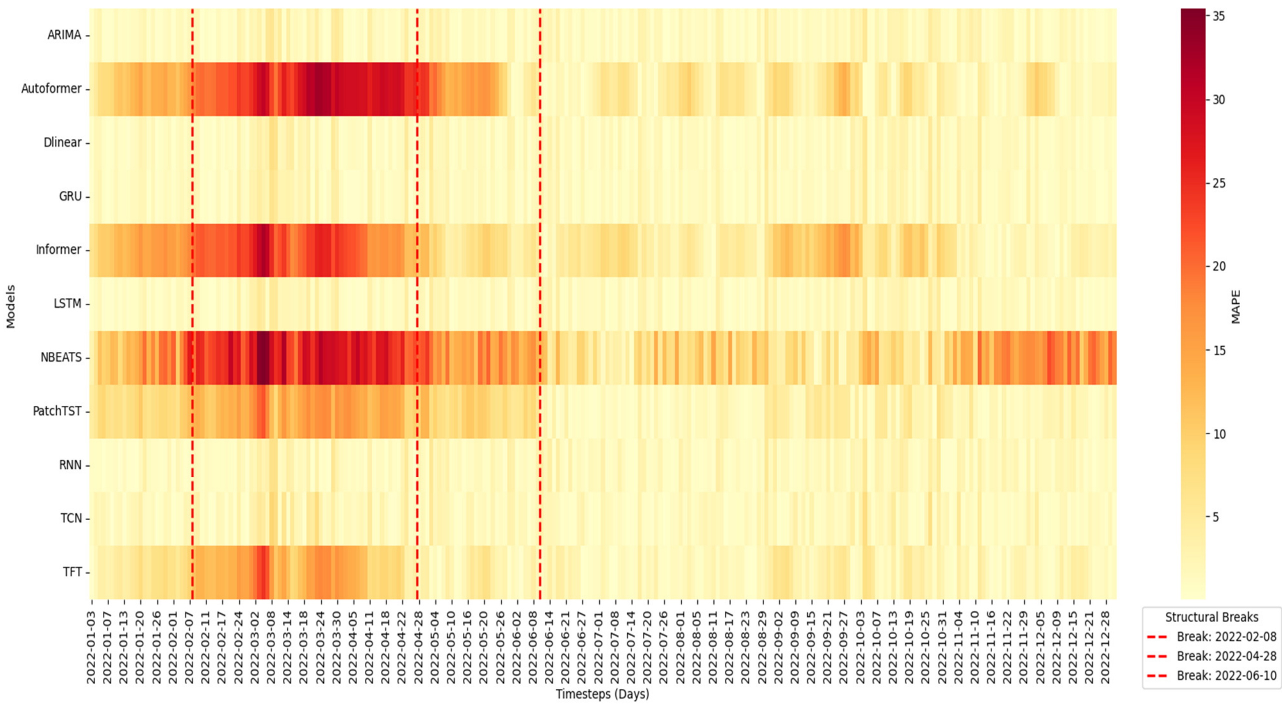

Fig. 4. Model-level errors over time (2022) with structural breaks

The average MAPE values for the one-month intervals before and after the breakpoints (Figure 5) reveal pronounced regime effects, which exhibit an architecture-dependent pattern. In the case of the February 8, 2022 break, a clear increase in error levels is observable across all models in the post-shock period, particularly for high-capacity models. For the Autoformer, Informer, and N-BEATS, the MAPE increases drastically (rising from 15.86% to 26.20% for N-BEATS), indicating significant generalization instability in the new market environment. In contrast, for ARIMA, DLinear, and the recurrent models, the increase in error levels is more moderate, suggesting a higher degree of short-term robustness.

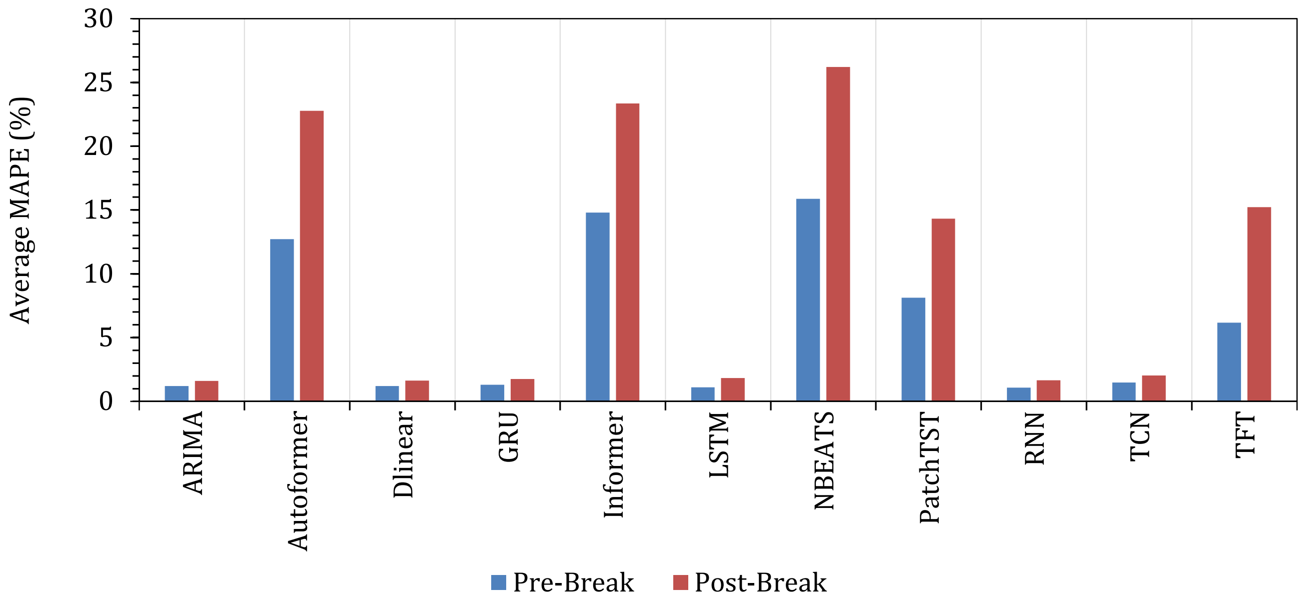

Fig. 5. Pre- and post-break average MAPE performance in Period 1 (1-month before and after 08.02.2022 break)

In the case of the second break (28 April 2022), however, an inverse trend is observable (Figure 6). Several complex models - specifically the Autoformer, Informer, N-BEATS, PatchTST, and particularly the TFT - exhibit a significant reduction in MAPE during the post-shock period (e.g., from 10.46% to 3.20% for the TFT). This suggests that following the initial turbulent phase, these models are more effective at internalizing the new trend and volatility structures. This phenomenon points to a regime-adaptive learning mechanism, where high model complexity becomes an advantage in a stabilizing yet structurally distinct environment.

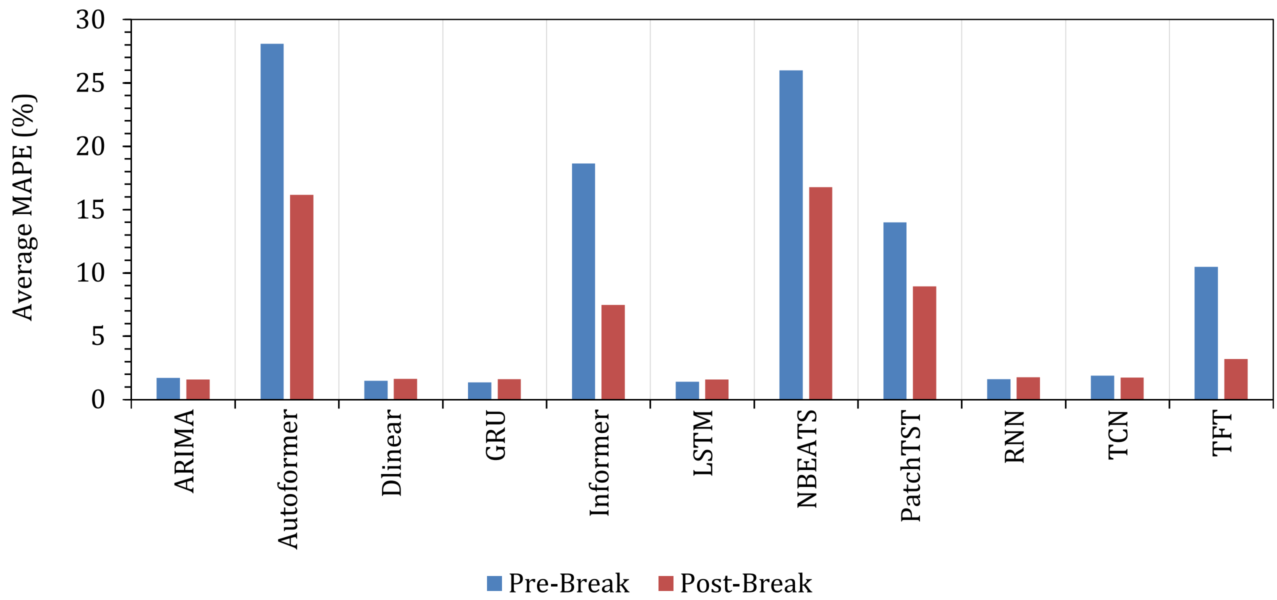

Fig. 6. Pre- and post-break average MAPE performance in Period 1 (1-month before and after 28.04.2022 break)

The pattern is further reinforced following the break on 10 June 2022 (Figure 7). An extreme reduction in error is observed for the Autoformer (dropping from 10.16% to 2.52%), while for the N-

BEATS, the MAPE is reduced by more than half (from 15.23% to 6.44%). The PatchTST and TFT models also exhibit substantial improvements. In contrast, classical and lower-complexity models (ARIMA, DLinear, GRU, TCN) show only moderate changes, pointing to a stable but less adaptive error profile.

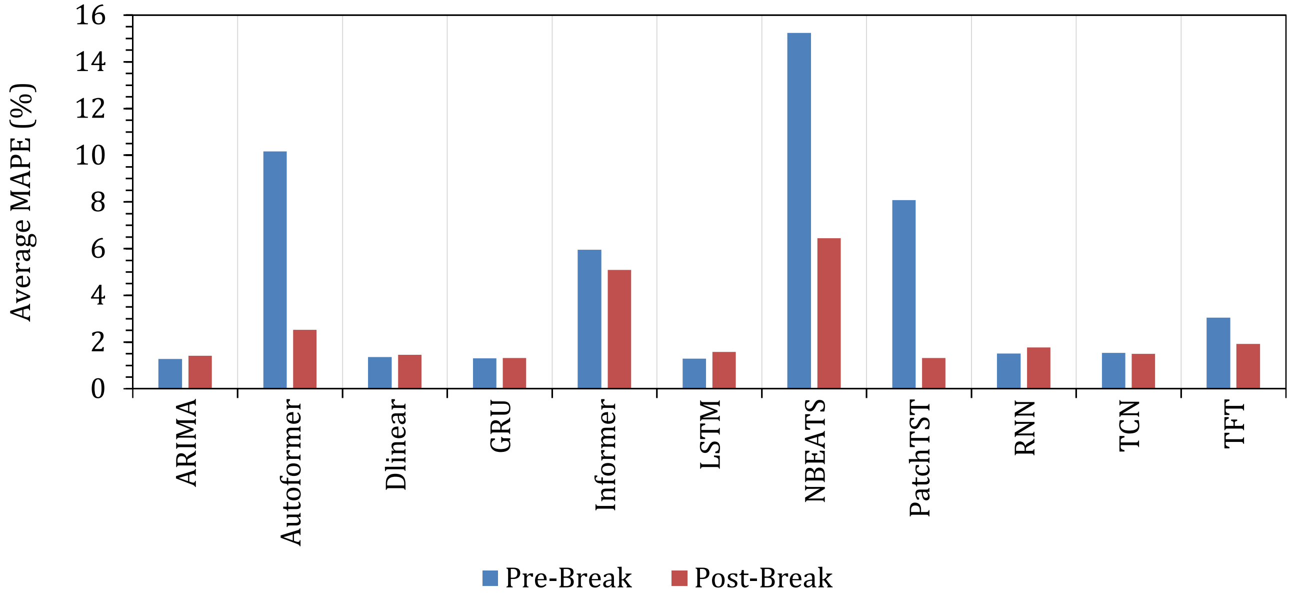

Fig. 7. Pre- and post-break average MAPE performance in Period 1 (1-month before and after 10.06.2022 break)

4.3.1 Theoretical integration and interpretation for the 2022 period

The error dynamics observed along the structural breaks are consistent in several aspects with the literature on non-stationary time series and concept drift; however, the correlations presented here serve primarily as an interpretative framework. The abrupt surge in MAPE following the February 8, 2022 break suggests that the models were calibrated to the statistical structure of the pre-shock regime, and adaptation to the new environment temporarily increased forecasting error. This phenomenon was more pronounced in complex architectures, which may indicate a higher degree of regime sensitivity within the examined sample. Conversely, the reduction in error following the April and June breaks is compatible with the possibility that certain models were partially able to adapt to the new distributional structure. This dynamics suggests a time-varying learning trajectory rather than static performance differences. The results can be plausibly linked to the Adaptive Markets Hypothesis: shocks may temporarily increase uncertainty and degrade predictability, while subsequent consolidation may lead to a partial reduction in error. However, it must be emphasized that due to the univariate design of this analysis, the direct identification of energy market, geopolitical, or other external mechanisms is not performed. Accordingly, the theoretical connections presented should be read not as direct proof, but as interpretative possibilities consistent with the empirical patterns. The results of Period 1 thus suggest that the relationship between model complexity and structural instability is non-linear, highlighting the significance of regime-dependent model selection and adaptive retraining in the forecasting of financial time series.

4.4 Structural Breaks in 2025

Based on our investigations, we conducted a detailed analysis of three specific breakpoints during this period. According to the results (Figure 8), a positive ΔMAPE is observable for nearly all models following the breakpoint on March 31, 2025, indicating that the initial shock significantly impaired forecasting accuracy. The most substantial error anomalies occurred in complex, Transformer-based, and deep neural models - particularly Autoformer, N-BEATS, and TFT - where the ΔMAPE exceeded the values of classical and simpler architectures by orders of magnitude. In contrast, ARIMA and the simpler recurrent and convolutional models (GRU, LSTM, RNN, TCN, DLinear, PatchTST) exhibited substantially lower sensitivity. This suggests that higher model complexity entails increased vulnerability during the initial phase of a regime shift. A marked turnaround was observed at the June 11, 2025 breakpoint, as a negative ΔMAPE appeared across all models. This indicates that, in the new regime, overall forecasting performance improved compared to the pre-shock state. The most significant improvements were again seen in complex deep learning models, especially Autoformer, Informer, TCN, and RNN. This pattern points to an adaptive learning and structural adjustment capability, meaning that after an initial performance degradation, these models effectively internalize the new dynamics. For ARIMA and more linear architectures, the improvement was more moderate, suggesting a stable but less aggressive adaptation mechanism. At the October 6, 2025 break, a mixed pattern emerged once again. Most models showed a positive ΔMAPE, indicating that this break posed a renewed challenge to forecasting accuracy. A significant decline in performance was observed particularly for N-BEATS and Autoformer, suggesting that complex models are more sensitive to multiple or recurring regime shifts in the context of univariate forecasting. Meanwhile, simpler recurrent and convolutional models (GRU, RNN, TCN, DLinear) continued to show relatively moderate shifts, signaling more robust but less flexible behavior. The results demonstrate that the shock sensitivity and adaptation dynamics of the models are heavily architecture-dependent. While complex deep learning models react with larger errors during initial regime shifts, they are capable of significant performance gains following stabilization. Conversely, classical and simpler neural models exhibit more balanced and stable behavior, but their adaptation potential is more limited, especially in the face of repeated shocks.

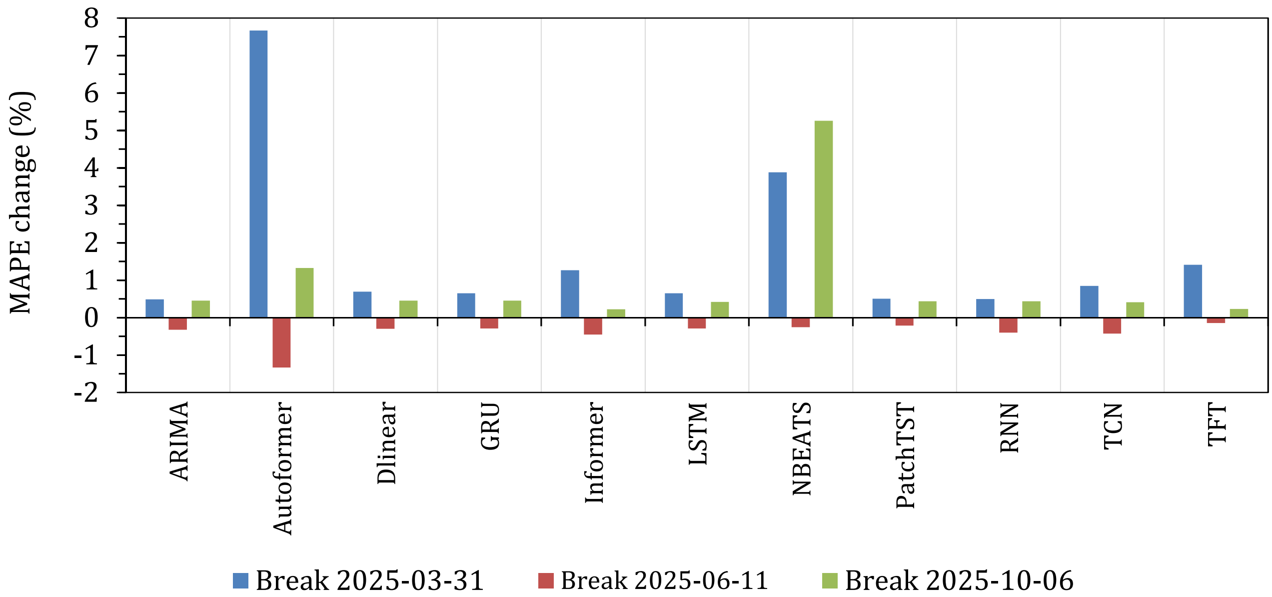

Fig. 8. Shock sensitivity analysis; Delta MAPE for each model around detected structural breaks in Period 2 (Year 2025)

The heatmap (Figure 9) reveals a markedly regime-dependent error dynamics that extends far beyond the differences suggested by aggregated MAPE metrics.

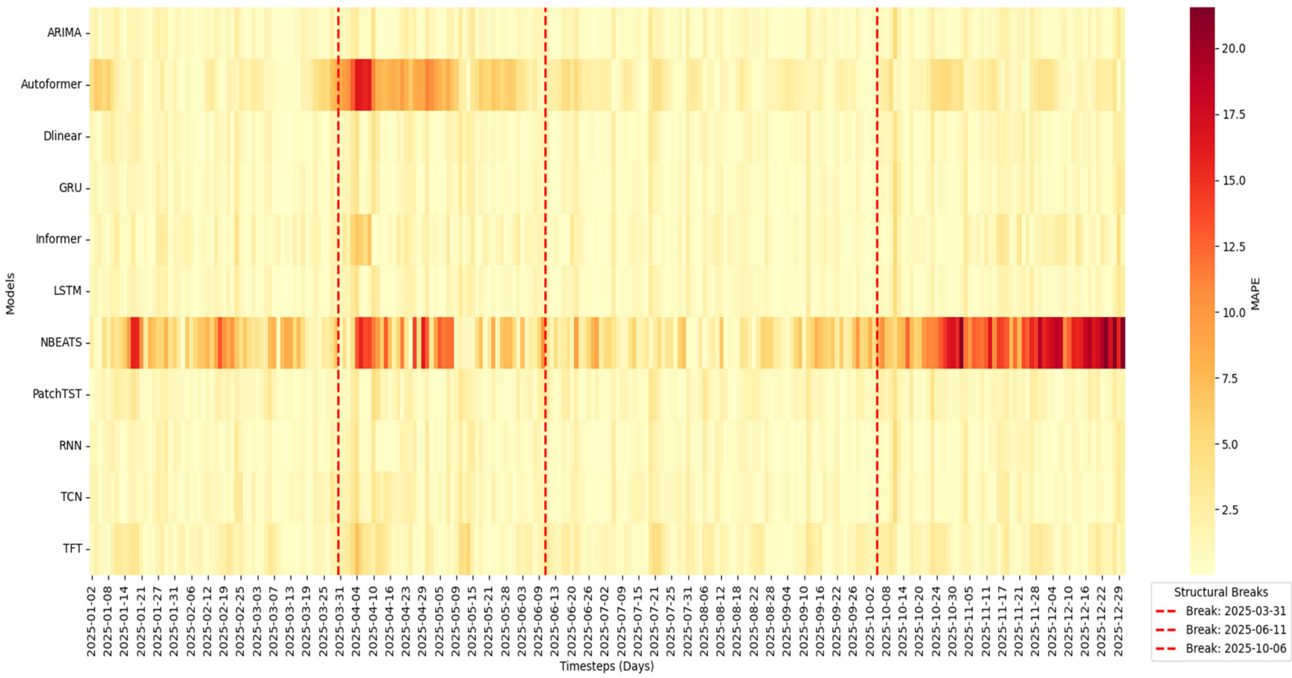

Fig. 9. Model-level errors over time (2025) with structural breaks

Following the structural break on March 31, 2025, a well-defined error clustering emerges for several models, particularly among high-capacity, deep learning architectures. For the Autoformer and N-BEATS models, an intense and prolonged surge in MAPE is observable after the breakpoint, indicating a highly unstable initial reaction to the regime shift. This suggests that the representation space of these models was closely fitted to the pre-break patterns; thus, the sudden structural change induced a generalization problem. This phenomenon can be interpreted as local overfitting within a quasi-stationary environment, where the learned temporal structures rapidly lose their validity due to the impact of an exogenous shock. In contrast, the error fluctuations for the Informer and TFT are less extreme and of shorter duration, which may point to divergent attention mechanisms and inductive biases, particularly in handling long-term dependencies. Following the second breakpoint on June 11, 2025, a clear error convergence is observable in several complex models. The heatmap visually confirms that for the Autoformer and Informer, MAPE levels gradually lighten, indicating an adaptive representation restructuring. In this case, the models are capable of internalizing the volatility and trend dynamics of the new regime; thus, adaptation is realized not merely through parameter fine-tuning, but through the rearrangement of internal temporal mappings. However, for the N-BEATS model, the higher error level persists longer, suggesting slower adaptation stemming from the model's highly deterministic, block-based structure. Conversely, classical and lower-complexity models, specifically ARIMA, DLinear, GRU, RNN, and TCN, do not exhibit pronounced error clustering around any of the breakpoints. The temporal progression of their errors is more homogeneous, and the color scale moves within a narrower range, indicating greater short-term robustness. At the same time, the heatmap shows that these models do not achieve the same degree of post-shock error reduction as the more complex architectures, supporting a dynamic interpretation of the bias–variance trade-off. Following the third breakpoint on October 6, 2025, the error structure overall becomes more balanced. Although higher error levels reappear for certain models - particularly N-BEATS - most architectures do not develop such a marked or persistent error cluster as seen after the first break. This suggests that the market, and in parallel, the models' predictive space, is converging toward a new, more stable equilibrium, in which the learned representations are already less vulnerable to further structural shifts. The heatmap thus empirically confirms that structural breaks do not merely cause temporary performance degradation but generate architecture-specific adaptation trajectories. Consequently, model complexity is not an inherent advantage or disadvantage, but a specifically regime-dependent performance factor interpretable through the interaction between learning dynamics and market structural instability.

The average MAPE values for the one-month intervals before and after the breakpoints (Figure 10) reveal pronounced regime effects in 2025 as well, exhibiting a clearly architecture-dependent pattern. Following the March 31, 2025 breakpoint, a distinct increase in error levels is observable across all models in the post-shock period, signaling the initial destabilizing effect of the regime shift. The most dramatic degradation again appears in the high-capacity models: for the Autoformer, the MAPE increases more than fourfold (from 2.49% to 10.16%), while the N-BEATS model also experiences a significant surge in error (from 4.28% to 8.15%). This points to a pronounced generalization instability, reflecting a lack of initial adaptation to the new market environment. In contrast, for ARIMA, DLinear, and the recurrent architectures (GRU, LSTM, RNN), the increase in MAPE is much more moderate. This pattern suggests short-term robustness, while also indicating that these models are less sensitive to the finer structural elements of regime shifts.

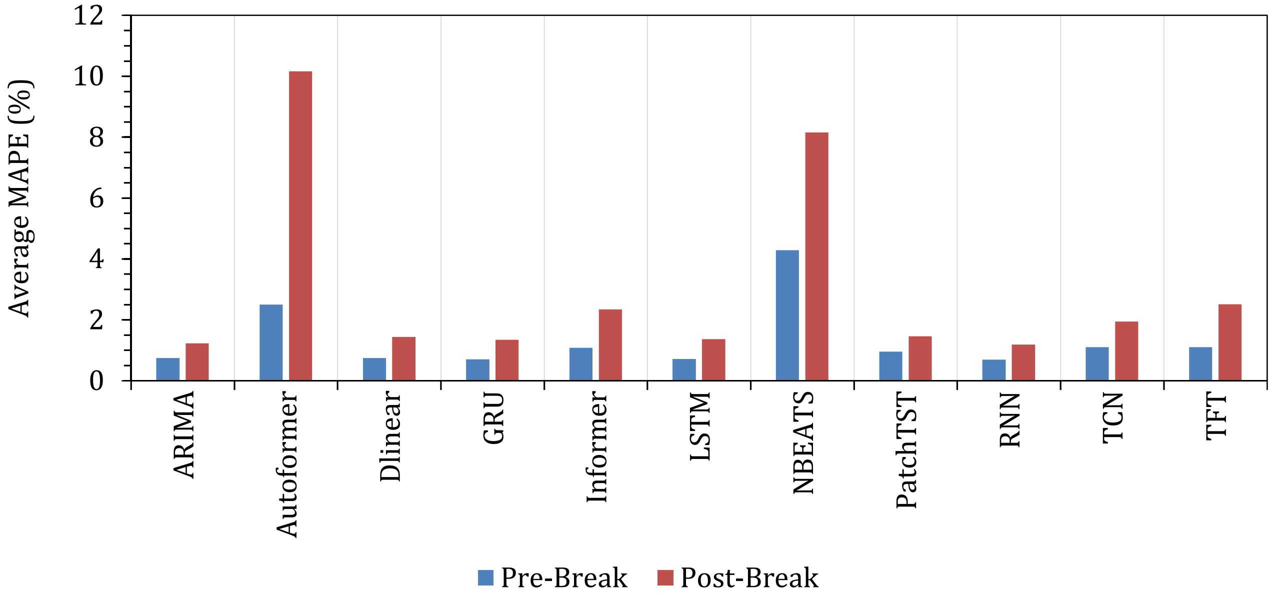

Fig. 10. Pre- and post-break average MAPE performance in Period 2 (1-month before and after 31.03.2025 break)

In the case of the second breakpoint on June 11, 2025 (Figure 11), an inverse trend is observable for most models. During the post-shock period, average MAPE values decreased for nearly all architectures, particularly among complex deep learning models. The error level for the Autoformer moderated significantly (from 3.91% to 2.58%), while for the Informer, GRU, LSTM, and RNN, the MAPE dropped below pre-break levels. This suggests that following the initial turbulent phase, the models, especially those with higher capacity, effectively internalize the trend and volatility structures of the new regime. However, for the N-BEATS model, the improvement was more moderate (from 4.02% to 3.76%), pointing to slower adaptation dynamics resulting from the model's highly structured, deterministic architecture. In the case of more classical models, the reduction in error was stable but limited in magnitude, indicating a balanced but less regime-sensitive behavior.

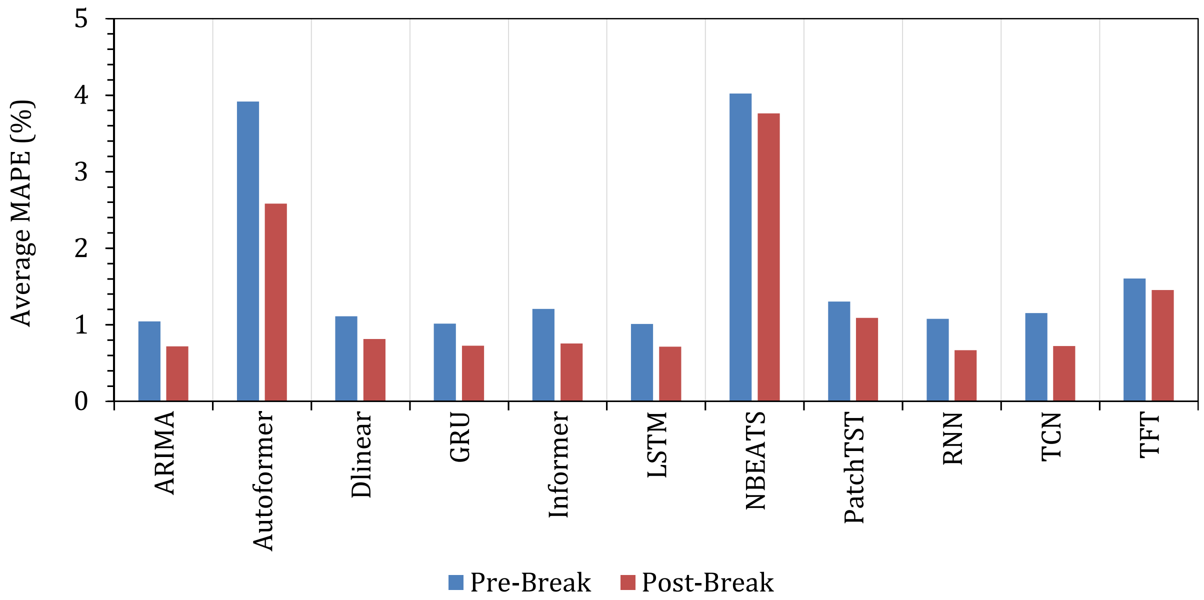

Fig. 11. Pre- and post-break average MAPE performance in Period 2 (1-month before and after 11.06.2025 break)

The third break on October 6, 2025 (Figure 12) again presents a heterogeneous picture. For several models - including ARIMA, DLinear, and the recurrent architectures - the post-break MAPE increases, pointing to forecasting difficulties caused by the renewed structural shift. However, the most pronounced degradation is once again observed in the case of N-BEATS, where the MAPE nearly doubles (rising from 5.41% to 10.66%). This suggests that the model is particularly sensitive to repeated or cumulative regime shifts, and its adaptive capacity is highly context-dependent. In contrast, for the Informer and partially the TFT, the increase in error levels is more moderate, suggesting that the representation mechanisms of these architectures are capable, to some extent, of buffering the impact of subsequent shocks.

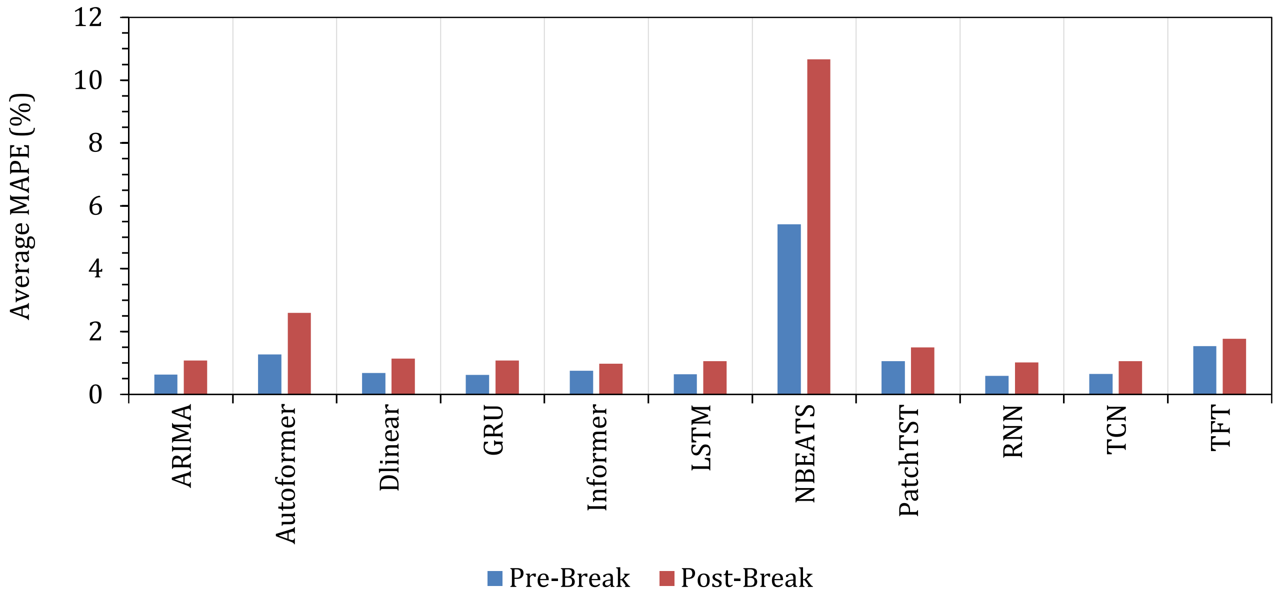

Fig. 12. Pre- and post-break average MAPE performance in Period 2 (1-month before and after 06.10.2025 break)

The 2025 results are consistent in several aspects with the literature on non-stationary time series and concept drift; however, the interpretations formulated here serve primarily as a plausible explanatory framework. The abrupt surge in MAPE following the March break suggests that the models were calibrated to the statistical structure of the pre-break regime, which the break modified discontinuously. This effect was more pronounced in complex architectures, which is compatible with the possibility of higher regime sensitivity within the examined sample. Conversely, the reduction in error following the June break may indicate that certain models were partially able to adapt to the new distributional structure. This dynamics suggests a time-varying learning trajectory rather than static performance differences. However, the univariate framework does not allow for the direct identification of individual external mechanisms. Accordingly, the theoretical connections presented should be treated not as direct proof, but as interpretations consistent with the empirical patterns. The 2025 results also confirm that the relationship between model complexity and structural instability is non-linear, emphasizing the importance of regime-dependent model selection and adaptive retraining.

4.5 Comparative Synthesis of Aluminum: 2022 vs. 2025

In the case of the aluminum time series, the 2022 and 2025 results exhibit qualitatively similar but quantitatively different regime dynamics. In 2022, the ΔMAPE analysis associated with the February break indicates a clear, widespread error shock, particularly among high-capacity deep learning models. On the heatmaps, this shock appears as a spatially and temporally concentrated pattern that gradually dissolves. The pre-post MAPE comparison confirms this: following the first break, the generalization instability of complex models dominates, while in the April and June breaks, error reduction is observed in several cases. In 2025, a similar mechanism emerges but with a moderate amplitude. The absolute magnitude of ΔMAPE values is lower, error peaks on the heatmaps are shorter-lived, and the pre-post average MAPE differences are more subdued. This suggests that the predictive instability associated with shocks was more moderate in 2025, rather than indicating that the market had reached a completely stable or "learned" state. The comparison between the two years shows that architecture-dependent shock-adaptation dynamics is a recurring phenomenon, while the magnitude and persistence of errors remain time-varying, market-state-dependent characteristics. This supports the conclusion that the 2022 analysis carries broader methodological lessons rather than being an isolated case.

5. Discussion

The findings of this study contribute to the resource policy discourse on critical raw materials and industrial metals in several ways. A decisive argument in favor of utilizing univariate models is operational responsiveness. While fundamental macroeconomic indicators are often available only with significant latency, aluminum futures prices can be processed immediately as high-frequency data. This enables real-time strategic adaptation, minimizing informational asymmetry arising from data reporting delays. This interpretation is consistent with the literature emphasizing that forecasting capability itself constitutes a strategic resource in industrial and resource-policy decision-making, particularly under conditions of elevated uncertainty and supply-chain vulnerability [12-14]. Reliable and rapidly updateable forecasts may therefore support more resilient procurement strategies, inventory management, and industrial planning.

One of the most significant findings is that predictive performance is regime-dependent, and model hierarchies may become unstable during structural breaks. The identification of breakpoints based on PELT and Bai-Perron tests empirically supports the view that the price dynamics of the aluminum market can be decomposed into several distinct structural phases. This observation is closely aligned with previous literature emphasizing that commodity markets are characterized by parameter instability, non-linear adjustment mechanisms, and regime-sensitive predictability [48,49,51,52]. Moreover, the post-2022 environment can be interpreted as a practical manifestation of the structural vulnerability discussed in recent resource-economics studies, where energy shocks, geopolitical fragmentation, and financial uncertainty fundamentally altered commodity market dynamics [56-59]. Our results (see Figure 1 and Figure 2) demonstrate that during turbulent periods, parsimonious models such as ARIMA or GRU provide a more stable forecasting foundation for strategic inventory decisions than more complex architectures. This result is broadly consistent with empirical findings suggesting that simpler or hybrid approaches may outperform highly parameterized deep learning architectures in noisy and structurally unstable environments [43-45]. Although recurrent neural architectures are capable of capturing non-linear dependencies, the present results indicate that excessive model complexity may amplify sensitivity to abrupt structural changes. The relatively stable performance of ARIMA, GRU, and LSTM across multiple horizons further supports the argument that robustness and generalization capacity may be more valuable than marginal improvements in in-sample optimization. This result transcends purely methodological questions; however, identifying deeper economic mechanisms within the current univariate empirical framework is inherently limited. Due to its energy-intensive production structure, aluminum prices may be particularly sensitive to energy market volatility and input-cost shocks, but the direct testing of these channels is outside the scope of this study. Nevertheless, the findings remain compatible with previous studies emphasizing the strong interdependence between industrial metals, energy markets, and macroeconomic cycles [10,25,34]. The structural breaks identified around 2022 and 2025 also coincide with the broader global environment characterized by energy crises, supply-chain disruptions, and heightened geopolitical uncertainty discussed extensively in the literature [61-63]. The results indicate that increasing model complexity did not yield a systematic advantage. In several cases, Transformer-based architectures underperformed relative to parsimonious benchmark models. Within the current data environment and configuration, this suggests that the structural uncertainty of commodity markets may limit the benefits of complex architectures. Similar conclusions appear in recent forecasting literature, where the predictive superiority of high-capacity models is found to be highly regime-sensitive and dependent on the stability of the underlying distribution [11,69,70]. In structurally unstable environments, the flexibility of attention-based architectures may come at the cost of increased variance and reduced short-term robustness. The strong performance of recurrent models in the Model Confidence Set analysis suggests that architectures balancing adaptability and parsimony may be particularly advantageous for industrial commodity forecasting. Consequently, the interpretation of predictive performance gains a resource-policy dimension. Supply security and industrial policy decisions, such as strategic stockpiling, long-term procurement contracts, or capacity expansion investments, are implicitly built upon future exchange rate trajectories. If model superiority is regime-sensitive, robustness and stability of dominance may be more important for decision-makers than marginal differences in accuracy. This interpretation aligns with the broader resource-policy literature emphasizing that predictive uncertainty itself should be treated as a strategic risk factor in industrial resilience planning [5,18,37]. The findings therefore support the argument that forecasting models should not be evaluated exclusively through conventional statistical metrics, but also through their capacity to maintain stable performance under structurally changing market conditions. Based on the results, simpler ARIMA-based models and recurrent neural networks (GRU, LSTM) were more stable members of the Model Confidence Set, indicating lower forecasting risk. This observation is compatible with the literature suggesting that recurrent architectures are particularly effective in capturing temporal dependencies and medium-term cyclical dynamics in commodity markets [41,42]. At the same time, the weaker relative performance of certain Transformer-based architectures indicates that the advantages of attention mechanisms may not automatically materialize in financial environments characterized by abrupt breaks, high noise levels, and limited sample sizes. The results therefore reinforce the importance of architecture-specific evaluation rather than assuming universal superiority of newer deep learning frameworks. The study's findings suggest that commodity market forecasting is not merely an algorithmic optimization problem but a task constrained by structural uncertainty. Integrating methodological discipline, model parsimony, and robustness testing is therefore crucial to ensuring resource-policy relevance. In this sense, the turbulent post-2022 period may be interpreted as a stress-test environment demonstrating that predictive performance itself becomes endogenous to the prevailing market regime. Consequently, future forecasting frameworks for industrial metals may benefit from adaptive retraining procedures, regime-sensitive model selection, and explicit treatment of uncertainty, particularly in the context of critical raw materials and industrial resilience strategies.

5.1 Theoretical Implications

The findings advance the theoretical understanding of regime dependence in forecasting performance. By showing that predictive dominance is not fixed but shifts with structural breaks, the results lend empirical support to the Adaptive Markets Hypothesis and to the concept-drift literature, and they suggest that model complexity should be theorized as a source of variance under distributional change rather than as a uniform predictor of accuracy.

5.2 Practical Implications

For practitioners engaged in inventory planning, procurement, and short-term hedging, the results recommend prioritizing robust, parsimonious models - ARIMA, GRU, and RNN - over more complex Transformer-based architectures, particularly when forecasts must remain reliable through periods of structural change rather than only under stable conditions.

5.3 Managerial Implications

Decisions on strategic stockpiles, long-term supply contracts, and capacity investments should incorporate an explicit assessment of model instability, not average accuracy alone. In this sense, choosing a forecasting model under regime uncertainty is itself a multi-criteria decision problem, in which accuracy must be weighed against stability, complexity, and interpretability rather than optimized in isolation; this framing is consistent with the broader body of work that pairs machine-learning-based predictive tools with structured multi-criteria decision-making frameworks to support transparent, data-driven choices in other applied decision domains [86]. Embedding this trade-off explicitly into procurement and risk-management practice can reduce the exposure of industrial decision-makers to the instability documented in this study.

6. Conclusion

This study compared the forecasting performance of eleven widely used time-series forecasting models for aluminum futures prices, including classical econometric approaches, recurrent neural networks, convolutional architectures, and recent Transformer-based models. The analysis was conducted using daily aluminum futures data and multiple forecasting horizons, allowing the evaluation of model behavior under different market conditions represented by the 2022 and 2025 forecasting periods. The empirical results reveal substantial heterogeneity in forecasting performance across model families. Recurrent neural network architectures, particularly GRU and RNN, consistently achieved the lowest forecasting errors across most horizons, while the classical ARIMA model remained highly competitive despite its relatively simple structure. In contrast, Transformer-based architectures such as Informer, Autoformer, and TFT generally failed to outperform simpler alternatives. Similarly, N-BEATS exhibited comparatively weak forecasting performance in both evaluation periods. These findings indicate that increasing model complexity does not automatically lead to improved forecasting accuracy in aluminum futures markets. The statistical comparisons based on Diebold–Mariano tests and Model Confidence Set procedures further support this conclusion. Although differences between the best-performing models were often limited, the results suggest that forecasting superiority is not determined solely by architectural sophistication. In several cases, relatively simple approaches produced forecasting accuracy comparable to, or better than, substantially more complex deep learning models. An important finding of the study is the relative stability of model rankings across the two evaluation periods. While forecasting errors differed between 2022 and 2025, the strongest-performing models remained largely unchanged. This suggests that recurrent neural network architectures and ARIMA-based approaches provide robust forecasting performance under a range of market conditions. At the same time, the results also demonstrate that no single model can be considered universally optimal across all forecasting horizons and environments. From a methodological perspective, the findings contribute to the growing literature questioning whether increasingly complex forecasting architectures necessarily provide superior predictive performance in financial and commodity markets. The results support previous evidence suggesting that simpler and more parsimonious models can remain highly competitive when forecasting noisy and structurally unstable time series.

The study is subject to several limitations. First, the analysis employs a univariate forecasting framework based exclusively on aluminum futures prices. Consequently, potentially informative exogenous variables such as energy prices, exchange rates, industrial production indicators, inventories, or measures of geopolitical uncertainty were not incorporated. Second, the evaluation focuses on a single commodity market, which may limit the generalizability of the findings to other raw materials. Third, although the selected periods capture different market environments, future research could employ rolling-window or regime-switching frameworks to examine forecasting performance under a broader range of market conditions.

Future research may extend the analysis by incorporating multivariate forecasting frameworks, probabilistic forecasting methods, ensemble approaches, and adaptive model-selection procedures. Such developments may provide further insights into the conditions under which advanced machine learning architectures can generate meaningful forecasting improvements in commodity markets. Overall, the results demonstrate that model complexity alone is not a sufficient predictor of forecasting success. For aluminum futures forecasting, parsimonious econometric and recurrent neural network approaches remain strong benchmarks and, in many cases, outperform considerably more sophisticated forecasting architectures.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgement

This research was carried out within the framework of the AI-Assisted Supply Chains and Marketing Research Group at the Hungarian University of Agriculture and Life Sciences (MATE).

References

[1] Miller, W. S., Zhuang, L., Bottema, J., Wittebrood, A., De Smet, P., Haszler, A., & Vieregge, A. J. M. S. (2000). Recent development in aluminium alloys for the automotive industry. *Materials Science and Engineering: A*, *280*(1), 37–49. https://doi.org/10.1016/S0921-5093(99)00653-X.

[2] Dursun, T., & Soutis, C. (2014). Recent developments in advanced aircraft aluminium alloys. *Materials and Design, 56*, 862–871. https://doi.org/10.1016/j.matdes.2013.12.002.

[3] Graf, A. (2021). Aluminum alloys for lightweight automotive structures, in: *Materials, Design and Manufacturing for Lightweight Vehicles.* Woodhead Publishing, pp. 97–123. https://doi.org/10.1016/B978-0-12-818712-8.00003-3.

[4] Virta. (2023). The Global Electric Vehicle Market Overview in 2023: *Statistics & Forecast*. https://www.virta.global/en/global-electric-vehicle-market.

[5] Renner, S., & Wellmer, F. W. (2020). Volatility drivers on the metal market and exposure of producing countries. *Mineral Economics, 33*(3), 311–340. https://doi.org/10.1007/s13563-019-00200-8.

[6] Ding, Y., Chen, S., Li, H., Sun, Q., Chen, H., & Yu, H. (2023). Causality inference among base metal, rare metal and precious metal markets. *Resources Policy, 85*, 104058. https://doi.org/10.1016/j.resourpol.2023.104058.

[7] Sun, H., Zhang, X., & Luo, C. (2025). A Review of Carbon Pricing Mechanisms and Risk Management for Raw Materials in *Low-Carbon Energy Systems. Energies, 18*(13), 3401. https://doi.org/10.3390/en18133401.

[8] Safarzynska, K., & Kryvyy, T. (2025). Integrating metals and minerals into climate-economic models: a review. *Climatic Change, 178*(7), 128. https://doi.org/10.1007/s10584-025-03966-9.

[9] Ernanto, Wiryono, S. K., & Faturohman, T. (2025). Macro and micro factors on global copper pricing: a historical data analysis. *Mineral Economics*, *38*(4), 845–867. https://doi.org/10.1007/s13563-025-00505-x.

[10] Marañon, M., & Kumral, M. (2020). Dynamics behind cycles and co-movements in metal prices: An empirical study using band-pass filters. *Natural Resources Research, 29*(3), 1487–1519. https://doi.org/10.1007/s11053-019-09535-z.

[11] Oikonomou, K., Damigos, D., & Dimitriou, D. (2025). Globality in the metal markets: Leveraging cross-learning to forecast aluminum and copper prices. *Resources Policy, 103*, 105558. https://doi.org/10.1016/j.resourpol.2025.105558.

[12] Liu, Q., Liu, M., Zhou, H., & Yan, F. (2022). A multi-model fusion based non-ferrous metal price forecasting. *Resources Policy, 77*, 102714. https://doi.org/10.1016/j.resourpol.2022.102714.

[13] Huang, Y. T., Bai, Y. L., Yu, Q. H., Ding, L., & Ma, Y. J. (2022). Application of a hybrid model based on the Prophet model. ICEEMDAN and multi-model optimization error correction in metal price prediction. *Resources Policy, 79*, 102969. https://doi.org/10.1016/j.resourpol.2022.102969.

[14] Tseng, S. H., & Nguyen, M. H. (2025). A metal price forecasting framework optimized with TLBO metaheuristic and selective for highest returns. *Expert Systems with Applications, 285*, 127937. https://doi.org/10.1016/j.eswa.2025.127937.

[15] Aïd, R., Basei, M., & Ferrari, G. (2025). A stationary mean-field equilibrium model of irreversible investment in a two-regime economy. *Operations Research, 73*(5), 2351–2374. https://doi.org/10.1287/opre.2023.0250.

[16] Litvinenko, V. S. (2020). Digital economy as a factor in the technological development of the mineral sector. *Natural Resources Research, 29*(3), 1521–1541. https://doi.org/10.1007/s11053-019-09568-4.

[17] Distefano, T., Lodi, L., & Biggeri, M. (2024). Material footprint and import dependency in EU27: Past trends and future challenges*. Journal of Cleaner Production, 472*, 143384. https://doi.org/10.1016/j.jclepro.2024.143384.

[18] Ramírez-Márquez, C., Posadas-Paredes, T., & Ponce-Ortega, J. M. (2025). From resource abundance to responsible scarcity: rethinking natural resource utilization in the age of hyper-consumption. *Resources, 14*(8), 118. https://doi.org/10.3390/resources14080118.

[19] Li, L., Peng, R., Lu, X., Yang, T., Li, H., & Chi, Y. (2025). Sustainable Supply of Critical Metal Resources During Power Sector's Low-Carbon Transition: A Systematic Literature Review. *Wiley Interdisciplonary Review. Energy Environ, 14*(2), e70005. https://doi.org/10.1002/wene.70005.

[20] U.S. Geological Survey. (2025). *Mineral Commodity Summaries 2025*. https://doi.org/10.3133/mcs2025.

[21] Umeaduma, C. M. G., & Dugbartey, A. N. (2023). Effect of exchange rate volatility on export competitiveness and national trade balances in emerging markets. *International Journal of Computer Applications Technology and Research, 12*(11), 57–71. https://doi.org/10.7753/IJCATR1211.1008.

[22] Liu, S., & Lin, M. (2024). Trade dynamics of ferrous metals in emerging and developing countries. *Resources Policy, 90*, 104742. https://doi.org/10.1016/j.resourpol.2024.104742.

[23] Gainetdinova, A., & Sohag, K. (2025). The dynamic response of Russian exchange rate to precious metals and minerals prices. *Mineral Economics, 38*(1), 59–75. https://doi.org/10.1007/s13563-024-00449-8.

[24] Islam, M. S., Islam, M. M., Ahmed, F., Rehman, A. U., Alam, M. F., & Islam, M. A. (2025). Exploring nexus between oil price shocks and copper production: analysing the role of mineral prices and geopolitical factors in Saudi Arabia. *Computational Economics*. https://doi.org/10.1007/s10614-025-10916-1.

[25] Månberger, A., & Stenqvist, B. (2018). Global metal flows in the renewable energy transition: Exploring the effects of substitutes, technological mix and development. *Energy Policy, 119*, 226–241. https://doi.org/10.1016/j.enpol.2018.04.056.

[26] Chen, J., Luo, Q., Tu, Y., Ren, X., & Naderi, N. (2023). Renewable energy transition and metal consumption: Dynamic evolution analysis based on transnational data. *Resources Policy, 85*, 104037. https://doi.org/10.1016/j.resourpol.2023.104037.

[27] Lin, B., & Zhang, Z. (2025). Geopolitics of renewable energy development: The role of energy metals. *Energy Policy 198,* 114482. https://doi.org/10.1016/j.enpol.2024.114482.

[28] Daehn, K., Basuhi, R., Gregory, J., Berlinger, M., Somjit, V., & Olivetti, E. A. (2022). Innovations to decarbonize materials industries. *Nature Reviews Materials, 7*(4), 275–294. https://doi.org/10.1038/s41578-021-00376-y.

[29] Patil, A. B., Struis, R. P., & Ludwig, C. (2023). Opportunities in critical rare earth metal recycling value chains for economic growth with sustainable technological innovations. *Circular Economy and Sustainability, 3*(2), 1127–1140. https://doi.org/10.1007/s43615-022-00204-7.

[30] DebRoy, T., & Elmer, J. W. (2024). Metals beyond tomorrow: Balancing supply, demand, sustainability, substitution, and innovations. *Materials Today, 80*, 737–757. https://doi.org/10.1016/j.mattod.2024.09.007.

[31] Pincheira, P., & Hardy, N. (2021). Forecasting aluminum prices with commodity currencies. *Resources Policy, 73*, 102066. https://doi.org/10.1016/j.resourpol.2021.102066.

[32] Dogah, K. E., Wu, Y., & Rognone, L. (2024). Commodity futures market conditions and climate policy risk: Evidence from energy and metals markets. *The Journal of Futures Market, 44*(10), 1694–1709. https://doi.org/10.1002/fut.22544.

[33] Oben, R. J., Zhakanova Isiksal, A., Assi, A. F., & Faisal, F. (2025). Volatility and return connectedness across decentralized finance assets, precious and industrial metals, green energy and technology markets. *Mineral Economics, 38*(4), 869–890. https://doi.org/10.1007/s13563-025-00507-9.

[34] Jia, L., Xu, R., Wu, J., Song, M., & Chen, X. (2023). Impacts of geopolitical risk and economic policy uncertainty on metal futures price volatility: Evidence from China. *Resources Policy, 87*, 104328. https://doi.org/10.1016/j.resourpol.2023.104328.

[35] Ghosh, B., & Ghosh, A. (2025). Impact of geopolitical risk & economic policy uncertainty among energy transition metals in extreme market conditions. *Mineral Economy, 38*(3), 551–567. https://doi.org/10.1007/s13563-025-00504-y.

[36] Chen, Y., & Li, J. (2025). The Determinants and Volatility of Prices in the Energy and Critical Mineral Markets, in: *Energy and Critical Mineral Security in China*. Springer Nature Singapore, Singapore, pp. 219–241. https://doi.org/10.1007/978-981-96-7147-2_9.

[37] Sarkheil, H., Salahjou, T., & Seyyedi Eshkiki, S. R. (2025). Analysis and forecasting Iran's copper market to move towards low-carbon in the face of climate change. *Mineral Economics, 38*(4), 961–983. https://doi.org/10.1007/s13563-025-00534-6.

[38] García-Gonzalo, E., García Nieto, P. J., Gracia Rodríguez, J., Sánchez Lasheras, F., & Fidalgo Valverde, G. (2023). A support vector regression model for time series forecasting of the COMEX copper spot price. *Logic Journal of the IGPL, 31*(4), 775–784. https://doi.org/10.1093/jigpal/jzac039.

[39] Huang, Y. T., Bai, Y. L., Ding, L., Zhu, Y. J., & Ma, Y. J. (2023). Application of a hybrid model based on ICEEMDAN, Bayesian hyperparameter optimization GRU and the ARIMA in nonferrous metal price prediction. *Cybernetics and Systems, 54*(1), 27–59. https://doi.org/10.1080/01969722.2022.2055383.

[40] Menéndez-García, L. A., García-Nieto, P. J., García-Gonzalo, E., & Lasheras, F. S. (2024). Time series analysis for COMEX platinum spot price forecasting using SVM, MARS, MLP, VARMA and ARIMA models: A case study. *Resources Policy, 95*, 105148. https://doi.org/10.1016/j.resourpol.2024.105148.

[41] Ozdemir, A. C., Buluş, K., & Zor, K. (2022). Medium-to long-term nickel price forecasting using LSTM and GRU networks. *Resources Policy, 78*, 102906. https://doi.org/10.1016/j.resourpol.2022.102906.

[42] Esangbedo, M. O., Taiwo, B. O., Abbas, H. H., Hosseini, S., Sazid, M., & Fissha, Y., (2024). Enhancing the exploitation of natural resources for green energy: An application of LSTM-based meta-model for aluminum prices forecasting. *Resources Policy, 92*, 105014. https://doi.org/10.1016/j.resourpol.2024.105014.

[43] Varshini, A., Kayal, P., & Maiti, M. (2024). How good are different machine and deep learning models in forecasting the future price of metals? Full sample versus sub-sample. *Resources Policy, 92*, 105040. https://doi.org/10.1016/j.resourpol.2024.105040.
[44] Kong, P., Sun, B., Yang, H., & Huang, X. (2024). Nonferrous metal price forecasting based on signal decomposition and ensemble learning. *Journal Process Control, 133*, 103146. https://doi.org/10.1016/j.jprocont.2023.103146.
[45] Nabavi, Z., Mirzehi, M., & Dehghani, H. (2024). Reliable novel hybrid extreme gradient boosting for forecasting copper prices using meta-heuristic algorithms: A thirty-year analysis. *Resources Policy, 90*, 104784. https://doi.org/10.1016/j.resourpol.2024.104784.
[46] Du, P., Gui, F., Shan, L., Xing, Q., & Wang, J. (2025). Probabilistic forecasting of non-ferrous metal prices based on outlier treatment algorithms, quantile regression based deep learning and two-phase multi-objective optimization. *Engineering Applications of Artificial Intelligence, 161*, 112277. https://doi.org/10.1016/j.engappai.2025.112277.
[47] Gouel, C., & Legrand, N. (2025). The role of storage in commodity markets: Indirect inference based on grain data. *Quantitative Economics, 16*(2), 705–747. https://doi.org/10.3982/QE2329.
[48] Sultana, N. (2025). Volatility regimes and structural shifts in geopolitical risk: Evidence from GARCH and breakpoint analysis. *International Review of Economics & Finance, 105*, 104803. https://doi.org/10.1016/j.iref.2025.104803.
[49] Pala, A. (2025). The Impact of Climate, Economic and Geopolitical Shocks on Energy Prices and Risk/Return Coefficient Behavior in Mid-and Long-Term: With Technical Analysis and Structural Break Test, in: *Clean Energy Research: Now and in the Future.* Springer Nature Switzerland, Cham, pp. 307–324. https://doi.org/10.1007/978-3-031-99312-1_29.
[50] Ahmed, F., & Sohag, K. (2025). Spillover effects of separated oil price shocks on regional financial stress amidst Russia–Ukraine and global geopolitical tensions: a novel GVAR approach. *Eurasian Economics Review, 15*(3), 593–639. https://doi.org/10.1007/s40822-025-00329-8.
[51] Herrera, P. A., Marazuela, M. A., & Hofmann, T. (2022). Parameter estimation and uncertainty analysis in hydrological modeling. *Wiley Interdisciplinary Reviews Water, 9*(1), e1569. https://doi.org/10.1002/wat2.1569.
[52] Ouyang, R., Chen, X., Fang, Y., & Zhao, Y. (2022). Systemic risk of commodity markets: A dynamic factor copula approach. *International Reviews of Financial Analysis, 82*, 102204. https://doi.org/10.1016/j.irfa.2022.102204.
[53] Anwer, Z., Khan, A., Naeem, M. A., & Tiwari, A. K. (2025). Modelling systemic risk of energy and non-energy commodity markets during the COVID-19 pandemic. *Annals of Operational Research, 345*(2), 1193–1227. https://doi.org/10.1007/s10479-022-04879-x.
[54] Zeng, H., Abedin, M. Z., Ahmed, A. D., & Lucey, B. (2025). Quantile and time–frequency risk spillover between climate policy uncertainty and grains commodity markets. *Journal of Futures Markets, 45*(6), 659–682. https://doi.org/10.1002/fut.22583.
[55] Gambhir, A., Albert, M. J., Doe, S. S., Donges, J. F., Farajalla, N., Giatti, L. L., Richardson, R. (2025). A systemic risk assessment methodological framework for the global polycrisis. *Nature Communications, 16*(1), 7382. https://doi.org/10.1038/s41467-025-62029-w.
[56] Gao, W., Zhang, H., Zhang, H., & Yang, S. (2024). The role of G7 and BRICS country risks on critical metals: Evidence from time-and frequency-domain approach. *Resources Policy, 88*, 104257. https://doi.org/10.1016/j.resourpol.2023.104257.
[57] Gupta, H., & Katoch, R. (2024). How do carbon emissions spillovers reshape metal market dynamics? Time–frequency insights on precious and non-precious metals. *Studies in Microeconomics,* 23210222241284389. https://doi.org/10.1177/23210222241284389.
[58] Woode, J. K., Adam, A. M., Owusu Junior, P., & Idun, A. A. A. (2024). Industrial metal and cryptocurrency market plummets: Interdependence, policy uncertainty, or investor sentiments?. *Economia e Politica Industriale: Journal of Industrial and Business Economics, 51*(4), 1001–1040. https://doi.org/10.1007/s40812-024-00315-2.
[59] Cresti, L., Mazzilli, D., Patelli, A., Sbardella, A., & Tacchella, A. (2025). Vulnerabilities and capabilities in the EU Automotive industry: Leveraging Input-Output Analysis and Economic Complexity. *arXiv Prepr.* arXiv2501.01781. https://doi.org/10.48550/arXiv.2501.01781.
[60] Saint Akadiri, S., & Ozkan, O. (2025). Critical minerals and structural oil shocks: Evidence from wavelet cross-quantile correlation. *Resources Policy, 103*, 105570. https://doi.org/10.1016/j.resourpol.2025.105570.
[61] Ushie, O. J., Nwokolo, S. C., Ohiero, P. O., Iwuji, P. C., Eyime, E. E., & Ogbulezie, J. C. (2025). Resilient strategies to mitigate the volatility of copper and aluminium critical mineral demand for net-zero electricity network production. *Sustainable Energy Research, 12*(1), 56. https://doi.org/10.1186/s40807-025-00178-8.
[62] Yin, L., Zhang, J., Cao, H., & Li, Y. (2025). The Role of Policy on Commodity Markets: From the Perspective of Narrative News Based on NLP. *Journal of Futures Markets, 45*(11), 1871–1910. https://doi.org/10.1002/fut.70017.
[63] Brühl, V. (2025). The economic rise of China–an analysis of China's growth drivers. *International Economics and Economic Policy, 22*(1), 16. https://doi.org/10.1007/s10368-024-00640-w.

[64] Zhang, Q., Chang, Y., & Chen, Z. (2026). From demand expansion to technology substitution: The structural impact of the energy transition on key metals markets. *Energy & Environment, 37*(1), 483–508. https://doi.org/10.1177/0958305X251389833.

[65] Maghyereh, A., & Ziadat, S. A. (2026). Dynamic Interactions in Futures Markets: Exploring Transitory and Persistent Intraday Volatility Linkages among Oil, Gold, Stocks, and Forex Markets. *Computational Economics*. https://doi.org/10.1007/s10614-025-11249-9.

[66] Yao, X., Almatooq, N., Askin, R. G., & Gruber, G. (2022). Capacity planning and production scheduling integration: improving operational efficiency via detailed modelling. *International Journal of Production Research. 60* (24), 7239–7261. https://doi.org/10.1080/00207543.2022.2028031.

[67] Kuruppu, S. C., Milne, M. J., & Tilt, C. A. (2024). Sustainability control systems in short-term operational and long-term strategic decision-making. *Meditari Accountancy Research, 32*(1), 234–265. https://doi.org/10.1108/MEDAR-12-2021-1548.

[68] Wang, M., Wang, Y., & Feng, C. (2025). Artificial intelligence, institutional environment, and corporate green transformation: Evidence from China's resource-based sector. *International Review of Economics Finance,* 104473. https://doi.org/10.1016/j.iref.2025.104473.

[69] Crespo Cuaresma, J., Fortin, I., Hlouskova, J., & Obersteiner, M. (2024). Regime-dependent commodity price dynamics: A predictive analysis. *Journal of Forecasting, 43*(7), 2822–2847. https://doi.org/10.1002/for.3152

[70] Cifarelli, G. (2025). Commodity Pricing through Crises: A Markov Switching Regime Probabilities Analysis. *International Journal of Finance & Banking Studies, 14*(3). https://doi.org/10.20525/ijfbs.v14i3.4057.

[71] Al-Haddad, A., & Abu-Rayash, A. (2026). Volatility modeling in energy commodity markets: A review of responses to shocks and the path towards resiliency and sustainability. *Applied Energy, 411*, 127559. https://doi.org/10.1016/j.apenergy.2026.127559.

[72] Kontopoulou, V. I., Panagopoulos, A. D., Kakkos, I., & Matsopoulos, G. K. (2023). A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks. *Future Internet, 15*(8), 255. https://doi.org/10.3390/fi15080255.

[73] Amalou, I., Mouhni, N., & Abdali, A. (2022). Multivariate time series prediction by RNN architectures for energy consumption forecasting. *Energy Reports, 8*, 1084–1091. https://doi.org/10.1016/j.egyr.2022.07.139.

[74] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

[75] Zheng, R., Bao, Y., Zhao, L., & Xing, L. (2023). Method to predict alloy yield based on multiple raw material conditions and a PSO-LSTM network. *Journals of Material Research and Technology, 27*, 3310–3322. https://doi.org/10.1016/j.jmrt.2023.10.046.

[76] Dorbane, A., Harrou, F., & Sun, Y. (2023). Exploring deep learning methods to forecast mechanical behavior of FSW aluminum sheets. *Journal of Material Engineering and Performance, 32*(9), 4047–4063. https://doi.org/10.1007/s11665-022-07376-1.

[77] Zhang, J., Li, Z., Wang, E., Yu, B., Li, J., & Ma, J. (2024). A Temporal Network Based on Characterizing and Extracting Time Series in Copper Smelting for Predicting Matte Grade. *Sensors, 24*(23), 7492. https://doi.org/10.3390/s24237492.

[78] Habib, A. B., Alam, M. G. R., & Uddin, M. Z. (2025). AUNET (Attention-based unified network): leveraging attention based N-BEATS for enhanced univariate time series forecasting. *IEEE Access*. https://doi.org/10.1109/ACCESS.2025.3574459.

[79] Oreshkin, B. N., Carpov, D., Chapados, N., & Bengio, Y. (2020). *N-BEATS: Neural basis expansion analysis for interpretable time series forecasting*. In *Proceedings of the International Conference on Learning Representations* (ICLR)).

[80] Zhou, R., Wang, S., Ren, Y., Liu, K., Peng, X., He, L., & Liu, H. (2025). Remaining useful life prediction of lithium-ion batteries based on informer model with multi-feature extraction and Bayesian optimization. *Chemical Engineering Science, 316*, 121878. https://doi.org/10.1016/j.ces.2025.121878.

[81] Jiang, Y., Gao, T., Dai, Y., Si, R., Hao, J., Zhang, J., & Gao, D. W. (2022). Very short-term residential load forecasting based on deep-autoformer. *Applied Energy, 328*, 120120. https://doi.org/10.1016/j.apenergy.2022.120120.

[82] Bao, G., Niu, Y., Cui, B., & Ji, W. (2025). Revealing the correlation between economic indicators and gold prices for forecasting: Medium term forecast framework with data patch. *Expert Systems with Applications,* 129594. https://doi.org/10.1016/j.eswa.2025.129594.

[83] Liu, J., Gong, C., Chen, S., & Zhou, N. (2023). Multi-step-ahead wind speed forecast method based on outlier correction, optimized decomposition, and dlinear model. *Mathematics, 11*(12), 2746. https://doi.org/10.3390/math11122746.

[84] Arrar, D., Ounoughi, C., Kamel, N., Kalvet, T., Tiits, M., & Ben Yahia, S. (2025). High-accuracy prediction of international raw material trade flows using temporal fusion transformer. *International Journal of Data Science and Analytics, 20*(6), 5631–5651. https://doi.org/10.1007/s41060-025-00781-4.

[85] Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020). A systematic review of fundamental and technical analysis of stock market predictions. *Artificial Intelligence Review, 53*(4), 3007–3057. https://doi.org/10.1007/s10462-019-09754-z.

[86] Tešić, D. (2025). A Machine Learning and Multi-Criteria Decision-Making Framework for Student Grade Prediction. *Smart Multi-Criteria Analytics and Reasoning Technologies, 1*(1), 22–32. https://doi.org/10.65069/smart1120253.