

Article

Do Supply-Chain Stress and Geopolitical Risk Predict Strategic Commodity and Clean Energy Market Returns? Evidence from Explainable Machine Learning

Nader Naifar 

Department of Finance, College of Business, Imam Mohammad Ibn Saud Islamic University (IMSIU),
Riyadh 11432, Saudi Arabia; naneifar@imamu.edu.sa

Highlights

What are the main findings?

- The forecasting value of supply-chain stress and geopolitical risk is strongly asset-specific, with the clearest gains observed for aluminum and nickel, while clean energy returns remain more closely linked to macro-financial predictors.
- Model comparison, Diebold-Mariano tests, Model Confidence Set analysis, and SHAP explainability show that nonlinear models can be useful, but no single forecasting algorithm uniformly dominates across all assets and loss functions.

What are the implications of the main findings?

- Forecasting models for strategic commodities and clean energy returns should be tailored to the specific asset rather than applied uniformly across markets.
- Explainable machine learning helps identify when disruption-related variables add meaningful predictive value beyond conventional macro-financial controls, which is useful for investors, risk managers, and policymakers.

Abstract

This study examines whether daily supply-chain stress and geopolitical risk improve the forecasting of strategic commodity and clean energy market returns. Using daily data on aluminum, copper, nickel, and clean energy from 10 February 2015 to 27 February 2026, the analysis compares a baseline forecasting model based on conventional market controls with augmented specifications that incorporate supply-chain stress, geopolitical risk, and their joint effects. The empirical framework combines multiple machine-learning algorithms with SHAP-based explainability to evaluate both forecast performance and the relative importance of predictors. Formal Diebold-Mariano tests are also used to assess whether the forecasting gains from augmented specifications are statistically significant. A Model Confidence Set analysis is further used to identify statistically superior model groups across the full set of algorithm-specification combinations. The results show that disruption-related predictors contain asset-specific forecasting information, while the comparison across algorithms indicates that no single model uniformly dominates across all assets and loss functions. The forecasting gains from disruption-related predictors, however, are strongly asset-specific and statistically uneven. For aluminum returns, augmented specifications that include supply-chain stress and/or geopolitical risk significantly improve forecast accuracy relative to the baseline. For copper returns, the evidence is weaker and mainly associated with geopolitical risk. For nickel returns, the joint inclusion of supply-chain stress and geopolitical risk provides the greatest improvement. By contrast, clean energy



Academic Editor: Luigi Grossi

Received: 5 June 2026

Revised: 12 July 2026

Accepted: 13 July 2026

Published: 15 July 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

returns remain more closely tied to conventional macro-financial conditions, with no statistically significant incremental gains from disruption-related variables. SHAP evidence further indicates that predictor importance is asset-specific rather than dominated by a single market factor across all assets. The findings highlight the importance of combining flexible forecasting methods with economically interpretable tools when evaluating disruption-sensitive commodity and clean energy markets.

Keywords: supply-chain stress; geopolitical risk; strategic commodities; clean energy; explainable machine learning; return forecasting; forecasting; commodity finance; energy finance

1. Introduction

The growing strategic importance of industrial metals and clean energy assets has significantly increased the need for more accurate and economically interpretable forecasting frameworks. Markets linked to industrial metals and clean energy have become increasingly central to debates on industrial resilience, decarbonization, and macro-financial stability. At the same time, recent research shows that forecasting performance in financial and energy markets can be improved by going beyond standard market variables and incorporating broader sources of uncertainty, volatility, and sentiment. For example, external volatility measures have been shown to improve stock-market forecasting, while sentiment and climate-related indicators have been found to enhance prediction in energy and clean energy markets [1–4]. Likewise, machine-learning-based evidence suggests that financial stress can improve the forecasting of clean energy, commodity, green bond, and ESG-related market dynamics [5].

These developments are particularly important in the current global environment, where real-economy disruptions and geopolitical instability increasingly interact with financial markets. The post-pandemic period, the Russia-Ukraine war, disruptions in global shipping routes, and the broader rise in geoeconomic fragmentation have all highlighted the vulnerability of strategic commodity and clean energy markets to shocks that originate outside conventional financial channels. Recent studies confirm that geopolitical risk materially affects energy market vulnerability and tail-risk forecasting [6,7]. Related evidence also shows that trade policy uncertainty and shipping risk significantly shape commodity-market dynamics, while supply-chain pressure affects traditional energy and metal markets [8,9]. More broadly, geopolitical shocks have been shown to alter commodity-market behavior during major crises such as the Russia-Ukraine conflict [10].

Theoretically, supply-chain stress and geopolitical risk may affect returns on strategic commodities and clean energy through distinct yet related channels. Supply-chain stress captures physical frictions in global trade and production networks, including shipping constraints, logistics bottlenecks, inventory pressures, and disruptions in the delivery of intermediate inputs. These mechanisms are especially relevant for aluminum, copper, and nickel because these metals are deeply embedded in industrial production, infrastructure, battery technologies, and low-carbon value chains. Geopolitical risk, by contrast, captures conflict risk, sanctions risk, trade restrictions, security-of-supply concerns, and broader political-strategic uncertainty. These forces may affect returns through changes in risk premia, expectations about global industrial demand, uncertainty over access to key inputs, and shifts in strategic supply conditions. For clean energy, the transmission mechanism is more indirect because clean energy indices consist of listed firms rather than physical commodities. Supply-chain and geopolitical shocks may therefore affect clean energy

returns through financing costs, policy uncertainty, investor sentiment, exchange-rate movements, and broader equity-market risk conditions. This distinction motivates the paper's asset-specific empirical design and the direct comparison between supply-chain stress and geopolitical risk as alternative forecasting channels.

Despite these advances, the literature remains fragmented in at least two respects. First, most studies examine either conventional financial assets, broad energy markets, or specific sustainability-related assets in isolation. Second, while recent work has explored the predictive roles of volatility, climate risk, sentiment, and geopolitical risk, much less attention has been paid to the joint forecasting of strategic commodity and clean energy prices under daily supply-chain stress and geopolitical risk. This gap is notable because disruptions in logistics and trade networks capture a distinct transmission channel from broader geopolitical tensions. Supply-chain stress reflects physical frictions in the movement of goods, shipping, and production networks, whereas geopolitical risk reflects political-strategic instability, conflict risk, and economic realignment. Although related, these two forces may carry different predictive content across assets and market regimes [6,8,9]. This distinction is especially relevant for markets connected to industrial transformation and the low-carbon transition. Strategic commodities such as base and industrial metals are highly exposed to supply-chain bottlenecks, transport stress, and input shortages, whereas clean energy market returns may also react strongly to uncertainty about global policy alignment, sanctions, security-of-supply concerns, and shifts in geopolitical risk. Existing clean energy forecasting studies already show that risk-related predictors can improve forecast performance, including global common volatility, geopolitical risk, climate risk, carbon transition risk, and financial stress [3–5,11]. However, the relative forecasting roles of daily supply-chain stress and geopolitical risk remain underexplored, particularly in a unified framework that compares strategic commodities with clean energy.

Against this background, this paper examines whether daily supply-chain stress and geopolitical risk improve the forecasting of strategic commodity and clean energy market returns. The study is designed to compare benchmark models based on standard market features with augmented models that include disruption-related predictors, thereby allowing us to test whether these indicators provide incremental forecasting value. In doing so, the paper speaks directly to the recent forecasting literature that uses machine learning and explainable artificial intelligence to evaluate the predictive contribution of uncertainty-related variables in financial and energy markets [3,7,11].

Methodologically, the paper adopts an explainable machine-learning framework. This choice is appropriate because the effects of disruption-related predictors are likely to be nonlinear, heterogeneous, and asset-specific. Recent studies show that explainable machine learning is increasingly useful in forecasting clean energy and energy market behavior, energy tail risk, and climate-related market behavior because it combines predictive flexibility with economic interpretability [3,7,11]. This is important in the present context, where the objective is not only to determine whether supply-chain stress and geopolitical risk improve forecasting performance, but also to identify which predictors matter most, for which assets, and under what modeling framework. To preserve the time-series nature of the forecasting exercise, the empirical design uses a chronological train-test split and expanding-window cross-validation. Forecast differences are further evaluated using Diebold-Mariano tests, allowing the analysis to distinguish statistically meaningful improvements from purely descriptive changes in forecast errors. In addition, the analysis implements a Model Confidence Set procedure to compare the full set of algorithm-specification combinations and to identify groups of models that are statistically indistinguishable from the best-performing alternative. This is important because daily return forecasting is characterized by low

signal-to-noise ratios, and small numerical differences in forecast errors may not imply statistically meaningful model superiority.

This study makes several contributions to the forecasting literature. First, it provides direct evidence on whether daily supply-chain stress and geopolitical risk contain incremental predictive information for strategic commodity and clean energy returns beyond conventional macro-financial controls. This is the central contribution of the paper. Second, it jointly examines returns on aluminum, copper, nickel, and clean energy within a unified framework, allowing the predictive content of disruption-related variables to be compared across physical commodity and transition-sensitive equity markets. Third, it distinguishes between physical trade and logistics frictions, captured by SCS, and broader political-strategic uncertainty, captured by GPR. Fourth, it uses multiple forecasting algorithms, Diebold-Mariano tests, Model Confidence Set analysis, and SHAP explainability to evaluate how different model classes exploit the information contained in these predictors and to determine whether observed forecasting gains are statistically meaningful and economically interpretable. These contributions complement recent work on financial stress, geopolitical risk, supply-chain pressure, and volatility-based predictors, while opening a new line of inquiry that directly links supply-chain disruption and geopolitical risk to the forecasting of strategic commodity and clean energy returns [5,6,8,9].

The empirical analysis is guided by three research questions. First, do daily supply-chain stress and geopolitical risk contain predictive information for strategic commodity and clean energy returns beyond conventional market variables? Second, which of these two disruption-related indicators provides stronger out-of-sample forecasting performance? Third, are their predictive effects that are uniform across aluminum, copper, nickel, and clean energy, or do they vary systematically across markets? By addressing these questions, the paper contributes to the forecasting literature through a framework that is both economically grounded and practically relevant under conditions of heightened global disruption.

The remainder of the paper is organized as follows. Section 2 reviews the related literature. Section 3 describes the data and empirical methodology. Section 4 presents the forecasting results and explainability analysis. Section 5 concludes with the main implications for forecasting practice, risk management, and policy.

2. Literature Review

The recent forecasting literature increasingly shows that commodity, energy, and transition-related markets are shaped not only by standard macro-financial variables but also by broader sources of disruption, uncertainty, and geopolitical instability. As global production systems become more fragmented and the energy transition raises the strategic importance of key industrial materials, forecasting frameworks must account for both financial and real-economy shocks. The relevant literature can be grouped into four related strands: forecasting under uncertainty in energy and clean energy markets, the effects of supply-chain stress and shipping-related disruptions, the influence of geopolitical risk on commodity and transition markets, and the rise of machine learning and explainable artificial intelligence in forecasting applications.

2.1. Forecasting Under Uncertainty in Energy and Clean Energy Markets

A first strand of research documents that uncertainty-related variables carry substantial predictive content for energy and clean energy markets. Ref. [4], for example, develop a carbon transition risk concern (CTRC) index and show that it significantly improves volatility prediction in China's clean energy market, outperforming a wide range of economic policy uncertainty indicators. Their results demonstrate that transition-specific risk measures can contain more useful forecasting information than broader macroeconomic

uncertainty proxies. In a related context, ref. [6] show that geopolitical risk materially increases energy market vulnerability and that forecasting models incorporating geopolitical variables outperform those based solely on conventional predictors. Similarly, ref. [7] find that geopolitical risk meaningfully contributes to predicting energy market tail risk, with machine-learning models delivering particularly strong performance.

This line of work is reinforced by studies focusing more specifically on clean energy and transition-related assets. Ref. [12] show that clean energy indices are more inefficient than traditional energy commodities and that their relationship with geopolitical risk changes markedly across crisis periods. Ref. [11] further demonstrate that external risk-based indicators can enhance clean energy forecasting, although their findings suggest that global common volatility outperforms geopolitical risk in predictive accuracy. These studies establish that external uncertainty indicators matter for forecasting, but they also imply that their relative importance depends on the market considered and the type of risk being measured.

2.2. Supply-Chain Stress, Shipping Disruptions, and Commodity-Market Dynamics

A second strand of the literature emphasizes the role of supply-chain stress, shipping conditions, and trade frictions in commodity and energy market behavior. Ref. [8] show that trade policy uncertainty affects commodity-market volatility both directly and indirectly through shipping risk, with the Baltic Dry Index capturing an important transmission channel. Their results suggest that shipping conditions are not merely auxiliary variables but can operate as active intermediaries in the propagation of uncertainty to commodity markets. Ref. [9] reach a similar conclusion, showing that global supply-chain pressure predicts returns in energy and metals markets, particularly under extreme supply-chain conditions. Their evidence also suggests that commodity responses are strongest in stress regimes rather than under normal market conditions, indicating a nonlinear, state-dependent transmission mechanism.

The relevance of supply-chain disruptions extends beyond commodity pricing into broader macroeconomic and energy outcomes. Ref. [13] show that energy uncertainty, supply-chain pressure, and oil-market volatility jointly weaken global economic resilience, especially during major crisis episodes. Ref. [14] also document asymmetric effects of supply-chain bottlenecks on energy returns, with the sign and strength of the relationship differing across countries and across the conditional distribution of energy returns. This literature indicates that supply-chain conditions should be treated as distinct forecasting signals that can affect markets through shipping costs, logistics constraints, and production-network disruptions.

2.3. Geopolitical Risk and the Behavior of Commodity Markets

A third body of work examines the role of geopolitical risk in shaping commodity, energy, and financial market behavior. Ref. [10] show that the Russia-Ukraine conflict generated immediate and heterogeneous effects across agricultural, energy, and financial markets, with particularly strong impacts on European natural gas and regional financial volatility. Ref. [15] find that the influence of geopolitical risk differs across transition-related commodities, with more established commodities displaying stronger and more persistent connections to geopolitical developments than newer transition materials. Ref. [16] likewise document heterogeneous responses of global commodities to geopolitical conflicts, emphasizing both resilience and asset-specific sensitivity under extreme geopolitical shocks.

This literature also highlights that geopolitical risk can alter transmission mechanisms across markets. Ref. [17] show that geopolitical risk predicts volatility spillovers among commodity, exchange-rate, and stock markets, with commodity markets shifting from

risk absorbers to risk transmitters under heightened geopolitical stress. Ref. [18] similarly demonstrate that machine-learning methods outperform traditional autoregressive models in forecasting stock-market volatility when disaggregated geopolitical-risk variables are included. The broader implication of this literature is that geopolitical risk is not only associated with contemporaneous market instability but also contains forward-looking information for forecasting volatility, spillovers, and broader market vulnerability.

2.4. Machine Learning, Deep Learning, and Explainable Forecasting Methods

A fourth strand of research concerns the growing use of machine learning and explainable artificial intelligence in forecasting applications. Ref. [19] show that the FEDformer architecture substantially outperforms recurrent neural networks and a standard econometric benchmark in forecasting commodity indices, highlighting the value of deep-learning methods in capturing complex temporal dependencies. Ref. [20] also find that artificial intelligence models can successfully forecast commodity prices during geopolitical crises, with dynamic Bayesian networks outperforming GRU models during the Russia-Ukraine war. Ref. [21] extend this literature by proposing an interpretable deep learning framework for crude oil futures uncertainty and showing that the drivers of predictive performance vary across short and long horizons.

From a methodological perspective, the choice of flexible machine-learning models is especially relevant for daily commodity and clean energy return forecasting. Daily market returns are typically characterized by weak average predictability, high noise, fat tails, volatility clustering, and abrupt changes in the information environment. This view is consistent with [22], who emphasize the low signal-to-noise ratio of asset returns and show that machine-learning methods can improve return prediction by capturing nonlinearities and complex interactions among predictors that are difficult to model with conventional linear approaches. These features pose challenges for conventional linear forecasting models because the predictive effects of external variables may be nonlinear, conditional on market states, and shaped by interactions among financial stress, exchange rates, oil market conditions, supply-chain stress, and geopolitical risk. Traditional econometric models, such as ARIMA-type specifications, serve as useful benchmarks for persistence and linear dynamics, while GARCH-type models are highly informative for volatility modeling. However, these approaches are less directly designed to capture complex nonlinear interactions among multiple predictors in short-horizon return forecasting. Similarly, MIDAS models are particularly useful when predictors are observed at different frequencies, but the present setting relies on daily predictors and daily returns, making nonlinear interaction modeling more central than mixed-frequency aggregation.

Tree-based ensemble models provide a useful alternative in this context because they can approximate nonlinear and threshold-dependent relationships without imposing a specific functional form. Models such as random forests, gradient boosting, XGBoost, and Extra Trees can capture higher-order interactions among predictors, accommodate heterogeneous effects across market states, and remain relatively robust to multicollinearity and outliers. Extra Trees is particularly attractive because it introduces additional randomization in selecting split points, which can reduce variance and improve generalization in noisy forecasting environments. This is important for daily financial and commodity data, where excessive model flexibility may otherwise lead to overfitting. In contrast, more complex deep-learning models such as LSTM and Transformer architectures can be powerful when long sequences, large sample sizes, and stable temporal dependencies are available, but they may require more tuning and greater computational complexity and can be more vulnerable to overfitting when the signal-to-noise ratio is low. Therefore, the

use of tree-based ensemble models provides a balanced forecasting strategy that combines nonlinear flexibility, robustness, and relatively transparent interpretation.

This theoretical rationale is closely related to the objective of the present study. The purpose is not only to maximize forecasting accuracy but also to identify whether supply-chain stress and geopolitical risk add meaningful information beyond conventional macro-financial controls. Tree-based ensembles are well-suited to this task because they allow the marginal contribution of disruption-related variables to depend on market conditions and interactions with other predictors. Extra Trees is also compatible with SHAP-based explainability, enabling examination of whether forecast gains are economically interpretable rather than purely statistical. Thus, the methodological choice is grounded in the structure of the data and the economic question: daily disruption-related signals may matter only under specific nonlinear conditions, and explainable tree-based models provide a suitable framework for detecting and interpreting such effects.

Recent work has also placed greater emphasis on interpretability and responsible forecasting. Ref. [23] show that higher predictive accuracy may come at the cost of interpretability and computational simplicity, thus emphasizing a governance-performance trade-off in energy forecasting. Ref. [24] demonstrate that random forests combined with SHAP analysis can produce both strong predictive performance and clear economic interpretability in a multi-energy forecasting context. Similar advances are observed in carbon-price forecasting, where refs. [25,26] show that hybrid deep learning models incorporating decomposition techniques, external variables, and sentiment indicators yield strong predictive gains. This literature suggests that the frontier of forecasting research is moving toward approaches that combine predictive power with transparent interpretation of predictor importance.

2.5. Strategic Commodities, Critical Materials, and Transition-Sensitive Markets

A related stream of the literature focuses on the forecasting and risk properties of strategic commodities and critical materials. Ref. [27] show that machine-learning models augmented with statistical dependency measures improve the prediction of price risk boundaries for industrial metals and critical minerals. Their findings underline the importance of nonlinear forecasting tools in markets characterized by interconnectedness and changing strategic significance. Ref. [15] similarly emphasize that transition-related commodities are not homogeneous in their responses to geopolitical dynamics, while Ref. [9] show that metals can serve hedging and diversification roles amid supply-chain and geopolitical stress. These contributions are especially relevant for aluminum, copper, and nickel, which are central to electrification, industrial production, and low-carbon infrastructure. At the same time, clean energy markets appear to respond differently from industrial commodity markets. Refs. [4,11] both confirm that uncertainty-related variables can improve clean energy forecasting, but their focus is on transition-risk concerns, global volatility, and geopolitical risk rather than on supply-chain stress. Ref. [12] further show that clean energy indices exhibit distinct efficiency patterns relative to traditional energy commodities. This body of work suggests that strategic commodities and clean energy should not be assumed to respond to the same predictors with the same intensity and that a joint forecasting analysis can help uncover such differences more clearly.

2.6. Research Gaps and Contributions of the Study

Although the existing literature provides important insights, several gaps remain. First, most studies examine either traditional energy markets, clean energy markets, or commodity markets separately, rather than jointly analyzing strategic commodity and clean energy returns within a unified forecasting framework. Second, while geopolitical risk has

attracted considerable attention, the role of daily supply-chain stress as a forward-looking predictor has received much less attention, especially in combination with geopolitical risk. Existing evidence on shipping risk, trade policy uncertainty, and supply-chain bottlenecks confirms their economic relevance, but these variables have rarely been integrated into a comparative daily forecasting design for aluminum, copper, nickel, and clean energy. Third, although machine learning and explainable AI are increasingly used in energy and transition-market forecasting, relatively few studies examine whether supply-chain stress and geopolitical risk provide incremental out-of-sample forecasting value beyond conventional macro-financial controls. Fourth, many forecasting studies emphasize descriptive accuracy metrics, whereas less attention is given to whether incremental forecasting gains are statistically significant under time-series-consistent validation procedures. This issue is particularly relevant in daily financial and commodity-market applications, where return predictability is typically weak, noisy, and sensitive to validation design.

Although this literature documents the relevance of several external predictors, the present study does not seek to incorporate all possible uncertainty, sentiment, or climate-related indicators. Instead, it focuses on a more specific question: whether daily supply-chain stress and geopolitical risk provide incremental forecasting information beyond a conventional macro-financial benchmark. This benchmark already includes a volatility proxy through the CBOE Volatility Index, together with policy uncertainty, broad equity-market performance, oil-market conditions, and exchange-rate movements. Sentiment and climate-related indicators are therefore treated as related but distinct extensions of the forecasting literature, while the empirical contribution of this paper lies in isolating the predictive roles of physical logistics stress and geopolitical risk across strategic commodities and clean energy returns.

This study addresses these gaps in three main ways. First, it jointly examines three strategic commodity indices, aluminum, copper, and nickel, with a clean energy index, thereby connecting industrial-transition commodities with transition-sensitive financial markets in one empirical framework. Second, it directly compares supply-chain stress, proxied by shipping and logistics conditions, with geopolitical risk, thereby distinguishing between physical trade frictions and broader political-strategic instability. Third, it adopts an explainable machine-learning framework that evaluates not only predictive accuracy across several forecasting algorithms but also the relative importance of the predictors through SHAP analysis. In addition, the forecasting design uses a chronological train-test split, expanding-window cross-validation, and Diebold-Mariano tests to evaluate whether the observed improvements over the baseline model are statistically meaningful. The paper contributes to the forecasting literature by clarifying when disruption-related variables materially improve predictive performance and when conventional macro-financial conditions remain the dominant source of forecast power.

3. Data Description and Research Methodology

This section outlines the data and methodological framework employed in the empirical analysis. It first describes the dataset's construction, the selected variables, their sources, and the economic rationale for their inclusion. It then presents the forecasting setup, including alternative model specifications, machine-learning algorithms, out-of-sample evaluation criteria, time-series validation procedures, formal forecast-comparison tests, and an explainable artificial intelligence approach for assessing predictor importance and interpreting the forecasting results.

3.1. Data Description

This study employs daily data to examine whether supply-chain stress and geopolitical risk improve the forecasting of strategic commodity and clean energy market returns. The sample spans the period from 10 February 2015 to 27 February 2026, yielding 2353 usable daily return observations after log-return transformation and lead-target construction. The analysis focuses on three strategic commodity indices, aluminum (ALU), copper (COP), and nickel (NIC), with a representative clean energy index (CEN). These markets were selected because they occupy a central position in contemporary industrial production, electrification, and energy transition processes. Aluminum, copper, and nickel are core industrial inputs for infrastructure, manufacturing, battery technologies, and low-carbon systems, while clean energy indices capture the financial market dimension of the transition toward renewable and low-emission energy systems. All variables are transformed into continuously compounded daily log returns, computed as $100 \times [\ln(X_t) - \ln(X_{t-1})]$. This return-based specification is appropriate for short-horizon financial forecasting because it focuses on daily market adjustments rather than common trends in price or index levels.

The empirical framework is built around two disruption-related predictors: supply-chain stress (SCS) and geopolitical risk (GPR). Supply-chain stress is included to capture disruptions in trade logistics, shipping conditions, and the broader functioning of global production networks. This variable is particularly relevant for strategic commodity markets, where returns may respond to changes in bottlenecks in transportation, inventories, and the delivery of intermediate inputs. Geopolitical risk is incorporated to reflect the role of conflict, international tensions, strategic fragmentation, and broader geoeconomic uncertainty in shaping market expectations and asset-price dynamics. Given the increasing importance of geopolitical realignment in global commodity and energy markets, the joint consideration of SCS and GPR allows the analysis to distinguish between physical trade frictions and political-strategic instability as alternative forecasting channels.

In addition to these key predictors, the study includes a set of conventional market controls commonly used in forecasting and asset-pricing applications. These comprise the CBOE Volatility Index (VIX), which proxies global financial market uncertainty and risk aversion; the Economic Policy Uncertainty Index (EPU), which captures uncertainty associated with economic policy conditions; the S&P 500 Index (SPX), which reflects broad equity-market performance and overall financial conditions; the S&P GSCI Crude Oil Total Return Index (OIL), which captures developments in energy market conditions; and the US Dollar Index (USD), which controls for exchange-rate effects that may influence internationally traded commodity and energy-related assets. Together, these variables provide a benchmark set of macro-financial predictors against which the incremental value of supply-chain stress and geopolitical risk can be assessed. Table 1 presents the variables used in the empirical analysis, their acronyms, and their role in the forecasting framework.

The dependent variable set consists of the daily log returns of the Dow Jones Commodity Aluminum, Copper, and Nickel indices, as well as the daily log return of the S&P Global Clean Energy Index. The explanatory variable set includes both disruption-related indicators and conventional macro-financial controls, all expressed as daily log changes or returns using the same transformation. The variable design is therefore intended to balance market-specific relevance with broader forecasting considerations. More specifically, the selected predictors allow the analysis to determine whether disruption-sensitive variables add forecasting value beyond standard financial conditions. The data were collected from multiple sources. Market-based indices, including the aluminum, copper, nickel, S&P 500, crude oil, and US dollar indices, were obtained primarily from Investing.com, while the S&P Global Clean Energy Index was sourced from S&P Global. The Economic Policy Uncertainty Index (EPU) was drawn from the official EPU database and captures uncertainty

related to economic policy conditions using a news-based methodology [28]. The daily version of the index was used and was synchronized with the financial market trading calendar used in the empirical analysis. The Geopolitical Risk Index (GPR) was obtained from the Caldara-Iacoviello database and measures adverse geopolitical events and associated risks using news-based information [29]. Supply-chain stress (SCS) is proxied by the Baltic Dry Index, a market-based shipping indicator that captures conditions in global dry-bulk freight markets, and is widely used as a proxy for stress in trade logistics and transport networks.

Table 1. Variable definitions, sources, and references.

Variable	Acronym	Operational Definition and Form	Role	Exact Source	Reference
Supply-chain stress proxy	SCS	Baltic Dry Index, transformed into daily log changes	Main disruption predictor	Investing.com/Baltic Exchange	[8,9]
CBOE Volatility Index	VIX	VIX index, transformed into daily log changes	Macro-financial control	Investing.com/CBOE	Standard market uncertainty proxy
Dow Jones Commodity Aluminum Index	ALU	Dow Jones Commodity Aluminum Index, transformed into daily log returns	Target asset	Investing.com/S&P Dow Jones Indices	Commodity return series
Dow Jones Commodity Copper Index	COP	Dow Jones Commodity Copper Index, transformed into daily log returns	Target asset	Investing.com/S&P Dow Jones Indices	Commodity return series
Dow Jones Commodity Nickel Index	NIC	Dow Jones Commodity Nickel Index, transformed into daily log returns	Target asset	Investing.com/S&P Dow Jones Indices	Commodity return series
Economic Policy Uncertainty Index	EPU	News-based economic policy uncertainty index, transformed into daily log changes	Macro-financial control	Economic Policy Uncertainty database	[28]
Geopolitical Risk Index	GPR	News-based geopolitical risk index, transformed into daily log changes	Main disruption predictor	Caldara-Iacoviello GPR database	[29]
S&P 500 Index	SPX	S&P 500 Index, transformed into daily log returns	Macro-financial control	Investing.com/S&P Global	Broad equity-market benchmark
S&P Global Clean Energy Index	CEN	S&P Global Clean Energy Index, transformed into daily log returns	Target asset	S&P Global	Clean energy equity benchmark
S&P GSCI Crude Oil Total Return Index	OIL	S&P GSCI Crude Oil Total Return Index, transformed into daily log returns	Macro-financial control	Investing.com/S&P Global	Energy market benchmark
US Dollar Index	USD	US Dollar Index, transformed into daily log changes	Macro-financial control	Investing.com	Exchange-rate benchmark

Note: All variables are transformed as $100 \times [\ln(X_t) - \ln(X_{t-1})]$. For readability, the original acronyms are preserved after transformation. Thus, ALU, COP, NIC, and CEN denote daily log returns of the target assets, while SCS, GPR, VIX, EPU, SPX, OIL, and USD denote the corresponding daily log changes or returns of the predictor variables.

The use of the Baltic Dry Index is particularly appropriate for the strategic metal markets because aluminum, copper, and nickel are closely linked to raw-material transport, dry-bulk shipping conditions, inventories, and industrial supply-chain bottlenecks.

However, its link to clean energy equities is more indirect. Clean energy indices consist of listed firms, and many clean energy components, such as solar panels, wind turbine components, inverters, and batteries, are transported through containerized or specialized logistics rather than dry-bulk carriers. For this reason, the BDI-based supply-chain stress proxy is interpreted as a broad global shipping and logistics indicator with a more direct economic connection to industrial metals and a weaker, indirect connection to clean energy returns. This distinction is considered explicitly when interpreting the asset-specific results.

Although container-shipping indicators such as the Freightos Baltic Index and the Drewry World Container Index may be more directly connected to the logistics of clean energy equipment, they are not used in the main analysis because a consistent daily series covering the full 2015–2026 sample and aligned with all market variables was not available in the dataset used for this study. Therefore, the supply-chain stress proxy should be interpreted as a broad indicator of dry-bulk shipping and logistics rather than a comprehensive measure of all supply-chain disruptions. More specifically, the BDI does not directly measure inventory pressures, port congestion, semiconductor shortages, container-shipping bottlenecks, or clean energy-specific logistics frictions. This limitation is particularly relevant for clean energy equities, whose supply chains involve containerized components, specialized equipment, semiconductors, inverters, batteries, and other manufactured inputs that are not fully captured by dry-bulk freight conditions.

A daily frequency is adopted for two reasons. First, the paper's objective is explicitly forecasting-oriented and therefore requires a sufficiently high-frequency setting to detect short-run predictive content in disruption-related variables. Second, supply-chain conditions, geopolitical tensions, and macro-financial variables can change rapidly during periods of market stress, implying that lower-frequency data may obscure the transmission of such shocks into strategic commodity and clean energy markets. The use of daily observations is therefore well-suited to the paper's goal of evaluating whether disruption-sensitive indicators improve out-of-sample forecasting performance for strategic commodity and clean energy returns. To construct the synchronized daily dataset, all series were aligned by calendar date, weekends and non-trading days were removed, and the final sample retained only common dates for which all variables were jointly available. Missing observations caused by holidays, market closures, or publication-calendar differences were handled through common-date matching before computing daily log returns or log changes. No future observations were used to fill missing values, thereby avoiding look-ahead bias.

Figure 1 illustrates the time-series evolution of the transformed target and predictor variables over the 2015–2026 sample period. Panel A reports the daily log returns of aluminum, copper, nickel, and clean energy, while Panel B reports the daily log changes or returns of the predictor variables. The figure highlights the high-frequency variation in the data and shows that volatility is not constant over time. Several episodes of intensified fluctuations are visible around major stress periods, including the COVID-19 shock, the Russia-Ukraine war, and the monetary-tightening period. This visual evidence supports the use of a return-based forecasting framework and motivates the inclusion of macro-financial, supply-chain, and geopolitical predictors.

Several patterns are evident from Figure 1. First, the target returns display clear volatility clustering, with stronger fluctuations during crisis and stress episodes. The commodity returns, especially aluminum and nickel, show pronounced spikes around the COVID-19 and Russia-Ukraine periods, suggesting that industrial metals were particularly sensitive to global disruption and transition-related uncertainty. Clean energy returns also exhibit heightened volatility around the COVID-19 period and the subsequent policy-tightening environment, consistent with the equity-like nature of clean energy indices.

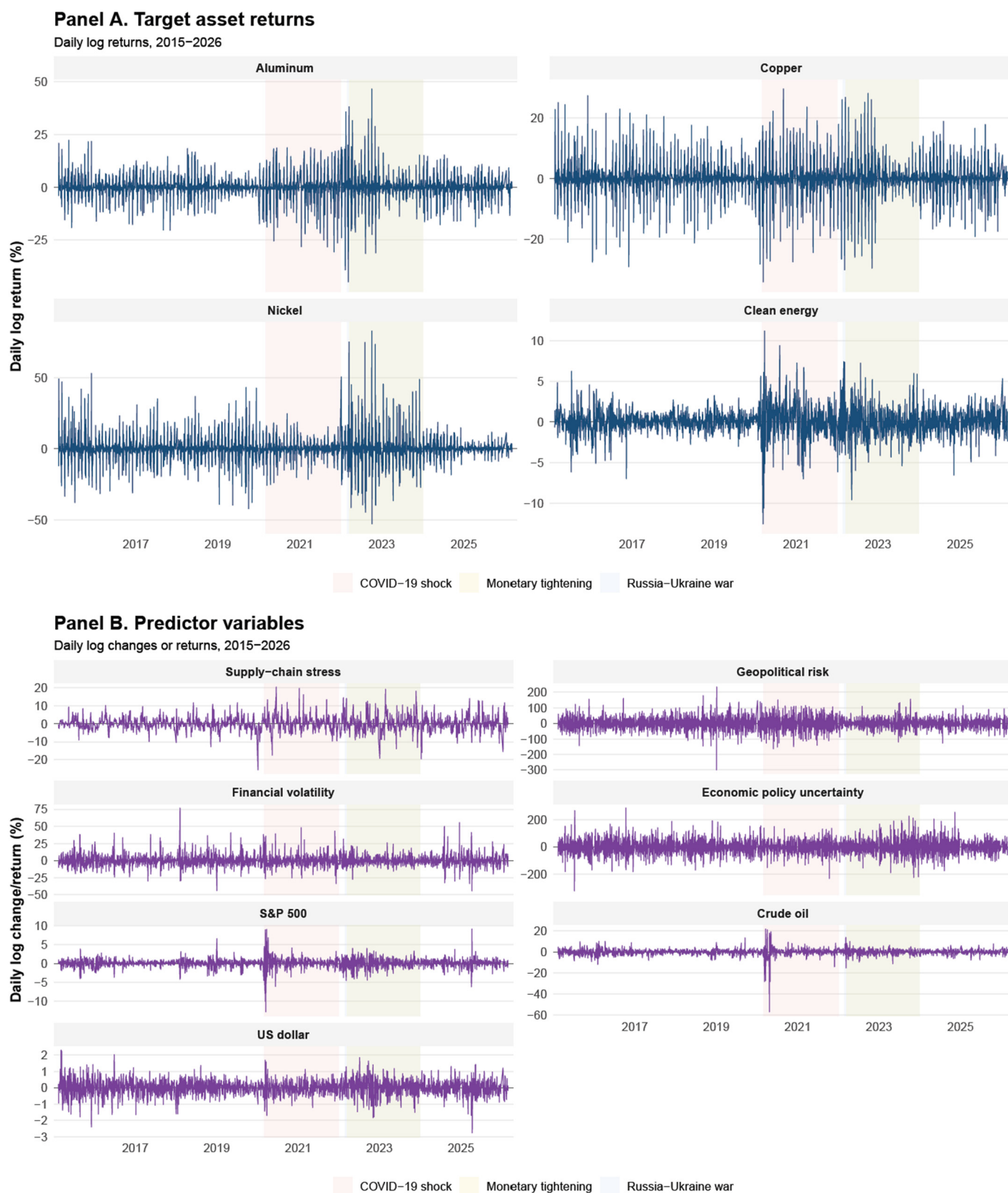


Figure 1. Time-series evolution of target returns and predictor variables. Note: **(Panel A)** reports the daily log returns of the target assets: aluminum, copper, nickel, and clean energy. **(Panel B)** reports the daily log changes or returns of the predictor variables, including supply-chain stress, geopolitical risk, financial volatility, economic policy uncertainty, the S&P 500, crude oil, and the US dollar.

Second, the predictor variables also display substantial time variation. Supply-chain stress becomes more volatile during the pandemic and logistics-stress period, while geopolitical risk records sharp movements around major geopolitical episodes. Financial volatility, economic policy uncertainty, crude oil prices, and the US dollar also exhibit episodic spikes, indicating that macro-financial conditions changed markedly over the sample period. These patterns reinforce the need for a forecasting framework that preserves the temporal ordering of the data and allows predictor importance to vary across market conditions.

Table 2 presents the descriptive statistics for the variables used in the forecasting models.

Table 2. Descriptive statistics.

Statistic	Observations	Minimum	Maximum	Median	Mean	SD
SCS	2353	−25.5780	20.3367	−0.0643	0.0568	3.7000
VIX	2353	−44.2449	76.8245	−0.7228	0.0033	8.7336
ALU	2353	−44.7602	46.5340	0.0106	0.0306	5.6636
COP	2353	−34.0988	29.5999	0.0328	0.0406	5.7298
NIC	2353	−52.8387	82.6623	−0.0403	0.0240	9.3230
EPU	2353	−325.0888	288.7114	0.0147	0.0677	58.1323
GPR	2353	−299.5883	234.4884	0.1407	0.0093	42.6816
SPX	2353	−12.7657	9.0895	0.0754	0.0513	1.2056
CEN	2353	−12.4971	11.1689	0.0490	0.0243	1.6868
OIL	2353	−56.8497	21.8263	0.1263	−0.0166	3.1521
USD	2353	−2.7588	2.3111	0.0000	0.0013	0.4717

Note: This table reports the descriptive statistics of the variables used in the empirical analysis over the sample period from 10 February 2015 to 27 February 2026. All variables are expressed in percentage daily log-return/log-change form, computed as $100 \times [\ln(X_t) - \ln(X_{t-1})]$. For each series, the table presents the number of observations, minimum, maximum, median, mean, and standard deviation. SCS denotes the supply-chain stress proxy based on the Baltic Dry Index; VIX is the CBOE Volatility Index; ALU, COP, and NIC represent the Dow Jones Commodity Aluminum, Copper, and Nickel indices; EPU denotes the Economic Policy Uncertainty Index; GPR is the Geopolitical Risk Index; SPX is the S&P 500 Index; CEN is the S&P Global Clean Energy Index; OIL is the S&P GSCI Crude Oil Total Return Index; and USD is the US Dollar Index.

Several features are noteworthy. First, the return-based descriptive statistics show that the transformed series are centered close to zero, as expected for daily log returns and log changes. The sample displays considerable variation in the disruption-related predictors, particularly in SCS, EPU, and GPR, indicating that the period under analysis includes substantial episodes of logistical and geopolitical stress. This is reflected in the relatively large standard deviations and wide minimum-maximum ranges of these variables. The commodity and clean energy indices also exhibit wide dispersion, especially those for nickel, copper, and aluminum, consistent with the high short-run volatility often observed in strategic commodity markets. Among the conventional controls, VIX and OIL display relatively high variation, while USD remains comparatively stable, reflecting the lower day-to-day volatility of broad exchange-rate movements.

From an economic standpoint, the selected variables are well aligned with the underlying research question. Strategic commodity returns may respond to disruptions in physical trade flows, logistics conditions, and industrial bottlenecks, making supply-chain stress a natural candidate predictor. At the same time, commodity and clean energy markets are increasingly embedded in a global environment shaped by sanctions, trade tensions, war-related uncertainty, and broader geopolitical risk, all of which justify the inclusion of the GPR index. The control variables further capture market-wide financial conditions,

energy price dynamics, policy uncertainty, and exchange-rate effects. Consequently, the final dataset provides a coherent empirical basis for testing whether supply-chain stress and geopolitical risk contain incremental forecasting information beyond conventional market predictors in explaining short-run movements in strategic commodity and clean energy returns.

3.2. Research Methodology

This study adopts a supervised machine-learning framework to evaluate whether supply-chain stress and geopolitical risk improve the prediction of strategic commodity and clean energy returns. The forecasting exercise is conducted separately for aluminum (ALU), copper (COP), nickel (NIC), and clean energy (CEN). For each asset, the dependent variable is the one-day-ahead daily log return, with the information set at time t used to predict the asset return at time $t + 1$. To assess the incremental value of disruption-related predictors, four model specifications are considered. The baseline specification includes only conventional market controls: VIX, EPU, SPX, OIL, and USD. The second specification augments the baseline model with supply-chain stress (SCS). The third specification adds geopolitical risk (GPR) to the baseline model. The fourth specification incorporates both SCS and GPR jointly. This structure allows us to distinguish between the predictive contributions of physical trade frictions, broader geopolitical instability, and their combined effects. The one-step-ahead forecasting model can be written as

$$r_{t+1}^{(m)} = f_m(X_t) + \varepsilon_{t+1} \quad (1)$$

where $r_{t+1}^{(m)}$ denotes the next-day log return of asset $m \in \{ALU, COP, NIC, CEN\}$, X_t is the vector of predictors observed at time t , $f_m(\cdot)$ is the forecasting function, and ε_{t+1} is the forecast error. Because the underlying relationship between disruption-related variables and asset returns may be nonlinear and asset-specific, the study compares several machine-learning algorithms with different functional forms and regularization structures.

3.2.1. Penalized Linear Models

Two penalized regression methods are employed: LASSO and Elastic Net. These models are useful in forecasting environments characterized by correlated predictors and potential overfitting. LASSO performs coefficient shrinkage and variable selection by minimizing:

$$\min_{\beta} \left\{ \frac{1}{N} \sum_{i=1}^N (y_i - X_i' \beta)^2 + \gamma \sum_{j=1}^p |\beta_j| \right\} \quad (2)$$

where y_i is the target variable, X_i is the predictor vector, β is the coefficient vector, and γ is the penalty parameter. By shrinking some coefficients toward zero, LASSO reduces model complexity and improves generalization. Elastic Net extends this idea by combining the L1 penalty of LASSO with the L2 penalty of Ridge regression:

$$\min_{\beta} \left\{ \frac{1}{N} \sum_{i=1}^N (y_i - X_i' \beta)^2 + \gamma \left[\alpha \sum_{j=1}^p |\beta_j| + \frac{1 - \alpha}{2} \sum_{j=1}^p \beta_j^2 \right] \right\} \quad (3)$$

where $\alpha \in [0, 1]$ governs the balance between variable selection and coefficient stabilization. Elastic Net is particularly attractive when explanatory variables are highly correlated, as is common in financial and commodity-market applications.

3.2.2. K-Nearest Neighbors Regression

The K-nearest neighbors (KNN) method is a nonparametric approach that predicts the value of a new observation by averaging the values of its K nearest neighbors in the predictor space. Let $N_k(x)$ denote the set of the k closest observations to the predictor vector x . The KNN forecast is given by

$$\hat{y}(x) = \frac{1}{k} \sum_{i \in N_k(x)} y_i \quad (4)$$

This method is useful when the association between predictors and the target variable is nonlinear but locally smooth. Since KNN is distance-based, the explanatory variables are standardized before estimation.

3.2.3. Gradient Boosting Methods

To capture nonlinearities, interactions, and complex threshold effects, the study also employs the gradient boosting machine (GBM) and XGBoost, two gradient-boosting algorithms widely used in forecasting applications. Both methods iteratively combine weak learners, typically shallow decision trees, to improve predictive performance. In general form, the boosting predictor can be expressed as

$$\hat{y}_i = \sum_{b=1}^B h_b(X_i) \quad (5)$$

where $h_b(\cdot)$ denotes the b -th regression tree and B is the number of boosting iterations. At each step, the algorithm updates the model to reduce a specified loss function, thereby improving the fit to the residual structure left unexplained by previous trees. GBM builds an additive ensemble of regression trees through sequential loss minimization, whereas XGBoost extends gradient boosting through regularization and scalable tree learning. Both are well suited to financial forecasting problems in which predictor effects may be nonlinear and highly interactive.

3.2.4. Extra Trees Regression

The main nonlinear benchmark in this study is the Extra Trees (ET) algorithm. ET is an ensemble tree method related to random forests, but it introduces additional randomness into the tree-building process by selecting split thresholds at random rather than exhaustively searching for the optimal threshold at each node. The final predictor is obtained by averaging across the ensemble of trees:

$$\hat{y}_i = \frac{1}{M} \sum_{m=1}^M T_m(X_i) \quad (6)$$

where $T_m(\cdot)$ denotes the m -th regression tree and M is the total number of trees. This structure reduces variance, limits overfitting, and allows the model to efficiently capture nonlinear relationships and higher-order interactions. As shown in the empirical results, ET provides a flexible nonlinear benchmark and is frequently competitive, but the full model comparison and MCS results indicate that its advantage is not uniform across all assets and loss functions.

3.2.5. Forecast Evaluation

The dataset is divided into training and test subsamples using a chronological 80:20 split, and model tuning is performed via expanding-window time-series cross-validation on the training subsample. This design preserves the temporal ordering of the data and ensures that future observations are not used to train or tune models that forecast earlier observations. To provide a comprehensive assessment of predictive accuracy,

four main forecast metrics are reported in the main tables: mean absolute error (MAE), root mean squared error (RMSE), coefficient of determination (R^2), and symmetric mean absolute percentage error (sMAPE). RMSE is retained instead of MSE in the main tables because RMSE is the square root of MSE and therefore produces the same model ranking while being easier to interpret in the units of the dependent variable. MSE is retained only in supplementary calculations and robustness outputs where needed. Lower values of MAE, RMSE, and sMAPE indicate better forecast performance, whereas higher values of R^2 indicate greater explanatory power. This multi-metric design is particularly useful in commodity and energy forecasting because different error criteria emphasize different aspects of predictive quality. For example, RMSE penalizes large forecast errors more heavily, while sMAPE provides a scale-adjusted error measure that is more suitable than MAPE when daily returns are close to zero.

To formally compare predictive accuracy across competing specifications, we apply the Diebold-Mariano test of equal forecast accuracy [30]. The test evaluates the null hypothesis that the two competing forecasts have equal predictive accuracy. In this study, Model 1 serves as the benchmark specification, while Models 2, 3, and 4 are compared against Model 1 for each asset and each forecasting algorithm. The test is implemented using squared forecast-error loss, so that the reported significance indicates whether the reduction in forecast errors from the augmented specifications is statistically meaningful rather than due to random variation.

3.2.6. Model Confidence Set

To complement the Diebold-Mariano tests, the study applies the Model Confidence Set (MCS) procedure of [31]. The MCS is used to compare the full set of 24 algorithm-specification combinations for each asset, consisting of six forecasting algorithms estimated under four model specifications. The procedure is implemented separately for each asset and for two loss functions: squared-error loss and absolute-error loss. The objective is to identify the subset of models that cannot be statistically distinguished from the best-performing model at the selected confidence level.

The MCS procedure starts from the full set of candidate models and sequentially eliminates inferior models until the null hypothesis of equal predictive ability among the remaining models can no longer be rejected. The analysis uses the range statistic based on pairwise loss differentials across competing models. Statistical inference is based on a block bootstrap procedure to preserve the time-series dependence in the out-of-sample loss differentials. The MCS is implemented at the 10% significance level, corresponding to a 90% confidence set, using 1000 bootstrap replications. Models retained in the final MCS are interpreted as statistically indistinguishable from the best-ranked model under the relevant loss function. This procedure is particularly useful in the present setting because daily return forecasts often differ only modestly across models, and numerical ranking alone may overstate differences in predictive performance.

3.2.7. Explainable Artificial Intelligence

To complement the forecasting results with an interpretable assessment of predictor importance, the study uses Shapley Additive exPlanations (SHAP). SHAP decomposes the model prediction into additive contributions from each explanatory variable and provides a game-theoretic measure of feature importance. For a given observation, the predicted value can be written as

$$g(x) = \varnothing_0 + \sum_{j=1}^p \varnothing_j x_j \quad (7)$$

where ϕ_0 is the baseline prediction and ϕ_j represents the contribution of predictor j . SHAP is particularly useful in the present context because it reveals not only which variables matter most for forecast performance, but also whether their effects differ across assets and model specifications. In this study, SHAP analysis is conducted for the full specification of the Extra Trees model because ET provides a flexible, nonlinear benchmark and is frequently retained among competitive models, while also enabling economically interpretable feature attribution. The resulting beeswarm and mean absolute SHAP plots enable identification of the dominant forecast drivers across aluminum, copper, nickel, and clean energy, thereby linking statistical performance to economically meaningful interpretation.

4. Empirical Results and Discussion

4.1. Preliminary Dependence Structure

Figure 2 presents the contemporaneous correlation structure among the variables included in the empirical analysis.

Several notable patterns emerge. First, the strategic commodity indices exhibit strong positive correlations with one another, especially aluminum and copper (0.76), while nickel is also positively associated with both aluminum and copper, with correlations of 0.54 in each case. These patterns suggest that industrial metal returns share an important common component, likely linked to global production conditions, industrial demand, and broader commodity-market sentiment. By contrast, their contemporaneous correlations with SPX are small, indicating that short-run movements in strategic commodity returns remain only weakly aligned with broad equity-market returns.

Second, the strongest cross-market relationship in the matrix is the negative correlation between VIX and SPX (−0.74), which is fully consistent with the conventional inverse link between market volatility and equity returns. VIX is also negatively correlated with clean energy returns (−0.41) and, to a lesser extent, with oil returns (−0.19), suggesting that increases in financial market stress tend to coincide with weaker performance in transition-sensitive and energy-related assets. Clean energy returns are positively associated with SPX (0.57) and OIL (0.22), highlighting the importance of broad financial conditions and energy market developments in this segment.

Third, supply-chain stress and geopolitical risk exhibit only weak contemporaneous correlations with most asset returns. SCS shows correlations close to zero with the financial and commodity return series, while GPR is similarly weakly correlated with the asset variables. This pattern suggests that the role of disruption-related variables in the forecasting framework may not operate primarily through strong contemporaneous linear associations. Instead, their predictive contribution may be more indirect, nonlinear, or conditional on specific market states, which supports the use of flexible machine-learning methods in the subsequent analysis. Finally, the US dollar index remains only weakly correlated with most variables, although it shows a modest negative correlation with clean energy returns (−0.20) and SPX (−0.09), consistent with the view that dollar appreciation can exert mild pressure on internationally traded assets and equity-sensitive market segments.

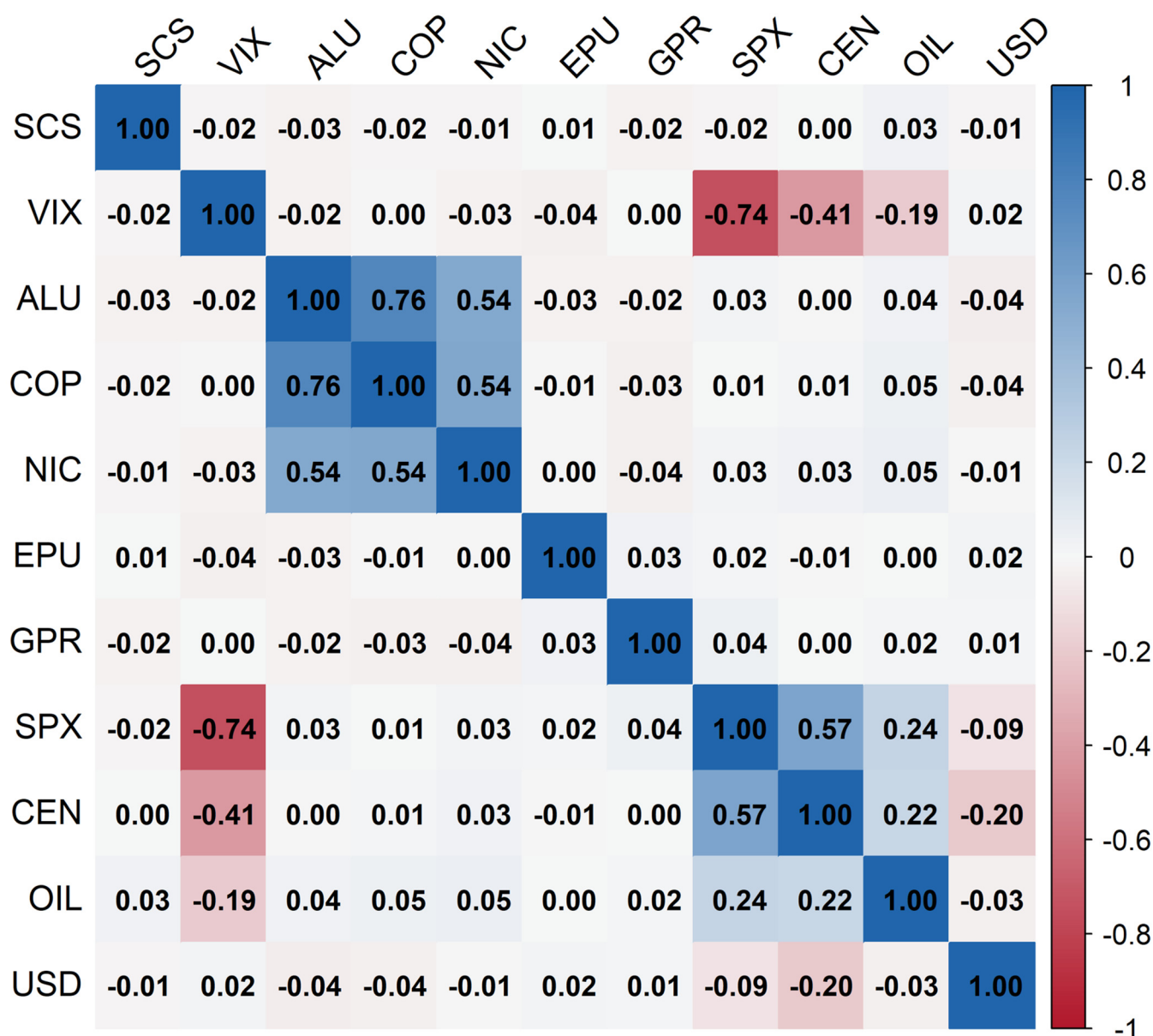


Figure 2. Correlation matrix. Note: The figure reports Pearson correlation coefficients among the variables included in the forecasting framework. All variables are expressed as daily log returns or log changes. Blue shades indicate positive correlations, red shades indicate negative correlations, and lighter colors indicate weaker associations. SCS denotes supply-chain stress; VIX is the CBOE Volatility Index; ALU, COP, and NIC are the Dow Jones Commodity Aluminum, Copper, and Nickel indices; EPU is the Economic Policy Uncertainty Index; GPR is the Geopolitical Risk Index; SPX is the S&P 500 Index; CEN is the S&P Global Clean Energy Index; OIL is the S&P GSCI Crude Oil Total Return Index; and USD is the US Dollar Index.

4.2. Forecasting Results for Aluminum Returns

Table 3 reports the out-of-sample forecasting performance for aluminum returns under the four model specifications.

Table 3. Forecasting results for aluminum returns under baseline and augmented model specifications.

Models	MAE_M1	RMSE_M1	R ² _M1	sMAPE_M1
ET	2.8521	4.7509	0.0004	1.5596
GBM	2.8425	4.7273	0.0003	1.5691
XGBoost	2.8118	4.6750	0.0140	1.6201
KNN	3.0864	4.8275	0.0001	1.5030
LASSO	2.7890	4.6930	0.0004	1.7399
EN	2.7893	4.6939	0.0004	1.7276
Models	MAE_M2	RMSE_M2	R ² _M2	sMAPE_M2
ET	2.8308	4.7137	0.0008	1.5780
GBM	2.8691	4.7409	0.0001	1.5817
XGBoost	2.7890	4.6397	0.0253	1.5507
KNN	3.0736	4.8025	0.0034	1.5319
LASSO	2.7890	4.6930	0.0004	1.7399
EN	2.7893	4.6939	0.0004	1.7276
Models	MAE_M3	RMSE_M3	R ² _M3	sMAPE_M3
ET	2.8146	4.6925	0.0034	1.5775
GBM	2.8189	4.7226	0.0011	1.6455
XGBoost	2.8146	4.6774	0.0121	1.6336
KNN	3.0829	4.7952	0.0021	1.5626
LASSO	2.7890	4.6930	0.0004	1.7399
EN	2.7893	4.6939	0.0004	1.7276
Models	MAE_M4	RMSE_M4	R ² _M4	sMAPE_M4
ET	2.8181	4.6909	0.0051	1.5581
GBM	2.7961	4.6959	0.0021	1.6381
XGBoost	2.8000	4.6470	0.0217	1.6055
KNN	3.0339	4.8029	0.0010	1.5172
LASSO	2.7890	4.6930	0.0004	1.7399
EN	2.7893	4.6939	0.0004	1.7276

Notes: This table reports out-of-sample forecasting results for aluminum returns across four model specifications. Model 1 includes the baseline controls: VIX, EPU, SPX, OIL, and USD. Model 2 augments the baseline model with supply-chain stress (SCS). Model 3 augments the baseline model with geopolitical risk (GPR). Model 4 includes both SCS and GPR in addition to the baseline controls. Forecasting performance is evaluated using mean absolute error (MAE), root mean squared error (RMSE), coefficient of determination (R²), and symmetric mean absolute percentage error (sMAPE). RMSE is retained instead of MSE to avoid redundant reporting, since RMSE is the square root of MSE and yields the same model ranking while being easier to interpret in the units of the dependent variable. Lower values of MAE, RMSE, and sMAPE indicate better forecast accuracy, whereas higher R² values indicate stronger predictive performance. Bold values indicate the best-performing model for each metric within each specification. All forecasts are based on daily log returns using a chronological train-test split and expanding-window cross-validation.

A clear pattern emerges across methods. For aluminum returns, forecast accuracy is generally similar across the competing models, consistent with the well-known difficulty of predicting daily financial and commodity returns. The Extra Trees (ET) model remains among the strongest performers across the four specifications, particularly in the augmented models, where it records competitive MAE, MSE, RMSE, and sMAPE values. This suggests that nonlinear and interaction effects are relevant for capturing short-run movements in aluminum returns, even though the overall degree of predictability remains modest. This result is broadly consistent with recent evidence showing that nonlinear and machine-learning methods tend to outperform conventional or penalized linear alternatives in

forecasting commodity and energy-related variables [11,19,20]. The specification-level results, however, are more nuanced. Under the ET model, Model 1 reports an MAE of 2.8521 and an RMSE of 4.7509. Adding supply-chain stress in Model 2 improves these values to 2.8308 and 4.7137, respectively. Model 3, which adds geopolitical risk, produces a further improvement, with MAE declining to 2.8146 and RMSE to 4.6925. Model 4, which includes both SCS and GPR, also improves over the baseline, with an MAE of 2.8181 and an RMSE of 4.6909. These results indicate that disruption-related predictors add incremental forecasting value for aluminum returns, although the magnitude of the improvement remains economically modest.

Economically, the results suggest that aluminum returns are driven not only by conventional macro-financial controls but also by short-run sensitivity to disruption-related information. The improvement associated with GPR is particularly relevant, implying that geopolitical risk may affect aluminum returns through uncertainty about global industrial demand, sanctions, trade restrictions, and strategic supply conditions. The contribution of SCS is also consistent with aluminum's exposure to shipping conditions, inventory levels, and physical supply chain bottlenecks. At the same time, the relatively low R^2 values across all specifications confirm that daily aluminum returns remain difficult to forecast and that the predictive contribution of SCS and GPR should be interpreted as incremental rather than dominant. This interpretation is broadly in line with [16], who show that aluminum and copper can display resilience to geopolitical shocks, and it is also compatible with [15], who argue that not all transition-related commodities respond to geopolitical forces with the same intensity. At the same time, the improvement observed after adding SCS and GPR is consistent with [9], who document stronger predictive effects of supply-chain pressure for metals in broader return-based settings, suggesting that the forecasting relevance of disruption variables may depend on the specific asset, prediction target, and model environment.

4.3. Forecasting Results for Copper Returns

Table 4 presents the forecasting results for copper returns.

As in the aluminum case, the copper return results confirm that nonlinear models remain useful, although performance differences across algorithms are generally modest. ET performs competitively across all four specifications, but the lowest error values are not concentrated in one model across all metrics. For example, XGBoost records the lowest RMSE in Model 4, while ET shows a relatively strong improvement in Model 3. This pattern indicates that copper return forecasting is a challenging setting in which different algorithms capture different aspects of short-run return dynamics. The relevance of nonlinear models is consistent with [11], who find that Extra Trees delivers strong predictive performance in clean energy forecasting, and more broadly with [27], who show that tree-based machine-learning methods are highly effective in modeling industrial metal risk dynamics.

Compared with aluminum, copper shows a more selective response to disruption-related predictors. Under the ET model, the baseline specification reports an MAE of 2.9076 and an RMSE of 4.9027. Adding supply-chain stress in Model 2 slightly improves MAE to 2.8795 but leaves RMSE almost unchanged at 4.9048. Model 3, which adds geopolitical risk, provides the clearest improvement for ET, reducing MSE from 24.0364 to 23.6400, RMSE from 4.9027 to 4.8621, and increasing R^2 from 0.0008 to 0.0089. By contrast, Model 4 does not dominate the ET results, although it still improves MSE relative to the baseline. This suggests that, for copper returns, geopolitical risk carries more incremental predictive information than supply-chain stress alone or the joint specification.

Table 4. Forecasting results for copper returns under baseline and augmented model specifications.

Models	MAE_M1	RMSE_M1	R ² _M1	sMAPE_M1
ET	2.9076	4.9027	0.0008	1.5903
GBM	2.8689	4.9108	0.0000	1.6249
XGBoost	2.8217	4.8758	0.0084	1.6482
KNN	3.1655	5.0166	0.0007	1.5162
LASSO	2.8242	4.8705	0.0026	1.8363
EN	2.8236	4.8701	0.0026	1.8270
Models	MAE_M2	RMSE_M2	R ² _M2	sMAPE_M2
ET	2.8795	4.9048	0.0010	1.5165
GBM	2.8333	4.8804	0.0018	1.6790
XGBoost	2.8238	4.8642	0.0111	1.6421
KNN	3.1582	5.0126	0.0008	1.5543
LASSO	2.8242	4.8705	0.0026	1.8363
EN	2.8236	4.8701	0.0026	1.8270
Models	MAE_M3	RMSE_M3	R ² _M3	sMAPE_M3
ET	2.8967	4.8621	0.0089	1.5868
GBM	2.8452	4.8933	0.0004	1.6931
XGBoost	2.8238	4.8848	0.0060	1.6487
KNN	3.2111	5.0285	0.0004	1.5874
LASSO	2.8242	4.8705	0.0026	1.8363
EN	2.8236	4.8701	0.0026	1.8270
Models	MAE_M4	RMSE_M4	R ² _M4	sMAPE_M4
ET	2.9052	4.8974	0.0013	1.5889
GBM	2.8519	4.8821	0.0021	1.6364
XGBoost	2.8209	4.8609	0.0111	1.6225
KNN	3.1916	5.0041	0.0012	1.5799
LASSO	2.8242	4.8705	0.0026	1.8363
EN	2.8236	4.8701	0.0026	1.8270

Notes: This table reports out-of-sample forecasting results for copper returns across four model specifications. Model 1 includes the baseline controls: VIX, EPU, SPX, OIL, and USD. Model 2 augments the baseline model with supply-chain stress (SCS). Model 3 augments the baseline model with geopolitical risk (GPR). Model 4 includes both SCS and GPR in addition to the baseline controls. Forecasting performance is evaluated using mean absolute error (MAE), root mean squared error (RMSE), coefficient of determination (R²), and symmetric mean absolute percentage error (sMAPE). RMSE is retained instead of MSE to avoid redundant reporting, since RMSE is the square root of MSE and yields the same model ranking while being easier to interpret in the units of the dependent variable. Lower MAE, RMSE, and sMAPE values indicate better forecast accuracy, whereas higher R² values indicate stronger predictive performance. Bold values indicate the best-performing model for each metric within each specification. All forecasts are based on daily log returns using a chronological train-test split and expanding-window cross-validation.

This result is economically intuitive. Copper is a core strategic metal for industrial production, electrification, and infrastructure investment, making it sensitive to global investment cycles, trade relations, geopolitical uncertainty, and expectations about future industrial demand. The stronger role of GPR suggests that copper returns may respond more to geopolitical uncertainty and strategic realignment than to daily changes in shipping-related supply-chain stress. Supply-chain stress still matters conceptually because copper is embedded in global mining, refining, and transportation networks, but the forecasting results indicate that its incremental contribution is weaker in the one-day-ahead return setting. The low R² values also indicate that daily copper returns remain difficult to predict, so the improvement from disruption-related variables should be interpreted as incremental rather than dominant.

This finding accords well with [9], who show that supply-chain pressure matters for metals, and with [6,12], who emphasize that geopolitical conditions can alter energy- and transition-related market behavior. It is also consistent with [15], who identify meaningful

links between geopolitical dynamics and transition commodities, and with [17], who show that geopolitical risk can quickly reshape market spillovers. In relation to these studies, our contribution is to show that copper return forecasting benefits most clearly from the geopolitical-risk channel, while the joint inclusion of SCS and GPR does not automatically yield the strongest predictive performance. This reinforces the broader conclusion that disruption-related predictors are asset-specific and should be evaluated separately, rather than assumed to uniformly improve forecasts across all strategic commodities.

4.4. Forecasting Results for Nickel Returns

Table 5 reports the corresponding results for nickel returns.

Table 5. Forecasting results for nickel returns under baseline and augmented model specifications.

Models	MAE_M1	RMSE_M1	R ² _M1	sMAPE_M1
ET	2.9553	5.0973	0.0001	1.5560
GBM	2.9185	5.0444	0.0013	1.5575
XGBoost	3.0559	5.1823	0.0011	1.5166
KNN	3.5477	5.4995	0.0011	1.5158
LASSO	2.7429	4.9654	0.0036	1.6977
EN	2.7399	4.9654	0.0036	1.7126
Models	MAE_M2	RMSE_M2	R ² _M2	sMAPE_M2
ET	2.9344	5.0787	0.0007	1.5772
GBM	2.8532	4.9878	0.0066	1.5728
XGBoost	3.0941	5.1542	0.0036	1.5017
KNN	3.6500	5.5927	0.0019	1.5475
LASSO	2.7389	4.9656	0.0035	1.7297
EN	2.7362	4.9660	0.0035	1.7486
Models	MAE_M3	RMSE_M3	R ² _M3	sMAPE_M3
ET	2.9217	5.0656	0.0000	1.5390
GBM	2.8477	4.9925	0.0045	1.5824
XGBoost	2.9621	5.1593	0.0004	1.5429
KNN	3.5724	5.5315	0.0004	1.4917
LASSO	2.7436	4.9640	0.0041	1.6988
EN	2.7405	4.9642	0.0041	1.7145
Models	MAE_M4	RMSE_M4	R ² _M4	sMAPE_M4
ET	2.8608	5.0229	0.0006	1.5639
GBM	2.8912	4.9976	0.0053	1.5444
XGBoost	2.9601	5.1451	0.0009	1.5398
KNN	3.5595	5.5051	0.0001	1.5204
LASSO	2.7390	4.9655	0.0036	1.7299
EN	2.7362	4.9660	0.0035	1.7486

Notes: This table reports out-of-sample forecasting results for nickel returns across four model specifications. Model 1 includes the baseline controls: VIX, EPU, SPX, OIL, and USD. Model 2 augments the baseline model with supply-chain stress (SCS). Model 3 augments the baseline model with geopolitical risk (GPR). Model 4 includes both SCS and GPR in addition to the baseline controls. Forecasting performance is evaluated using mean absolute error (MAE), root mean squared error (RMSE), coefficient of determination (R²), and symmetric mean absolute percentage error (sMAPE). RMSE is retained instead of MSE to avoid redundant reporting, since RMSE is the square root of MSE and yields the same model ranking while being easier to interpret in the units of the dependent variable. Lower MAE, RMSE, and sMAPE values indicate better forecast accuracy, whereas higher R² values indicate stronger predictive performance. Bold values indicate the best-performing model for each metric within each specification. All forecasts are based on daily log returns using a chronological train-test split and expanding-window cross-validation.

Table 5 shows that nickel return forecasting is more mixed across algorithms than the earlier commodity cases, but nonlinear and ensemble-based methods still provide useful evidence on the role of disruption-related predictors. ET performs competitively

across the four specifications, although LASSO and Elastic Net record slightly lower error values in several specifications, while GBM performs strongly when supply-chain stress is included. This indicates that nickel return forecasting exhibits weak, noisy predictability, with different algorithms capturing distinct aspects of the return process. At the same time, the relatively strong performance of LASSO and Elastic Net in selected nickel specifications may also indicate that these regularized linear models appropriately shrink disruption-related predictors when their incremental linear forecasting contribution is weak. In this sense, their performance should not be interpreted only as evidence against nonlinear modeling, but also as evidence that the predictive contribution of SCS and GPR is selective and depends on the asset, model structure, and loss function considered. Nevertheless, the ET results remain particularly informative because they provide the basis for the SHAP analysis and allow a consistent comparison of the four model specifications. This finding remains consistent with the broader machine-learning forecasting literature, especially the work of [19,20,27], all of whom show that flexible nonlinear models are particularly well suited to complex commodity-market dynamics.

The more important result concerns the relative performance of the specifications. For nickel returns, the strongest ET-based improvement is observed in Model 4, which includes both supply-chain stress and geopolitical risk. Under ET, the baseline model reports an MAE of 2.9553 and an RMSE of 5.0973. Adding supply-chain stress in Model 2 improves these values to 2.9344 and 5.0787, respectively. Model 3, which adds geopolitical risk alone, produces a weaker improvement, with an MAE of 2.9469 and an RMSE of 5.0605. Model 4 provides the clearest gain, reducing MAE to 2.8607 and RMSE to 5.0234, while increasing R^2 to 0.0051. This pattern suggests that nickel returns are especially sensitive to the combined effect of physical supply-chain frictions and geopolitical risk rather than to either channel in isolation.

This result is economically intuitive and reflects the structural characteristics of the nickel market. Compared with aluminum and copper, nickel is more closely connected to battery supply chains and transition-related demand, while also remaining important for stainless steel production. Its supply and processing chains are geographically concentrated and exposed to country-specific risks in production, refining, export, and policy. This creates a setting in which physical supply-chain stress and geopolitical risk are likely to interact. Supply-chain stress captures disruptions in shipping, logistics, delivery conditions, and global trade frictions, while geopolitical risk captures uncertainty related to resource access, export restrictions, sanctions, strategic competition, and security-of-supply concerns. Because nickel is both an industrial metal and a battery-related transition material, the joint inclusion of SCS and GPR can capture complementary information that is less visible when each predictor is used separately. In other words, for nickel, supply-chain stress and geopolitical risk appear to act as complementary rather than competing forecasting signals. The low R^2 values again indicate that daily return predictability is limited, but the error reductions under Model 4 show that disruption-related variables contain incremental information for nickel.

This result is strongly consistent with [8], who highlight the role of shipping risk as a transmission channel, and with [9], who show that supply-chain pressure has predictive power in metals markets. It is also compatible with [13,14], both of whom stress the broader economic relevance of supply-chain bottlenecks. Compared with the more mixed geopolitical findings in [15,16], the nickel results suggest that geopolitical risk becomes more informative when combined with direct supply-chain stress, rather than when introduced as a standalone predictor.

4.5. Forecasting Results for Clean Energy Returns

Table 6 presents the out-of-sample forecasting performance for clean energy returns.

Table 6. Forecasting results for clean energy returns under baseline and augmented model specifications.

Models	MAE_M1	RMSE_M1	R ² _M1	sMAPE_M1
ET	1.1097	1.4537	0.0024	1.6264
GBM	1.1183	1.4658	0.0002	1.7295
XGBoost	1.1137	1.4561	0.0004	1.6735
KNN	1.1360	1.4875	0.0012	1.5319
LASSO	1.0946	1.4338	0.0157	1.7176
EN	1.0953	1.4340	0.0159	1.7308
Models	MAE_M2	RMSE_M2	R ² _M2	sMAPE_M2
ET	1.1112	1.4518	0.0017	1.6501
GBM	1.1124	1.4514	0.0011	1.7247
XGBoost	1.1218	1.4665	0.0004	1.6687
KNN	1.1432	1.4920	0.0001	1.5393
LASSO	1.0994	1.4376	0.0150	1.8049
EN	1.0996	1.4378	0.0149	1.8074
Models	MAE_M3	RMSE_M3	R ² _M3	sMAPE_M3
ET	1.1041	1.4488	0.0045	1.6259
GBM	1.1191	1.4659	0.0001	1.6727
XGBoost	1.1167	1.4606	0.0000	1.6877
KNN	1.1295	1.4813	0.0022	1.5442
LASSO	1.0953	1.4346	0.0141	1.7196
EN	1.0961	1.4350	0.0137	1.7322
Models	MAE_M4	RMSE_M4	R ² _M4	sMAPE_M4
ET	1.1065	1.4465	0.0034	1.6665
GBM	1.1257	1.4711	0.0023	1.7307
XGBoost	1.1235	1.4681	0.0010	1.7060
KNN	1.1352	1.4796	0.0017	1.5516
LASSO	1.0994	1.4376	0.0150	1.8049
EN	1.0996	1.4378	0.0149	1.8074

Notes: This table reports out-of-sample forecasting results for clean energy returns across four model specifications. Model 1 includes the baseline controls: VIX, EPU, SPX, OIL, and USD. Model 2 augments the baseline model with supply-chain stress (SCS). Model 3 augments the baseline model with geopolitical risk (GPR). Model 4 includes both SCS and GPR in addition to the baseline controls. Forecasting performance is evaluated using mean absolute error (MAE), root mean squared error (RMSE), coefficient of determination (R²), and symmetric mean absolute percentage error (sMAPE). RMSE is retained instead of MSE to avoid redundant reporting, since RMSE is the square root of MSE and yields the same model ranking while being easier to interpret in the units of the dependent variable. Lower MAE, RMSE, and sMAPE values indicate better forecast accuracy, whereas higher R² values indicate stronger predictive performance. Bold values indicate the best-performing model for each metric within each specification. All forecasts are based on daily log returns using a chronological train-test split and expanding-window cross-validation.

Table 6 shows that clean energy return forecasting exhibits modest predictability and relatively small differences across model specifications. ET remains competitive across the four specifications, but the strongest error values are often obtained by LASSO and Elastic Net, particularly for MAE, MSE, and RMSE. This suggests that, unlike the commodity return series, clean energy returns contain a stronger linear or regularized component that can be captured by penalized models. This also suggests that LASSO and Elastic Net may be appropriately shrinking disruption-related predictors when their incremental linear forecasting contribution is limited. Nevertheless, ET remains important because it provides a flexible nonlinear benchmark and forms the basis of the SHAP explainability analysis. This result partly mirrors [11], who identify Extra Trees as a strong forecasting model for

clean energy returns, and it is consistent with the broader evidence in [4] that nonlinear predictors and model structures can improve clean energy forecasting performance.

The relative performance of the model specifications differs from that of copper and nickel. Under ET, the baseline specification reports an MAE of 1.1097 and an RMSE of 1.4537. Adding supply-chain stress in Model 2 produces a marginally lower RMSE of 1.4518, but MAE increases slightly to 1.1112. Model 3, which adds geopolitical risk, provides the best ET-based improvement, reducing MAE to 1.1041, MSE to 2.0991, and RMSE to 1.4488, while increasing R^2 to 0.0045. Model 4 further reduces RMSE to 1.4465, but it does not improve MAE relative to Model 3 and raises sMAPE. These results suggest that the incremental contribution of disruption-related variables to clean energy return forecasting is limited and metric-dependent.

Economically, the findings indicate that clean energy returns are driven primarily by broader financial market conditions rather than by the direct inclusion of disruption-related variables. More specifically, clean energy equity returns appear to be more sensitive to changes in discount rates, proxied by VIX and USD, and to expectations about future cash flows, proxied by SPX, OIL, and EPU, than to dry-bulk freight stress. This interpretation is consistent with the return-correlation evidence, where clean energy is positively correlated with SPX and OIL and negatively correlated with VIX and USD. Supply-chain stress retains some relevance, but its contribution remains weaker than that observed for selected industrial metals. Geopolitical risk provides some incremental information in the ET specification, but the improvement is small and does not indicate that GPR dominates conventional macro-financial predictors. Thus, in the clean energy market, conventional macro-financial and market-wide predictors appear to play a larger role than disruption-specific indicators.

This result should also be interpreted in light of the BDI-based supply-chain proxy's measurement scope. The Baltic Dry Index captures dry-bulk freight conditions and is therefore more directly related to raw materials and industrial commodity flows than to the logistics of clean energy equities. Clean energy firms are exposed to supply-chain conditions through a longer, more indirect transmission chain that involves manufactured components, container shipping, specialized equipment, semiconductor inputs, inverters, batteries, financing costs, and policy expectations. As a result, a dry-bulk shipping proxy may fail to capture the specific supply-chain frictions most relevant for clean energy firms. The weak contribution of SCS in clean energy forecasting should therefore not be interpreted as evidence that supply-chain disruptions are irrelevant for clean energy firms. Rather, it suggests that the BDI may be an incomplete proxy for the types of supply-chain frictions most relevant to clean energy equities. A container-shipping indicator, port-congestion measure, semiconductor supply-chain proxy, or battery-material logistics indicator could plausibly capture clean energy-specific bottlenecks more directly than the BDI. However, because a consistent daily series covering the full sample was not available in the present dataset, this possibility is left for future research. This helps explain why SCS contributes little to clean energy return forecasts, whereas macro-financial predictors such as USD, EPU, VIX, SPX, and OIL remain more influential.

This result partially aligns with [11], who also find that broad financial conditions dominate clean energy forecasting, and with [12], who show that clean energy markets are strongly influenced by evolving market inefficiencies and broader energy market dynamics. However, our evidence differs from that of [4], in which transition-related risk indicators have stronger predictive power, suggesting that the relevance of external risk measures may depend on whether the forecasting target is market returns, volatility, or transition-risk exposure, as well as on the institutional and geographical setting of the market under study.

4.6. Incremental Forecasting Value of Disruption-Related Predictors

Table 7 reports the percentage improvement of Models 2–4 over Model 1 for aluminum returns.

Table 7. Percentage improvement of augmented models over the baseline model in forecasting aluminum returns.

Models	Comparison	MAE	MSE	RMSE	R ²	ΔR^2	sMAPE
ET	Model 2 vs. Model 1	−0.75 *	−1.56 *	−0.78 *	115.81 *	0.0004	1.18
GBM	Model 2 vs. Model 1	0.94	0.58	0.29	−70.51	−0.0002	0.81
XGBoost	Model 2 vs. Model 1	−0.81 **	−1.50 **	−0.75 **	80.14 **	0.0113	−4.28 **
KNN	Model 2 vs. Model 1	−0.41	−1.03	−0.52	2274.93	0.0033	1.92
LASSO	Model 2 vs. Model 1	0.00	0.00	0.00	0.00	0.0000	0.00
EN	Model 2 vs. Model 1	0.00	0.00	0.00	0.00	0.0000	0.00
ET	Model 3 vs. Model 1	−1.32 ***	−2.44 ***	−1.23 ***	790.24 ***	0.0030	1.15
GBM	Model 3 vs. Model 1	−0.83	−0.20	−0.10	283.77	0.0008	4.87
XGBoost	Model 3 vs. Model 1	0.10	0.10	0.05	−13.67	−0.0019	0.83
KNN	Model 3 vs. Model 1	−0.11	−1.33	−0.67	1383.14	0.0020	3.96
LASSO	Model 3 vs. Model 1	0.00	0.00	0.00	0.00	0.0000	0.00
EN	Model 3 vs. Model 1	0.00	0.00	0.00	0.00	0.0000	0.00
ET	Model 4 vs. Model 1	−1.19 **	−2.51 **	−1.26 **	1228.35 **	0.0047	−0.10 **
GBM	Model 4 vs. Model 1	−1.63	−1.32	−0.66	640.27	0.0018	4.40
XGBoost	Model 4 vs. Model 1	−0.42	−1.19	−0.60	54.50	0.0077	−0.90
KNN	Model 4 vs. Model 1	−1.70	−1.02	−0.51	589.02	0.0009	0.94
LASSO	Model 4 vs. Model 1	0.00	0.00	0.00	0.00	0.0000	0.00
EN	Model 4 vs. Model 1	0.00	0.00	0.00	0.00	0.0000	0.00

Note: This table reports the percentage changes in out-of-sample forecasting performance for aluminum returns relative to Model 1. Model 2 augments the baseline controls with supply-chain stress (SCS), Model 3 augments the baseline controls with geopolitical risk (GPR), and Model 4 includes both SCS and GPR in addition to the baseline controls. For MAE, MSE, RMSE, and sMAPE, negative values indicate improved forecast accuracy relative to Model 1, whereas positive values indicate weaker performance. For R², positive values indicate improved predictive performance. Percentage changes are computed relative to Model 1. Negative values for MAE, MSE, RMSE, and sMAPE indicate reductions in forecast error, whereas positive values for R² indicate improved explanatory performance. Because baseline R² values are often close to zero in daily return forecasting, the table also reports ΔR^2 , defined as the absolute difference between the augmented specification's R² and Model 1's R². This facilitates interpretation and avoids overstating economically small changes in predictive performance. Statistical significance is assessed using Diebold–Mariano tests based on squared forecast-error loss. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

The percentage-improvement results in Table 7 provide clearer evidence on the incremental forecasting value of disruption-related predictors for aluminum returns. Unlike the raw forecasting metrics in Table 3, Table 7 directly compares each augmented specification with the baseline model and reports whether the observed differences are statistically significant. Within the ET specifications, the results show that adding disruption-related variables improves forecasts in all three augmented specifications. Under ET, Model 2 reduces MAE by 0.75%, MSE by 1.56%, and RMSE by 0.78% relative to the baseline, with the improvement statistically significant at the 10% level. Model 3 provides the greatest improvement, reducing MAE by 1.32%, MSE by 2.44%, and RMSE by 1.23%, with significance at the 1% level. Model 4 also improves forecast accuracy, lowering MAE by 1.19%, MSE by 2.51%, and RMSE by 1.26%, with significance at the 5% level. These results suggest that aluminum returns contain incremental predictive information from both supply-chain stress and geopolitical risk, although the magnitude of the improvements is economically modest. The strongest evidence is associated with Model 3, indicating that geopolitical risk provides particularly relevant information for forecasting aluminum returns. Model 4 also performs better than the baseline, but the improvement is not clearly greater than that of Model 3, suggesting that adding SCS to GPR does not yield a substantial additional gain for ET. This

pattern implies that disruption-related predictors are useful for aluminum returns, but their contribution is selective rather than uniformly additive.

Across the other algorithms, the results are more heterogeneous. XGBoost benefits from adding SCS in Model 2, with statistically significant reductions in MAE, MSE, and RMSE, whereas GBM improves in Model 4 primarily under absolute-error loss rather than squared-error loss. KNN shows some reductions in error metrics, but these gains are not statistically significant. LASSO and Elastic Net remain largely unchanged, indicating that the augmented variables do not materially alter the performance of the penalized linear specifications. The evidence from Table 7 suggests that disruption-related variables improve aluminum return forecasting primarily when used in nonlinear models that capture interaction effects and threshold-type relationships.

Economically, the findings indicate that aluminum returns are not fully explained by conventional macro-financial controls. Geopolitical risk appears to contain forward-looking information about strategic supply conditions, trade restrictions, sanctions risk, and uncertainty surrounding global industrial demand. Supply-chain stress also contributes, but its role is weaker and less persistent across the augmented specifications. In this respect, the evidence is consistent with [16], who report that some commodities remain relatively resilient to geopolitical shocks, and with [9], who document predictive gains from supply-chain pressure in metals markets. The aluminum results therefore support a balanced interpretation: disruption-related predictors improve forecasts, but their role is incremental and asset-specific rather than dominant.

Table 8 provides much stronger evidence of incremental predictive value in the copper market.

The percentage-improvement results for copper returns show that the gains from disruption-related predictors are modest and method-dependent. For the ET model, Model 3 provides the clearest improvement over the baseline, reducing MAE by 0.37%, MSE by 1.65%, RMSE by 0.83%, and sMAPE by 0.22%, with the Diebold-Mariano test indicating statistical significance at the 10% level. Model 2 improves MAE and sMAPE for ET but slightly worsens MSE and RMSE, whereas Model 4 yields only small changes relative to the baseline. This suggests that, for ET, geopolitical risk contains more useful incremental information for copper return forecasting than supply-chain stress alone or the joint specification. Across the other algorithms, the results are mixed. GBM benefits from adding SCS in Model 2, with statistically significant reductions in MAE, MSE, and RMSE, whereas the joint specification improves some error metrics but does not produce a statistically significant gain. XGBoost and KNN show limited and inconsistent changes across specifications, while LASSO and Elastic Net remain unchanged. These findings indicate that the forecasting value of disruption-related predictors for copper returns depends on both the model class and the specific disruption channel considered.

Economically, the results suggest that copper returns are exposed to geopolitical risk, but the evidence does not support a uniform conclusion that combining physical and geopolitical disruption channels always produces the strongest forecast gains. Copper remains a strategically important metal for electrification, infrastructure, and industrial production, but its one-day-ahead return dynamics appear to respond more clearly to geopolitical uncertainty than to daily changes in supply-chain stress. This result resonates with [9], who show that supply-chain pressure matters for metals, while also complementing [6,12], who show that geopolitical conditions predict energy-related market dynamics. In contrast to a one-size-fits-all disruption channel, the copper evidence suggests that geopolitical risk provides the more informative short-horizon signal, whereas the joint SCS-GPR specification does not systematically dominate.

Table 8. Percentage improvement of augmented models over the baseline model in forecasting copper returns.

Models	Comparison	MAE	MSE	RMSE	R ²	ΔR^2	sMAPE
ET	Model 2 vs. Model 1	−0.96	0.09	0.04	35.57	0.0002	−4.64
GBM	Model 2 vs. Model 1	−1.24 *	−1.23 *	−0.62 *	44,519.18 *	0.0018	3.34
XGBoost	Model 2 vs. Model 1	0.07	−0.48	−0.24	31.91	0.0027	−0.37
KNN	Model 2 vs. Model 1	−0.23	−0.16	−0.08	7.82	0.0001	2.51
LASSO	Model 2 vs. Model 1	0.00	0.00	0.00	0.00	0.0000	0.00
EN	Model 2 vs. Model 1	0.00	0.00	0.00	0.00	0.0000	0.00
ET	Model 3 vs. Model 1	−0.37 *	−1.65 *	−0.83 *	1085.34 *	0.0081	−0.22 *
GBM	Model 3 vs. Model 1	−0.83	−0.71	−0.36	10,097.85	0.0004	4.20
XGBoost	Model 3 vs. Model 1	0.07	0.37	0.18	−28.18	−0.0024	0.03
KNN	Model 3 vs. Model 1	1.44	0.48	0.24	−48.26	−0.0003	4.70
LASSO	Model 3 vs. Model 1	0.00	0.00	0.00	0.00	0.0000	0.00
EN	Model 3 vs. Model 1	0.00	0.00	0.00	0.00	0.0000	0.00
ET	Model 4 vs. Model 1	−0.08	−0.22	−0.11	66.13	0.0005	−0.09
GBM	Model 4 vs. Model 1	−0.59	−1.16	−0.58	52,469.16	0.0021	0.71
XGBoost	Model 4 vs. Model 1	−0.03	−0.61	−0.31	32.01	0.0027	−1.56
KNN	Model 4 vs. Model 1	0.82	−0.50	−0.25	60.05	0.0005	4.20
LASSO	Model 4 vs. Model 1	0.00	0.00	0.00	0.00	0.0000	0.00
EN	Model 4 vs. Model 1	0.00	0.00	0.00	0.00	0.0000	0.00

Note: This table reports the percentage changes in out-of-sample forecasting performance for copper returns relative to Model 1. Model 2 augments the baseline controls with supply-chain stress (SCS), Model 3 augments the baseline controls with geopolitical risk (GPR), and Model 4 includes both SCS and GPR in addition to the baseline controls. For MAE, MSE, RMSE, and sMAPE, negative values indicate improved forecast accuracy relative to Model 1, whereas positive values indicate weaker performance. For R², positive values indicate improved predictive performance. Percentage changes are computed relative to Model 1. Negative values for MAE, MSE, RMSE, and sMAPE indicate reductions in forecast error, whereas positive values for R² indicate improved explanatory performance. Because baseline R² values are often close to zero in daily return forecasting, the table also reports ΔR^2 , defined as the absolute difference between the augmented specification's R² and Model 1's R². This facilitates interpretation and avoids overstating economically small changes in predictive performance. Statistical significance is assessed using Diebold-Mariano tests based on squared forecast-error loss. * denote significance at the 10% levels, respectively.

From Table 9, the percentage-improvement results show that the incremental forecasting value of disruption-related predictors for nickel returns is most evident in the joint specification rather than in the supply-chain-only specification. Under ET, Model 2 reduces MAE by 0.71%, MSE by 0.73%, and RMSE by 0.37%, but these gains are not statistically significant. Model 3 also improves MAE, MSE, RMSE, and sMAPE, although the change in R² is negative. The strongest ET-based evidence is observed in Model 4, where the joint inclusion of SCS and GPR reduces MAE by 3.20%, MSE by 2.90%, and RMSE by 1.46%, while increasing R² by 342.38%. These gains are statistically significant at the 10% level according to the Diebold-Mariano test. Across other algorithms, the results are more heterogeneous. GBM benefits most from Model 3, in which the inclusion of GPR reduces MAE by 2.42%, MSE by 2.04%, and RMSE by 1.03%, with significance at the 5% level. XGBoost also shows reductions in MAE under Models 3 and 4, while LASSO and Elastic Net display only very small changes. KNN generally does not benefit from the augmented specifications. This heterogeneity reinforces the idea that the predictive value of disruption-related variables depends on both the market being forecast and the algorithm used.

Table 9. Percentage improvement of augmented models over the baseline model in forecasting nickel returns.

Models	Comparison	MAE	MSE	RMSE	R ²	ΔR^2	sMAPE
ET	Model 2 vs. Model 1	−0.71	−0.73	−0.37	430.64	0.0006	1.36
GBM	Model 2 vs. Model 1	−2.24	−2.23	−1.12	400.46	0.0053	0.98
XGBoost	Model 2 vs. Model 1	1.25	−1.08	−0.54	231.20	0.0025	−0.98
KNN	Model 2 vs. Model 1	2.88	3.42	1.69	74.83	0.0008	2.09
LASSO	Model 2 vs. Model 1	−0.14	0.01	0.01	−1.07	−0.0001	1.89
EN	Model 2 vs. Model 1	−0.14	0.02	0.01	−1.14	−0.0001	2.10
ET	Model 3 vs. Model 1	−1.14	−1.24	−0.62	−99.07	−0.0001	−1.09
GBM	Model 3 vs. Model 1	−2.42 **	−2.04 **	−1.03 **	238.77 **	0.0032	1.60
XGBoost	Model 3 vs. Model 1	−3.07	−0.88	−0.44	−66.21	−0.0007	1.73
KNN	Model 3 vs. Model 1	0.70	1.17	0.58	−67.48	−0.0007	−1.59
LASSO	Model 3 vs. Model 1	0.02	−0.06	−0.03	16.03	0.0005	0.06
EN	Model 3 vs. Model 1	0.02	−0.05	−0.02	14.55	0.0005	0.11
ET	Model 4 vs. Model 1	−3.20 *	−2.90 *	−1.46 *	342.38 *	0.0005	0.51
GBM	Model 4 vs. Model 1	−0.94	−1.84	−0.93	304.28	0.0040	−0.84
XGBoost	Model 4 vs. Model 1	−3.13	−1.43	−0.72	−14.04	−0.0002	1.53
KNN	Model 4 vs. Model 1	0.33	0.21	0.10	−87.96	−0.0010	0.31
LASSO	Model 4 vs. Model 1	−0.14	0.01	0.00	0.28	0.0000	1.90
EN	Model 4 vs. Model 1	−0.14	0.02	0.01	−1.14	−0.0001	2.10

Note: This table reports the percentage changes in out-of-sample forecasting performance for nickel returns relative to Model 1. Model 2 augments the baseline controls with supply-chain stress (SCS), Model 3 augments the baseline controls with geopolitical risk (GPR), and Model 4 includes both SCS and GPR in addition to the baseline controls. For MAE, MSE, RMSE, and sMAPE, negative values indicate improved forecast accuracy relative to Model 1, whereas positive values indicate weaker performance. For R², positive values indicate improved predictive performance. Percentage changes are computed relative to Model 1. Negative values for MAE, MSE, RMSE, and sMAPE indicate reductions in forecast error, whereas positive values for R² indicate improved explanatory performance. Because baseline R² values are often close to zero in daily return forecasting, the table also reports ΔR^2 , defined as the absolute difference between the augmented specification's R² and Model 1's R². This facilitates interpretation and avoids overstating economically small changes in predictive performance. Statistical significance is assessed using Diebold-Mariano tests based on squared forecast-error loss. **, and * denote significance at the 5%, and 10% levels, respectively.

From a forecasting standpoint, the nickel results are important because they show that the economic value of disruption-related variables is not uniform across strategic metals. For nickel, the greatest ET-based improvement occurs when supply-chain stress and geopolitical risk are jointly included, suggesting that physical trade frictions and geopolitical instability act as complementary forecasting signals. This is economically plausible because nickel is deeply embedded in global battery, stainless steel, and energy-transition value chains, where shipping constraints, inventory disruptions, sanctions risk, and strategic industrial policies may interact. Thus, nickel appears more responsive to the combined disruption environment than to a single channel alone.

This market-specific asymmetry is highly consistent with [8], who identify shipping-related transmission channels, and with [14], who show that the effects of supply bottlenecks can be nonlinear and state-dependent. It also complements [15], whose results imply that not all transition-sensitive commodities react equally to geopolitical signals. Our evidence adds that, for nickel returns, supply-chain stress and geopolitical risk become most informative when modeled jointly, highlighting the importance of combining physical and political-strategic disruption indicators in short-horizon forecasting.

Table 10 reports the corresponding percentage improvements for clean energy returns.

Table 10. Percentage improvement of augmented models over the baseline model in forecasting clean energy returns.

Models	Comparison	MAE	MSE	RMSE	R ²	ΔR^2	sMAPE
ET	Model 2 vs. Model 1	0.13	−0.26	−0.13	−25.98	−0.0007	1.45
GBM	Model 2 vs. Model 1	−0.53 *	−1.96 *	−0.98 *	585.50 *	0.0009	−0.28 *
XGBoost	Model 2 vs. Model 1	0.72	1.43	0.71	−10.36	0.0000	−0.29 **
KNN	Model 2 vs. Model 1	0.63	0.61	0.30	−92.27	−0.0011	0.49
LASSO	Model 2 vs. Model 1	0.44	0.54	0.27	−4.79	−0.0007	5.08
EN	Model 2 vs. Model 1	0.39	0.54	0.27	−6.56	−0.0010	4.43
ET	Model 3 vs. Model 1	−0.51	−0.67	−0.33	89.82	0.0021	−0.03
GBM	Model 3 vs. Model 1	0.07	0.01	0.01	−64.90	−0.0001	−3.29
XGBoost	Model 3 vs. Model 1	0.26	0.62	0.31	−97.88	−0.0004	0.85
KNN	Model 3 vs. Model 1	−0.58	−0.84	−0.42	93.73	0.0010	0.80
LASSO	Model 3 vs. Model 1	0.06	0.12	0.06	−10.10	−0.0016	0.12
EN	Model 3 vs. Model 1	0.08	0.15	0.08	−13.64	−0.0022	0.08
ET	Model 4 vs. Model 1	−0.29	−0.99	−0.50	42.30	0.0010	2.46
GBM	Model 4 vs. Model 1	0.66	0.72	0.36	1332.10	0.0021	0.07
XGBoost	Model 4 vs. Model 1	0.87	1.65	0.82	112.38 ***	0.0006	1.94
KNN	Model 4 vs. Model 1	−0.07	−1.06	−0.53	44.95	0.0005	1.29
LASSO	Model 4 vs. Model 1	0.44	0.54	0.27	−4.79	−0.0007	5.08
EN	Model 4 vs. Model 1	0.39	0.54	0.27	−6.56	−0.0010	4.43

Note: This table reports the percentage changes in out-of-sample forecasting performance for clean energy returns relative to Model 1. Model 2 augments the baseline controls with supply-chain stress (SCS), Model 3 augments the baseline controls with geopolitical risk (GPR), and Model 4 includes both SCS and GPR in addition to the baseline controls. For MAE, MSE, RMSE, and sMAPE, negative values indicate improved forecast accuracy relative to Model 1, whereas positive values indicate weaker performance. For R², positive values indicate improved predictive performance. Percentage changes are computed relative to Model 1. Negative values for MAE, MSE, RMSE, and sMAPE indicate reductions in forecast error, whereas positive values for R² indicate improved explanatory performance. Because baseline R² values are often close to zero in daily return forecasting, the table also reports ΔR^2 , defined as the absolute difference between the augmented specification's R² and Model 1's R². This facilitates interpretation and avoids overstating economically small changes in predictive performance. Statistical significance is assessed using Diebold-Mariano tests based on squared forecast-error loss. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

The percentage-improvement results for clean energy returns are mixed and, in general, less favorable for the augmented specifications than those observed for the strategic metal markets. For ET, Model 2 slightly improves MSE and RMSE but worsens MAE, R², and sMAPE. Model 3 improves MAE, MSE, RMSE, and sMAPE, but the gains are modest and statistically insignificant. Model 4 improves MSE and RMSE but worsens sMAPE. These results indicate that adding SCS and GPR does not generate a robust or statistically significant improvement for clean energy return forecasting under the ET model.

Across the other algorithms, the evidence is also heterogeneous. GBM benefits from the inclusion of SCS in Model 2, with statistically significant reductions in MAE, MSE, RMSE, and sMAPE, while XGBoost shows a statistically significant improvement only in sMAPE under Model 2 and in R² under Model 4. KNN records some improvements in selected error metrics under Models 3 and 4, but these gains are not statistically significant. LASSO and Elastic Net generally deteriorate when disruption-related variables are added. Overall, the evidence suggests that the incremental predictive role of supply-chain stress and geopolitical risk in clean energy returns is limited, model-dependent, and less consistent than in aluminum or nickel.

Economically, this pattern suggests that clean energy returns are shaped more strongly by conventional macro-financial and market-wide factors than by direct disruption-related indicators. This is consistent with the correlation evidence showing that clean energy returns are more strongly related to SPX, VIX, OIL, and USD than to SCS or GPR. Supply-chain stress can occasionally improve forecasts, particularly in GBM, but its effect is not

broad enough to establish a systematic forecasting advantage across model classes. Geopolitical risk alone also provides limited incremental information, and the joint specification does not consistently outperform the baseline.

This conclusion is broadly aligned with [11], who show that not all external risk indicators add equally strong forecasting value in clean energy markets, and with [12], who emphasize the dominant role of broader market conditions in shaping clean energy dynamics. At the same time, the findings differ from [4], in which transition-related concerns play a stronger predictive role. This difference suggests that the forecasting relevance of external risk variables depends on the selected clean energy measure, the type of risk indicator, and whether the target variable is market returns, volatility, or transition-risk exposure.

4.7. SHAP-Based Interpretability Analysis

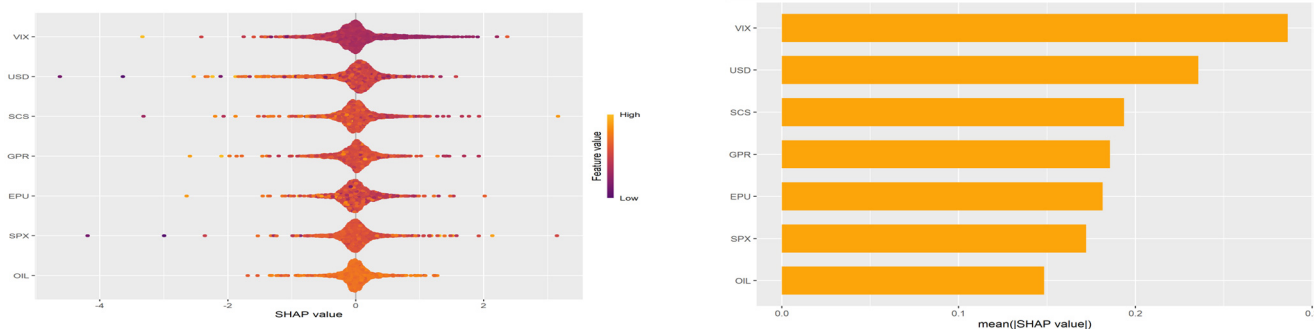
Figure 3 presents the SHAP-based interpretation of the Extra Trees model for aluminum, copper, nickel, and clean energy returns, thereby clarifying the relative importance of the predictors included in the full specification.

The SHAP results indicate substantial heterogeneity in the drivers of return forecasts across the four markets. Unlike in a setting where a single predictor uniformly dominates across all assets, the return-based SHAP evidence shows that predictor importance is asset-specific. VIX, USD, EPU, SPX, GPR, and SCS each play different roles depending on the market being forecast. This finding indicates that broad financial market conditions remain important, but they do not operate through a single dominant predictor for all strategic commodity and clean energy returns. Instead, the forecasting structure combines market-wide risk conditions, exchange-rate effects, policy uncertainty, geopolitical risk, and supply-chain stress in different proportions across assets. This evidence is consistent with [11], who emphasize the importance of external financial predictors in clean energy forecasting, and it also accords with the broader market-centered evidence reported by [12].

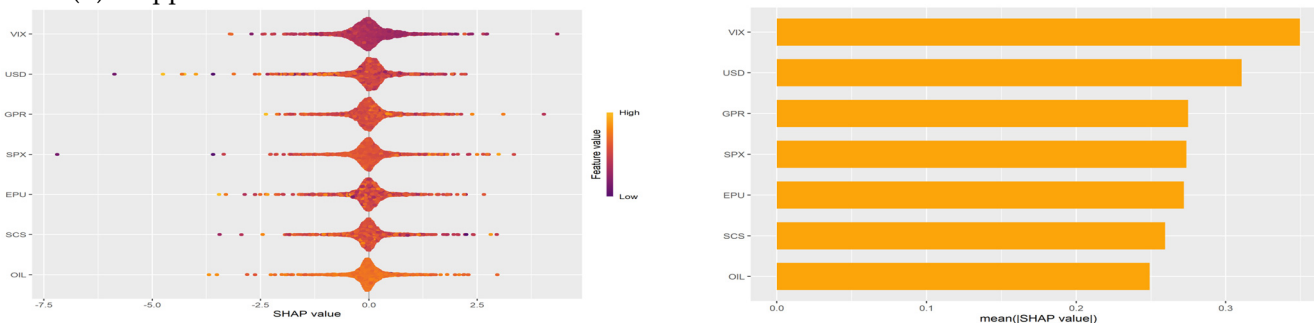
For aluminum returns, VIX is the most influential predictor in the full ET specification, followed by USD, SCS, GPR, EPU, SPX, and OIL. This suggests that aluminum return forecasts are primarily shaped by financial market stress and exchange-rate movements, while disruption-related variables provide additional, albeit secondary, predictive power. The relatively high ranking of SCS is economically intuitive because aluminum is closely linked to global trade flows, shipping conditions, and industrial supply chains. GPR also contributes meaningfully, suggesting that geopolitical uncertainty may affect aluminum returns through trade restrictions, sanctions risk, strategic supply conditions, and expectations about global industrial demand. This ranking is consistent with the earlier finding that SCS and GPR improve aluminum return forecasts, although their contribution remains incremental rather than dominant.

For copper returns, VIX again emerges as the leading predictor, followed by USD, GPR, SPX, EPU, SCS, and OIL. The prominent role of the VIX indicates that copper return forecasts are strongly conditioned by global risk aversion and financial market stress. The importance of USD is consistent with copper's status as an internationally traded, dollar-denominated commodity, while GPR's relatively high ranking supports the forecasting evidence that geopolitical risk provides the clearest incremental signal for copper returns. SCS contributes less than GPR but remains part of the forecasting structure, indicating that copper is exposed to both physical and geopolitical disruption channels, although not with equal intensity. This result is in line with [8,9], who emphasize the forecasting relevance of supply-chain and shipping-related conditions, and it complements [15], who show that transition-relevant commodities can respond to geopolitical developments in heterogeneous ways.

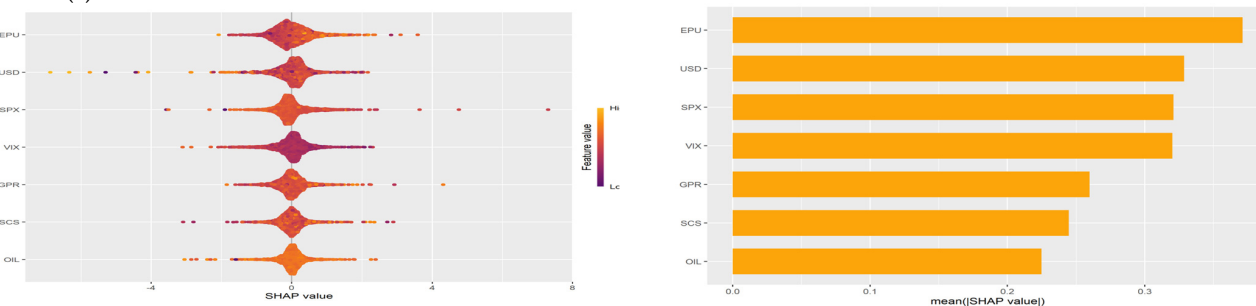
Panel (a): Aluminum



Panel (b): Copper



Panel (c): Nickel



Panel (d): Clean Energy

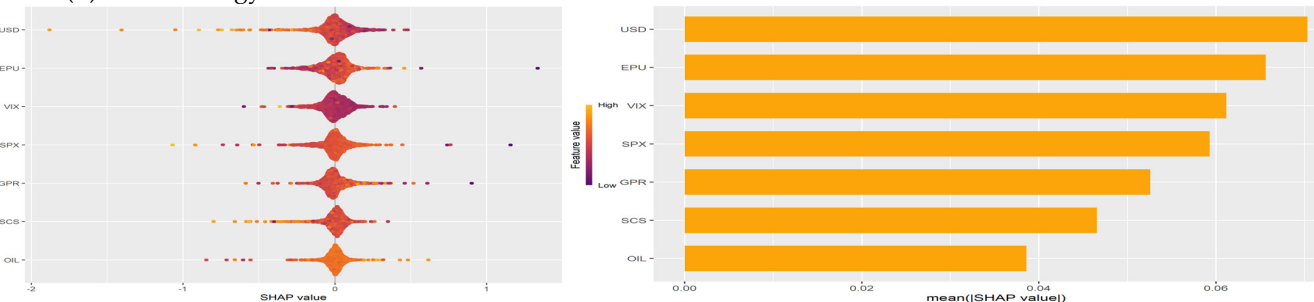


Figure 3. SHAP-based interpretation of the Extra Trees forecasting model. Note: The figure reports SHAP-based interpretability results for the Extra Trees forecasting model. For each asset, the left panel displays the SHAP beeswarm plot and the right panel shows mean absolute SHAP importance. The plots illustrate both the direction and magnitude of each predictor's contribution to forecasts, as well as the relative ranking of predictors across returns in the aluminum, copper, nickel, and clean energy markets.

Nickel displays a broader predictor profile. EPU is the most influential predictor for nickel returns, followed by USD, SPX, VIX, GPR, SCS, and OIL. This ranking suggests that nickel return forecasts are especially sensitive to economic policy uncertainty and broad macro-financial conditions. This is economically plausible because nickel is deeply embedded in global industrial production, battery supply chains, and energy-transition investment, all of which can be affected by uncertainty about trade policy, industrial policy,

and macroeconomic regulation. The presence of both GPR and SCS among the relevant predictors is also important. Although neither dominates the SHAP ranking, their joint contribution is consistent with the forecasting results, which show that the combined SCS-GPR specification provides the greatest ET-based improvement for nickel returns. The nickel results are consistent with the supply-chain-based narrative emphasized by [8,9,13], and also suggest that geopolitical risk is more informative when combined with direct supply-chain stress.

For clean energy returns, USD is the most influential predictor, followed by EPU, VIX, SPX, GPR, SCS, and OIL. This pattern indicates that clean energy returns are more closely tied to exchange-rate movements, policy uncertainty, financial market volatility, and broad equity-market conditions than to direct disruption-specific indicators. The leading role of USD and the high ranking of EPU are consistent with the equity-like nature of clean energy indices, which are sensitive to global capital flows, discount-rate conditions, policy uncertainty, and investor sentiment. GPR and SCS contribute to the forecasting process but remain secondary, which is consistent with the forecasting results showing that augmented specifications do not generate robust or statistically significant ET-based improvements for clean energy returns. Thus, clean energy behaves more like a transition-sensitive equity segment than a physical commodity market in the one-day-ahead return forecasting framework. This result mirrors the market-dominance pattern highlighted by [11], while also aligning with [12], who emphasize the importance of broader financial and energy market conditions in clean energy dynamics.

The SHAP evidence reinforces two central conclusions. First, broad macro-financial conditions remain important across all examined assets, but their influence is transmitted through different variables rather than through a single common predictor. VIX and USD are especially important for aluminum and copper, EPU is central for nickel, and USD and EPU dominate clean energy. Second, the influence of disruption-related variables is clearly asset-specific. SCS plays a relatively stronger role for aluminum and remains relevant for nickel, whereas GPR is particularly informative for copper and contributes to nickel when combined with SCS. Clean energy returns, by contrast, are more strongly shaped by conventional macro-financial variables than by disruption-specific indicators. These findings further support the use of flexible machine-learning methods that can capture nonlinear and heterogeneous predictive relationships across markets. More broadly, the SHAP results complement the responsible and interpretable forecasting literature, including [21,23–26], by showing that explainability tools do not merely improve transparency but also help reveal how disruption-related predictors operate differently across strategic commodity and clean energy markets.

4.8. Robustness Checks

To assess the stability of the baseline findings, we conduct several robustness checks. First, we examine the time-series properties of the variables using Augmented Dickey-Fuller (ADF) unit-root tests. The results, reported in Appendix A Table A1, show that several variables in levels do not reject the unit-root null, whereas all transformed daily log-return/log-change series reject the unit-root null at the 1% level. This supports the return-based forecasting framework and reduces the risk that the results are driven by persistence or common trends in level series. Second, we examine whether the main findings are sensitive to the forecasting horizon by extending the prediction target from one-day-ahead returns to five-day-ahead returns. Third, we use an alternative chronological 70:30 train-test split. Fourth, we augment the forecasting specification with the lagged dependent variable to account for short-run return persistence. Fifth, we compare the Extra Trees full specification with conventional forecasting benchmarks, including a zero-return

benchmark, historical mean, AR(1), and ARIMA models. The corresponding benchmark results are reported in Appendix A Table A4, with Diebold-Mariano statistics and p -values under squared-error and absolute-error loss. Finally, we examine regime-specific SHAP values to assess whether predictor importance is stable across different macro-financial environments. The regimes are defined using major global events over the sample period, and the regime-specific SHAP values are computed by re-estimating the Extra Trees full specification within each sub-period. These checks preserve the temporal ordering of the data and provide additional evidence on whether the predictive role of supply-chain stress and geopolitical risk is robust across alternative validation designs, forecast horizons, and benchmark models.

The five-day-ahead robustness results are reported in Appendix A Tables A2 and A3. The five-day-ahead target is constructed by leading each transformed return series by five trading days, so that predictors observed at time t are used to forecast the asset return at time $t + 5$. The same chronological train-test split and expanding-window validation structure are preserved, and the Diebold-Mariano tests are implemented using the five-day forecast horizon. This test is particularly relevant in the present setting because the effects of supply-chain stress and geopolitical risk may not be fully reflected in returns within a single trading day. A slightly longer horizon therefore provides a useful way to examine whether the predictive relevance of disruption-related variables persists beyond the ultra-short run. Consistent with the main analysis, the five-day-ahead robustness exercise is implemented using the Extra Trees (ET) model, which provides a flexible nonlinear benchmark for detecting interaction effects and nonlinearities. Specifically, the five-day-ahead target is constructed by leading each return series by five trading days, so that the information set available at time t is used to predict the return observed at time $t + 5$. The same chronological train-test structure and expanding-window validation logic are preserved to avoid look-ahead bias. This choice is consistent with the broader machine-learning forecasting literature, including [11,19,20], which highlights the usefulness of flexible nonlinear models in complex market environments.

Generally, the five-day-ahead results confirm that the forecasting contribution of disruption-related variables is asset-specific rather than uniform across all markets. For aluminum, the augmented specifications improve several error-based metrics. Model 3, which adds geopolitical risk, produces the lowest MAE and sMAPE, while Model 4, which includes both SCS and GPR, produces the lowest MSE and RMSE and the highest R^2 . This suggests that disruption-related information remains relevant for aluminum at a slightly longer horizon, although the gains remain economically modest. This interpretation is consistent with the more cautious evidence of [15,16], both of whom indicate that industrial and transition-related commodities do not always react uniformly or instantaneously to external geopolitical shocks.

For copper, the five-day-ahead evidence is weaker. The baseline specification remains competitive for MSE and RMSE, while Model 3 and Model 4 generate only small improvements in selected metrics. Model 3 slightly improves MAE, whereas Model 4 raises R^2 but does not consistently reduce the main forecast-error measures. This confirms that the predictive role of disruption-related variables for copper is more limited and less stable at the five-day horizon than for aluminum or nickel. Nevertheless, the fact that Model 4 improves selected metrics remains consistent with the broader view that copper is exposed to both physical trade disruptions and geopolitical uncertainty. This result is broadly compatible with [9], who document the predictive importance of supply-chain pressure for metals, and with [6,12], who emphasize the relevance of geopolitical conditions in energy and transition-related markets.

For nickel, the five-day-ahead robustness check provides clearer support for the joint disruption channel. Model 4 reduces MAE from 2.9268 to 2.9034, MSE from 25.9485 to 25.3049, and RMSE from 5.0940 to 5.0304, while increasing R^2 from 0.0003 to 0.0030. These findings reinforce the interpretation that nickel returns are sensitive to the combined information content of supply-chain stress and geopolitical risk. Given nickel's role in battery production, stainless steel, and transition-related supply chains, this result is economically plausible because physical bottlenecks and geopolitical uncertainty may interact in shaping market expectations. This pattern is closely aligned with [8,9], who stress the role of shipping risk and supply-chain pressure in metals markets, while also complementing [15], whose evidence suggests that transition-sensitive commodities can display delayed and heterogeneous responses to geopolitical conditions.

For clean energy returns, the evidence on robustness remains more mixed. Model 4 slightly improves MSE and RMSE relative to the baseline, but it does not improve MAE, R^2 , or sMAPE. Model 2 and Model 3 also do not produce systematic gains. This confirms the main finding that clean energy returns are more strongly driven by conventional macro-financial conditions than by direct disruption-related indicators. In other words, clean energy behaves more like a transition-sensitive equity segment than a physical commodity market in this short-horizon forecasting environment. This result is broadly compatible with [11], who show that not all external risk indicators contribute equally to clean energy forecasting, and with [12], who highlight the dominant influence of broader market conditions in clean energy dynamics.

The alternative 70:30 chronological split and lagged-dependent-variable checks further support a cautious interpretation. The strongest evidence of robustness is observed for aluminum, where SCS and GPR continue to improve forecast accuracy across alternative validation designs. In the 70:30 split, Model 2 and Model 3 significantly reduce MAE, MSE, and RMSE relative to the baseline. In the lagged-dependent-variable specification, Models 2 and 4 yield statistically significant improvements, indicating that disruption-related predictors retain incremental value even after accounting for lagged aluminum returns. Copper and nickel show more selective evidence, with improvements varying by model specification and the error metric. Clean energy shows occasional gains, especially under the alternative split, but these gains do not consistently persist across all robustness designs.

The econometric benchmark comparison, reported in Appendix A Table A4, provides additional evidence on the benchmark-adjusted performance of the Extra Trees full specification. The model is compared with four conventional forecasting benchmarks: a zero-return forecast, a historical-mean forecast, an AR(1) model, and an ARIMA model. The results show that the performance advantage of the Extra Trees full specification is not uniform across all metrics and assets, which is consistent with the weak predictability and high noise typically observed in daily return forecasting. In terms of MAE, MSE, and RMSE, the conventional benchmarks are often numerically close to, and in some cases slightly better than, the Extra Trees full specification. However, the Extra Trees full specification yields lower sMAPE values across all four assets, suggesting it provides more stable relative forecast accuracy when prediction errors are scaled by the magnitudes of actual and predicted returns. The Diebold-Mariano tests indicate that most benchmark differences are not statistically significant under squared-error loss. Under absolute-error loss, the clearest evidence appears for nickel, where the Extra Trees full specification differs significantly from the zero-return, historical mean, AR(1), and ARIMA benchmarks. These results support a cautious interpretation: the explainable machine-learning framework offers useful benchmark-adjusted gains for selected assets and metrics, especially nickel, but it should not be interpreted as uniformly dominating simple conventional benchmarks across all return-forecasting criteria.

To further evaluate the relative ranking of all model-specification combinations, Appendix A Table A5 reports the Model Confidence Set results for each asset across the 24 algorithm-specification combinations considered. The MCS analysis complements the Diebold-Mariano tests by identifying the subset of models that are statistically indistinguishable from the best-performing specification under squared-error and absolute-error loss. The results provide a more cautious interpretation of the comparative forecasting evidence. For aluminum and copper, all models are retained under squared-error loss, indicating that no specification statistically dominates the full set of alternatives. Under absolute-error loss, the retained sets become smaller, but Extra Trees specifications remain statistically competitive, especially ET_M3 and ET_M4 for copper. For nickel, Extra Trees specifications are retained under squared-error loss, but penalized linear models dominate the superior set under absolute-error loss. For clean energy, Extra Trees models are retained under both loss functions, but baseline and penalized specifications are also statistically comparable. The MCS evidence confirms that Extra Trees is often competitive and frequently retained in the superior model set, but it does not uniformly dominate all alternatives. This supports a more balanced conclusion: nonlinear and interaction-based methods are useful in this forecasting setting, but their statistical advantage is asset- and loss-function-specific rather than universal.

Finally, the regime-specific SHAP analysis, reported in Appendix A Tables A6 and A7, provides additional evidence that predictor importance is not time-invariant. Table A5 formally defines the four macro-financial regimes used to partition the sample, while Table A6 reports the three leading predictors for each asset and regime based on mean absolute SHAP values from the Extra Trees full specification. The results show that predictor importance varies meaningfully across regimes and assets. During the pre-COVID-19 period, geopolitical risk, policy uncertainty, oil-market conditions, exchange-rate movements, and equity-market conditions were among the leading predictors, depending on the asset. During the COVID-19 and logistics-stress period, volatility, policy uncertainty, and supply-chain stress became more visible, especially for industrial metals, with SCS appearing among the top three predictors for nickel. During the Russia-Ukraine and monetary-tightening period, equity-market and exchange-rate variables gained in importance, particularly SPX and USD, reflecting the role of broad financial conditions and dollar movements in short-run return dynamics. In recent periods, oil-market conditions, volatility, supply-chain stress, and geopolitical risk have regained relevance depending on the asset. These results confirm that the full-sample SHAP rankings should be interpreted as average effects rather than time-invariant relationships and provide direct empirical support for the regime-specific interpretation.

4.9. Discussion and Synthesis of Findings

The empirical results yield four principal conclusions. First, the forecasting exercise confirms the usefulness of flexible machine-learning methods, although model performance is more modest and more heterogeneous than would be expected in highly persistent level-based forecasting. Extra Trees (ET) remains an important nonlinear benchmark across the analysis because it captures interaction effects and provides the basis for the SHAP explainability exercise. However, the return-based results also show that no single model uniformly dominates all assets and all metrics. This is consistent with the low signal-to-noise ratio of daily return forecasting and supports a more cautious interpretation of predictive performance. The usefulness of nonlinear and tree-based models aligns closely with that reported by [11,19,20,27], all of whom document the value of flexible machine-learning methods in complex forecasting environments. At the same time, the benchmark comparison with zero-return, historical mean, AR(1), and ARIMA models indicates that

machine-learning gains are strongest for selected markets rather than universal across all assets.

Second, the incremental forecasting value of disruption-related predictors is strongly asset-specific. For aluminum returns, the augmented specifications provide statistically significant improvements over the baseline, especially when geopolitical risk is included and when SCS and GPR are modeled jointly. For copper returns, the evidence is weaker and mainly associated with the geopolitical risk channel rather than the joint specification. For nickel returns, the joint SCS-GPR specification provides the greatest ET-based improvement, indicating that physical supply-chain frictions and geopolitical uncertainty act as complementary forecasting signals. Clean energy returns, by contrast, show no robust ET-based improvement from the augmented specifications and remain more closely tied to conventional macro-financial variables. The robustness analysis reinforces this heterogeneity. The 70:30 chronological split and lagged-dependent-variable checks provide the strongest support for aluminum, while the five-day-ahead results provide clearer evidence for the joint disruption channel in nickel. Clean energy remains the least robust case, confirming that its return dynamics are less directly linked to SCS and GPR. This heterogeneity is consistent with [12,15,16], who point to asset-specific responses to geopolitical and transition-related shocks, but our findings extend this literature by showing how these differences translate into statistically tested out-of-sample forecasting performance.

Third, the results imply that broad macro-financial conditions remain central to forecasting strategic commodity and clean energy returns, but the SHAP evidence no longer supports the view that a single predictor dominates all markets. Instead, predictor importance differs sharply across assets. For aluminum and copper, VIX and USD are the leading predictors, highlighting the importance of financial market stress and exchange-rate conditions. For nickel, EPU is the most influential predictor, followed by USD, SPX, VIX, GPR, and SCS, indicating that policy uncertainty and macro-financial conditions are particularly relevant for nickel return forecasting. For clean energy, USD, EPU, VIX, and SPX dominate the SHAP ranking, which is consistent with the equity-like nature of clean energy indices. This pattern is also consistent with the correlation structure in Figure 2, where clean energy is more closely associated with broad equity-market and financial-risk variables than with SCS or GPR. Disruption-related variables therefore do not replace conventional macro-financial controls; rather, they enhance forecasts for specific assets and horizons where the relevant transmission channels are sufficiently strong. The weaker incremental role of SCS and GPR for aluminum and clean energy also has a clear economic interpretation. Aluminum is a globally traded industrial metal whose short-run returns are rapidly repriced in response to broad financial conditions, exchange-rate movements, oil-market dynamics, and global risk sentiment. As a result, supply-chain stress and geopolitical risk may influence aluminum returns indirectly through volatility, the US dollar, energy costs, and global demand expectations, rather than through a dominant, standalone channel. Clean energy returns are even more strongly linked to conventional macro-financial conditions because clean energy indices are composed of listed firms rather than physical commodities. Their short-run returns are therefore highly sensitive to discount-rate expectations, policy uncertainty, equity-market sentiment, exchange-rate movements, and financial volatility. This helps explain why clean energy returns show limited statistically significant improvement from the augmented SCS-GPR specifications: the relevant information is partly absorbed by broader macro-financial variables. Thus, weak incremental gains in forecasting do not imply that supply-chain stress and geopolitical risk are economically irrelevant; rather, their effects are indirect, conditional, and more likely to emerge through interactions with broader financial conditions or during specific stress regimes. In this sense, our evidence refines prior studies: it supports the relevance of supply-chain and geopolitical variables,

but shows that their predictive contribution is conditional on market structure, forecasting horizon, and macro-financial regime.

Fourth, the forecasting evidence carries direct practical implications. The results suggest that model specification should be tailored to the asset class rather than imposed uniformly across markets. For aluminum, monitoring SCS and GPR can improve short-horizon risk assessment, especially when nonlinear models are used. For copper, GPR appears to provide the most relevant incremental signal, but the gains are modest and should be interpreted cautiously. For nickel, the joint monitoring of supply-chain stress and geopolitical risk is particularly important because the two channels appear to interact in shaping return predictability. For clean energy, forecasting models should place greater emphasis on exchange rates, policy uncertainty, financial volatility, and broader equity-market conditions rather than relying primarily on disruption-specific indicators. More generally, the results demonstrate that explainable machine learning offers a useful framework for identifying when disruption-related variables genuinely enhance forecast performance and when conventional macro-financial variables remain sufficient. This final point connects the present study to the recent literature on interpretable and responsible forecasting, particularly [23,24], by demonstrating that predictive accuracy and economic interpretability can be combined in a way that is directly useful to investors, risk managers, and policymakers.

5. Conclusions

This paper examines whether daily supply-chain stress and geopolitical risk improve the forecasting of strategic commodity and clean energy market returns within an explainable machine-learning framework. Using daily data on aluminum, copper, nickel, and clean energy from 10 February 2015 to 27 February 2026, the analysis compares a baseline model with conventional market controls to augmented specifications that incorporate supply-chain stress, geopolitical risk, and their joint effects. The empirical framework is based on one-day-ahead log-return forecasting, chronological validation, formal Diebold-Mariano tests, and SHAP-based explainability. The empirical design is motivated by the growing importance of disruption-sensitive markets in an environment characterized by supply-chain fragility, geopolitical risk, and heightened macro-financial uncertainty. The central contribution of the paper is to document when and for which assets daily supply-chain stress and geopolitical risk provide incremental predictive information beyond conventional macro-financial controls.

Several findings emerge. In direct response to the first research question, the results show that daily supply-chain stress and geopolitical risk do contain predictive information beyond conventional market variables, but this information is not equally strong across all assets. Aluminum returns benefit from augmented specifications that include SCS and/or GPR, copper returns show weaker evidence mainly through the geopolitical-risk channel, and nickel returns benefit most from the joint SCS-GPR specification. Clean energy returns, by contrast, do not show robust ET-based improvements from the augmented specifications and remain more closely linked to conventional macro-financial variables. Thus, the answer to the first research question is affirmative but conditional: disruption-related variables add forecasting value in selected markets rather than uniformly across all strategic commodity and clean energy returns.

In response to the second research question, the results show that neither supply-chain stress nor geopolitical risk is universally superior as a forecasting signal. GPR provides the clearest incremental signal for copper; SCS and GPR jointly matter most for nickel; and SCS contributes meaningfully to aluminum but does not dominate the full forecasting structure. This indicates that the relative forecasting strength of the two disruption-related indicators

depends on the asset being forecast. Supply-chain stress is more informative when physical trade and logistics frictions are more relevant, whereas geopolitical risk is more informative when political-strategic uncertainty directly affects market expectations.

In response to the third research question, the evidence confirms substantial heterogeneity in assets. The predictive effects of disruption-related variables vary across aluminum, copper, nickel, and clean energy, rather than applying uniformly across markets. Aluminum displays statistically significant but economically modest improvements from augmented specifications. Copper shows weaker gains, mainly associated with geopolitical risk. Nickel benefits most from the joint inclusion of supply-chain stress and geopolitical risk, suggesting complementarity between physical disruption and political uncertainty. Clean energy returns remain the least responsive to disruption-related predictors and are more strongly associated with macro-financial variables. Overall, the findings show that the forecasting role of SCS and GPR is asset-specific, nonlinear, and conditional on market structure.

The model comparison results also indicate that flexible machine-learning methods are useful in this forecasting environment, but their advantage is not universal across all assets and metrics. The model comparison results show that flexible machine-learning methods are useful in this forecasting environment, but their advantage is not universal. The MCS results confirm that several competing models are often statistically indistinguishable from the best-ranked specification, depending on the asset and loss function. Therefore, the main conclusion is not that one algorithm dominates, but that supply-chain stress and geopolitical risk contain selective predictive information that different model classes exploit with varying degrees of success. The benchmark comparison with zero-return, historical mean, AR(1), and ARIMA models further shows that ET provides the clearest value added for nickel and partial evidence for copper, while its gains for aluminum and clean energy are more limited. Thus, the paper supports a cautious interpretation of machine-learning predictability: nonlinear models are useful, but they do not deliver universal forecasting dominance across all markets.

The SHAP-based interpretability analysis provides additional economic insight. Unlike in a setting where a single predictor dominates all markets, the return-based SHAP evidence shows that predictor importance is asset-specific. For aluminum and copper, VIX and USD are among the leading predictors, indicating the importance of financial market stress and exchange-rate conditions. For nickel, EPU is the most influential predictor, followed by USD, SPX, VIX, GPR, and SCS, suggesting that policy uncertainty and macro-financial conditions are central to forecasting nickel returns. For clean energy, USD, EPU, VIX, and SPX dominate the ranking, confirming that clean energy behaves more like a transition-sensitive equity segment than a physical commodity market. SCS and GPR remain relevant in selected cases, but their role is generally incremental and conditional rather than dominant.

These findings carry several implications for different stakeholders. For investors and portfolio managers, the results suggest that forecasting models should be tailored to the specific market under consideration rather than applied uniformly across strategic metals and clean energy assets. For aluminum, monitoring both supply-chain stress and geopolitical risk can improve short-horizon risk assessment. For copper, geopolitical risk appears to provide the more informative incremental signal, although the gains remain modest. For nickel, jointly monitoring supply-chain stress and geopolitical risk is particularly useful because the two channels appear to interact in shaping return predictability. For clean energy forecasting, models should emphasize macro-financial predictors such as exchange rates, policy uncertainty, financial volatility, and broad equity-market conditions. For commodity traders and risk managers, the results indicate that disruption-related indicators

contain economically meaningful information beyond that captured by standard financial controls in selected strategic metal markets. For policymakers, the findings highlight that disruptions in logistics systems and broader geopolitical risk can contain measurable predictive information for strategically important markets, implying that resilience-enhancing policies in trade infrastructure, shipping networks, and industrial supply chains may have indirect stabilizing effects on market uncertainty. More broadly, the evidence suggests that the transmission of external shocks is not uniform across markets, with implications for industrial policy, strategic stockpiling, and transition-related market surveillance.

The study also opens several avenues for future research. First, future studies could examine whether disruption-related variables have stronger predictive content for volatility, downside risk, or tail-risk measures than for mean returns. Second, future work could explore more regime-dependent or crisis-specific forecasting designs, particularly during episodes such as the COVID-19 shock, the Russia-Ukraine war, and periods of severe shipping disruption. Third, future research could expand the asset universe to include additional strategic minerals, renewable-energy equities, battery-related assets, or regional clean energy indices to test whether the present results generalize across broader transition-sensitive markets. Fourth, future research could extend the present framework by incorporating indicators of investor sentiment, climate policy risk, or climate transition concern to assess whether these variables provide additional forecasting value, particularly for clean energy returns. Finally, future studies could examine whether clean energy return predictability improves when the BDI is replaced or complemented by supply-chain indicators more closely aligned with clean energy production networks, such as container-freight rates, port-congestion measures, semiconductor-shortage indicators, or battery-material logistics proxies. Such extensions would help determine whether the weak role of BDI-based supply-chain stress in clean energy forecasting reflects the limited relevance of supply-chain shocks for clean energy equities or the use of a dry-bulk proxy that does not fully capture clean energy-specific logistics channels.

Funding: This work was supported and funded by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) (grant number IMSIU-DDRSP2604).

Data Availability Statement: The data supporting the findings of this study are detailed in Table 1 (definitions and sources).

Conflicts of Interest: The author declares no conflicts of interest.

Appendix A

Table A1. Augmented Dickey-Fuller unit-root tests for levels and transformed series.

Variable	ADF (Level)	ADF (Return/Log Change)
SCS	−3.9598 **	−12.4049 ***
VIX	−5.6673 ***	−15.1134 ***
ALU	−3.1754 *	−22.2357 ***
COP	−2.6346	−21.1832 ***
NIC	−3.1166	−21.6507 ***
EPU	−3.9965 **	−21.9536 ***
GPR	−7.0392 ***	−19.3844 ***
SPX	−1.9327	−13.2516 ***
CEN	−2.2469	−11.6223 ***
OIL	−2.8136	−12.7512 ***
USD	−2.4202	−13.8151 ***

Note: This table reports Augmented Dickey-Fuller test statistics for the original level series and the transformed daily log-return/log-change series. The null hypothesis is that the series contains a unit root. ***, **, and * denote rejection of the unit-root null at the 1%, 5%, and 10% significance levels, respectively.

Table A2. Five-day-ahead forecasting results using the Extra Trees model.

<i>Panel A. Aluminum (ALU)</i>					
Model	MAE_M1	MSE_M1	RMSE_M1	R ² _M1	sMAPE_M1
ET	2.9860	22.7166	4.7662	0.0001	1.6208
	MAE_M2	MSE_M2	RMSE_M2	R ² _M2	sMAPE_M2
ET	2.9359	22.4657	4.7398	0.0000	1.6352
	MAE_M3	MSE_M3	RMSE_M3	R ² _M3	sMAPE_M3
ET	2.8945	22.4955	4.7429	0.0001	1.5820
	MAE_M4	MSE_M4	RMSE_M4	R ² _M4	sMAPE_M4
ET	2.8970	22.2054	4.7123	0.0014	1.6223
<i>Panel B. Copper (COP)</i>					
Model	MAE_M1	MSE_M1	RMSE_M1	R ² _M1	sMAPE_M1
ET	2.9793	23.8698	4.8857	0.0061	1.5508
	MAE_M2	MSE_M2	RMSE_M2	R ² _M2	sMAPE_M2
ET	2.9986	24.2850	4.9280	0.0001	1.6222
	MAE_M3	MSE_M3	RMSE_M3	R ² _M3	sMAPE_M3
ET	2.9656	24.0466	4.9037	0.0011	1.6089
	MAE_M4	MSE_M4	RMSE_M4	R ² _M4	sMAPE_M4
ET	2.9741	23.9196	4.8908	0.0035	1.6101
<i>Panel C. Nickel (NIC)</i>					
Model	MAE_M1	MSE_M1	RMSE_M1	R ² _M1	sMAPE_M1
ET	2.9268	25.9485	5.0940	0.0003	1.4929
	MAE_M2	MSE_M2	RMSE_M2	R ² _M2	sMAPE_M2
ET	3.0087	26.2450	5.1230	0.0000	1.5434
	MAE_M3	MSE_M3	RMSE_M3	R ² _M3	sMAPE_M3
ET	2.9121	25.6095	5.0606	0.0019	1.4992
	MAE_M4	MSE_M4	RMSE_M4	R ² _M4	sMAPE_M4
ET	2.9034	25.3049	5.0304	0.0030	1.5367
<i>Panel D. Clean Energy (CEN))</i>					
Model	MAE_M1	MSE_M1	RMSE_M1	R ² _M1	sMAPE_M1
ET	1.1125	2.1188	1.4556	0.0022	1.6546
	MAE_M2	MSE_M2	RMSE_M2	R ² _M2	sMAPE_M2
ET	1.1168	2.1216	1.4566	0.0009	1.6892
	MAE_M3	MSE_M3	RMSE_M3	R ² _M3	sMAPE_M3
ET	1.1205	2.1328	1.4604	0.0001	1.6945
	MAE_M4	MSE_M4	RMSE_M4	R ² _M4	sMAPE_M4
ET	1.1156	2.1139	1.4539	0.0012	1.6987

Note: This table reports the out-of-sample five-day-ahead forecasting performance of the Extra Trees model for aluminum (ALU), copper (COP), nickel (NIC), and clean energy (CEN) returns. Model 1 includes the baseline controls, namely, VIX, EPU, SPX, OIL, and USD. Model 2 adds supply-chain stress (SCS), Model 3 adds geopolitical risk (GPR), and Model 4 includes both SCS and GPR. The five-day-ahead target is constructed by leading each transformed return series by five trading days, so that predictors observed at time t are used to forecast the asset return at time $t + 5$. Lower values of MAE, MSE, RMSE, and sMAPE indicate better forecast accuracy, whereas a higher R^2 indicates stronger predictive performance.

Table A3. Percentage improvement of augmented models in five-day-ahead forecasting.

Asset	Comparison	MAE	MSE	RMSE	R ²	sMAPE
ALU	Model 2 vs. Model 1	−1.68	−1.10	−0.55	−99.10	0.88
ALU	Model 3 vs. Model 1	−3.07	−0.97	−0.49	−58.42	−2.40
ALU	Model 4 vs. Model 1	−2.98	−2.25	−1.13	866.46	0.09
COP	Model 2 vs. Model 1	0.65	1.74	0.87	−98.19	4.61
COP	Model 3 vs. Model 1	−0.46	0.74	0.37	−82.73	3.75
COP	Model 4 vs. Model 1	−0.17	0.21	0.10	−43.38	3.83
NIC	Model 2 vs. Model 1	2.80	1.14	0.57	−98.65	3.38
NIC	Model 3 vs. Model 1	−0.50	−1.31	−0.66	549.92	0.42
NIC	Model 4 vs. Model 1	−0.80	−2.48	−1.25	903.85	2.93
CEN	Model 2 vs. Model 1	0.39	0.13	0.07	−59.20	2.09
CEN	Model 3 vs. Model 1	0.72	0.66	0.33	−95.48	2.41
CEN	Model 4 vs. Model 1	0.28	−0.23	−0.12	−43.04	2.66

Note: This table reports the percentage change in five-day-ahead forecasting performance when Models 2, 3, and 4 are compared with Model 1. For MAE, MSE, RMSE, and sMAPE, negative values indicate improved forecast accuracy relative to Model 1, whereas positive values indicate weaker performance. For R², positive values indicate improved predictive performance. Model 2 adds supply-chain stress (SCS), Model 3 adds geopolitical risk (GPR), and Model 4 includes both SCS and GPR.

Table A4. Benchmark comparison between the Extra Trees full specification and conventional forecasting benchmarks.

Asset	Model	MAE	MSE	RMSE	R ²	sMAPE	DM Stat., Squared Loss	DM <i>p</i> -Value, Squared Loss	DM stat., Absolute Loss	DM <i>p</i> -Value, Absolute Loss
Aluminum	ET full specification, Model 4	2.8213	22.1043	4.7015	0.0029	1.5488	-	-	-	-
Aluminum	Zero-return benchmark	2.7976	22.0267	4.6933	-	1.9958	−0.1950	0.8455	−0.9112	0.3627
Aluminum	Historical mean	2.7966	22.0240	4.6930	-	1.9198	−0.2013	0.8406	−0.9506	0.3423
Aluminum	AR(1)	2.7970	22.0247	4.6931	0.0001	1.9202	−0.1994	0.8420	−0.9339	0.3509
Aluminum	ARIMA	2.7973	22.0266	4.6933	0.0000	1.9869	−0.1951	0.8454	−0.9210	0.3575
Copper	ET full specification, Model 4	2.8610	23.8797	4.8867	0.0011	1.5899	-	-	-	-
Copper	Zero-return benchmark	2.8300	23.7567	4.8741	-	2.0000	−0.3633	0.7165	−1.3085	0.1914
Copper	Historical mean	2.8296	23.7522	4.8736	-	1.8826	−0.3751	0.7078	−1.3222	0.1867
Copper	AR(1)	2.8304	23.7528	4.8737	0.0000	1.8823	−0.3733	0.7091	−1.2894	0.1979
Copper	ARIMA	2.8315	23.7600	4.8744	0.0000	1.9920	−0.3532	0.7241	−1.2371	0.2167
Nickel	ET full specification, Model 4	2.8294	25.1166	5.0116	0.0014	1.5271	-	-	-	-
Nickel	Zero-return benchmark	2.7249	24.7430	4.9742	-	2.0000	−0.8720	0.3836	−2.9065	0.0038 ***
Nickel	Historical mean	2.7264	24.7431	4.9742	-	1.9133	−0.8768	0.3810	−2.8689	0.0043 ***
Nickel	AR(1)	2.7268	24.7435	4.9743	0.0001	1.9042	−0.8769	0.3810	−2.8571	0.0045 ***
Nickel	ARIMA	2.7417	24.8626	4.9862	0.0005	1.9927	−0.5269	0.5985	−2.3458	0.0194 **
Clean energy	ET full specification, Model 4	1.1050	2.0855	1.4441	0.0048	1.6667	-	-	-	-
Clean energy	Zero-return benchmark	1.1092	2.0871	1.4447	-	2.0000	0.0547	0.9564	0.4979	0.6188
Clean energy	Historical mean	1.1087	2.0856	1.4442	-	1.9260	0.0031	0.9975	0.4388	0.6610

Table A4. Cont.

Asset	Model	MAE	MSE	RMSE	R ²	sMAPE	DM Stat., Squared Loss	DM <i>p</i> -Value, Squared Loss	DM stat., Absolute Loss	DM <i>p</i> -Value, Absolute Loss
Clean energy	AR(1)	1.1092	2.0868	1.4446	0.0020	1.9263	0.0428	0.9659	0.4891	0.6250
Clean energy	ARIMA	1.1097	2.0882	1.4451	0.0020	2.0000	0.0941	0.9251	0.5485	0.5836

Note: This table compares the out-of-sample forecasting performance of the Extra Trees full specification, Model 4, with conventional forecasting benchmarks, including the zero-return benchmark, historical mean forecast, AR(1), and ARIMA models. Lower values of MAE, MSE, RMSE, and sMAPE indicate better forecast accuracy, whereas higher R² indicates stronger predictive performance. Diebold-Mariano statistics and *p*-values compare each benchmark model with the Extra Trees full specification. The squared-loss test is based on squared forecast errors, while the absolute-loss test is based on absolute forecast errors. *** and ** denote statistical significance at the 1% and 5% levels, respectively. A dash indicates that the Extra Trees full specification is the reference model.

Table A5. Model Confidence Set results across algorithm-specification combinations.

Asset	Loss Function	Best-Ranked Model by Average Loss	Number of Retained Models	Retained Models	Interpretation
Aluminum	Squared-error loss	XGBoost_M2	24	ET_M1; GBM_M1; XGBoost_M1; KNN_M1; LASSO_M1; EN_M1; ET_M2; GBM_M2; XGBoost_M2; KNN_M2; LASSO_M2; EN_M2; ET_M3; GBM_M3; XGBoost_M3; KNN_M3; LASSO_M3; EN_M3; ET_M4; GBM_M4; XGBoost_M4; KNN_M4; LASSO_M4; EN_M4	No model is statistically eliminated; ET is competitive but not dominant.
Aluminum	Absolute-error loss	XGBoost_M2	19	ET_M1; GBM_M1; XGBoost_M1; LASSO_M1; EN_M1; ET_M2; XGBoost_M2; LASSO_M2; EN_M2; ET_M3; GBM_M3; XGBoost_M3; LASSO_M3; EN_M3; ET_M4; GBM_M4; XGBoost_M4; LASSO_M4; EN_M4	ET specifications remain in the superior set, but several alternatives are statistically comparable.
Copper	Squared-error loss	XGBoost_M2	24	ET_M1; GBM_M1; XGBoost_M1; KNN_M1; LASSO_M1; EN_M1; ET_M2; GBM_M2; XGBoost_M2; KNN_M2; LASSO_M2; EN_M2; ET_M3; GBM_M3; XGBoost_M3; KNN_M3; LASSO_M3; EN_M3; ET_M4; GBM_M4; XGBoost_M4; KNN_M4; LASSO_M4; EN_M4	No model is statistically eliminated; ET is competitive but not dominant.
Copper	Absolute-error loss	XGBoost_M2	17	GBM_M1; LASSO_M1; EN_M1; GBM_M2; XGBoost_M2; LASSO_M2; EN_M2; ET_M3; GBM_M3; XGBoost_M3; LASSO_M3; EN_M3; ET_M4; GBM_M4; XGBoost_M4; LASSO_M4; EN_M4	ET_M3 and ET_M4 are retained, indicating statistical competitiveness of ET augmented specifications.
Nickel	Squared-error loss	LASSO_M3	20	ET_M1; GBM_M1; XGBoost_M1; LASSO_M1; EN_M1; ET_M2; GBM_M2; XGBoost_M2; LASSO_M2; EN_M2; ET_M3; GBM_M3; XGBoost_M3; LASSO_M3; EN_M3; ET_M4; GBM_M4; XGBoost_M4; LASSO_M4; EN_M4	ET specifications are retained under squared-error loss, but linear penalized models also remain competitive.
Nickel	Absolute-error loss	EN_M2	6	EN_M1; LASSO_M2; EN_M2; EN_M3; LASSO_M4; EN_M4	ET is not retained; penalized linear models dominate under absolute-error loss.
Clean energy	Squared-error loss	LASSO_M1	17	ET_M1; GBM_M1; XGBoost_M1; LASSO_M1; EN_M1; ET_M2; XGBoost_M2; LASSO_M2; EN_M2; ET_M3; GBM_M3; LASSO_M3; EN_M3; ET_M4; GBM_M4; LASSO_M4; EN_M4	ET is retained, but baseline and penalized specifications are also statistically comparable.

Table A5. *Cont.*

Asset	Loss Function	Best-Ranked Model by Average Loss	Number of Retained Models	Retained Models	Interpretation
Clean energy	Absolute-error loss	LASSO_M1	17	ET_M1; GBM_M1; XGBoost_M1; LASSO_M1; EN_M1; ET_M2; LASSO_M2; EN_M2; ET_M3; GBM_M3; XGBoost_M3; LASSO_M3; EN_M3; ET_M4; GBM_M4; LASSO_M4; EN_M4	ET is retained, but simpler specifications remain statistically comparable.

Note: This table reports Model Confidence Set results following [31]. The candidate set includes 24 algorithm-specification combinations for each asset: six algorithms, namely, Extra Trees (ET), gradient boosting machine (GBM), XGBoost, K-nearest neighbors (KNN), LASSO, and Elastic Net (EN), estimated under four specifications. Model 1 includes the baseline controls; Model 2 adds supply-chain stress; Model 3 adds geopolitical risk; and Model 4 includes both supply-chain stress and geopolitical risk. The MCS is computed separately for each asset using squared-error loss and absolute-error loss. Retained models are statistically indistinguishable from the best-ranked model at the selected confidence level. A larger retained set indicates weaker statistical separation among competing forecasting models.

Table A6. Definition of macro-financial regimes used in the regime-specific SHAP analysis.

Regime	Period	Economic Interpretation
Pre-COVID-19 period	10 February 2015–28 February 2020	Normal pre-pandemic market environment before the global COVID-19 shock
COVID-19 and logistics-stress period	1 March 2020–31 December 2021	Pandemic shock, supply-chain disruption, and global logistics stress
Russia-Ukraine and monetary-tightening period	24 February 2022–31 December 2023	Geopolitical shock following the Russia-Ukraine war and the global monetary-tightening cycle
Recent period	1 January 2024–27 February 2026	Post-shock adjustment period with continuing geopolitical, energy market, and financial uncertainty

Note: The regimes are defined using major global macro-financial and geopolitical events over the sample period. The regime-specific SHAP analysis re-estimates the Extra Trees full specification within each sub-period where sufficient observations are available and computes mean absolute SHAP values to identify the most influential predictors in each regime.

Table A7. Top SHAP predictors by asset and macro-financial regime.

Asset	Regime	Period	Top Predictor	Second Predictor	Third Predictor
ALU	Pre-COVID-19	10 February 2015 to 28 February 2020	GPR (0.2454)	EPU (0.2190)	VIX (0.1978)
ALU	COVID-19 and logistics stress	1 March 2020 to 31 December 2021	VIX (0.6774)	EPU (0.5390)	GPR (0.5208)
ALU	Russia-Ukraine and monetary tightening	24 February 2022 to 31 December 2023	SPX (0.5570)	USD (0.5083)	VIX (0.3776)
ALU	Recent period	1 January 2024 to 27 February 2026	OIL (0.2807)	SPX (0.2357)	EPU (0.2232)
COP	Pre-COVID-19	10 February 2015 to 28 February 2020	EPU (0.3151)	OIL (0.2805)	USD (0.2586)
COP	COVID-19 and logistics stress	1 March 2020 to 31 December 2021	VIX (0.6520)	OIL (0.5216)	SPX (0.5214)
COP	Russia-Ukraine and monetary tightening	24 February 2022 to 31 December 2023	SPX (0.4893)	USD (0.3422)	VIX (0.2818)
COP	Recent period	1 January 2024 to 27 February 2026	VIX (0.3052)	SPX (0.2571)	OIL (0.2131)
NIC	Pre-COVID-19	10 February 2015 to 28 February 2020	EPU (0.4023)	USD (0.3991)	OIL (0.3403)
NIC	COVID-19 and logistics stress	1 March 2020 to 31 December 2021	EPU (0.5515)	VIX (0.5236)	SCS (0.4427)
NIC	Russia-Ukraine and monetary tightening	24 February 2022 to 31 December 2023	SPX (1.5246)	EPU (1.0750)	OIL (0.6801)
NIC	Recent period	1 January 2024 to 27 February 2026	OIL (0.2374)	EPU (0.2356)	SCS (0.1746)

Table A7. Cont.

Asset	Regime	Period	Top Predictor	Second Predictor	Third Predictor
CEN	Pre-COVID-19	10 February 2015 to 28 February 2020	SPX (0.0838)	VIX (0.0695)	EPU (0.0572)
CEN	COVID-19 and logistics stress	1 March 2020 to 31 December 2021	EPU (0.2216)	USD (0.2030)	GPR (0.1554)
CEN	Russia-Ukraine and monetary tightening	24 February 2022 to 31 December 2023	USD (0.1890)	VIX (0.1088)	SPX (0.1034)
CEN	Recent period	1 January 2024 to 27 February 2026	GPR (0.0863)	VIX (0.0763)	OIL (0.0652)

Note: This table reports the three most important predictors in each asset-regime pair based on mean absolute SHAP values from the Extra Trees full specification. The full specification includes supply-chain stress (SCS), geopolitical risk (GPR), VIX, EPU, SPX, OIL, and USD. Higher mean absolute SHAP values indicate a stronger average contribution to the model's forecasts within a given regime. ALU, COP, NIC, and CEN denote aluminum, copper, nickel, and clean energy returns, respectively. The regimes are defined in Appendix A Table A6.

References

1. Korkusuz, B. Beyond the S&P 500: Examining the role of external volatilities in market forecasting. *Rev. Econ. Des.* **2025**, *29*, 767–794. [\[CrossRef\]](#)
2. Nguyen, B.N. A sentiment-based approach to predict energy price volatility using distilRoBERTa and GARCH models. *Energy Econ.* **2025**, *149*, 108646. [\[CrossRef\]](#)
3. Basher, S.A.; Sadorsky, P. How important are climate change risks for predicting clean energy stock prices? Evidence from machine learning predictive modeling and interpretation. *J. Clim. Financ.* **2025**, *10*, 100058. [\[CrossRef\]](#)
4. Chen, W.; Chen, Z.; Luo, Q. Predicting volatility in China's clean energy sector: Advantages of the carbon transition risk. *Financ. Res. Lett.* **2025**, *72*, 106534. [\[CrossRef\]](#)
5. Soltani, H.; Taleb, J.; Ben Hamadou, F.; Boujelbène, M. Using machine learning to forecast clean energy, commodities, green bonds and ESG index prices: How important is financial stress? *EuroMed J. Bus.* **2024**, *21*, 318–344. [\[CrossRef\]](#)
6. Liu, Z.; Wang, Y.; Yuan, X.; Ding, Z.; Ji, Q. Geopolitical risk and vulnerability of energy markets. *Energy Econ.* **2025**, *141*, 108055. [\[CrossRef\]](#)
7. Chowdhury, M.A.F.; Abdullah, M.; Abakah, E.J.A.; Tiwari, A.K. Geopolitical risk and energy market tail risk forecasting: An explainable machine learning approach. *J. Commod. Mark.* **2025**, *39*, 100478. [\[CrossRef\]](#)
8. Shang, M.; Zhang, L.; Duan, H.; Wang, L.; Xiao, N. Trade policy uncertainty, shipping risk, and commodity markets. *Financ. Res. Lett.* **2025**, *73*, 106604. [\[CrossRef\]](#)
9. Elsayed, A.H.; Gozgor, G.; Khalfaoui, R.; Tarchella, S. Impact of supply chain pressure on traditional energy and metal markets: A wavelet-based quantile-on-quantile perspective. *J. Commod. Mark.* **2025**, *38*, 100472. [\[CrossRef\]](#)
10. Aizenman, J.; Lindahl, R.; Stenvall, D.; Uddin, G.S. Geopolitical shocks and commodity market dynamics: New evidence from the Russia-Ukraine conflict. *Eur. J. Political Econ.* **2024**, *85*, 102574. [\[CrossRef\]](#)
11. Ben Jabeur, S.; Bakkar, Y.; Cepni, O. Do global COVOL and geopolitical risks affect clean energy prices? Evidence from explainable artificial intelligence models. *Energy Econ.* **2025**, *141*, 108112. [\[CrossRef\]](#)
12. Almeida, D.; Ferreira, P.; Dionísio, A.; Aslam, F. Exploring the connection between geopolitical risks and energy markets. *Energy Econ.* **2025**, *141*, 108113. [\[CrossRef\]](#)
13. Yang, S.; Fu, Y. Interconnectedness among supply chain disruptions, energy crisis, and oil market volatility on economic resilience. *Energy Econ.* **2025**, *143*, 108290. [\[CrossRef\]](#)
14. Zhou, Y.; Li, B.; Zhao, M. Asymmetric association between supply chain bottlenecks and consumer energy prices: Evidence from quantile-on-quantile approach. *Energy Econ.* **2025**, *147*, 108556. [\[CrossRef\]](#)
15. Doroshenko, L.; De Crescenzo, I.; Mastroeni, L.; Mazzocchi, A. Geopolitical risks, critical materials and energy transition: Insights from wavelet analysis. *Resour. Policy* **2025**, *108*, 105666. [\[CrossRef\]](#)
16. Khan, N.; Mejri, S.; Hammoudeh, S. How do global commodities react to increasing geopolitical risks? New insights into the Russia-Ukraine and Palestine-Israel conflicts. *Energy Econ.* **2024**, *138*, 107812. [\[CrossRef\]](#)
17. Hao, X.; Ma, Y.; Pan, D. Geopolitical risk and the predictability of spillovers between exchange, commodity and stock markets. *J. Multinatl. Financ. Manag.* **2024**, *73*, 100843. [\[CrossRef\]](#)
18. Niu, Z.; Wang, C.; Zhang, H. Forecasting stock market volatility with various geopolitical risks categories: New evidence from machine learning models. *Int. Rev. Financ. Anal.* **2023**, *89*, 102738. [\[CrossRef\]](#)
19. Ge, L.; Huang, Q.; Zhu, F.; Chen, S. Advanced time series forecasting for commodities: Insights from the FEDformer model. *Energy Econ.* **2025**, *147*, 108513. [\[CrossRef\]](#)

20. Abid, A.; Souissi, N.; Béjaoui, A.; Jeribi, A. Predicting commodity prices amidst Russia-Ukraine war: Some evidences from artificial intelligence models. *Financ. Res. Open* **2025**, *1*, 100023. [[CrossRef](#)]
21. Mao, Y.; Yu, X.; Zhu, J.; Wang, F. Forecasting crude oil futures price uncertainty: An explainable deep learning framework integrating news sentiment. *Expert Syst. Appl.* **2025**, *297*, 129419. [[CrossRef](#)]
22. Gu, S.; Kelly, B.; Xiu, D. Empirical asset pricing via machine learning. *Rev. Financ. Stud.* **2020**, *33*, 2223–2273. [[CrossRef](#)]
23. Su, Y.; Feng, X. Machine learning approaches to predicting energy price correlation: From a responsible AI perspective. *Technol. Forecast. Soc. Change* **2026**, *225*, 124515. [[CrossRef](#)]
24. Zhao, Z.; Wang, M.; Wei, J.; Cen, X.; Du, S.; Wu, Z.; Liu, H.; Wang, W. Interpretable Machine Learning for Multi-Energy Supply Station Revenue Forecasting: A SHAP-Driven Framework to Accelerate Urban Carbon Neutrality. *Energies* **2025**, *18*, 1624. [[CrossRef](#)]
25. Li, Y.; Yang, N.; Bi, G.; Chen, S.; Luo, Z.; Shen, X. Carbon price forecasting using a hybrid deep learning model: TKMixer-BiGRU-SA. *Symmetry* **2025**, *17*, 962. [[CrossRef](#)]
26. Yang, X.; Chen, H.; Li, X.; Liu, X. Carbon Price Forecasting via a CNN-BiLSTM Model Integrating VMD and Classified News Sentiment. *Big Data Cogn. Comput.* **2026**, *10*, 82. [[CrossRef](#)]
27. Choi, I.; Kim, W.C. Practical forecasting of risk boundaries for industrial metals and critical minerals via statistical machine learning techniques. *Int. Rev. Financ. Anal.* **2024**, *94*, 103252. [[CrossRef](#)]
28. Baker, S.R.; Bloom, N.; Davis, S.J. Measuring economic policy uncertainty. *Q. J. Econ.* **2016**, *131*, 1593–1636. [[CrossRef](#)]
29. Caldara, D.; Iacoviello, M. Measuring geopolitical risk. *Am. Econ. Rev.* **2022**, *112*, 1194–1225. [[CrossRef](#)]
30. Diebold, F.X.; Mariano, R.S. Comparing predictive accuracy. *J. Bus. Econ. Stat.* **1995**, *13*, 253–263. [[CrossRef](#)]
31. Hansen, P.R.; Lunde, A.; Nason, J.M. The model confidence set. *Econometrica* **2011**, *79*, 453–497. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.