



SCHOOL OF  
ECONOMICS AND  
MANAGEMENT

# Feature Selection for Rare Earth Element Price Forecasting

A Hybrid Econometric and Machine Learning Approach

by

Jonathan Evaldsson

&

Jóhann Ari Sigfússon

Department of Economics

Lund University School of Economics and Management

**Supervisor:** Luca Margaritella

DABN01 – Master Thesis in Data Analytics and Business Economics, 15 ECTS

## Abstract

This thesis addresses the need for robust price forecasting in the volatile Rare Earth Elements (REEs) markets. With a specific focus on neodymium, given it is the most utilized REEs, we investigate how feature selection can be effectively performed to guide REE price forecasting by proposing and evaluating three distinct methodologies: Adaptive Lasso, Post-Double Selection (PDS), and an integrated approach combining PDS with Granger Causality testing. After applying these feature selection methods, a Random Forest is then employed in a rolling window forecasting strategy with dynamic selection. Our empirical analysis reveals that no single feature selection method consistently optimizes forecasting accuracy across all horizons. While no feature selection method demonstrates superiority in every forecasting horizon, the PDS with Granger Causality approach proves most effective for short-term forecasting, suggesting that statistical inference can enhance short-term accuracy. Conversely, the PDS method generally underperforms due to its tendency to retain an overly dense set of predictors. The study highlights the inherent challenges in forecasting REE prices, particularly due to frequent periods of high volatility. This research contributes valuable insights into developing more accurate and reliable REE price prediction tools, emphasizing the importance of tailoring feature selection strategies to specific forecasting horizons.

**Keywords:** Rare Earth Elements, Price Forecasting, Feature Selection, Adaptive Lasso, Post-Double Selection, Granger Causality, Lag Augmentation

## **Acknowledgements**

We extend our sincere gratitude to Luca Margaritella, our supervisor, for his invaluable guidance, generosity with his time, and insightful feedback throughout the course of this thesis. His profound knowledge and dedication were instrumental in shaping our research direction and overcoming challenges.

# CONTENTS

|          |                                                                         |           |
|----------|-------------------------------------------------------------------------|-----------|
| <b>1</b> | <b>INTRODUCTION</b>                                                     | <b>1</b>  |
| 1.1      | Market Development . . . . .                                            | 1         |
| 1.2      | Transaction Dynamics and Price Risk in the REE Market . . . . .         | 3         |
| 1.3      | Research Question and Thesis Structure . . . . .                        | 4         |
| <b>2</b> | <b>LITERATURE REVIEW</b>                                                | <b>5</b>  |
| 2.1      | Forecasting rare earth elements . . . . .                               | 5         |
| 2.2      | Forecasting commodities . . . . .                                       | 6         |
| 2.3      | Research Gap and Contribution . . . . .                                 | 8         |
| <b>3</b> | <b>DATA</b>                                                             | <b>9</b>  |
| 3.1      | Data Acquisition . . . . .                                              | 9         |
| 3.2      | Data Handling . . . . .                                                 | 10        |
| <b>4</b> | <b>METHODOLOGY</b>                                                      | <b>12</b> |
| 4.1      | Correlated predictors . . . . .                                         | 12        |
| 4.2      | Applying Adaptive Lasso for feature selection . . . . .                 | 12        |
| 4.3      | Post-double selection . . . . .                                         | 15        |
| 4.4      | Granger causality . . . . .                                             | 16        |
| 4.5      | Forecasting model - Random Forest . . . . .                             | 19        |
| 4.6      | Forecasting method and model evaluation . . . . .                       | 20        |
| 4.7      | Dynamic feature selection . . . . .                                     | 22        |
| <b>5</b> | <b>EMPIRICAL ANALYSIS</b>                                               | <b>23</b> |
| 5.1      | Exploratory Analysis . . . . .                                          | 23        |
| 5.2      | Model Selection . . . . .                                               | 25        |
| 5.3      | Random Forest Application and Evaluation . . . . .                      | 27        |
| 5.4      | Model 1: One-Shot Adaptive Lasso . . . . .                              | 28        |
| 5.5      | Model 2: Post Double Selection . . . . .                                | 29        |
| 5.6      | Model 3: Post Double Selection with Granger Causality testing . . . . . | 30        |
| 5.7      | Comparison of the models . . . . .                                      | 31        |
| <b>6</b> | <b>CONCLUSION</b>                                                       | <b>34</b> |

|                                             |            |
|---------------------------------------------|------------|
| 6.1 Limitations & Future Research . . . . . | 35         |
| <b>References</b>                           | <b>35</b>  |
| <b>Appendices</b>                           | <b>i</b>   |
| <b>A Tables</b>                             | <b>i</b>   |
| <b>B Figures</b>                            | <b>iii</b> |
| <b>C Generative AI Usage</b>                | <b>vi</b>  |

# List of Figures

|    |                                                                      |     |
|----|----------------------------------------------------------------------|-----|
| 1  | Rare Earth Elements Production (1990 to 2024) . . . . .              | 2   |
| 2  | Data Split Visualised . . . . .                                      | 21  |
| 3  | Line Plot of Neodymium Price per Metric Ton (2005–2024) . . . . .    | 24  |
| 4  | Selection Frequencies Across Models and Forecast Stages . . . . .    | 27  |
| 5  | Forecast Comparison of Neodymium Prices ( $h = 1, 3, 12$ ) . . . . . | 33  |
| B1 | Rolling Evaluation Framework Across Forecast Horizons . . . . .      | iii |
| B2 | Correlation Matrix of Model Variables . . . . .                      | iii |
| B3 | Time Series Plots of all Variables (2005-2024) . . . . .             | iv  |
| B4 | Granger Test Outcomes Across Forecast Horizons . . . . .             | v   |

## List of Tables

|    |                                                                            |    |
|----|----------------------------------------------------------------------------|----|
| 1  | Variables . . . . .                                                        | 10 |
| 2  | Descriptive statistics for Neodymium . . . . .                             | 25 |
| 3  | Model 1: One-Shot Alasso Selection . . . . .                               | 28 |
| 4  | Model 2: PDS Selection . . . . .                                           | 29 |
| 5  | Model 3: PDS-GC Selection . . . . .                                        | 30 |
| A1 | The 17 rare-earth elements (REEs) and their principal applications . . . . | i  |
| A2 | Grid search results for Random Forest models . . . . .                     | ii |
| A3 | Grid Search Hyperparameter Values for Random Forest . . . . .              | ii |

# 1 INTRODUCTION

Rare earth elements (REEs) are a group of 17 chemically similar metals that are critical to modern technologies because of their unique magnetic, luminescent, and catalytic properties. Despite their name, REEs are relatively abundant in the Earth’s crust, yet occur in dispersed minerals that are costly to extract and process (Goodenough et al., 2018). During the past two decades, the global REE market has experienced dramatic price swings, most notably the 2010-2011 REE crisis, when China halved its export quotas, causing a dramatic price surge for REEs (Chen et al., 2024). This inherent price instability, coupled with the strategic importance of REEs, underscores a critical need for robust price forecasting. Such foresight is vital for informed investment decisions, effective policy formulation, and resilient supply chain management. This chapter provides an overview of the Rare Earth Element market’s structure and highlights the compelling need for robust price forecasting to inform investment decisions, policy formulation, and resilient supply chain management.

REEs are embedded in virtually every flagship green-energy technology. Neodymium and praseodymium are used to create the high-performance permanent magnets that drive the electric vehicle motors and the light-weight generators used in offshore wind turbines. As shown in Table A1 in Appendix A, the REEs serve a wide range of high-tech markets. The International Energy Agency (IEA) calculates that the demand for REE used in the manufacture of permanent magnets alone could increase threefold under current policies and could increase more than sevenfold in a net zero policy path by 2040 (IEA, 2021). Other rare earth elements such as europium, terbium, and yttrium are crucial for the functionality of high-efficiency LEDs and flat-panel displays, whereas erbium is essential for long-distance fibre-optic cables. It is clear that rare earth elements (REEs) are of significant importance for the advancement of green energy technologies and digital transformation (U.S. Geological Survey, 2025).

## 1.1 Market Development

The global REE market is defined by a densely concentrated supply network and opaque trading mechanisms. Primary production comes from the mining of a handful of different deposit types, namely certain calcium-rich rocks known as carbonatites, unique volcanic ore bodies called peralkaline igneous, and ion-adsorption clay deposits where metals cling to the soil (Goodenough et al., 2018). For much of the past three decades, China has dominated global REE sourcing and processing, its dominance in the rare-earth market is due to a combination of geological endowment, cost advantages, and state-led policy. China has a vast number of ion-absorption clay deposits, but the leading factor in its

dominance stems from its perfection of low-cost hydrometallurgical leaching techniques that could strip REEs from weathered soils with simple reagents and minimal energy input. Its strengths in natural resources and processing were further supported by public policy and inexpensive labour (Depraite & Goutte, 2023).

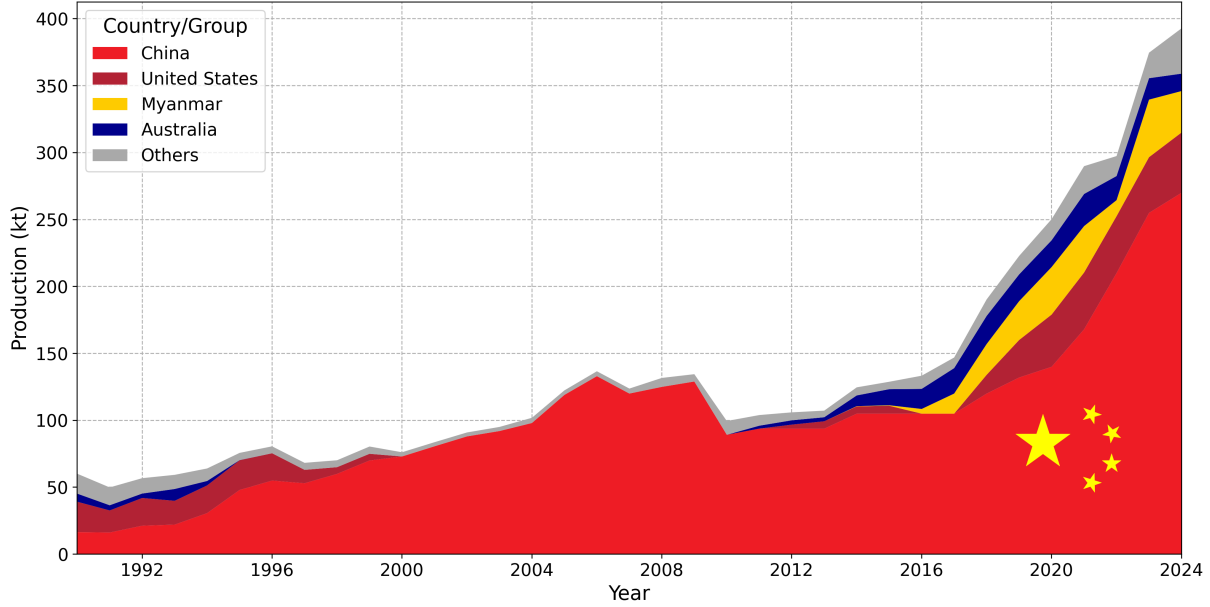


Figure 1: Global Rare Earth Elements Production by Country Between 1990 and 2024 in Rare Earth Oxide Equivalent Kilotons. *Data source:* U.S. Geological Survey

By 2005, China was responsible for more than 90% of the global supply of rare-earth oxides (REOs), which are the end products obtained from processing deposits rich in rare earth elements. The 2010-2011 "rare-earth crisis" exposed the risks inherent in such concentration. In the wake of a maritime dispute with Japan, China abruptly halved its export quotas and effectively embargoed shipment to key customers, triggering a supply squeeze that led to an 11-fold increase in prices for the majority of REOs. The crisis highlighted the inherent risks associated with such concentration in the market, and sparked a diverse set of responses around the world. In the United States, the long-dormant Mountain Pass mine was acquired by Molycorp in 2012, refurbished and brought back into production by 2017. Australia moved in parallel with the Mt Weld deposit and the Browns Range project. Canada and Europe launched their own upstream campaign, and today Sweden is readying its Per Geijer deposit for development, which is set to become Europe's first modern REE mine. Meanwhile, a global exploration boom has taken hold: with potential projects beginning in Africa, Southeast Asia, and South America (Depraite & Goutte, 2023). Collectively, these moves represent the serious effort that has taken place to dilute China's dominance in the REO market. Governments are also backing this momentum with new industrial policies, Europe's Critical Raw Materials Act, the U.S. Inflation Reduction Act, Australia's Critical Minerals Strategy and Canada's Crit-

ical Minerals Strategy all aim to secure supply chains for green technologies to mitigate single-country dominance (IEA, 2021). Since the 2010-2011 REE crisis, China’s dominance in mining production has weakened somewhat, but despite international efforts to diversify sources, China continues to play a central role, accounting for about 60% of global rare earths mining production, down from a previous high of 92% in 2010. Despite the shift in mining production, China’s strength in processing and refinement remains dominant at 90% as of 2021 due to its advanced hydrometallurgical refinement techniques (Goodenough et al., 2018; IEA, 2021). The diversification of rare-earth mine production has seen China transition from primarily an exporter to becoming increasingly reliant on REE imports from other nations, notably the United States, to feed its extensive refining capacity. Although growing interdependence in the REE sector has contributed to some stabilisation of global supply chains (Chen et al., 2024), this equilibrium is frequently tested by geopolitical dynamics. Recent trade tensions, for example, have seen China reportedly leverage its dominant market position once more. Specifically, China’s decision in April 2025 to implement export controls on select rare earth materials in response to U.S. tariffs serves as a stark reminder that REEs continue to be a significant geopolitical instrument (Jackson et al., 2025). Although the full market impact of these latest export limitations remains to be seen, such policy-driven interventions underscore the critical need for forecasting methodologies that can effectively account for these abrupt structural changes and the inherent geopolitical volatility within the REE market.

### *1.2 Transaction Dynamics and Price Risk in the REE Market*

The rare earth elements market is characterised by unique transaction mechanisms and significant barriers to traditional financial hedging. Primarily, REEs are traded through opaque bilaterally negotiated Over-the-Counter (OTC) contracts between producers, processors, and end-users, where terms are set privately (Depraeter & Goutte, 2023). This is in contrast to many other commodities that are traded on transparent exchanges, where firms can hedge themselves against price risk. However, emerging exchanges such as the Baotou Rare Earth Exchange (BREX) and the Ganzhou Rare Earth Exchange are expanding their presence within the Chinese market. Recently, China Northern Rare Earth, a major producer, announced its plans to shift from its own OTC list prices to BREX’s real-time quotes for long-term contracts starting January 2025, and a ”Baotou Index” based on actual transactions is anticipated around August 2026 (Shanghai Metal Market, 2025). While these new exchanges signal a potential shift towards greater transparency, the market still largely relies on opaque OTC dealings, and effective financial hedging instruments remain scarce. In this environment, market participants have limited tools for immediate price risk mitigation, thereby heightening the importance of robust forecasting methods to anticipate future price trajectories and inform strategic decision-making.

In summary, the rare-earth elements market is defined by its strategic importance to modern technologies and the green energy transition, yet remains characterised by a complex global structure and persistent vulnerabilities. The market is highly concentrated in terms of production and processing, which makes it susceptible to geopolitical disruptions and dramatic price swings, as historically demonstrated by events like the 2010-2011 REE crisis. Despite ongoing efforts to diversify supply chains and establish exchanges, structural risks continue to underscore the need for robust forecasting methods. Consequently, accurate and adaptable price forecasting models are not only beneficial but critical for supporting informed investment, policy, and supply chain decisions.

### *1.3 Research Question and Thesis Structure*

The research presented in this thesis is specifically driven by this imperative: to develop forecasting methods capable of addressing the unique challenges facing the REE sector. Addressing this, our main research question is: How can feature selection be performed effectively to guide the forecasting of rare earth element prices? In this study, we will concentrate on neodymium as our response variable, given its prominent role as the most extensively utilized REE. Our work proposes and evaluates three different feature selection methods that build on each other, including Adaptive Lasso, Post-Double Selection (PDS), and the integration of Granger Causality testing with lag augmentation to handle potentially non-stationary data. These selected features are then utilised within a Random Forest framework, known for its ability to capture complex, non-linear relationships, to generate multiple steps ahead forecasts. By systematically comparing these distinct feature selection approaches, this thesis aims to enhance the accuracy and reliability of REE price predictions.

The rest of the thesis is structured as follows: [Section 2](#) summarises the previous literature. [Section 3](#) presents the data acquisition and handling. [Section 4](#) details the empirical strategy. [Section 5](#) presents the empirical analysis and results, including their discussion. Finally, [section 6](#) concludes the thesis.

## 2 LITERATURE REVIEW

This chapter provides a comprehensive review of the academic literature relevant to Rare Earth Element (REE) price forecasting and, more broadly, commodity price prediction. Given the critical importance of REEs in modern technologies and the inherent volatility of their market, effective forecasting methodologies are paramount. This review explores existing research on REE-specific price prediction and examines general forecasting techniques for commodities. The insights derived will highlight current research frontiers and identify specific gaps that this thesis aims to address.

### *2.1 Forecasting rare earth elements*

Over the last 10 years, research on REE price forecasting has evolved from basic time-series models towards more specialised and hybrid approaches. Univariate models have been and remain in the foreground as the method of choice. Early works often employed ARIMA-family models with some data pre-processing. For instance, Riesgo García et al. (2018) developed a novel “transgenic time series” approach to represent and forecast REE prices. This approach entailed silencing certain genes (volatility spikes). They applied this method to short-term forecasts of several REE prices and they outperformed baseline ARIMA models.

Proelss et al. (2020) examined REE price volatility using short-memory ARMA, long-memory ARFIMA, and GARCH models, finding strong evidence of long memory in REE returns. Their simple ARFIMA(0, d, 0) model outperformed GARCH in out-of-sample volatility forecasts. This result indicates that REE markets may exhibit persistent volatility clusters and mean-reversion that standard short-memory models miss.

In recent years, the incorporation of machine learning (ML) and hybrid methods into commodity forecasting has influenced REE studies as well, especially given the potentially non-linear drivers of REE prices (e.g., technology demand, geopolitical tensions). A notable example utilising ML methods to predict rare earth equity prices is that of Henriques and Sadorsky (2023), who applied a suite of ML algorithms to predict the daily price direction of rare earth stock indices. Their results showed that non-linear models like random forests, Extra Trees ensembles, recurrent neural networks (RNNs), and support vector machines (SVMs) achieved higher prediction accuracy than linear benchmarks.

Guo and Mai (2024) explored the impact that seasonality, significant events, finance, and policies have on China’s rare earth export prices and how to accurately forecast them. Their approach was to divide the metals into two groups, light metals and heavy metals, to see if they respond differently to external changes. This is because heavy REE is

used in more cutting-edge applications compared to light REE, such as optical fibers, and thus has greater susceptibility to extreme events. Light REEs, on the other hand, are more commonly used in traditional industries, and are therefore more susceptible to the influence of business cycles, and thus exhibit a seasonality pattern.

To rank variable importance, the Temporal Fusion Transformers (TFT) attention mechanism was utilised. Their variables were divided into two groups: prior time-varying variables and posterior time-varying variables. The encoder part of the TFT model contained the importance of variables in the past, while the decoder reflects the importance of variables in future predictions; hence, only prior time-varying variables are included in the decoder variable importance. For light REEs, China's quarterly GDP has the most profound impact on the export price during their investigated time period. Another significant impact on the light REEs' price is the proportion of the world's REE reserves that China controls. As more deposits outside of China are exploited, China's influence on the REE price will decrease.

Their influence analysis also demonstrates the impact of the Shanghai and Shenzhen 300 Index on the price, highlighting a clear connection between financial markets and commodity prices, as also shown in the paper by Batten et al. (2010). The importance of financial markets is also highlighted by the influence that the commodity price index (CGPI) and the US dollar index have on the prices of the heavy metal REEs. For the prediction part of the paper, the TFT model was used along with SVR models and the traditional time series model ARIMA. The evaluation was based on six common accuracy indicators; among the models, the TFT model outperformed all other models on all indicators.

## *2.2 Forecasting commodities*

Gargano and Timmermann (2014) investigated the predictability of commodity spot prices for both in-sample and out-of-sample forecasts. The authors used 16 variables as predictors, seven of which are commonly used in the literature to predict stock returns, e.g., the dividend-price ratio, treasury bill rate, and inflation. The remaining variables were chosen to measure the broad state of the economy, such as industrial production growth and the log growth in the monthly M1 money stock. The paper examined the mean squared forecast error of both univariate and multivariate models. For the multivariate models, the authors considered shrinkage methods such as ridge regression and subset combinations, both designed to reduce the effects of parameter estimation error on the forecast.

They achieved the highest predictability for commodity prices based on out-of-sample

forecasts at the quarterly horizon. Another conclusion is that the predictability of commodity prices is closely related to the economic cycle. In particular, during recessions, commodity prices are most predictable.

Regarding whether macroeconomic factors influence the price of commodities, Batten et al. (2010) investigated the macroeconomic determinants of volatility in precious metals markets. They modelled the monthly price volatilities of four precious metals (gold, silver, platinum, and palladium) and investigated the effects of macroeconomic determinants (business cycle, monetary environment, and financial market sentiment). Among their findings was that gold volatility responds only to monetary variables, a discrepancy from the other metals. For example, the volatility of platinum and palladium varies depending on the S&P 500 and World indices, indicating that these metals are more likely to act as financial market instruments than gold. In contrast, neither monetary nor financial market variables had a significant effect on the volatility of silver. However, they found significant volatility spillover between the precious metals markets.

Another main metal on the commodity market is copper. Sánchez Lasheras et al. (2015) examined the applicability of forecasting the copper spot price from the New York Commodity Exchange (COMEX) using neural networks and ARIMA models. The parameter selection of the ARIMA model was based on the Hyndman-Khandakar algorithm that resulted in an ARIMA (1,1,0) model. The algorithm tests for unit roots and minimisation of AIC and MLE to obtain the optimal parameters for the model. The choice of neural network model was a multilayer perceptron (MLP), which is a feedforward artificial neural network. The architecture contains 16 input layers, corresponding to the closing spot prices in the 16 previous working days. Hence, both models are univariate models, but with a different number of lags selected. Their findings were that the MLP model outperformed the ARIMA model when comparing the RMSE values. Despite this, the authors emphasised that the ARIMA model is sufficient for forecasting, especially if the computational time and the black-box perspective of neural networks are problematic for the user.

As with copper, gold and mineral commodities prices overall are essential to make mineral investment decisions. Madziwa et al. (2022) paper tested univariate models as ARIMA and stochastic geometric Brownian motion to forecast gold prices. The authors also applied the multivariate Autoregressive Distributed Lag (ARDL) model. Their results found that the ARDL model outperformed the other models based on all evaluation criteria. The authors highlighted the fact that the ARDL model was able to capture the long-term relationship between gold prices, gold demand and interest rates. Since long-term forecasting is important for mine planning the multivariate ARDL model is advantageous compared to the univariate models.

Many researchers have also pursued hybrid models that combine traditional time-series methods with machine learning. Ji et al. (2019) used an ARIMA-CNN-LSTM model to forecast carbon prices where the ARIMA component first models the linear trend and seasonality, and the CNN-LSTM component then models the non-linear residual structure. They achieved better results with the hybrid model than with each counterpart.

### *2.3 Research Gap and Contribution*

The preceding sections have summarised the evolving landscape of REE price forecasting, from traditional time-series models to hybrid and machine learning approaches. Despite this progress and the unique characteristics of REE markets, particularly their geopolitical and economic significance, a clear research gap remains. Existing literature lacks an explicit and systematic focus on advanced feature selection methods specifically tailored for REE price forecasting across multiple horizons.

This thesis directly addresses this gap. Our work contributes to the existing literature by proposing and systematically evaluating advanced feature selection methodologies, including Adaptive Lasso, Post-Double Selection (PDS), and the integration of Granger Causality testing, within a Random Forest framework to enhance REE price predictability. This research aims to provide valuable insights into developing more accurate and reliable forecasting tools, thereby laying the groundwork for answering our main research question.

### 3 DATA

#### 3.1 Data Acquisition

The bulk of our macro and financial series came from Bloomberg Terminal. Rare-earth oxide spot prices (99% purity, CNY per metric ton) were obtained from Asian Metal database via Bloomberg Terminal. The data for China’s REE production, as well as China’s production share, is sourced from U.S. Geological Survey ([n.d.](#)), and the data for China’s imports and exports of REEs is sourced from United Nations, Statistics Division ([2025](#)). All Bloomberg-sourced series report the final traded price at each calendar month close, which may not capture intra-month price dynamics. Nevertheless, using end-of-month prices is standard in monthly-frequency analyses and all of the financial pricing variables have the same reporting.

Our analysis includes thirty-five variables spanning several domains and are listed in Table 1. The first set of variables comprises monthly closing spot prices for eight rare-earth elements out of the seventeen known rare-earth elements. These eight elements were selected because their records cover the full January 2005 to December 2024 period, whereas the nine remaining elements were either not quoted in the Asia Metal database or were only quoted beginning in 2011 and were therefore excluded rather than imputed. Specifically, elements like Scandium were not quoted at all, while others such as Holmium and Lutetium were only quoted starting in 2011. The second set consists of major global equity indices such as the S&P 500, Hang Seng Index, Nikkei 225 and more. A third set consists of macroeconomic indicators for both China and the United States, including consumer price indices, monetary aggregates, and real GDP growth rates. A fourth set features commodity prices for precious metals and crude oil to reflect broader market trends. The final set comprises annual figures for China’s rare earth imports and exports.

The selection of features in this paper is based on prior research on forecasting REE prices as well as commodity prices more broadly. For instance, Guo and Mai ([2024](#)) incorporated a range of financial and supply-and-demand factors to assess their influence on REE export prices. Similarly, Gargano and Timmermann ([2014](#)) and Batten et al. ([2010](#)) used macroeconomic and financial variables in their forecasting models for commodities and precious metals, respectively. These categories of variables were chosen based on established literature showing their influence on commodity markets; financial indices reflect global economic sentiment and investment flows, macroeconomic indicators capture broader supply and demand dynamics, and other commodity prices serve as proxies for general market trends and input costs.

Table 1: Variables

| Variable Name                       | Ticker    | Frequency | Unit        | Data Coverage       |
|-------------------------------------|-----------|-----------|-------------|---------------------|
| Neodymium                           |           | Monthly   | CNY/mt      | Jan 2005 – Dec 2024 |
| Cerium                              |           | Monthly   | CNY/mt      | Jan 2005 – Dec 2024 |
| Dysprosium                          |           | Monthly   | CNY/mt      | Jan 2005 – Dec 2024 |
| Europium                            |           | Monthly   | CNY/mt      | Jan 2005 – Dec 2024 |
| Lanthanum                           |           | Monthly   | CNY/mt      | Jan 2005 – Dec 2024 |
| Praseodymium                        |           | Monthly   | CNY/mt      | Jan 2005 – Dec 2024 |
| Terbium                             |           | Monthly   | CNY/mt      | Jan 2005 – Dec 2024 |
| Yttrium                             |           | Monthly   | CNY/mt      | Jan 2005 – Dec 2024 |
| S&P 500                             | SPX Index | Monthly   | USD         | Jan 2005 – Dec 2024 |
| NASDAQ Composite Index              | CCMP      | Monthly   | USD         | Jan 2005 – Dec 2024 |
| The Hang Seng Index                 | HSI       | Monthly   | HKD         | Jan 2005 – Dec 2024 |
| The Nikkei-225 Stock Average        | NKY       | Monthly   | JPY         | Jan 2005 – Dec 2024 |
| U.S. 10 Year Treasury               | USGG10YR  | Monthly   | USD         | Jan 2005 – Dec 2024 |
| USA M1 Money Supply                 | M1        | Monthly   | USD         | Jan 2005 – Dec 2024 |
| USA M2 Money Supply                 | M2        | Monthly   | USD         | Jan 2005 – Dec 2024 |
| China M1 Money Supply               | CNMSM1    | Monthly   | CNY         | Jan 2005 – Dec 2024 |
| China M2 Money Supply               | CNMSM2    | Monthly   | CNY         | Jan 2005 – Dec 2024 |
| USD to CNY exchange rate            | USDCNY    | Monthly   | CNY         | Jan 2005 – Dec 2024 |
| Index performance for China CPI YoY | CNCPIYOY  | Monthly   |             | Jan 2005 – Dec 2024 |
| China GDP Constant Price YoY        | CNGDPYOY  | Monthly   |             | Jan 2005 – Dec 2024 |
| U.S. Consumer Price Index (CPI) YoY | CPI YOY   | Monthly   |             | Jan 2005 – Dec 2024 |
| US Industrial production            | IP        | Monthly   | USD         | Jan 2005 – Dec 2024 |
| S&P Global Semiconductor Index      | SPGSMCUP  | Monthly   | USD         | Jan 2005 – Dec 2024 |
| China Exports to US                 | CHEXUS    | Monthly   | USD         | Jan 2005 – Dec 2024 |
| Palladium                           | PALL      | Monthly   | USD         | Jan 2005 – Dec 2024 |
| Silver                              | SILV      | Monthly   | USD         | Jan 2005 – Dec 2024 |
| Gold                                | GOLDS     | Monthly   | USD         | Jan 2005 – Dec 2024 |
| Copper (Comex)                      | HG1       | Monthly   | USD         | Jan 2005 – Dec 2024 |
| Natural Gas                         | NG1       | Monthly   | USD         | Jan 2005 – Dec 2024 |
| Brent Crude                         | CO1       | Monthly   | USD         | Jan 2005 – Dec 2024 |
| WTI Crude Oil (Nymex)               | CL1       | Monthly   | USD         | Jan 2005 – Dec 2024 |
| China REE Export                    |           | Annual    | Weight (mt) | Jan 2005 – Dec 2023 |
| China REE Import                    |           | Annual    | Weight (mt) | Jan 2005 – Dec 2023 |
| China REE Production                |           | Annual    | Weight (mt) | Jan 2005 – Dec 2024 |
| China REE Production Share          |           | Annual    | Percentage  | Jan 2005 – Dec 2024 |

### 3.2 Data Handling

The sample covers the period from January 2005 to December 2024, to the exception of China’s REE trade data that had not been updated with 2024 figures. Missing 2024 trade observations were filled by carrying forward the last available annual value, and all production and trade series were similarly forward filled to produce a monthly frequency. This approach maintains the integrity of the reported annual values across the monthly time frame, assuming the annual figure represents the state for the entire period until the next annual update. We merged and aligned all series on a common calendar index and performed cleaning operations to address any remaining inconsistencies using the Pandas library in Python. No stationarity tests or variable transformations (e.g., differencing or log-scaling) were performed. This is because our forecasting methodology, which

we'll detail in the next chapter, is specifically designed to operate on data in levels. All data extraction, transformation, and merging steps were implemented in Python version 3.13 using Pandas version 2.2 for time-series manipulation and NumPy 2.2 for numerical operations.

## 4 METHODOLOGY

This chapter presents the empirical strategy and analytical framework employed to address the research question concerning the effective performance of feature selection in guiding Rare Earth Element (REE) price forecasting. Given the high-dimensional nature of the dataset and the complex dynamics of REE markets, this chapter meticulously outlines the steps taken in model construction and evaluation. The chapter is structured as follows: we begin by presenting the initial data pre-processing and the rationale behind addressing multicollinearity among predictors. Subsequently, the chapter elaborates on the three distinct feature selection methodologies: Adaptive Lasso, Post-Double Selection, and Granger Causality testing, each building upon the previous one to generate different model specification. Following this, we describe the selection and characteristics of the primary forecasting algorithm, the Random Forest. Finally, the chapter concludes by detailing the adopted forecasting strategy, including the rolling window approach and dynamic feature selection, along with the metrics used to assess model performance.

### *4.1 Correlated predictors*

Due to the high number of features included in the dataset, we sought to investigate and mitigate potential issues related to multicollinearity among the predictors; therefore, a correlation matrix was constructed. Features with pairwise correlation coefficients exceeding the chosen threshold were removed. According to Dormann et al. (2013), selecting variables based on the absolute value of correlation coefficients is the most commonly applied method across different scientific disciplines, although threshold values vary, with 0.7 being the most frequently used.

In this study, a more conservative threshold of 0.95 was adopted, based on the visual observation that many of the variables exhibited trending behaviour. To avoid removing potentially important variables due to spurious correlations, a more conservative threshold was chosen. At this early stage, we prioritised retaining features over aggressive dimensionality reduction, since additional feature selection techniques were applied later in the process. This initial step of feature selection was applied consistently across all models presented in this chapter.

### *4.2 Applying Adaptive Lasso for feature selection*

To model the interdependence between the remaining features we applied a vector autoregressive (VAR) model. A VAR is an  $n$ -equation,  $n$ -variable linear model in which each variable is explained by its own lagged values as well as the current and past values of the remaining  $n - 1$  variables (Stock & Watson, 2001).

Up to this point, no formal unit root testing (such as the Augmented Dickey-Fuller test) had been conducted on the predictors. However, based on visual inspection of the features (see Figure B3 in Appendix B), we concluded that several of them exhibit non-stationary behaviour. The decision to forgo the ADF-test was motivated by the fact that later stages of the methodology employ techniques that do not require the data to be stationary, thereby circumventing the need for traditional differencing or transformation of the features. Hence, based on the remaining features, a VAR model in levels was constructed.

We model the  $K$ -dimensional multivariate time series  $\mathbf{Y}_t \in \mathbb{R}^K$  as a VAR( $p$ ) process where  $\mathbf{Y}_t = \{(Y_{t,1}, Y_{t,2}, \dots, Y_{t,K})'\}$

$$\mathbf{Y}_t = \mathbf{A}_0 + \mathbf{A}_1 \mathbf{Y}_{t-1} + \dots + \mathbf{A}_p \mathbf{Y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad (1)$$

where:  $\mathbf{A}_0 \in \mathbb{R}^K$  is a vector of intercepts,  $\mathbf{A}_1, \dots, \mathbf{A}_p \in \mathbb{R}^{K \times K}$  are autoregressive coefficient matrices. We assume that error terms  $\boldsymbol{\varepsilon}_t$  are independent across time and follow a multivariate Gaussian distribution with zero mean and covariance matrix  $\boldsymbol{\Sigma}$ . This specification allows the potential presence of unit roots and cointegration among the features.

Modelling the VAR in levels is appropriate here because the primary objective is not inference, such as impulse response analysis, but rather to organize the features into a VAR model and estimate them using a dimensionality reduction technique. Therefore, the standard assumption of stationarity can be relaxed. The selection of  $p$  was based on information criteria. The most common are Schwarz (1978)'s Bayesian Information Criterion (BIC) and Akaike (1998)'s Akaike Information Criterion (AIC). Since BIC is more conservative and consistent in larger samples than AIC, we opted for the former to prevent the VAR model to be highly over-parameterised.

Since the VAR is a multivariate time series model, it captures interdependencies among multiple components. As a result, the VAR model is heavily parameterised, and the number of autoregressive coefficients increases quadratically with the dimensionality of the system. For large (or even moderate) dimensions, this can result in noisy parameter estimates and obscure the interpretation of temporal relationships (Davis et al., 2016).

To address this issue a shrinkage method exploiting sparsity was used to filter out non-contributing features. One of the most common approaches is the lasso, originally proposed by Tibshirani (1996), which adds a penalty on the absolute values of the models coefficient estimates. However, using plain lasso was problematic in our context, as the VAR model was specified in levels. Lee et al. (2022) investigated the limitations and potential of various lasso methods in predictive regression frameworks involving a mix of stationary, local unit root, and cointegrated predictors. The study highlights that

standard lasso and its standardized variant fail to consistently estimate parameters in such mixed-root settings. Specifically, plain lasso applies uniform penalization across all variables, which poses a critical limitation: it does not adapt to the differing statistical properties of the predictors. This is evident in the behaviour of the variable screening process. As the penalty term increases, coefficients associated with slowly converging predictors (with a convergence rate of  $\sqrt{n}$ ) are overly shrunk, while those with fast convergence (rate of  $n$ ) remain largely unaffected. In contrast, adaptive lasso is shown to outperform plain lasso in mixed-root environments. Based on these findings, the adaptive lasso was selected as the regularization technique.

To ensure that the adaptive lasso penalised coefficients consistently across predictors, we applied standardisation (z-score normalisation). This was done after constructing the VAR model to avoid distorting the inherent time dependencies present in the time series data. As noted by Ogasawara et al. (2010), applying normalisation techniques to non-stationary time series can be problematic, as the assumptions of constant mean and variance over time are violated. However, we argue that although the predictors may exhibit mixed properties (stationary, unit root, and cointegrated), normalisation was applied solely to prevent the adaptive lasso from introducing scale-dependent bias in the regularisation process. Furthermore, the VAR framework in this study is not used for statistical inference or direct forecasting. Its sole purpose is to model lagged interdependencies and serve as a basis for the application of the adaptive lasso. After applying the adaptive lasso the standardisation was reverted.

The adaptive lasso adds a weighted penalty to the existing lasso penalty on the estimated coefficients of the model. Zou (2006) shows that a root- $n$  consistent estimator of the true parameter vector  $\hat{\mathbf{a}}_k^*$ , such as the ordinary least squares estimator  $\hat{\mathbf{a}}_k^{\text{OLS}}$ , can be used to define the weight vector. A value  $\gamma > 0$  is then chosen, and the weight vector is defined as  $\hat{\mathbf{w}} = 1/|\hat{\mathbf{a}}_k^{\text{ols}}|^\gamma$ . The Adaptive Lasso estimator  $\hat{\mathbf{a}}_k^{\text{alasso}}$  is then given by Equation (2) for the  $k$ -th equation in the VAR model:

$$\hat{\mathbf{a}}_k^{\text{alasso}} = \arg \min_{\mathbf{a}_k} \left\{ \|\mathbf{y}_k - \mathbf{X}\mathbf{a}_k\|^2 + \lambda_n \sum_{j=1}^{pK} \hat{w}_j |a_{k,j}| \right\}, \quad (2)$$

where  $\mathbf{y}_k$  is the response vector for variable  $k$ , and  $\mathbf{X}$  is the common design matrix of lagged predictors. The parameters  $\lambda_n$  and  $\gamma$  are the two tuning parameters of the Adaptive Lasso. In practice,  $\gamma$  is often fixed at either 1 or 2, while  $\lambda_n$  is selected as the primary tuning parameter Lee et al. (2022).

For  $\gamma$ , we fixed the value at 1, following common practice. For the hyperparameter  $\lambda_n$ , we

chose to use information criteria rather than cross-validation (CV) to preserve the time series structure. Since BIC is known to choose a more parsimonious model than AIC, we opted for the former to prevent the dimension reduction from being too liberal and retaining too many features in the final subset.

This paper focuses on forecasting REEs prices, with particular emphasis on the price of neodymium, as mentioned in the introduction. Consequently, attention is restricted to a single equation within the VAR system, specifically the one corresponding to neodymium.

Thus, the first feature selection process leading up to Model 1 is as follows: the adaptive lasso was applied to the neodymium equation. All non-zero coefficients identified by the regularisation were retained in the model, while those that shrunk to zero were excluded. This dimensionality reduction step constituted the final phase of the variable selection process. Combined with the earlier correlation threshold, this defines the full procedure used to select relevant predictors for Model 1 and is summarised in algorithm 1.

---

**Algorithm 1** Model 1: One-step feature selection procedure

---

- 1: Remove variables with absolute pairwise correlation coefficients exceeding 0.95.
  - 2: Organize a VAR( $p$ ) model as in Equation (1) on the remaining variables, selecting lag order  $p$  using the BIC.
  - 3: Estimate the neodymium ( $k = \text{Nt}$ ) equation using Adaptive Lasso (Equation (2)), where the penalty parameter  $\lambda_n$  is selected via the BIC.
  - 4: Select direct predictors of neodymium:  $\mathcal{S}_1 \leftarrow \{\hat{a}_{\text{Nt},j} \neq 0\}$ .
  - 5: Define final predictor subset:  $\mathcal{S} = \mathcal{S}_1$
- 

### 4.3 Post-double selection

To mitigate the risk of prematurely excluding relevant features in Model 1, we employed the post-double-selection (PDS) method proposed by Belloni et al. (2014). This approach addresses a key shortcoming of single-selection methods: their inability to adequately control for omitted variable bias.

The framework results in that we first focused on the coefficients that were reduced to zero in the neodymium equation, indicating features that did not have a direct influence on neodymium. For features with zero coefficients, their corresponding equations in the VAR were excluded from further analysis. Next, we shifted focus to the equations of the features that remained in the neodymium equation (i.e. those with non-zero coefficients). Adaptive Lasso was then applied to all those equations in the VAR model, this step helped us identify whether previously excluded features might have a indirect effects.

For example, if the U.S. 10-year treasury yield was not selected in the neodymium equation

(i.e. deemed irrelevant) but is selected as a predictor in the gold price equation and if gold is a significant predictor of neodymium, then the treasury yield is retained in the model due to its indirect relevance. This two-step procedure differs from a standard one-shot lasso by reducing the likelihood of false negatives, ensuring that indirectly influential features are not prematurely excluded.

As mentioned earlier, BIC was used for selecting the optimal value of  $\lambda_n$  for the one-shot Adaptive Lasso; however, in the double-selection process, different information criteria were applied. AIC was applied in the first selection round and BIC in the second.

The rationale for choosing the less conservative AIC in the first step was to reduce the aggressiveness of the regularisation and allow more predictors to be retained. This ensures that the second step is not overly constrained by a stringent initial selection. If BIC were used in the first step, its stricter penalty could eliminate relevant predictors too early, thereby preventing other variables from benefiting from their potential indirect contributions.

However, BIC was used in the second step for deciding the optimal value of  $\lambda_n$ . This was done to mitigate the risk of retaining all initial features due to selecting the union of features. As AIC is more liberal than BIC it would have been a higher probability of this happening if the former was chosen as the information criteria. This defines the full procedure used to select relevant predictors for Model 2 and is summarised in algorithm 2.

---

**Algorithm 2** Model 2: Post-Double-Selection Procedure

---

- 1: Remove variables with absolute pairwise correlation coefficients exceeding 0.95.
  - 2: Organize a VAR( $p$ ) model as in Equation (1) on the remaining variables, selecting lag order  $p$  using the BIC.
  - 3: Estimate the neodymium ( $k = \text{Nt}$ ) equation using Adaptive Lasso (Equation (2)), where the penalty parameter  $\lambda_n$  is selected via the AIC.
  - 4: Select direct predictors of neodymium:  $\mathcal{S}_1 \leftarrow \{\hat{a}_{\text{Nt},j} \neq 0\}$ .
  - 5: For each  $i \in \mathcal{S}_1$ , estimate its VAR equation using Adaptive Lasso with  $\lambda_n$  selected by BIC.
  - 6: Let  $\mathcal{S}_2$  include all variables  $j \notin \mathcal{S}_1$  such that: there exists  $i \in \mathcal{S}_1$  with  $\hat{a}_{i,j} \neq 0$
  - 7: Define final predictor subset:  $\mathcal{S}_{\text{final}} = \mathcal{S}_1 \cup \mathcal{S}_2$
- 

#### 4.4 Granger causality

The feature selection processes used for Models 1 and 2 did not involve testing whether the retained predictors were statistically significant. Typically, traditional econometric

inference methods, such as the Granger causality test are applied within a VAR framework to determine which variables significantly explain the response variable’s behaviour. Therefore, in the third and final feature selection process, we followed the same procedure as in Algorithm 2, but added an inferential layer by applying the Granger causality test to assess the statistical significance of the final set of retained predictors. The Granger causality test specifically assesses whether or not the lagged versions of the predictors Granger-cause neodymium. It is important to note, however, that Granger causality does not imply true causality. Rather, it is a statistical tool for assessing the predictive power of a variable, conditional on the given information set.

As mentioned earlier in the paper, the VAR was specified in levels. As a result, the retained variables may include a mixture of stationary processes, stochastic trends, and cointegrated relationships. To avoid the need for traditional pre-estimation diagnostic tests, such as unit root and cointegration tests, we adopted the lag-augmented approach proposed by Toda and Yamamoto (1995) and extended by Hecq et al. (2023). This technique involves adding additional lags to the VAR model beyond the optimal lag length selected by information criteria, in our case by the BIC. Doing so enables valid statistical inference without requiring prior transformation of the variables (e.g., differencing to achieve stationarity), which would otherwise be necessary under standard VAR inference assumptions. We decided to add two extra lags to the model to account for the possibility that some predictors may be integrated of order 2. The lags were added on top of the existing lags.

Hecq et al. (2023) point out the necessity of not under-specifying the number of lags as it would break the lag augmentation and must therefore be avoided. In the paper they point out that both AIC and BIC succeed in almost never under-specifying the number of lags. Based on this we decided to use information criteria rather than to decide the number of lags ourselves.

The lag-augmentation method proceeds as follows: each predictor is individually tested for Granger causality by adding two additional augmented lags of that predictor. A t-test (or an F-test if multiple original lags are present) is used to assess the statistical significance of the original lag coefficients. If the null hypothesis ( $\beta = 0$ ) is rejected, we conclude that the predictor has a statistically significant influence on the response variable. This process is then repeated for the next predictor in the subset. Importantly, the augmented lags in each step apply only to the current treatment variable; they are not included simultaneously for all predictors in the model.

The mathematical proof of adding extra lags to circumvent the need for stationarity and to enable inference is beyond the scope of this paper; we therefore refer to the cited

literature that discusses the topic.

Since we are conducting multiple tests using a single-inference procedure, the risk of making a Type I error increases substantially. To address this issue, we applied the Benjamini–Hochberg procedure (Benjamini & Hochberg, 1995) to gain control of the False Discovery Rate (FDR). The choice of the significance level  $\alpha$  then becomes a trade-off between rejecting more hypotheses and tolerating a higher FDR. In our application, it also becomes a parameter of how sparse the feature selection is, a higher value leads to more variables passing the Granger Causality test.

For the purposes of forecasting, we selected an FDR of 0.10 to allow for a broader inclusion of potentially predictive variables while maintaining reasonable control over false discoveries. While the FDR level could, in principle, be treated as a hyperparameter to be optimised through various techniques, we found that the chosen value was reasonable for the objectives of our study and therefore did not pursue this avenue further.

Each individual Granger causality test evaluates the following hypotheses:

**Null Hypothesis:** All lagged values of  $z_t$  have no effect on neodymium:

$$H_0 : a_1^z = a_2^z = \dots = a_p^z = 0$$

**Alternative Hypothesis:** At least one lag of  $z_t$  affects neodymium:

$$H_1 : a_k^z \neq 0 \text{ for some } k \in \{1, \dots, p\}$$

The Benjamini–Hochberg procedure was then applied to the resulting set of  $p$ -values from the hypothesis testing. The features that pass the Benjamini–Hochberg threshold are then retained as the final set of predictors for Model 3. This concludes the whole feature selection process for Model 3 and is summarised in algorithm 3.

---

**Algorithm 3** Model 3: Feature selection via PDS with Granger Causality

---

- 1: Remove variables with absolute pairwise correlation coefficients exceeding 0.95.
  - 2: Organize a VAR( $p$ ) model as in Equation (1) on the remaining variables, selecting lag order  $p$  using the BIC.
  - 3: Estimate the neodymium ( $k = \text{Nt}$ ) equation using Adaptive Lasso (Equation (2)), where the penalty parameter  $\lambda_n$  is selected via the AIC.
  - 4: Select direct predictors of neodymium:  $\mathcal{S}_1 \leftarrow \{\hat{a}_{\text{Nt},j} \neq 0\}$ .
  - 5: For each  $i \in \mathcal{S}_1$ , estimate its VAR equation using Adaptive Lasso with  $\lambda_n$  selected by BIC.
  - 6: Let  $\mathcal{S}_2$  include all variables  $j \notin \mathcal{S}_1$  such that: there exists  $i \in \mathcal{S}_1$  with  $\hat{a}_{i,j} \neq 0$
  - 7: Define predictor subset:  $\mathcal{S}_{\text{final}} = \mathcal{S}_1 \cup \mathcal{S}_2$
  - 8: For each  $j \in \mathcal{S}_{\text{final}}$ , augment with  $p + 2$  lags and test for Granger causality toward neodymium.
  - 9: Apply the Benjamini–Hochberg procedure to the resulting Granger  $p$ -values.
  - 10: Select all variables  $j \in \mathcal{S}_{\text{final}}$  that Granger cause neodymium based on the Benjamini–Hochberg procedure with a FDR threshold of 10%.
- 

#### 4.5 Forecasting model - Random Forest

So far in this chapter, we have presented three different feature selection processes used to define the final set of predictors for our forecasting models. Each process builds upon the previous one. Despite these differences in feature selection, all forecasting models employ the same algorithm: the Random Forest (RF). Introduced by Breiman (2001), Random Forest is an ensemble learning method that combines multiple decision trees. It is a powerful tool for prediction, and due to the Law of Large Numbers, it tends to be robust against overfitting.

Lin et al. (2017) provided the following explanation for regression tasks using the RF algorithm. First, generate  $n$  bootstrap sample sets from the original dataset. Second, for each bootstrap sample set, grow an unpruned regression tree with the following modification: at each node, randomly sample  $p$  features from all the input features and choose the best split from those  $p$  features, where  $p < m$  and  $m$  is the total number of input features. Third, predict the output for a new observation by averaging the predictions of the  $n$  regression trees when the new input is fed into the RF.

In the paper by Henriques and Sadorsky (2023), Random Forests were among the models with the highest accuracy in forecasting REEs stock prices. However, their approach differed from ours in both objective (stock prices versus spot prices) and forecasting horizon, as they focused on short-term forecasts of up to 20 days. This raised the question

of whether Random Forest is an appropriate choice for long-term forecasting. To explore this, we considered the findings of Herrera et al. (2019), who found that Random Forests outperformed deep learning and traditional time series models in forecasting commodity prices over longer horizons. Additionally, Random Forests offer the advantage of capturing non-linear relationships among lagged predictors, something traditional linear time series models may fail to adequately address.

Up to this point in the paper, we have applied methods to address the potential presence of non-stationary features and cointegration, enabling effective regularisation and valid statistical inference without requiring formal testing or transformation of the features. Since the RF model does not impose assumptions about the underlying data structure, we chose to maintain this approach and therefore did not test each predictor or transform the variables to ensure stationarity.

The model specification for the RF was determined using a grid search (see A3 in Appendix A) combined with time series cross-validation on the training dataset. In addition to tuning the standard hyperparameters, we also performed a grid search to determine the optimal number of lags for all predictors for each forecast horizon. This approach provides insight into how the forecasting algorithm evaluates the trade-off between including more or fewer lags for the predictors to improve accuracy across both short- and long-term horizons. The combination that yielded the lowest Root Mean Square Error (RMSE) on the validation set within the training data was then fixed and applied during the forecasting phase on the out-of-sample period.

#### *4.6 Forecasting method and model evaluation*

As stated in Gargano and Timmermann (2014), a recursive rolling estimation window was used as the forecasting strategy. They emphasised the distinction between in-sample and out-of-sample forecasting: in-sample forecasts are not truly ex-ante, as they rely on information from the full sample, which would not be available in real-time to decision-makers. The recursive approach addresses this issue by using only past information to estimate parameters and generate forecasts.

For one-step-ahead forecasting, the procedure is relatively straightforward: estimate the model using the training data, forecast one period ahead, compute the forecast error, update the model by adding the next actual observed value, and drop the oldest observation from the rolling window. Since the actual observation becomes available for each 1-ahead step forecast, it can replace the forecasted value, making this approach essentially equivalent to a direct forecasting strategy.

However, applying recursive rolling estimation in a multivariate setting for multi-step forecasting introduces additional complexity. This paper aims to forecast both short- and long-term horizons—specifically: 1-month, 3-month, and 12-month ahead forecasts. Forecasting  $h$ -steps ahead in a multivariate context requires generating forecasts for all predictors at each future time point, so they can be included in subsequent forecast steps. This means that the forecasting model must be run recursively for each horizon and each variable, which increases computational complexity and introduces the risk of error accumulation along the forecast horizon.

Due to these challenges, and for the sake of simplicity, we chose to apply the direct forecasting strategy across all forecast horizons. We still employed a rolling window approach, after each direct forecast was made, the estimation window advanced by one step, incorporating the latest observed value while dropping the oldest. This procedure was repeated until the end of the test period, forecast errors were recorded at each step to compute the RMSE. The size of the rolling window was set to 10 years of the most recent data points in the training set and used to forecast into the holdout-set (test-data).

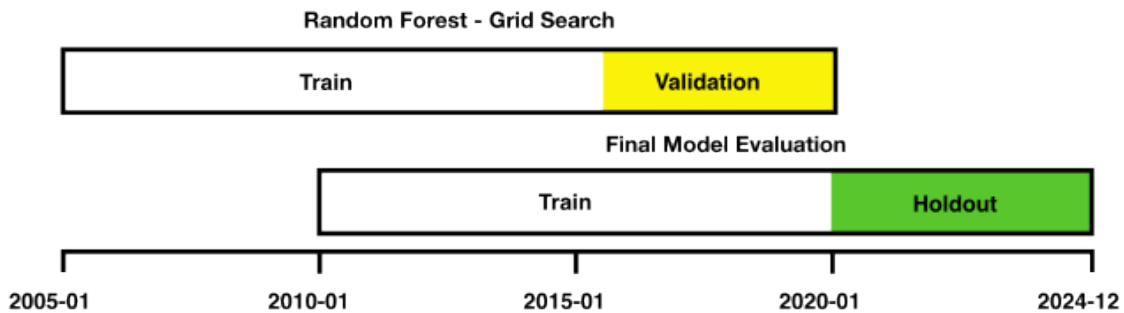


Figure 2: Training-validation-holdout split for time-based model development. The full dataset is divided chronologically to preserve temporal structure. The initial portion is used as the training set to fit a Random Forest model (2005-01 until 2015-06), followed by a validation set (2015-07 until 2019-12 marked in yellow) used for hyperparameter tuning via grid search. After model selection, the final model is retrained on the extended training window (2010-01 until 2019-12) and evaluated on a temporally disjoint holdout set (2020-01 until 2024-12 marked in green), ensuring an unbiased assessment on unseen future data.

The data splitting strategy for model development and evaluation is visually explained in [Figure 2](#). The full 20-year dataset (January 2005 to December 2024) was chronologically divided into a 75% portion for model development and a 25% holdout set for final, unbiased evaluation. Within the model development phase, the initial portion of the dataset was further split for hyperparameter tuning via grid search. Specifically, the period from January 2005 to June 2015 served as the initial training set to fit the Random Forest model, followed by a validation set from July 2015 to December 2019.

After the optimal model specification was determined through this tuning process, the

final model was continuously retrained using a 10-year rolling window using the same forecasting approach as in the model development, but now employing the Random Forest model with its selected hyperparameters. This rolling window started from January 2010 and extended up to December 2019 for the initial training window, and then advanced chronologically. This rigorous approach ensured its forecasting performance was evaluated on a truly unseen holdout set from January 2020 until December 2024, thereby providing a realistic assessment of out-of-sample performance on future data.

The main purpose of using a rolling window was to obtain a representative evaluation metric of how the model performs over time, rather than generating a single forecast for a specific point in the test set. The latter approach would only provide insight into model performance at one instance, whereas our rolling evaluation reflects the model’s average forecasting accuracy across the entire test horizon. The size of the rolling window was chosen to ensure that enough observations remained in the sample to conduct lag selection using information criteria within the VAR framework. This was important because, in order to identify the optimal number of lags, the VAR model must be able to consider lag lengths greater than one ( $p > 1$ ). The reason for this will become apparent in the next section. [Figure B1](#) in [Appendix B](#) provides a visual explanation of the forecasting scheme and its execution across different horizons.

As described in the forecasting process, RMSE is one of the evaluation metrics to compare the model’s performance over time along with the Mean Absolute Percentage Error (MAPE). These metrics are commonly used in the literature which enables the ability to compare this paper’s results with other works.

#### *4.7 Dynamic feature selection*

As mentioned in the discussion on determining the rolling window size, the window needed to be large enough to allow estimation of a VAR model with sufficient number of lags. This was necessary to enable dynamic feature selection and to observe how the selection process evolves over time.

Under this approach, after each  $h$ -step forecast, the rolling window adds the latest true observation and removes the oldest one. Thus, at time  $t + h$ , before generating the next forecast, the feature selection algorithms were re-run (excluding step 1). This ensured that the most relevant predictors were selected based on the most up-to-date information. The main motivation for this procedure was to simulate a real-world forecasting environment as closely as possible, rather than fixing the feature set based solely on the training data.

## 5 EMPIRICAL ANALYSIS

This chapter presents the empirical analysis conducted to address the research question. It includes both the results and their corresponding discussion. The chapter begins with an exploratory analysis of the response variable and the correlation matrix of all included variables, followed by the results and discussion of the feature selection process and the forecasting performance of the different models across the various forecasting horizons.

### 5.1 Exploratory Analysis

Figure B2 in Appendix B shows the Pearson correlation heatmap for all predictors, sorted by each variable’s domain. We observed strong correlations (above 0.95) among certain metal prices, like lanthanum and cerium; between equity indices such as NASDAQ and Nikkei; and among monetary aggregates, including US M2 and China’s M1 and M2 measures.

To reduce the potential issue of multicollinearity, we omitted the following variables: Lanthanum, Yttrium, NASDAQ Composite Index (CCMP), Nikkei-225 Index (NKY), US M2 Money Supply (M2), China’s M1 and M2 Money Supply (CNMSM1 and CNMSM2) and WTI Crude Oil 1st Generic Futures (CL1). This exclusion left us with 27 variables. While variance inflation factors (VIFs) are typically used for further multicollinearity testing, we didn’t calculate them. This is because non-stationary variables can yield high  $R^2$  values solely from shared trajectories rather than true collinearity, making VIFs unreliable.

Figure 3 displays the 20-year price trajectory of neodymium, showing a pronounced spike around 2011. This spike can be attributed to China halving its export quotas, as discussed in subsection 1.1. While this sharp movement might suggest a structural break, Proelss et al. (2020) argue that the series does not exhibit a formal break when examining neodymium’s absolute returns during the same period. They also advocate using rolling-window estimates to mitigate the need to account for the potential of structural breaks.

Although traditional econometric structural break tests (e.g., Chow, Bai-Perron) were not performed, given our forecasting methodology’s mixed-root design, we conducted an exploratory change-point analysis using the PELT algorithm. We applied the PELT algorithm with an RBF cost function (Gaussian kernel), following the methodology outlined by Truong et al. (2020), which detects distributional shifts in variance and higher-order moments rather than only mean changes. Using a penalty parameter of 10 and enforcing no minimum segment length, the algorithm returned a single breakpoint at the final

observation (December 2024). This indicates no statistically significant internal change points in the neodymium price series under these settings, reinforcing the view that, aside from the end-of-sample boundary, the series does not exhibit pronounced distributional shifts that would warrant explicit regime-based segmentation within our model. However, it's important to note that this change-point analysis on the raw series serves an exploratory purpose. The effective handling of any structural shifts within the data is ultimately reflected in the robustness and forecasting accuracy of our model.

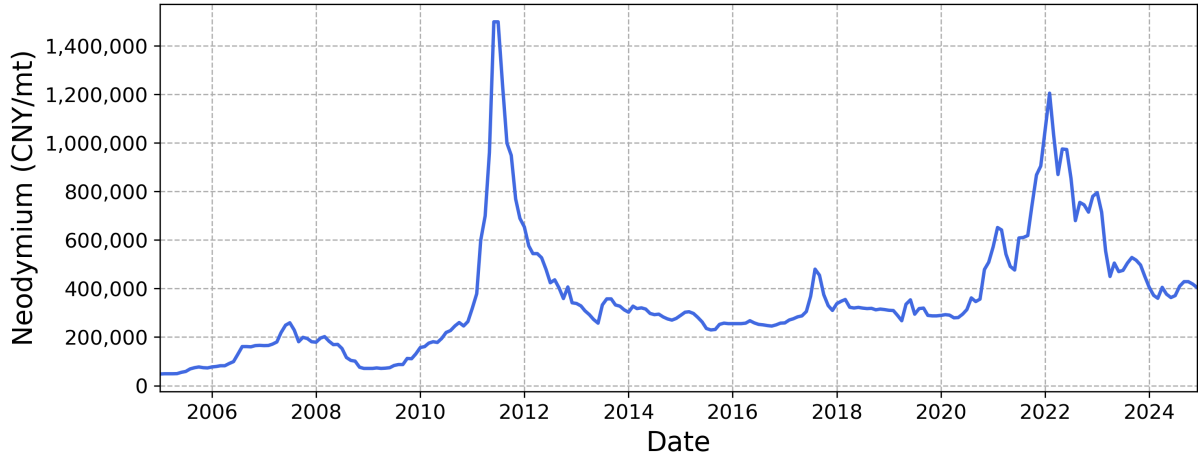


Figure 3: Neodymium price per metric ton of 99% pure oxide, shipped in plastic bags within steel barrels. Pricing data from January 2005 to December 2024. *Data source: Asia Metal database via Bloomberg Terminal*

Table 2 presents a summary of descriptive statistics and diagnostic tests for the neodymium 99% oxide spot price series. The average price of 354,227 CNY/mt is accompanied by a high standard deviation of 240,726 CNY/mt, reflecting substantial volatility over the sample period. Observations range from a low of 48,000 CNY/mt to a peak of 1,499,000 CNY/mt. The skewness of 1.78 indicates a right-tailed distribution driven by occasional price spikes, while the kurtosis of 4.13 suggests heavier tails than a normal distribution. A first-order autocorrelation coefficient of 0.97 confirms the the series evolves with strong persistence. The Augmented Dickey-Fuller (ADF) test returns a statistic of -2.68 and a p-value of 0.077, which narrowly avoids rejecting the null hypothesis of a unit root at the 5% significance level.

Table 2: Descriptive statistics for Neodymium

| Statistic                | Value     |
|--------------------------|-----------|
| Count                    | 240       |
| Mean                     | 354,227.1 |
| Std. Dev.                | 249,726.3 |
| Minimum                  | 48,000    |
| 25th percentile          | 198,250   |
| Median (50th percentile) | 297,500   |
| 75th percentile          | 419,875   |
| Maximum                  | 1,499,000 |
| Skewness                 | 1.78      |
| Kurtosis                 | 4.13      |
| Autocorrelation (lag 1)  | 0.97      |
| ADF Statistic            | -2.68     |
| ADF p-value              | 0.077     |

Note: The ADF p-value is based on a 0.05 significance level, and it was modelled with a constant.

## 5.2 Model Selection

Figure 4 presents a heatmap illustrating variable selection frequencies across the forecast horizons:  $h = 1$ ,  $h = 3$ , and  $h = 12$ . As detailed in the methodology chapter, the selection algorithms are re-executed each time the rolling window advances one step (with the exception of step 1 in the algorithms). Consequently the outcome of the selection process varies each window, meaning there isn't a single, definitive subset of features consistently employed for all forecasts generated by the models. This dynamic approach mimics real-world applications where one would add the most recent observation to the sample and re-run the model selection algorithm. Importantly, the underlying VAR model, foundational to all Adaptive Lasso (AdLasso) procedures, unanimously selected a lag order of one according to the BIC at every stage.

The heatmap, Figure 4, offers a detailed view of how frequently each potential predictor was selected at different stages of the Adaptive Lasso (AdLasso) process, which captures both direct and indirect effects, and subsequent Granger causality testing. Each cell represents the percentage of forecast origins in which a specific variable was chosen. For instance, for a forecast horizon of  $h = 1$ , "Terbium" was selected at every forecasting origin in Model 1 and Model 2, but in 91.7% of the windows for Model 3. The variables are then sorted by their average selection frequencies across all stages and the three horizons.

Examining the selections for each model and horizon, we observed distinct patterns. For

instance, at a forecast horizon of  $h = 1$ , praseodymium is only selected into Model 1 in 48.3% of the forecasting origins, but in 80% of windows for Model 3 after the second stage of AdLasso for indirect effects and Granger causality testing. A similar increase in selection is observed in forecasting horizons  $h = 3$  and  $h = 12$ . Similarly, the same increase is observed in the selection of USA M1 Money Supply. This disparity indicates that Model 1 (One-Shot AdLasso) appears less effective at identifying certain predictive effects, particularly those revealed through the sequential AdLasso stages and Granger causality testing implemented in Model 3.

It is worth noting that Model 2, which uses a PDS procedure to account for indirect effects, consistently selects a very high proportion of available variables across nearly all forecasting origins. This observation reinforces our suspicion that, while adept at identifying comprehensive sets of relevant predictors, the PDS method may inherently prioritise inclusivity over the parsimony typically associated with effective dimensionality reduction.

This contrasts with Model 1 (One-Shot AdLasso), suggesting a potential under-identification of valuable indirect effects when relying solely on a single-stage selection process and also with Model 3, which integrates Granger Causality testing to reintroduce sparsity and ensure, with statistical certainty, that no significant predictive power is omitted. This outcome is, of course, conditional on the underlying assumptions of the Granger causality test and the characteristics of the observed data. The specific results of these Granger causality tests for each rolling window in Model 3 are further detailed in [Figure B4](#) in [Appendix B](#). Notably, the VAR models used for these Granger causality tests also consistently selected an optimal lag order of 1 across all rolling windows according to the BIC.

Collectively, these observations highlight a crucial trade-off in variable selection: while single-stage methods (like Model 1) risk omitting relevant predictors, comprehensive methods (like Model 2), using PDS, may lead to models that are overly dense. Model 3, by integrating PDS with Granger causality testing, aims to mitigate these issues, providing a statistically sound and parsimonious set of features that are linked to the target variable’s future behaviour.

Due to the nature of dynamic feature selection, the ability to identify a definitive set of important predictors is more limited compared to a static feature selection framework. However, based on the frequency with which certain variables are selected, financial market indicators appear to exhibit predictive power, consistent with the findings of Guo and Mai (2024). In our case, this effect is not uniform across markets: for example, the S&P 500 is frequently selected by Model 3, whereas the Hang Seng Index is never included.

The interconnection between various REEs is also reflected in the frequency with which they are selected as predictors. A similar pattern to what Batten et al. (2010) observed, as they found a significant volatility spillover between the precious metals markets. Additionally, the results from the variable selection indicate that some precious metals, such as Palladium, are considered relevant for forecasting REEs prices.

|                             | <b>h = 1</b>   |              |              | <b>h = 3</b>   |              |              | <b>h = 12</b>  |              |              |
|-----------------------------|----------------|--------------|--------------|----------------|--------------|--------------|----------------|--------------|--------------|
| Neodymium                   | 100.0          | 100.0        | 100.0        | 100.0          | 100.0        | 100.0        | 100.0          | 100.0        | 100.0        |
| Terbium                     | 100.0          | 100.0        | 91.7         | 100.0          | 100.0        | 91.4         | 100.0          | 100.0        | 89.8         |
| S&P 500                     | 100.0          | 100.0        | 75.0         | 100.0          | 100.0        | 74.1         | 100.0          | 100.0        | 69.4         |
| China export to US          | 96.7           | 100.0        | 70.0         | 96.6           | 100.0        | 69.0         | 95.9           | 100.0        | 63.3         |
| Palladium Spot Price        | 86.7           | 100.0        | 75.0         | 86.2           | 100.0        | 74.1         | 83.7           | 100.0        | 69.4         |
| China REEs Production Share | 96.7           | 100.0        | 56.7         | 98.3           | 100.0        | 55.2         | 98.0           | 100.0        | 46.9         |
| Copper (Comex)              | 88.3           | 100.0        | 50.0         | 87.9           | 100.0        | 51.7         | 85.7           | 100.0        | 61.2         |
| U.S. 10 Year Treasury       | 95.0           | 100.0        | 43.3         | 94.8           | 100.0        | 44.8         | 93.9           | 100.0        | 40.8         |
| Dysprosium                  | 91.7           | 100.0        | 41.7         | 91.4           | 100.0        | 43.1         | 89.8           | 100.0        | 51.0         |
| China REEs Production       | 76.7           | 96.7         | 60.0         | 75.9           | 96.6         | 58.6         | 75.5           | 100.0        | 65.3         |
| Europium                    | 90.0           | 98.3         | 36.7         | 91.4           | 98.3         | 37.9         | 91.8           | 98.0         | 44.9         |
| Praseodymium                | 48.3           | 100.0        | 80.0         | 48.3           | 100.0        | 79.3         | 55.1           | 100.0        | 75.5         |
| China REEs Imports          | 81.7           | 100.0        | 41.7         | 81.0           | 100.0        | 43.1         | 77.6           | 100.0        | 51.0         |
| China GDP YoY               | 98.3           | 100.0        | 21.7         | 98.3           | 100.0        | 22.4         | 98.0           | 100.0        | 26.5         |
| Silver Spot Price           | 88.3           | 100.0        | 28.3         | 87.9           | 100.0        | 29.3         | 89.8           | 100.0        | 34.7         |
| China CPI YoY               | 100.0          | 100.0        | 15.0         | 100.0          | 100.0        | 15.5         | 100.0          | 100.0        | 18.4         |
| U.S. CPI YoY                | 98.3           | 100.0        | 13.3         | 98.3           | 100.0        | 13.8         | 98.0           | 100.0        | 16.3         |
| Cerium                      | 71.7           | 98.3         | 33.3         | 72.4           | 98.3         | 34.5         | 77.6           | 98.0         | 40.8         |
| Brent Crude                 | 86.7           | 100.0        | 20.0         | 86.2           | 100.0        | 20.7         | 83.7           | 100.0        | 24.5         |
| USD to CNY exchange rate    | 96.7           | 100.0        | 10.0         | 96.6           | 100.0        | 10.3         | 95.9           | 100.0        | 12.2         |
| Gold Spot Price             | 80.0           | 100.0        | 23.3         | 79.3           | 100.0        | 24.1         | 77.6           | 100.0        | 28.6         |
| USA M1 Money Supply         | 41.7           | 100.0        | 50.0         | 41.4           | 100.0        | 51.7         | 49.0           | 100.0        | 61.2         |
| US Industrial production    | 98.3           | 100.0        | 0.0          | 98.3           | 100.0        | 0.0          | 98.0           | 100.0        | 0.0          |
| Natural Gas                 | 95.0           | 100.0        | 3.3          | 94.8           | 100.0        | 3.4          | 93.9           | 100.0        | 4.1          |
| CSI 300 Index               | 88.3           | 96.7         | 8.3          | 87.9           | 96.6         | 6.9          | 91.8           | 100.0        | 8.2          |
| The Hang Seng Index         | 55.0           | 100.0        | 0.0          | 55.2           | 100.0        | 0.0          | 57.1           | 100.0        | 0.0          |
| China REEs Exports          | 48.3           | 78.3         | 0.0          | 50.0           | 79.3         | 0.0          | 59.2           | 89.8         | 0.0          |
|                             | Selected (%)   | Selected (%) | Selected (%) | Selected (%)   | Selected (%) | Selected (%) | Selected (%)   | Selected (%) | Selected (%) |
|                             | <b>Model 1</b> |              |              | <b>Model 1</b> |              |              | <b>Model 1</b> |              |              |
|                             | <b>Model 2</b> |              |              | <b>Model 2</b> |              |              | <b>Model 2</b> |              |              |
|                             | <b>Model 3</b> |              |              | <b>Model 3</b> |              |              | <b>Model 3</b> |              |              |

Figure 4: Comparison of variable selection frequencies across forecast horizons  $h = 1$ ,  $h = 3$ , and  $h = 12$ . The heatmap illustrates the percentage of rolling-window forecast origins for which each variable was selected. For each forecast horizon, the results are presented for three distinct models (Model 1, Model 2, and Model 3). Within each model's section, the columns represent the selection frequencies at different stages of the variable selection process: the first column shows selection during AdLasso Stage 1 (Direct Effects), the second column shows selection during AdLasso Stage 2 (Indirect Effects), and the final column indicates variables chosen after Granger causality testing. Variables are labelled by their descriptive names and are presented grouped by their corresponding forecast horizon.

### 5.3 Random Forest Application and Evaluation

Random Forests serve as the common forecasting engine for all three models. Hyperparameters were selected through a grid search combined with time-series cross validation on a rolling window that was updated at every forecast origin. The search included different parameters for maximum tree depth, minimum samples per leaf, the number of trees in the ensemble, and the optimal lag length for each forecast horizon  $h \in \{1, 3, 12\}$ . Details of this procedure can be found in [subsection 4.5](#).

Table A2 in Appendix A presents the settings that minimised the in-sample RMSE for each model and horizon. Several patterns can be observed.

First, horizon length influences the preferred lag structure where a longer horizon seems to perform better with a higher lag order, except in Model 2 that performed best with 12 lags on forecasting horizon  $h = 12$ .

Second, Model 3 performs better on shallower trees (a maximum depth of 4) at the shorter horizons ( $h = 1$  and  $h = 3$ ) while the other two models favour deeper or unrestricted depth.

Third, all three models share the characteristic of performing better when the minimum number of observations per leaf is moderate (3 to 5 samples per leaf) at longer horizons. This is an expected outcome, as moderate minimum samples per leaf helps in averaging out noise which can build up at longer horizons, and improves the generalization capability of the trees by preventing them from fitting too closely to individual noisy observations.

Overall, Model 3 demonstrated superior performance in the in-sample grid search, yielding the most accurate forecasts for 2 out of the 3 horizons (1-step and 12-step). Notably, for the longest 12-step horizon, Model 3 achieved an in-sample RMSE of 42,967.32 CNY/mt, representing a substantial reduction of 23.9% compared to the next best performing model at that horizon.

#### 5.4 Model 1: One-Shot Adaptive Lasso

Table 3: Model 1: One-Shot Alasso Selection

| Horizon | Out-of-Sample |          | In-Sample |          |
|---------|---------------|----------|-----------|----------|
|         | RMSE          | MAPE (%) | RMSE      | MAPE (%) |
| 1-step  | 116,184       | 15.07    | 35,528    | 8.04     |
| 3-step  | 209,454       | 24.80    | 49,803    | 8.84     |
| 12-step | 322,851       | 45.21    | 70,996    | 19.17    |

*Note:* RMSE is measured in CNY/mt.

Table 3 summarizes the forecasting performance of Model 1 on the hold-out sample. The model demonstrates its highest accuracy for the 1-step horizon, with an out-of-sample MAPE of 15.07% and an RMSE of 116,184 CNY/mt. As the forecasting horizon increases, the model’s accuracy declines significantly. For the quarterly forecast (3-step horizon), the out-of-sample MAPE more than doubles to 24.8% and the RMSE rises to 209,454 CNY/mt. This trend continues for the annual forecast (12-step horizon), where the out-of-sample MAPE further increases to 45.21%, and the RMSE reaches 322,851 CNY/mt.

This observed decline in accuracy with longer horizons is consistent with the inherent challenges of generating accurate long-term forecasts compared to short-term ones.

The performance decline may also be influenced by the choice of using direct forecasting rather than a recursive approach. In direct forecasting, no intermediate data between  $t$  and  $t + h$  is utilized. However, it can be argued that using recursively forecasted values might lead to even worse performance due to the potential for error accumulation over time.

A notable observation from [Table 3](#) is the substantial discrepancy between the in-sample and out-of-sample performance metrics across all horizons. For instance- the 1-step in-sample MAPE is 8.04% compared to 15.07% out-of-sample. Similar differences are evident for the 3-step and 12-step horizons. While some decrease in performance is to be expected, the magnitude of these discrepancies suggests potential overfitting. Additionally, significant price increases observed within the hold-out window could contribute to this discrepancy and we will discuss this in [subsection 5.7](#). To mitigate overfitting issues, future research could also explore more conservative hyperparameters for the Random Forest model. It’s possible that the less conservative hyperparameter setting for the Adaptive Lasso in Model 1 contributed to this behaviour, suggesting the model might benefit from more conservative regularization.

### 5.5 Model 2: Post Double Selection

Table 4: Model 2: PDS Selection

| Horizon | Out-of-Sample |          | In-Sample |          |
|---------|---------------|----------|-----------|----------|
|         | RMSE          | MAPE (%) | RMSE      | MAPE (%) |
| 1-step  | 112,714       | 14.78    | 33,911    | 7.44     |
| 3-step  | 197,995       | 26.53    | 49,198    | 8.80     |
| 12-step | 318,118       | 49.04    | 63,757    | 14.06    |

*Note:* RMSE is measured in CNY/mt.

[Table 4](#) summarises the forecasting performance of Model 2, which utilizes PDS for feature selection, on the hold-out sample. Model 2 performs most accurately for the one-month-ahead (1 step) forecast, but its forecasting performance deteriorates significantly at higher forecasting horizons. This observed deterioration in accuracy at longer forecast horizons mirrors the patterns noted in Model 1.

A notable observation from [Table 4](#) is the significant discrepancy between the in-sample and out-of-sample performance metrics across all horizons. While some performance drop is expected in out-of-sample evaluation, the magnitude of these differences suggests

potential overfitting to the training data. However, this also illustrates the challenges inherent in longer-horizon forecasts, as the significant price increases observed in the out-of-sample window are difficult to predict.

### 5.6 Model 3: Post Double Selection with Granger Causality testing

Table 5: Model 3: PDS-GC Selection

| Horizon | Out-of-Sample |          | In-Sample |          |
|---------|---------------|----------|-----------|----------|
|         | RMSE          | MAPE (%) | RMSE      | MAPE (%) |
| 1-step  | 108,439       | 14.22    | 32,908    | 5.73     |
| 3-step  | 196,729       | 27.23    | 49,615    | 8.44     |
| 12-step | 328,326       | 46.85    | 44,154    | 10.40    |

*Note:* RMSE is measured in CNY/mt.

Table 5 presents the forecasting performance of Model 3, which incorporates PDS combined with Granger Causality (GC) testing for feature selection. Mirroring the first two models, Model 3 also performs best on one-step ahead forecasts with accuracy declining at longer horizons.

Despite Model 3’s design for statistical rigour, Table 5 reveals a substantial discrepancy between in-sample and out-of-sample performance metrics across all horizons. This gap widens dramatically for the 12-step-ahead forecast, where the in-sample MAPE of 10.4% contrasts sharply with the out-of-sample MAPE of 46.85%. This significant divergence underscores that even a statistically robust model, aimed at selecting predictive features, struggles to generalize well to highly volatile, unseen market conditions. The substantial price increases observed within the hold-out window, as noted for other models, present an inherent difficulty that challenges the predictive capabilities of any model. While Granger Causality identifies statistical predictive content, it does not necessarily equip the model to precisely forecast dramatic market shifts.

The observed high volatility during the out-of-sample period aligns with the findings of Proelss et al. (2020), who emphasize that the volatility of REE prices remains persistently high over time and therefore warrants explicit modelling. While our approach does not directly model volatility, the use of RF models allows for the capture of non-linear patterns in the data, some of which may be attributable to volatility. The discrepancy observed between in-sample and out-of-sample performance suggests that explicitly modelling volatility could have improved forecasting accuracy.

### 5.7 Comparison of the models

As expected, the models perform better during the in-sample period than in the out-of-sample period. Since the primary objective of this study is to evaluate how well the models forecast future REE prices - specifically Neodymium, all comparisons hereafter will be based solely on the out-of-sample period.

Comparing the model's performance for the monthly forecasts, Model 3 performs the best on both evaluation metrics, favouring the most sparse model among the three for the short-term horizon. This suggests that applying statistical inference through Granger causality testing has a positive effect on the model's predictive accuracy at this horizon. Given that the lagged response variable is the only predictor consistently selected in every iteration, it is regarded as the most optimal feature throughout the out-of-sample period. Thus, it appears to carry the most predictive power, placing greater emphasis on it relative to other predictors. Combined with the fact that the autocorrelation of neodymium is 0.97 at one lag, this supports the conclusion that fewer exogenous variables are required for short-term forecasting the price of neodymium. These results demonstrate the applicability of introducing regularisation methods together with statistical testing to reduce model dimensionality and improve the forecasting accuracy on a short-term horizon.

For the quarterly forecasts, Model 1 performs best in terms of MAPE (24.80%), while simultaneously yielding the highest RMSE value of 209,454 CNY/mt. This discrepancy between the metrics becomes more apparent when examining [Figure 5](#). For instance, at the start of 2022, Model 1 shows a larger absolute error compared to the other models, which contributes to the elevated RMSE value. The model's behaviour suggests it struggles to adapt quickly to sharp changes in certain periods, while accurately capturing the underlying trend in others. This highlights the challenge of forecasting during the out-of-sample period, which features greater volatility compared to previous periods in the data.

Since Model 1, on average, retains the second-highest number of features. This suggests that including too few features leads the model to lose important information for forecasting medium-term horizons, whereas including too many features reduces predictability. In contrast to the monthly forecasts, where statistical significance testing was favoured, the quarterly forecasts benefit from a simpler approach. Here, Model 1 removes a sufficient number of features using Adaptive Lasso in a single step, resulting in improved forecasting performance when evaluated using MAPE.

More interestingly, comparing Model 2 to Model 3 shows that the additional dimensionality reduction using Granger causality tests has little to no effect on improving forecasting accuracy at the 3-month horizon. Since Model 2 frequently retains nearly all predict-

ors, it effectively functions as a model without dimensionality reduction, in contrast to Model 3, which is frequently the most sparse model. Although the improvement in forecast accuracy is not significant, the exclusion of several variables may still be favourable, particularly if one of the objectives is to obtain a more parsimonious model.

For the annual forecasts, Model 1 once again yields the lowest MAPE value (45.21%) but has the second-highest RMSE value of 322,851 CNY/mt. However, due to the larger difference in MAPE between Model 1 and Model 2 (49.04%) compared to their RMSE difference (Model 2: 318,118 CNY/mt), we argue that Model 1 should be regarded as the better choice, rather than Model 2 based solely on its lower RMSE. By examining [Figure 5](#), it can be observed that Model 1's forecasts are more adaptive to changes in the true observations, whereas the other models tend to produce smoother, more generalized forecasts.

Overall, Model 2 does not perform best in any of the forecast horizons based on MAPE, and only once based on RMSE. As noted in the model selection results, Model 2 frequently retains all features, meaning that no effective dimensionality reduction is applied prior to the Random Forest. Therefore, the forecasting results suggest that more conservative regularization techniques are generally preferable. However, Model 2 is not consistently outperformed by both of the other models. For instance, in the monthly forecasts, Model 2 performs better than Model 1 but worse than Model 3, indicating that the type of dimensionality reduction applied has a meaningful impact. The purpose of applying the PDS procedure was to guard against omitted variable bias; however, given Model 1's performance on the medium- and long-term horizons, this does not appear to be a major concern in this case.

Granger causality testing was used to assess whether applying statistical significance to exclude non-relevant predictors would improve forecasting accuracy. Compared to Model 2, it does so in two-thirds of the horizons, but only in one-third when compared to Model 1. This suggests that Granger causality mostly adds value over Model 2, but not over the simpler process used in Model 1. Hence, this more extensive procedure isn't necessarily justified when the primary goal is to improve forecasting performance.

Since each model builds upon the feature selection applied in the preceding models, the effect of Granger causality testing may be influenced by this dependency. For instance, if Model 2 had used BIC to determine the hyperparameter for the adaptive Lasso in both stages, the union of retained features might have been smaller. In that case, applying inference on an already sparser model could have altered the outcome of the Granger causality testing.

Another aspect to consider, as mentioned in the methodology chapter, is that the threshold

value for the false discovery rate could be further investigated to optimise the effect of feature selection based on statistical inference.

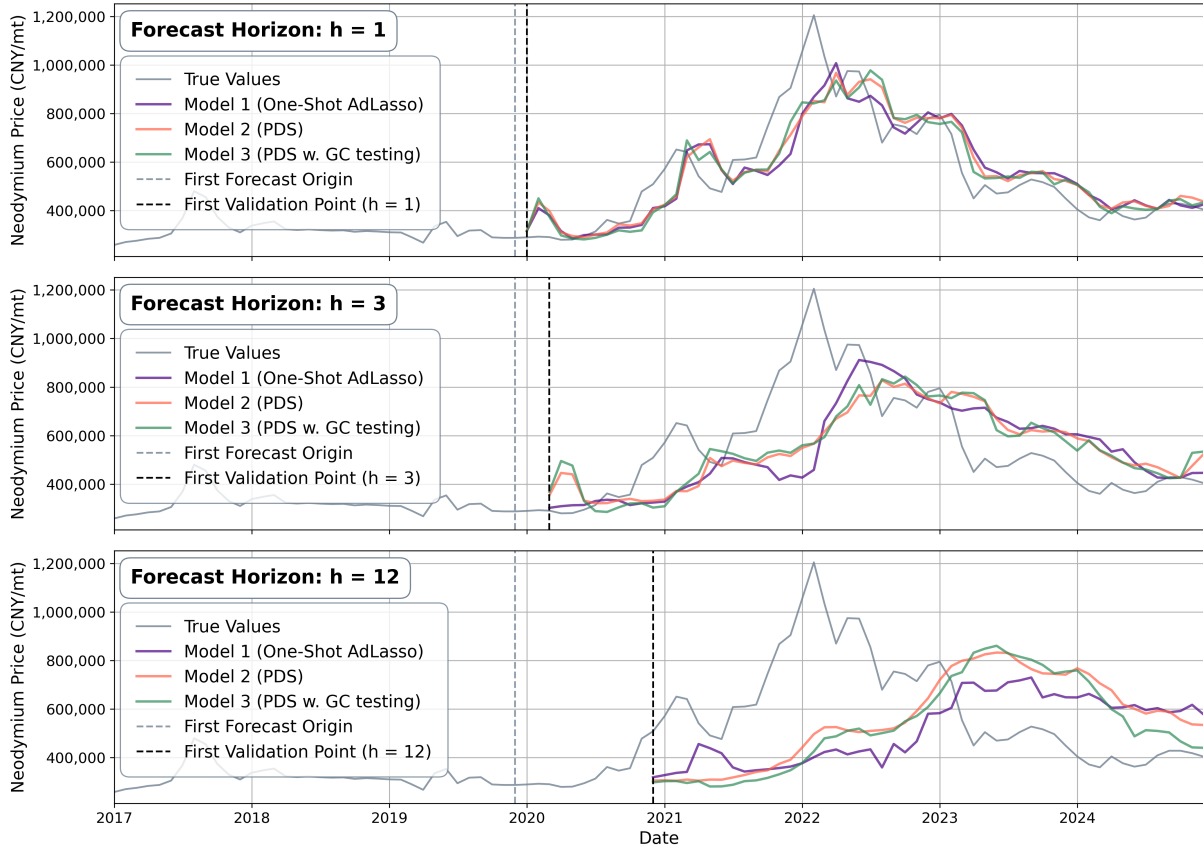


Figure 5: Forecasts of neodymium prices at horizons  $h = 1$ ,  $h = 3$ , and  $h = 12$  using both the PDS and One-Shot AdLasso selection methods. The true price series appears in gray, with Random Forest forecasts overlaid in green (PDS) and blue (AdLasso) lines. Dashed vertical lines mark the first forecast origin and the associated validation points. The plot highlights the alignment and comparative accuracy of multi-step-ahead forecasts relative to observed price dynamics.

It's apparent from Figure 5, as well as from the evaluation metrics of each model, that the 12-month long-term forecast deviates considerably from the true values. In particular, this is evident during the sharp price increase of neodymium from the second half of 2021 to early 2022. This discrepancy is likely due to the use of direct forecasting, where the absence of intermediate values in the forecasting process appears to affect the model's performance. However, the use of the alternative recursive approach, which incorporates forecasted values from  $t$  to  $t + 12$ , is not guaranteed to improve the forecasting accuracy and may therefore warrant further investigation.

Based on the evolution of the 12-month forecasts, it's clear that the behaviour of the true observations affects the forecasts with approximately the same lag period. For example, the period previously mentioned appears in the forecast one year later, with the forecasted trend increasing while the actual trend decreases. This pattern is not unique to the

annual forecasts but is also observed in the quarterly and monthly forecasts, a common phenomenon in time series forecasting.

## 6 CONCLUSION

This paper has proposed a set of sequentially built methods aimed at identifying the most effective approach for feature subset selection in forecasting the price of REEs. The analysis began with the application of Adaptive Lasso, which was then extended into a post-double-selection framework and further enhanced through the incorporation of Granger causality testing. These dimensionality reduction techniques were implemented using methods that avoid the need to transform all variables into stationarity, thereby preserving the underlying data structure.

Overall, our findings based on the out-of-sample period indicate that there is no single feature selection method that consistently enhances forecasting accuracy across all time horizons. When evaluating performance by forecast horizon, Model 3, which builds upon the feature selection strategies of the previous models and incorporates Granger causality testing, achieves the best results on both evaluation metrics for the short-term horizon. This suggests that incorporating traditional econometric significance tests prior to applying machine learning models may improve short-term forecasting accuracy for REEs prices. However, these conclusions are conditional on the specific time period and data used and should not be generalized to other contexts or forecast horizons without further validation.

For the quarterly and annual horizons, the feature selection results are indecisive to draw any conclusions on. The post-double-selection method does not yield consistently superior forecasting performance and fails to demonstrate clear advantages over the other approaches. Additionally, because it frequently retains all variables, it does not effectively fulfil its primary purpose of dimensionality reduction.

In closing, this study has presented a sequential framework for feature selection in REEs forecasting. While acknowledging that no single method consistently outperforms across all horizons, our findings highlight that using Granger causality testing in conjunction with PDS outperforms the AdLasso and the PDS on its own. However, the mixed performance observed at the quarterly and annual horizons underscores the need to scrutinise the applicability of these methods in different scenarios.

### *6.1 Limitations & Future Research*

Several limitations warrant consideration when interpreting the findings of this study, and these naturally lead to avenues for further research. Firstly, the reliance on a direct forecasting strategy, rather than a recursive approach for multi-step forecasts, may have influenced the observed performance, particularly at longer horizons where the absence of intermediate forecasted values could limit the model’s adaptability. Future work could compare these two forecasting methods in this context. Secondly, a notable discrepancy was observed in the in-sample and out-of-sample performance of the forecasts. This could be due to overfitting of the Random Forest models, or it could also be due to the high volatility observed in the hold-out window. To address this, future research could explore strategies to reduce overfitting in the Random Forest model, or investigate alternative regularisation techniques. Lastly, the generalizability of the findings is limited by the scope of the robustness checks performed. Specifically, the methodologies were not tested on other REEs beyond neodymium, nor were they applied to a broader range of commodity types, which could reveal different optimal feature selection strategies.

Building upon PDS with Granger causality testing, further exploration can be made of tuning the False Discovery Rate (FDR) threshold within the framework as a hyperparameter, potentially leading to more refined feature sets. Furthermore, investigating the performance of these feature selection methods on higher frequency data could provide valuable insights into their applicability for short-term forecasting. Additionally, a comparative analysis into the sensitivity of the results to window size could prove fruitful in shedding light on the trade-offs between coefficient adaptability and information retention. Also, further research could look into comparing static vs dynamic selection of features. Finally, extending the application of these methodologies to different types of data, beyond REEs, would contribute to a broader understanding of their effectiveness in diverse forecasting contexts.

# References

Akaike, Hirotugu (1998). Information Theory and an Extension of the Maximum Likelihood Principle. *Selected Papers of Hirotugu Akaike*. Ed. by Emanuel Parzen, Kunio Tanabe & Genshiro Kitagawa. New York, NY: Springer New York, pp. 199–213. ISBN: 978-1-4612-7248-9. DOI: [10.1007/978-1-4612-1694-0\\_15](https://doi.org/10.1007/978-1-4612-1694-0_15).

Batten, Jonathan A., Cetin Ciner & Brian M. Lucey (2010). The Macroeconomic Determinants of Volatility in Precious Metals Markets. *Resources Policy* 35.2, pp. 65–71. DOI: [10.1016/j.resourpol.2009.12.002](https://doi.org/10.1016/j.resourpol.2009.12.002).

Belloni, A., V. Chernozhukov & C. Hansen (2014). Inference on Treatment Effects after Selection among High-Dimensional Controls. *The Review of Economic Studies* 81.2, pp. 608–650. DOI: [10.1093/restud/rdt044](https://doi.org/10.1093/restud/rdt044).

Benjamini, Yoav & Yosef Hochberg (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57.1, pp. 289–300. DOI: [10.2307/2346101](https://doi.org/10.2307/2346101).

Breiman, Leo (2001). Random Forests. *Machine Learning* 45.1, pp. 5–32. DOI: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324).

Chen, Wei-Qiang, Matthew J. Eckelman, Benjamin Sprecher, Wei Chen & Peng Wang (2024). Interdependence in Rare Earth Element Supply between China and the United States Helps Stabilize Global Supply Chains. *One Earth* 7.2, pp. 242–252. DOI: [10.1016/j.oneear.2024.01.011](https://doi.org/10.1016/j.oneear.2024.01.011).

Davis, Richard A., Pengfei Zang & Tian Zheng (2016). Sparse Vector Autoregressive Modeling. *Journal of Computational and Graphical Statistics* 25.4, pp. 1077–1096. DOI: [10.1080/10618600.2015.1092978](https://doi.org/10.1080/10618600.2015.1092978).

Depraiter, Lisa & Stephane Goutte (2023). The Role and Challenges of Rare Earths in the Energy Transition. *Resources Policy* 86, p. 104137. DOI: [10.1016/j.resourpol.2023.104137](https://doi.org/10.1016/j.resourpol.2023.104137).

Dormann, Carsten F., Jane Elith, Sven Bacher, Carsten Buchmann, Gudrun Carl, Gabriel Carré, Jaime R. García Marquéz, Bernd Gruber, Bruno Lafourcade, Pedro J. Leitão, Tamara Münkemüller, Colin McClean, Patrick E. Osborne, Björn Reineking, Boris Schröder, Andrew K. Skidmore, Damaris Zurell & Sven Lautenbach (2013). Collinearity: A Review

of Methods to Deal with It and a Simulation Study Evaluating Their Performance. *Eco-  
graphy* 36.1, pp. 27–46. DOI: [10.1111/j.1600-0587.2012.07348.x](https://doi.org/10.1111/j.1600-0587.2012.07348.x).

Gargano, Antonio & Allan Timmermann (2014). Forecasting Commodity Price Indexes Using Macroeconomic and Financial Predictors. *International Journal of Forecasting* 30.3, pp. 825–843. DOI: [10.1016/j.ijforecast.2013.09.003](https://doi.org/10.1016/j.ijforecast.2013.09.003).

Goodenough, Kathryn M., Frances Wall & David Merriman (2018). The Rare Earth Elements: Demand, Global Resources, and Challenges for Resourcing Future Generations. *Natural Resources Research* 27.2, pp. 201–216. DOI: [10.1007/s11053-017-9336-5](https://doi.org/10.1007/s11053-017-9336-5).

Guo, Qing & Zishan Mai (2024). How Do Seasonal, Significant Events, and Policies Affect China’s REE Export Prices? Based on Deep Learning Perspective. *Resources Policy* 96, p. 105205. DOI: [10.1016/j.resourpol.2024.105205](https://doi.org/10.1016/j.resourpol.2024.105205).

Hecq, Alain, Luca Margaritella & Stephan Smeekes (2023). *Inference in Non-stationary High-Dimensional VARs*. Preprint. arXiv: [2302.01434](https://arxiv.org/abs/2302.01434) [econ.EM]. <https://arxiv.org/abs/2302.01434> [Accessed 04/14/2025].

Henriques, Irene & Perry Sadorsky (2023). Forecasting Rare Earth Stock Prices with Machine Learning. *Resources Policy* 86, p. 104248. DOI: [10.1016/j.resourpol.2023.104248](https://doi.org/10.1016/j.resourpol.2023.104248).

Herrera, Gabriel Paes, Michel Constantino, Benjamin Miranda Tabak, Hemerson Pistori, Jen-Je Su & Athula Naranpanawa (2019). Long-Term Forecast of Energy Commodities Price Using Machine Learning. *Energy* 179, pp. 214–221. DOI: [10.1016/j.energy.2019.04.077](https://doi.org/10.1016/j.energy.2019.04.077).

IEA (2021). *The Role of Critical Minerals in Clean Energy Transitions*. International Energy Agency. <https://www.iea.org/reports/the-role-of-critical-minerals-in-clean-energy-transitions/mineral-requirements-for-clean-energy-transitions> [Accessed 05/09/2025].

Jackson, Lewis, Amy Lv, Eric Onstad, Ernest Scheyder, Lewis Jackson & Ernest Scheyder (2025). *China Hits Back at US Tariffs with Export Controls on Key Rare Earths*. <https://www.reuters.com/world/china-hits-back-us-tariffs-with-rare-earth-export-controls-2025-04-04/> [Accessed 05/11/2025].

Ji, Lei, Yingchao Zou, Kaijian He & Bangzhu Zhu (2019). Carbon Futures Price Forecasting Based with ARIMA-CNN-LSTM Model. *7th International Conference on Information Technology and Quantitative Management (ITQM 2019): Information tech-*

*nology and quantitative management based on Artificial Intelligence* 162, pp. 33–38. DOI: [10.1016/j.procs.2019.11.254](https://doi.org/10.1016/j.procs.2019.11.254).

Lee, Ji Hyung, Zhentao Shi & Zhan Gao (2022). On LASSO for Predictive Regression. *Journal of Econometrics* 229.2, pp. 322–349. DOI: [10.1016/j.jeconom.2021.02.002](https://doi.org/10.1016/j.jeconom.2021.02.002).

Lin, Lin, Fang Wang, Xiaolong Xie & Shisheng Zhong (2017). Random Forests-Based Extreme Learning Machine Ensemble for Multi-Regime Time Series Prediction. *Expert Systems with Applications* 83, pp. 164–176. DOI: [10.1016/j.eswa.2017.04.013](https://doi.org/10.1016/j.eswa.2017.04.013).

Madziwa, Lawrence, Mallikarjun Pillalamarry & Snehamoy Chatterjee (2022). Gold Price Forecasting Using Multivariate Stochastic Model. *Resources Policy* 76, p. 102544. DOI: [10.1016/j.resourpol.2021.102544](https://doi.org/10.1016/j.resourpol.2021.102544).

Ogasawara, Eduardo, Leonardo C. Martinez, Daniel De Oliveira, Geraldo Zimbrao, Gisele L. Pap & Marta Mattoso (2010). Adaptive Normalization: A Novel Data Normalization Approach for Non-Stationary Time Series. *The 2010 International Joint Conference on Neural Networks (IJCNN)*. 2010 International Joint Conference on Neural Networks (IJCNN). Barcelona, Spain: IEEE, pp. 1–8. ISBN: 978-1-4244-6916-1. DOI: [10.1109/IJCNN.2010.5596746](https://doi.org/10.1109/IJCNN.2010.5596746).

Proelss, Juliane, Denis Schweizer & Volker Seiler (2020). The Economic Importance of Rare Earth Elements Volatility Forecasts. *International Review of Financial Analysis* 71, p. 101316. DOI: [10.1016/j.irfa.2019.01.010](https://doi.org/10.1016/j.irfa.2019.01.010).

Riesgo García, María Victoria, Alicja Krzemień, Miguel Ángel Manzanedo del Campo, Carmen Escanciano García-Miranda & Fernando Sánchez Lasheras (2018). Rare Earth Elements Price Forecasting by Means of Transgenic Time Series Developed with ARIMA Models. *Sustainable management and exploitation of extractive waste: towards a more efficient resource preservation and waste recycling* 59, pp. 95–102. DOI: [10.1016/j.resourpol.2018.06.003](https://doi.org/10.1016/j.resourpol.2018.06.003).

Sánchez Lasheras, Fernando, Francisco Javier De Cos Juez, Ana Suárez Sánchez, Alicja Krzemień & Pedro Riesgo Fernández (2015). Forecasting the COMEX Copper Spot Price by Means of Neural Networks and ARIMA Models. *Resources Policy* 45, pp. 37–43. DOI: [10.1016/j.resourpol.2015.03.004](https://doi.org/10.1016/j.resourpol.2015.03.004).

Schwarz, Gideon (1978). Estimating the Dimension of a Model. *The Annals of Statistics* 6.2. DOI: [10.1214/aos/1176344136](https://doi.org/10.1214/aos/1176344136).

- Shanghai Metal Market (2025). *Major Change! China Northern Rare Earth Cancels Listing Prices and Adopts Real-Time Quotes from the Rare Earth Exchange. Is the "Baotou Index" for Rare Earth Price Trading on the Agenda?* - Shanghai Metal Market. <https://www.metal.com/en/newscontent/103123341> [Accessed 05/10/2025].
- Stock, James H & Mark W Watson (2001). Vector Autoregressions. *Journal of Economic Perspectives* 15.4, pp. 101–115. DOI: [10.1257/jep.15.4.101](https://doi.org/10.1257/jep.15.4.101).
- Tibshirani, Robert (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 58.1, pp. 267–288. DOI: [10.1111/j.2517-6161.1996.tb02080.x](https://doi.org/10.1111/j.2517-6161.1996.tb02080.x).
- Toda, Hiro Y. & Taku Yamamoto (1995). Statistical Inference in Vector Autoregressions with Possibly Integrated Processes. *Journal of Econometrics* 66.1, pp. 225–250. DOI: [10.1016/0304-4076\(94\)01616-8](https://doi.org/10.1016/0304-4076(94)01616-8).
- Truong, Charles, Laurent Oudre & Nicolas Vayatis (2020). Selective Review of Offline Change Point Detection Methods. *Signal Processing* 167, p. 107299. DOI: [10.1016/j.sigpro.2019.107299](https://doi.org/10.1016/j.sigpro.2019.107299).
- U.S. Geological Survey (2025). *Mineral Commodity Summaries 2025*. Report 2025 V1.2. Reston, VA: United States Geological Survey, p. 212. <https://pubs.usgs.gov/publication/mcs2025>.
- U.S. Geological Survey (n.d.). *Rare Earths Statistics and Information*. <https://www.usgs.gov/centers/national-minerals-information-center/rare-earths-statistics-and-information> [Accessed 05/18/2025].
- United Nations, Statistics Division (2025). *UN Comtrade Database: HS 280530 (Inorganic Chemicals)*. [https://data.un.org/Data.aspx?d=ComTrade&f=\\_11Code%3A29%3BcmdCode%3A280530](https://data.un.org/Data.aspx?d=ComTrade&f=_11Code%3A29%3BcmdCode%3A280530) [Accessed 05/10/2025].
- Zou, Hui (2006). The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association* 101.476, pp. 1418–1429. DOI: [10.1198/016214506000000735](https://doi.org/10.1198/016214506000000735).

## Appendix A: Tables

Table A1: The 17 rare-earth elements (REEs) and their principal applications

| Rare-earth element | Primary applications                                                   |
|--------------------|------------------------------------------------------------------------|
| Scandium (Sc)      | Ceramics; fuel cells; metallurgy                                       |
| Yttrium (Y)        | Catalysts; ceramics; lasers; metallurgy; phosphors                     |
| Lanthanum (La)     | Batteries; catalysts; ceramics; glass; metallurgy                      |
| Cerium (Ce)        | Catalytic converters; ceramics; glass; metallurgy; polishing compounds |
| Praseodymium (Pr)  | Aerospace alloys; batteries; ceramics; colorants; permanent magnets    |
| Neodymium (Nd)     | Catalysts; lasers; permanent magnets                                   |
| Promethium (Pm)    | Nuclear batteries; luminous paint                                      |
| Samarium (Sm)      | Cancer treatments; nuclear; permanent magnets                          |
| Europium (Eu)      | Nuclear control rods; phosphors                                        |
| Gadolinium (Gd)    | Medical imaging; metallurgy; permanent magnets                         |
| Terbium (Tb)       | Fiber optics; lasers; permanent magnets; solid-state devices           |
| Dysprosium (Dy)    | Data storage devices; lasers; permanent magnets                        |
| Holmium (Ho)       | Lasers; nuclear control rods; permanent magnets                        |
| Erbium (Er)        | Fiber optics; glass colorant; lasers; optical amplifiers               |
| Thulium (Tm)       | Lasers; metallurgy                                                     |
| Ytterbium (Yb)     | Catalysts; lasers; metallurgy; scintillators                           |
| Lutetium (Lu)      | Cancer therapies; electronics; medical imaging                         |

*Source: Adapted from the U.S. Geological Survey's Mineral Commodity Summaries 2025*

Table A2: Grid search results for Random Forest models

| Model          | Parameter            | h=1       | h=3       | h=12      |
|----------------|----------------------|-----------|-----------|-----------|
| <b>Model 1</b> | Lags                 | 1         | 1         | 12        |
|                | RMSE                 | 35,528.22 | 49,506.69 | 60,897.31 |
|                | Max depth            | 8         | 8         | 8         |
|                | Min samples per leaf | 1         | 5         | 5         |
|                | Estimators           | 300       | 1000      | 300       |
| <b>Model 2</b> | Lags                 | 1         | 12        | 1         |
|                | RMSE                 | 33,911.28 | 47,796.31 | 56,485.38 |
|                | Max depth            | None      | 8         | 8         |
|                | Min samples per leaf | 1         | 3         | 5         |
|                | Estimators           | 300       | 300       | 300       |
| <b>Model 3</b> | Lags                 | 1         | 1         | 12        |
|                | RMSE                 | 33,115.37 | 49,867.57 | 42,967.32 |
|                | Max depth            | 4         | 4         | 8         |
|                | Min samples per leaf | 1         | 5         | 5         |
|                | Estimators           | 1000      | 300       | 800       |

*Note: RMSE values are based on in-sample data.*

Table A3: Grid Search Hyperparameter Values for Random Forest

| Hyperparameter | Description                                                     | Values              |
|----------------|-----------------------------------------------------------------|---------------------|
| Estimators     | Defines the number of trees in the model                        | 300, 500, 800, 1000 |
| Depth          | Controls the maximum depth of each tree                         | None, 4, 8          |
| Features       | Limits the number of features to consider when splitting a node | "sqrt"              |
| Samples leaf   | The minimum number of samples required to be at a leaf node     | 1, 3, 5             |
| Lags           | The number of feature lags to include                           | 1, 3, 6, 12         |

## Appendix B: Figures

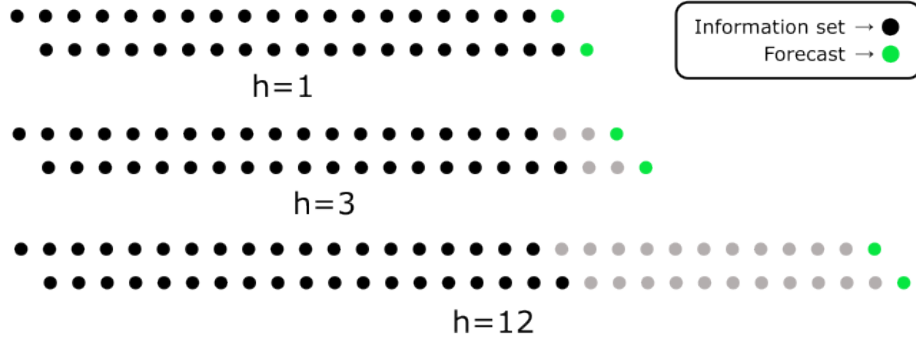


Figure B1: Forecasting scheme for horizons  $h = 1$ ,  $h = 3$ , and  $h = 12$ , where each row pair represents a rolling evaluation window. Black dots denote the information set used for model estimation, and green dots mark the forecast targets at each horizon. Gray dots indicate data, outside the information set, that lies beyond the forecast origin but within the forecast horizon. This setup illustrates how the forecasting challenge increases with longer horizons.

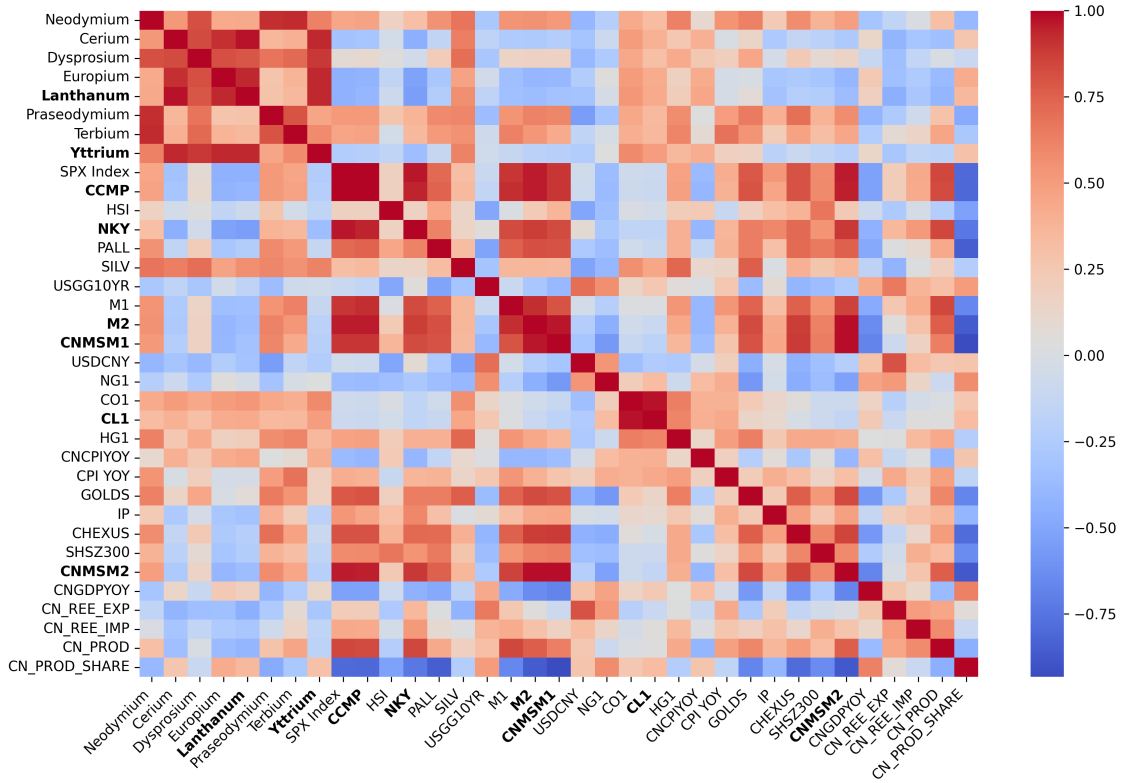


Figure B2: Pearson correlation matrix of all model variables, ordered by hierarchical clustering, with darker reds indicating stronger positive correlations and darker blues stronger negative correlations; variables shown in bold exhibit pairwise correlations above 0.95 and are therefore excluded from subsequent analyses.

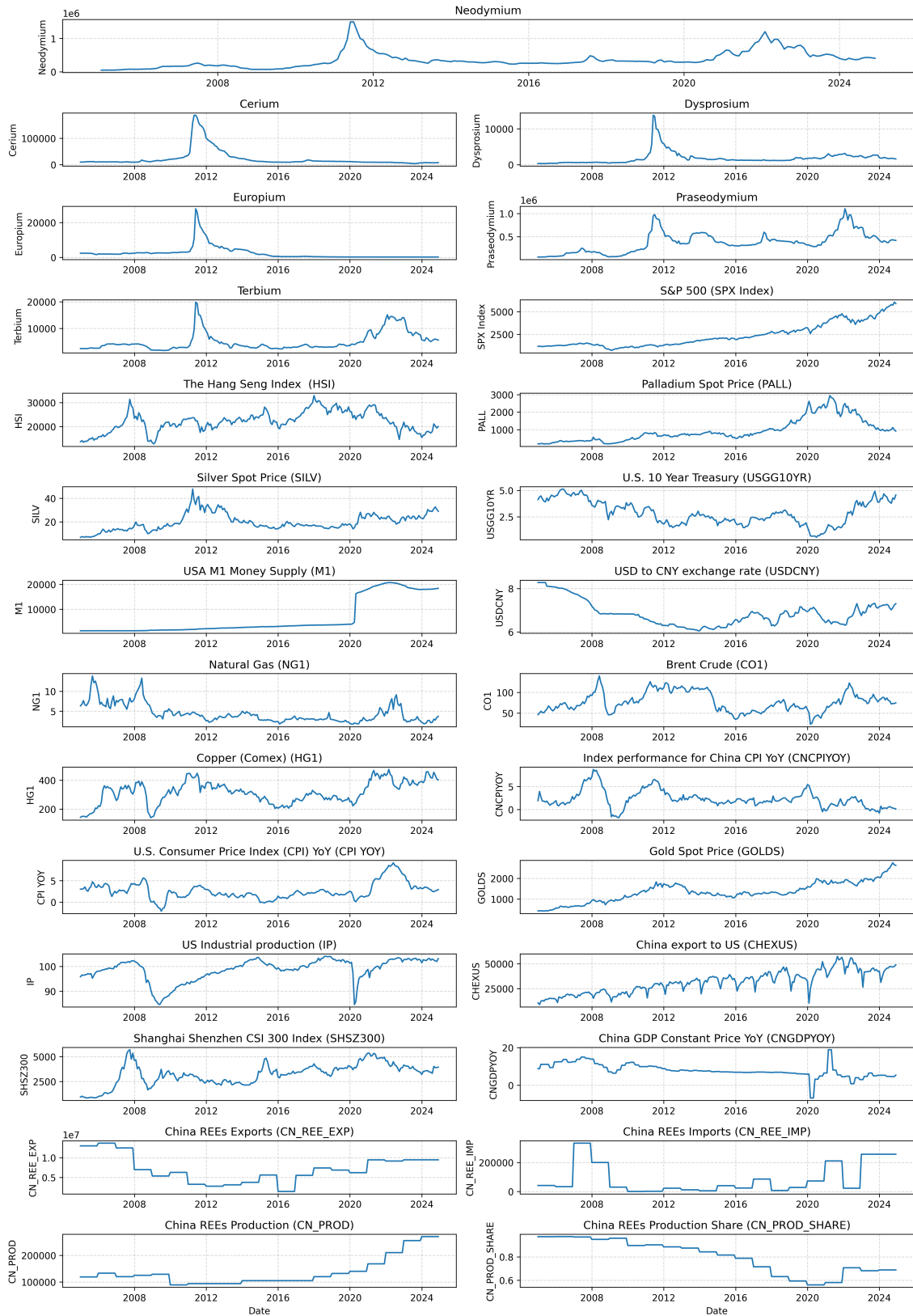


Figure B3: Time series of all variables used in the study (2005–2024). Each line corresponds to a model variable, illustrating trends and changes over time.

*Data Sources: Bloomberg Terminal, U.S. Geological Survey, U.N. Data*



Figure B4: Granger causality test results for Model 3 across rolling estimation windows at forecast horizons  $h = 1$  (top),  $h = 3$  (middle), and  $h = 12$  (bottom). Cells shaded green denote variables for which the null hypothesis of no Granger causality is rejected at a false discovery rate of 0.10, after applying the Benjamini–Hochberg procedure; red cells denote non-significant results.

## Appendix C: Generative AI Usage

This document details the use of artificial intelligence tools in this thesis. This use aligns with guidelines from the Department of Economics, Lund University School of Economics and Management.

Artificial Intelligence Tools Used: ChatGPT, Claude, Gemini.

These tools were used in the following way: Generative AI tools supported practical tasks in this thesis. Responsibility for all content, decisions, and interpretations remained with the authors.

ChatGPT and Claude provided aid with coding of the Python scripts used in the research. This included help with fixing code errors, generating template code for further work, and explaining programming concepts relevant to empirical analysis and data handling. All code and solutions were reviewed, understood, and adjusted by the authors. This support affected Section 4 (Methodology) and Section 5 (Empirical Analysis), specifically in implementing feature selection algorithms and forecasting models.

Gemini was used for grammar and writing improvements. Its application involved making sentences clearer, improving word choice, and checking spelling throughout the thesis. The authors retained full control over the final text, using suggestions to enhance clarity and academic precision.

No confidential or copyrighted material was transmitted to these tools. All analysis, findings, and interpretations were performed solely by the authors.