UNIVERSITY OF TARTU
Faculty of Social Sciences
School of Economics and Business Administration

Xinran Zhang

FORECASTING CHINA'S TOTAL ENERGY CONSUMPTION USING MACHINE LEARNING MODELS

Master's Thesis
Supervisor: Luca Alfieri, PhD

Tartu 2026

I have written this Research paper independently. Any ideas or data taken from other authors or other sources have been fully referenced.

Table of contents

Abstract

In the context of China's "dual carbon" goals and energy transition, forecasting energy consumption has become an important issue for policy analysis. This study develops a forecasting framework for China's annual macroeconomic data by combining machine learning methods with explainable analysis techniques. Variables are collected from multiple areas, including the economy, industry, energy, and international trade. Lagged variables are further introduced to capture time dependence in macroeconomic series, while mutual information is used for feature screening. Several machine learning models, including Ridge, Lasso, SVR, Random Forest, XGBoost, and KNN, are compared with a benchmark AR(1) model. Model performance is evaluated through rolling cross-validation and pseudo-out-of-sample forecasting. The empirical results indicate that, under small-sample macroeconomic conditions, regularized linear models perform more steadily than more flexible machine learning approaches. Among them, Ridge achieves forecasting accuracy comparable to the AR(1) benchmark and exhibits relatively stable performance across different forecasting settings. To examine model interpretability, SHAP is employed to analyze the contribution of variables from both feature and temporal perspectives. The results suggest that energy supply conditions, openness-related indicators, and real economic activity are closely associated with changes in energy consumption. Most models also rely more heavily on short-term lagged information, highlighting the importance of recent observations in forecasting. Overall, the combination of forecasting models and SHAP analysis helps clarify the mechanisms behind model predictions and provides an additional perspective for energy policy analysis in China.

Keywords: energy consumption; machine learning; forecasting

JEL:C41, C45, C53

Introduction

Energy plays an important role in economic growth and social development (Akintande et al., 2020). Previous studies suggest that the relationship between energy consumption and economic growth is not stable across countries or periods, and its mechanisms may differ under different stages of economic development (Esso, 2010). In China, the carbon peaking and carbon neutrality goals (J. Xi, 2020), have further increased the importance of energy consumption analysis, since energy use is closely related not only to economic activity but also to emission reduction and energy structure adjustment. Against this background, improving the accuracy of energy consumption forecasting has become increasingly relevant for policy analysis.

Research on energy consumption forecasting has undergone noticeable changes over the past decades. Earlier studies mainly focused on traditional time-series and econometric approaches. For example, Nowicka-Zagrajek and Weron (2002) use an Autoregressive Moving Average (ARMA) model to forecast electricity load. More recently, machine learning methods have been introduced into energy forecasting studies. Neubauer et al. (2025), for instance, incorporate multiple exogenous variables and apply models such as neural networks and decision trees to model nonlinear relationships and interactions among variables.

Despite ongoing advances in forecasting methods, existing studies still face two main limitations. First, many machine learning models are often treated as black boxes and lack systematic interpretability mechanisms, making it difficult to compare the roles of variables across different models (Lecuivre, 2022). Second, although lagged variables are widely used in modeling, the extent to which models rely on information from different time periods remains unclear (Leung et al., 2023).

To address these issues, this study aims to develop an integrated framework for forecasting and interpretability of China's energy consumption using machine learning methods. Specifically, this study evaluates the forecasting performance of several machine learning models in a pseudo-out-of-sample framework and further examines how the models leverage information from both variable and temporal dimensions using explainable machine learning techniques.

Mutual information is used for feature selection (Malik et al., 2025), and the pseudo-out-of-sample forecasting framework is employed to evaluate model performance (Campbell & Thompson, 2008). On the interpretability side, SHapley Additive exPlanations (SHAP) is used to analyze the model across both variable and temporal dimensions (Neubauer et al., 2025). The global SHAP analysis identifies variables that consistently rank as important across models, capturing key factors that robustly influence changes in China’s energy consumption. SHAP values of lagged variables are grouped by lag order to examine how models depend on information from different time periods.

Based on the framework above, this study differs from existing research in two aspects. First, rather than focusing on a single forecasting model (Zhao & Liu, 2019), it compares several machine learning methods within the same rolling forecasting setting using China’s annual macroeconomic data. Given the limited sample size of macro-level time-series data, the comparison also helps evaluate the stability of different forecasting approaches under small-sample conditions. Second, the study introduces explainable machine learning methods into the analysis of China’s energy consumption forecasting. Existing studies mainly discuss the importance of explanatory variables (Ma, 2019; C. Liu & Qian, 2023), whereas this study further examines how models use information across different lag periods. This allows observing the relative importance of short- and long-term information in the forecasting process.

The empirical results suggest that, for China’s annual macroeconomic data, regularized linear models tend to perform more steadily in out-of-sample forecasting than more flexible machine learning models. Recent lagged information appears to play a larger role in prediction than longer-term historical information. These results provide additional evidence for the application of interpretable machine learning methods in macroeconomic forecasting under limited-data conditions and offer a useful perspective for energy policy analysis in China.

The rest of the study is organized as follows. Section 1 reviews the literature. Section 2 introduces the data. Section 3 presents the methodology. Section 4 shows the results and analysis. Section 5 concludes the study.

1. Literature Review

1.1 Determinant of Energy Consumption

From a macro-level perspective, energy consumption is shaped by a range of interrelated factors. Existing literature can be broadly categorized into three dimensions: economic, industrial, and socio-demographic factors, which together influence energy consumption across countries and over time. The following sections review these determinants in turn.

Many studies have investigated the relationship between economic factors and energy consumption, yet no clear consensus has emerged. First, studies on the causal relationship between energy consumption and economic growth report mixed results. Kouakou (Kouakou, 2011) concludes that, in the short run, an increase in electricity consumption promotes economic growth. However, in the long run, electricity consumption is found to Granger-cause economic growth. Furthermore, structural changes in the economy can be treated as long-run driving variables explaining electricity demand in Ghana (Adom et al., 2012). Lee & Chien (2010) focus on the G-7 countries and examine the dynamic linkages among energy consumption, capital stock, and real income (real GDP per capita). Their results show a unidirectional relationship from energy consumption to real income in Canada, Italy, and the UK, while causality runs from real income to energy consumption in France and Japan, and no causality is found in Germany and the UK, and the impacts of real income on energy consumption are only significant in Canada and the UK. Second, the results vary not only across countries but also over time. Esso (2010) finds that 1988 is a turning point for seven Sub-Saharan African countries. Specifically, the significant positive long-run impact of economic growth turned negative after 1988 in Ghana and South Africa. Third, cross-country heterogeneity plays an important role. In studies on China, Bloch et al. (2015) show that energy consumption and economic growth influence each other, and similar findings are also reported for UFM countries in the study by Esseghir & Haouaoui Khouni (2014). Esso (2010) also finds that bidirectional causality between energy consumption and real GDP is observed in Côte d'Ivoire, while unidirectional causality from real GDP to energy consumption is observed in Congo and Ghana.

In addition, other economic variables also influence energy consumption. Dokas et al. (2022) conclude that trade openness has a negative impact on energy consumption in developed countries and a positive impact in developing countries. More interestingly, Canh et al. (2021) show, using a global sample, that the shadow economy—a higher level of the informal sector—

increased energy consumption. Its impact is stronger in low- and middle-income countries, but weaker in high-income countries. In contrast, its effect on renewable energy consumption is weaker in low-, lower-middle-, and high-income countries, but stronger in upper-middle-income countries.

The impact of financial development on energy consumption is also important (Sadorsky, 2010). Omri & Kahouli (2014) manage to highlight that FDI exerts a stronger effect on energy consumption in lower-income countries than in high-income countries. However, evidence from Pakistan indicates that FDI's contribution to electricity consumption is smaller than that of income and population growth (Zaman et al., 2012).

Existing studies suggest that changes in energy consumption can be decomposed into activity, structure, and intensity effects (Akyürek, 2020). The activity effect refers to the expansion of industrial output, which typically increases energy consumption. For example, Steenhof (2006) finds that, between 1998 and 2002, rising industrial activity and fuel shifts in China contributed to higher industrial electricity demand, although part of this increase was offset by improvements in energy efficiency. The structure effect captures changes in the composition of industrial sectors, which can either increase or reduce energy consumption depending on the relative share of energy-intensive industries. Wang et al. (2019), using the Logarithmic Mean Divisia Index (LMDI) method, confirm that industrial structure is one of the key drivers of energy consumption. The intensity effect reflects changes in energy efficiency. Improvements in energy productivity are often associated with lower energy consumption. However, Li et al. (2020) show that energy use may instead increase because of the rebound effect originally discussed by Jevons & Flux (1965).

Socio-demographic factors are also closely related to energy consumption patterns. Population growth is recognized as a primary driver of energy consumption. Akintande et al. (2020) provide empirical evidence by developing a renewable energy consumption model using annual data from 1996 to 2016 for Ethiopia, South Africa, Nigeria, the DR Congo, and Egypt. Urbanization is another key determinant. Mahalik & Mallick (2014) focus primarily on how significantly the proportion of the urban population in India affects long-run energy consumption. In contrast, financial development, economic growth rate, and industrialization have negative impacts. Similarly, when Adom et al. (2012)analyze the situation in Ghana, they conclude that the degree of urbanization is the main driver of aggregate domestic electricity demand. Human capital also affects energy consumption, although its impact is more complex. Yao et al. (2019) examine selected OECD economies over the period 1965–2014 and find that human capital reduces energy consumption and promotes the use of clean energy through income, technology, and factor complementarity. The effects show cross-country heterogeneity and are mainly observed in the long run.

Despite population growth and urbanization, behavioral factors may be another significant driver of energy consumption. Thomas & Rosenow (2020) summarize that energy demand is greatly influenced by people's lifestyles, spending habits, and transportation choices. For example, households with higher consumption levels often use more electrical appliances, occupy larger living spaces, and travel more frequently, all of which are associated with higher energy use. Transport choices also matter. Using private vehicles generally consumes more energy than public transportation.

Table 1 - Summary of Energy Consumption Determinants

Category	Variables Suggested	Representative Studies
Economic factors	GDP, economic growth, real income, capital stock, trade openness, exports, shadow economy, FDI, financial development	Kouakou (2011); Lee & Chien (2010); Dokas et al. (2022); Sadorsky (2010); Omri & Kahouli (2014); Canh et al. (2021)
Industrial factors	Industrial output, industrial activity, industrial structure, fuel shifts, energy intensity, energy efficiency, energy productivity, rebound effect	Steenhof (2006); Wang et al. (2019); Li et al. (2020)
Socio-demographic factors	Population growth, urbanization, human capital, lifestyle, spending habits, transportation choices, consumption behavior	Akintande et al. (2020); Mahalik & Mallick (2014); Yao et al. (2019); Thomas & Rosenow (2020)

In summary, the existing literature suggests that energy consumption is influenced by a broad set of economic, industrial, and socio-demographic factors. Taken together, these variables can form a core set of exogenous inputs for energy forecasting models, providing a theoretical basis for variable selection in this field. Table 1 summarizes the main categories of variables identified in the existing literature.

1.2 Forecasting Methods

Previous studies on energy consumption forecasting have mainly relied on three approaches: traditional statistical methods, machine learning techniques, and hybrid models. Their performance often differs depending on data characteristics and model structure. The next sections discuss these approaches separately.

1.2.1 Traditional Method

Existing methods for energy consumption forecasting can be broadly classified into three categories: time series models, econometric models, and simulation and scenario analysis methods. Time-series models mainly rely on historical observations and autocorrelation patterns for forecasting. Sen et al. (2016) examines energy consumption and greenhouse gas (GHG) emissions in the Indian iron and steel industry using ARIMA models. Their results suggest that different indicators may require different model specifications. For energy consumption, ARIMA(1,0,0)×(0,1,1) performs relatively well, whereas ARIMA(0,1,4)×(0,1,1) is more suitable for GHG emissions. Nowicka-Zagrajek & Weron (2002) develop a short-term load forecasting approach based on a heavy-tailed ARMA model and data preprocessing techniques. Compared with conventional models that rely on normality assumptions, their approach finds that hyperbolic distributions may better describe the data and improve forecasting accuracy. In addition, De Oliveira & Cyrino Oliveira (2018) show that combining ARIMA with methods such as bagging and exponential smoothing can improve electricity demand forecasting across different countries.

Econometric approaches mainly focus on the relationship between energy consumption and economic variables. Earlier studies commonly used methods such as multiple linear regression (MLR), transfer function models, exponential smoothing techniques, and expert systems. These methods are generally effective for relatively simple linear relationships but may be less suitable for more complex nonlinear problems. Moghram & Rahman (1989) compare several forecasting methods using the same dataset and report noticeable differences in forecasting performance across models. Clements et al. (2016) propose a multiple-equation time-series framework in which different periods within a day are modeled separately. The model incorporates daily and weekly seasonal patterns as well as lagged load information to describe dynamic relationships. Although the framework remains linear and is estimated using ordinary least squares (OLS), the results suggest that it can still achieve relatively stable forecasting performance.

Simulation and scenario-based approaches are commonly applied in long-term energy forecasting research. These methods usually consider policy adjustment, industrial linkages, and structural changes in the economy. Zhong et al. (2021) construct a dynamic input–output simulation model to examine household consumption-driven indirect energy consumption (IEC) under urbanization. The framework incorporates energy, material, and economic balances to describe interactions across different sectors. By examining changes in consumption structure from 2005 to 2017 and simulating trends from 2018 to 2030, the results show that urban consumption-driven IEC increases significantly, and the gap between urban and rural areas continues to grow. Fang et al. (2017) consider the relationships among energy, the economy, and the environment, including their interactions and feedback effects, and treat economic development as the main driver of energy demand. The energy policy control factor is also included. Based on this, a system dynamics simulation model is built to predict and give early warning of China's energy demand. The model outputs include total energy consumption, different types of energy consumption, CO2 emissions, and several energy warning indicators.

1.2.2 Machine Learning and Deep Learning Methods

With the development of data-driven techniques, machine learning and deep learning methods have been widely applied in energy consumption forecasting. Shirzadi et al. (2021) conduct several machine models support vector machine (SVM) and Random Forest (RF), and deep learning methods non-linear auto-regressive exogenous (NARX) neural network and recurrent neural networks (Long Short-Term Memory), to compare the district level models for predicting the electrical load demand with a dataset containing nine years of historical load demand for Bruce County, Ontario, Canada. The results reveal that the deep learning model is more competitive than the machine learning model, with an R-Squared of about 0.93–0.96 and a Mean Absolute Percentage Error (MAPE) of about 4–10%. Bagherzadeh et al. (2021) predict the energy consumption of East Melbourne wastewater treatment plants. Gradient Boosting Machine (GBM) performs best among neural networks and decision trees, demonstrating its ability to model nonlinear and irregular patterns.

Rahman et al. (2018) use a deep recurrent neural network based on Long Short-Term Memory (LSTM) to predict electricity consumption in commercial and residential buildings over the medium- and long-term. The method can capture nonlinear patterns and long-term dependencies in time series and can improve prediction accuracy to some extent. More importantly, it can handle missing data. Other neural network models have also been applied in energy forecasting. For example, Wang & Li (2010) propose a new energy consumption prediction model based on the least squares support vector machine (LS-SVM) regression model, and verify its effectiveness using input variables such as GDP, population, industrial structure, imports and exports, and government expenditure of China. Karatasou et al. (2006) use an Artificial Neural Network (ANN) to predict building energy consumption. By including time and weather variables, the model can capture nonlinear relationships in energy demand. In addition, Szoplik (2015) applies an ANN-based multilayer perceptron (MLP) model to estimate natural gas consumption. The model is trained on historical data with time information and meteorological variables as input to identify nonlinear relationships in energy demand.

1.2.3 Hybrid Model

Hybrid model combines two or more modeling approaches to improve predictive performance and leverage the strengths of different methods to enhance model precision. Depending on the chosen core models, the hybrid model can be an improved traditional method, a machine learning model, a deep learning model, or a combination of different models.

In recent years, various hybrid models have been proposed for energy forecasting. One common approach is decomposition-based hybrid models. N. Liu et al.(2014) characterize the features of micro-grid load, such as small capacity, strong fluctuation, and high randomness, and propose a hybrid forecasting model combining Empirical Mode Decomposition (EMD), Extended Kalman Filter (EKF), Extreme Learning Machine with Kernel (KELM), and Particle Swarm Optimization (PSO). First, the original load data are decomposed into several IMF components by EMD. Then, EKF and KELM are used to predict different component types. PSO is used to optimize the model parameters. In addition, given the limited computing resources in practice, they propose a method that combines offline optimization, periodic updating, and online forecasting.

Another type of hybrid model combines different modeling approaches to capture both linear and nonlinear relationships. X. Liu et al. (2016) combine grey model, neural network, and input–output analysis to predict energy consumption and CO2 emissions for Spain's 36 sub-sectors under three GDP growth scenarios (optimistic, baseline, and pessimistic) from 2010 to 2015. The grey model is mainly used to describe the overall trend in the data, while neural networks are introduced to model nonlinear relationships and improve forecasting results. Input–output models have also been applied to examine industrial structures and estimate emissions.

In addition, hybrid models which combine linear models and neural networks have also developed. Pao (2009) for example, uses historical data on electricity and oil consumption in Taiwan to develop a hybrid forecasting framework based on linear models and artificial neural networks (ANN). The seasonal exponential generalized autoregressive conditional heteroskedasticity (SEGARCH) model is used to describe seasonal volatility, whereas the Winters exponential smoothing–EGARCH (WARCH) model incorporates trend, seasonality, and variance changes. ANN methods are further employed to model nonlinear relationships. The empirical results suggest that the WARCH-ANN framework performs relatively well, with MAPE remaining below 5% in the sample analysis.

1.3 Summary and Relevant Gap in the Literature

Existing studies on energy consumption forecasting have gradually expanded from traditional time-series models to more data-driven and hybrid approaches. Although forecasting accuracy has improved in many cases, several limitations remain, especially in studies using macro-level energy data.

First, many studies focus primarily on forecasting performance, while paying less attention to model interpretability. As a result, it is often difficult to identify how individual variables influence predictions or how their effects differ across models. Another issue concerns the use of lagged information. While lagged variables are commonly included in forecasting models, relatively few studies systematically examine the role of information from different time periods, particularly in the context of China's structural changes and limited annual sample sizes. The role of different lag structures and their contribution to model predictions remains insufficiently explored, thereby constraining understanding of the dynamic mechanisms underlying changes in energy consumption.

Based on these gaps, this study applies SHAP values to enhance model transparency by jointly analyzing variable importance and temporal structure. Specifically, it develops an integrated framework that combines prediction and explanation, enabling a more comprehensive interpretation of how models use information across different lag periods. This framework aims to provide more reliable and interpretable insights to support China's energy carbon-reduction goals.

2. Data Description and Data Treatment

2.1 Data Description

In this study, the total energy consumption of China is selected as the dependent variable, which is obtained from the China Statistical Yearbook 2025 by the National Bureau of Statistics of China(2025). The unit of measurement is ten thousand tons of standard coal. The indicator refers to the total energy consumed by all sectors of the economy and households in a given region over a given period, including both primary and secondary energy consumption, without double-counting(National Bureau of Statistics of China, 2025). According to the statistical definition in the China Statistical Yearbook 2025, total energy consumption comprises three components: final energy consumption, energy transformation losses, and energy losses. Since this indicator has already been calculated using a unified standard, the original values are used directly without further recalculation.

Based on the determinants identified in the literature review in section 1.1 and the statistical classification of the China Statistical Yearbook 2025(National Bureau of Statistics of China, 2025), a total of 50 independent variables are selected. These variables are drawn from 10 categories: National Accounts, Population, Transportation, Postal Services and Software Industry, Energy, International Trade, Industry, Agriculture, Prices, Fixed Asset Investment, and Retail Trade.

The dataset covers the period from 1990 to 2024, yielding 35 observations before preprocessing. Table 2 reports the descriptive statistics of the variables used in this study, including the mean, standard deviation, minimum, and maximum values. The numbering, names, units, and classifications of these variables are reported in Table A in the Appendix.

Table 2 - Descriptive statistics of variables

Variable	Variable Name	Mean	SD	Min	Max
Y	Total energy consumption of China	304,544.06	161,057.17	98,703.00	596,000.00
X1	Per Capita GDP	32,426.80	30,080.71	1,666.00	95,749.00
X2	Civil Aviation Freight Volume	446,032.39	427,402.38	18,909.60	1,349,083.50
X3	Manufactured Goods Exports	37,347.10	27,650.59	5,017.20	91,413.90
X4	Total Imports and Exports	178,960.75	158,313.85	7,744.10	492,087.10
X5	Railway Petroleum Freight	148,023.02	132,572.58	6,904.50	405,442.10
X6	Railway Major Freight Volume	229,038.26	241,758.45	6,148.30	765,562.50
X7	Railway Major Freight Volume	131,138.89	8,315.34	114,333.00	141,260.00

Variable	Variable Name	Mean	SD	Min	Max
X8	Railway Major Freight Volume	61,519.09	21,337.72	30,195.00	94,350.00
X9	Railway Major Freight Volume	69,619.80	13,306.04	46,478.00	85,947.00
X10	Per Capita GDP	6.35	3.81	-1.48	14.39
X11	Per Capita GDP	176,791.17	96,043.20	93,308.00	431,239.73
X12	Per Capita GDP	1,507,422.72	768,893.73	354,642.80	3,557,010.00
X13	Per Capita GDP	22,797.22	4,148.66	11,627.48	28,300.34
X14	Per Capita GDP	24,572.19	21,324.80	1,660.00	73,021.33
X15	Total Freight Volume	2,852,631.17	1,653,235.77	970,602.00	5,783,625.00
X16	Railway Freight Volume	299,174.66	120,420.03	150,681.00	517,477.00
X17	Highway Freight Volume	2,126,884.89	1,222,657.37	724,040.00	4,188,016.00
X18	Waterway Freight Volume	379,845.86	290,532.50	80,094.00	981,060.00
X19	Civil Aviation Freight Volume	396.74	261.73	37.00	898.00
X20	Total Retail Sales of Consumer Goods	168,632.58	161,422.77	8,300.10	483,344.70
X21	Total Fixed Asset Investment	190,775.24	180,462.86	4,517.00	520,915.86
X22	Electricity Generation	39,126.91	29,367.45	6,212.00	100,868.81
X23	Hydropower Generation	6,493.57	4,697.11	1,250.90	14,256.82
X24	Thermal Power Generation	28,744.26	19,504.86	4,944.80	63,742.63
X25	Natural Gas Production	890.55	725.17	152.98	2,464.51
X26	Coke Production	29,404.46	16,023.91	7,328.30	49,260.00
X27	Crude Oil Production	18,043.62	2,391.03	13,830.60	21,455.58
X28	Fuel Exports	22,997.05	17,609.02	4,069.00	64,163.00
X29	Manufactured Goods Exports	1,268,851.02	1,128,988.23	46,205.00	3,415,049.80
X30	Fuel Imports	162,938.48	169,084.57	1,272.00	535,666.00
X31	Manufactured Goods Imports	733,089.24	573,121.48	43,492.00	1,710,512.00
X32	Total Sown Area of Crops	158,569.65	7,506.44	147,740.70	173,000.08
X33	Agricultural Machinery Power	74,620.95	30,121.04	28,707.70	115,888.42
X34	Grain Output	55,196.84	9,456.57	43,069.53	70,649.89
X35	Total Energy Production	261,515.23	125,242.09	103,922.00	498,000.00
X36	Consumer Price Index	505.30	141.89	216.40	709.40

Variable	Variable Name	Mean	SD	Min	Max
X37	Producer Price Index	103.03	6.47	94.60	124.00
X38	Retail Price Index	104.49	8.31	92.10	135.10
X39	Total Imports and Exports	166,101.52	138,390.39	5,560.10	438,233.92
X40	Foreign Direct Investment	34,770.03	15,240.40	7,273.00	83,437.00
X41	Per Capita Raw Coal Production	2.00	0.84	0.94	3.39
X42	Per Capita Crude Oil Production	137.48	10.59	121.84	155.49
X43	Per Capita Electricity Generation	2,874.02	2,042.28	547.22	7,159.02
X44	Electricity Production Elasticity	1.01	0.35	0.04	1.61
X45	Energy Production Elasticity	0.67	0.53	0.03	2.81
X46	Rural Electricity Consumption	5,142.84	3,036.84	844.50	9,524.42
X47	Railway Operating Mileage	7.96	3.80	5.34	16.21
X48	Highway Mileage	314.76	166.72	102.83	549.04
X49	Railway Petroleum Freight	10,532.02	2,580.67	3,765.00	13,322.00
X50	Railway Major Freight Volume	260,391.12	104,437.53	146,210.00	517,379.00

2.2 Data Cleaning

This study cleans the raw data to improve data quality and enhance the reliability of subsequent modeling. The data cleaning includes three aspects: duplicate values, missing values, and outlier detection.

Duplicate values can cause biased results or model overfitting. In this study, no exact duplicates or inconsistent repeats were found, indicating that the data are mostly unique. The missing values of the independent variables are interpolated using a time-series-based linear interpolation method: connecting the two known values with a straight line and estimating the missing value based on its position along that line. Bhavsar & Khushbu(2025) find that the overall characteristics of the data remain almost unchanged after linear interpolation, including the average level and variability. And data distribution and the original trend are also well preserved without breaking the time structure. Finally, six missing values are imputed: Energy Production Elasticity *(X45)* for 1998, 2015, and 2016, and Rural Electricity Consumption *(X46)* for 2015–2017.

For outlier treatment, the Interquartile Range (IQR) method is used to identify unusual observations. For the detected outliers, different strategies are applied: for a few variables with clear economic meaning or structural features, the original outliers are kept; for the others, outliers are treated as missing values and filled using the same linear interpolation method. This approach helps smooth extreme fluctuations while preserving the general trend of the series.

It should be noted that, China's energy consumption data have experienced several structural shifts over time, including economic transformation, external shocks, and the COVID-19 pandemic. These observations are therefore not regarded as abnormal values and are retained in the dataset, since they reflect actual changes in the economy. To account for possible structural variation over time, this study adopts a rolling window forecasting framework. Under this setting, model estimation is updated using the most recent observations, which helps maintain forecasting stability when the data-generating environment changes. In the outlier treatment process, five variables, Highway Passenger Traffic *(X12)*, Producer Price Index *(X37),* Retail Price Index *(X38),* Foreign Direct Investment *(X40)*, and Energy Production Elasticity *(X45),* are retained, while other outliers are removed. The reasons for retaining these observations are discussed in Appendix 1.

2.3 Statistical Transformation and Stationarity

Statistical transformations and stationarity tests are conducted before model estimation. To reduce heteroskedasticity and skewness in the data, logarithmic transformations are applied to variables with strictly positive values. Some variables, however, remain in their original forms because they are expressed as indices or ratios and are not suitable for log transformation. These include Natural Population Growth Rate (X10), Consumer Price Index (X36), Producer Price Index (X37), Retail Price Index (X38), Electricity Production Elasticity (X44), and Energy Production Elasticity (X45).

Stationarity is examined using the Augmented Dickey–Fuller (ADF) test. A trend term is included in the test specification, and the lag length is selected according to the Akaike Information Criterion (AIC). For non-stationary variables, differencing is applied and the ADF test is repeated. The procedure continues until the variable becomes stationary or reaches second-order differencing.

Only variables that pass the ADF test at the 10% significance level are retained in the final dataset. The detailed test results are reported in Appendix Table B. After preprocessing, including outlier treatment and stationarity testing, non-stationary variables are removed, leaving 47 stationary variables for the subsequent analysis.

3. Methodology

3.1 Feature Construction and Selection

3.1.1 Lagged Feature Construction

To capture the dynamic dependence among time series variables, this study constructs lag features for all variables. The current value of a time series usually depends on its historical observations. By introducing lag terms, it is possible to effectively describe the temporal dependence structure and transform the time series forecasting problem into a supervised learning problem (Bontempi et al., 2013). In the machine learning framework, this type of feature construction is referred to as time-delay embedding, in which values from previous time steps are used as input features to improve the model's ability to capture temporal dependence (Karmaker, 2025).

Specifically, this study treats the time dimension explicitly as a feature and constructs first-, second-, and third-order lag terms for each variable. And all the periods are included in the feature set to form multi-order lag features. Given that this study uses annual data with a relatively small sample size, to avoid the problems of dimensional expansion and overfitting caused by too many lag terms, the maximum lag order is set to three to ensure the model captures temporal dependence while maintaining stability and generalization under small-sample conditions.

During the modeling process, only lagged values of explanatory variables are used as input features to further explore the dynamic impact of exogenous variables on the dependent variable. To ensure robustness, alternative model specifications, including the inclusion of lagged dependent variables, are considered in supplementary analyses. Since lag features cannot be constructed for the first few observations, these observations are removed. Finally, after introducing up to three lag orders, the available sample period is adjusted to 1995–2024.

3.1.2 Mutual Information-Based Feature Selection

This study adopts a mutual information (MI)-based feature selection method. The mutual information between each candidate feature and the dependent variable is calculated to measure their relevance(Huang et al., 2016). A higher MI value indicates a stronger dependency between the feature and the dependent variable, and thus a higher contribution to prediction. Malik et al.(2025) show that mutual information can effectively analyze the relationship between input features and electricity load data and can capture nonlinear relationships between dependent and independent variables, making it suitable for complex time-series problems such as energy forecasting.

Mutual information measures the degree of information dependence between two random variables, X and Y. Let $p(x)$ and $p(y)$ denote the marginal probability distributions of X and Y, respectively, and let $p(x, y)$ represent their joint probability distribution. The mathematical expression of mutual information is given in Equation (1).

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right) \qquad (1)$$

Here, log denotes the logarithm, commonly taken with base 2 in information theory. According to Equation (1), mutual information measures the reduction in joint uncertainty between two random variables. If X and Y are completely independent, then $p(x, y)=p(x)p(y)$ the mutual information $I(X,Y)=0$, indicating that there is no information shared between the two variables. If X and Y are dependent, then $\frac{p(x,y)}{p(x)p(y)} \neq 1$, which implies $I(X;Y)>0$. A larger value of mutual information indicates a stronger information dependence between the variables.

After computing the MI values for all features, they are ranked in descending order of importance. To balance model performance and complexity, the top 30% of features with the highest MI values are selected as the final input variables for model training. This approach reduces redundant variables while retaining variables with relatively high information content.

3.2 Model Estimation and Validation Framework

3.2.1 Rolling Cross-Validation and Hyperparameter Tuning

The dataset is split chronologically into training and test samples to preserve the time-series structure of the data. Specifically, this study uses the first 80% as the training set and the remaining 20% as the test set to evaluate out-of-sample forecasting performance (Hyndman & Athanasopoulos, 2018). The training set includes data from 1995 to 2018, and the test set covers the period from 2019 to 2024. During model training and tuning, this study adopts a rolling cross-validation approach within the time-series cross-validation framework of (Hyndman & Athanasopoulos, 2018). In the context of rolling estimation, excessively long windows may incorporate outdated information, increasing forecast bias and deteriorating predictive accuracy (Inoue et al., 2017). Therefore, this study adopts a heuristic approach, setting the window size to 50% of the training data to balance the amount of historical information and model complexity. In each fold, observations within the window are used as the training set, and the subsequent observation is used as the validation set.

Feature selection is also repeated throughout the rolling forecasting procedure rather than conducted only once. Instead, it is performed separately for each rolling fold using the corresponding training subset. This ensures that feature selection only uses information available at that time and avoids information leakage.

Within each rolling training subset, feature selection and hyperparameter tuning are further implemented. The grid search procedure is applied, and the optimal parameter combination is selected based on the lowest validation RMSE. This approach strictly follows the time-series information structure and enhances the robustness of model estimation. The same procedure is applied to all machine learning models. Finally, the optimal parameters from all folds are combined to determine the final parameter setting for each model.

3.2.2 Pseudo-Out-of-Sample Forecasting and Model Selection

This study uses a pseudo-out-of-sample forecasting method to evaluate the predictive performance of the models during the test period. Specifically, at each forecasting point, the model is re-estimated using only the information available up to the previous period, and a one-step-ahead forecast is made for the current period. This procedure follows the out-of-sample forecasting framework described in Campbell & Thompson (2008). It is designed to simulate a real rolling forecasting situation. Regarding the time window setting, this study uses a rolling window approach for pseudo-out-of-sample forecasting. Following Alfieri (2025), the window size is set to 50% of the available observations to balance stability and adaptability. At the same time, to check the robustness of the results, an expanding window is also used as a comparison setting.

Regarding feature specification, in addition to lagged terms of the independent variables, this study also includes lagged terms of the dependent variable as an extended model. This is done to examine the effect of endogenous dynamic structure on the forecasting results and to conduct robustness analysis.

Model performance is evaluated using root mean squared error (RMSE) and mean absolute error (MAE). Forecasting results during the test period are then compared across different models. In addition, a first-order autoregressive model (AR(1)) is included as a benchmark model for comparison. Under the rolling forecasting framework, the forecasting results of different machine learning models are compared with those of the autoregressive benchmark.

3.3 SHAP-Based Model Interpretation and Analysis

This study introduces Shapley Additive Explanations (SHAP) after completing the model forecasting evaluation. Neubauer et al. (2025) systematically apply SHAP in a time series forecasting context and show that SHAP can be used not only to measure overall feature importance, but also to identify how inputs at different time points affect predictions, thereby revealing the model's dependence on lagged information. Following this idea, this study also treats SHAP as a key tool linking forecasting performance with model interpretability.

Although model comparison in the previous section is based on a pseudo-out-of-sample (pseudo-OOS) forecasting framework, the SHAP analysis in this section relies on static models. During the recursive forecasting process, feature selection is updated for each training window, so both the estimation sample and selected variables may change over time. As a result, SHAP values calculated from different re-estimated models would be based on different model specifications and may not be directly comparable. To avoid this issue, pseudo-OOS forecasting is used only for forecasting evaluation, while SHAP analysis is conducted using the final static models. This setup separates forecasting assessment from model interpretation and helps keep the SHAP results more stable across models.

The SHAP analysis is carried out in three stages. SHAP values are first calculated for the test samples using the final static models. These values are then aggregated to evaluate the overall importance of different variables. In addition, SHAP contributions are grouped by lag order to examine the relative importance of information from different time periods. Because the main purpose of the SHAP analysis is to interpret machine learning models and their variable selection patterns, the analysis focuses on six models: Lasso, Ridge, SVR, Random Forest, XGBoost, and KNN.

3.3.1 SHAP Value Estimation

SHAP is an interpretation method based on the Shapley value concept from game theory. It assigns a numerical value to each feature to quantify its contribution to the model output.

For a given sample x, the Shapley value of feature i is defined as:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!} [f(S \cup \{i\}) - f(S)] \quad (2)$$

Here, N represents the set of all features, and $|N|$ is the total number of features. S is a subset of features that does not include feature i, and $|S|$ is the number of features in subset S. The function $f(S)$ represents the model output using only the feature subset *S,* while $f(S \cup \{i\})$ represents the model output after adding feature *i.* The value ϕ_I indicates the contribution of feature *i* to the prediction. The weighting term $\frac{|S|!(|N|-|S|-1)!}{|N|!}$ is used to ensure fair contribution calculation across all feature combinations.

In practice, exact computation of SHAP values is infeasible, thus, approximation methods are used. Because the exact computation of Shapley values requires evaluating all possible feature combinations, the computational cost increases exponentially with the number of features. This makes it difficult to apply in high-dimensional settings. Therefore, this study adopts a Monte Carlo sampling approach to approximate SHAP values (Štrumbelj & Kononenko, 2014)

Specifically, SHAP values are estimated by randomly sampling feature inclusion orders and measuring the change in model predictions before and after adding each feature. In this way, the marginal contribution of each feature can be obtained. By repeating this process many times and averaging the results, an unbiased estimate of the Shapley value can be achieved. In this study, the number of simulations is set to 2000, as the accuracy of Monte Carlo estimation improves with increasing sample size (Štrumbelj & Kononenko, 2014).

3.3.2 Global Feature Importance

To measure the overall importance of features in the prediction process, this study constructs global feature importance based on SHAP values (Neubauer et al., 2025). Unlike local explanations at the individual observation level, global feature importance focuses on the overall data distribution and reflects each variable's average contribution to the model.

Specifically, after determining the optimal hyperparameters by using rolling cross-validation, MI is applied again on the training set to select features, and the final static model is then trained. After obtaining predictions for the test set, SHAP values are computed for each feature. These SHAP values are further aggregated across different lag orders at the variable level. The corresponding formula is given as follows.

$$Global\ SHAP_k = \sum_{l=1}^{3} \left(\frac{1}{N} \sum_{i=1}^{N} \left| \varphi_{i,k,l} \right| \right) \quad (3)$$

Here, $\phi_{i,k,l}$ represents the SHAP value of the k-*th* base variable at lag l for the i-*th* observation, and Ndenotes the total number of observations. Following Neubauer et al. (2025) to avoid canceling positive and negative contributions, this study uses the absolute values of SHAP and averages across observations to measure feature importance. This indicator reflects the average impact of each variable on the model predictions during the test period.

3.3.3 Temporal Importance Analysis

Time series models not only select which variables to use for prediction, but also specify which lag periods the model depends on. Therefore, it is necessary to further analyze the relative importance of different lag orders under the SHAP framework.

Following the approach of Neubauer et al. (2025), to analyze the influence of input time steps, the importance along the time dimension can be described by aggregating SHAP contributions across different input time points. In their local SHAP analysis, SHAP values are assigned to different input hours, and the cumulative share is used to examine which recent input time points are more important for prediction, revealing the model's dependence on the time structure.

Specifically, the average absolute SHAP value of each lagged feature is first calculated on the test set. Then, within the same lag order, these average absolute SHAP values are summed across all variables to obtain the overall temporal importance for that lag. Based on this, the temporal importance of each lag is further normalized to a share to reflect the relative weight of different lag orders in the overall temporal explanation. The corresponding formula is given as follows.

$$Temporal\ Importance_l = \sum_{k=1}^{K} \left(\frac{1}{N} \sum_{i=1}^{N} \left| \varphi_{i,k,l} \right| \right) \quad (4)$$

$$TemporalShare_l = \frac{Temporal\ Importance_l}{\sum_{j=1}^{3} TemporalImportance_j} \quad (5)$$

Here, for each base variable k, the average absolute SHAP value at lag l is calculated across the test samples. Then, these values are summed across all variables to obtain the overall importance level for that lag. To facilitate comparison across different lag orders, the temporal importance is further normalized to get a relative share. This ensures that the total importance across all lags equals one, so that the contributions of different lagged information are comparable. A larger Temporal Share value indicates that the model relies more on information from that lag in the overall prediction process.

It should be noted that the analysis of temporal importance is also based on the final static model and its selected feature set. Therefore, the Temporal Importance and Temporal Share here reflect the model's relative dependence on different lag orders within the selected variable space, rather than the absolute ranking of time effects based on the full set of original variables.

4. Results

4.1 Baseline Forecasting Performance under Rolling-CV Tuning

Under the baseline setting that includes only lagged exogenous variables (X lags), this study compares the performance of different models across two frameworks: rolling cross-validation and pseudo-out-of-sample forecasting. Rolling cross-validation is mainly used for model tuning and performance evaluation during the training stage, while pseudo-out-of-sample forecasting is used to test the model's generalization ability in a real-time series forecasting setting.

First, based on the results of rolling cross-validation (see Table C in the appendix), there are clear differences in the models' in-sample performance. Among them, XGBoost performs the best, followed by RF and Ridge. SVR and KNN show moderate performance, while Lasso performs relatively poorly. This indicates that a simple linear structure has clear limitations in this problem.

Table 3 -Pseudo-Out-of-Sample Forecasting under Expanding and Rolling Window Schemes

Model	Exp. RMSE	Exp. MAE	Roll. RMSE	Roll. MAE	Rank (Exp.)	Rank (Roll.)
Ridge	0.0206	0.0178	0.0203	0.0163	1	1
XGBoost	0.0222	0.0196	0.0217	0.0169	2	5
RF	0.0246	0.0224	0.0207	0.0168	3	2
SVR	0.0251	0.0231	0.0220	0.0191	4	6
Lasso	0.0259	0.0217	0.0209	0.0172	5	3
KNN	0.0287	0.0268	0.0211	0.0167	6	4

Notes. Exp. and Roll. denote expanding and rolling window schemes, respectively. Rankings are based on RMSE.

However, under the pseudo-out-of-sample forecasting framework (see Table 3), the model performance changes significantly. In the rolling window setting, Ridge performs best and ranks highest on both RMSE and MAE, indicating strong generalization. RF and Lasso have medium performance. In contrast, SVR and KNN have higher prediction errors and perform poorly overall. It is worth noting that XGBoost, which performs the best in the rolling cross-validation stage, shows a clear decline in out-of-sample performance. The comparison between the rolling and expanding window settings shows that XGBoost's performance varies widely across different window sizes, indicating that its predictive ability is sensitive to the sample window. In comparison, Ridge maintains stable performance across both settings, demonstrating greater robustness. Therefore the model with the best in-sample performance does not always give the best out-of-sample results. This suggests that more flexible models may overfit in small samples, whereas simpler models, such as Ridge, can provide more stable predictions.

In conclusion, Ridge performs best overall across different settings. This indicates that regularized models with regularization are more suitable for small-sample energy forecasting.

4.2 Benchmark comparison and robustness analysis

4.2.1 Comparison with AR(1) Benchmark

Table 4 - Relative Forecast Performance of Ridge against AR(1)

Window	Benchmark	Relative RMSE	RMSE Gain (%)	Relative MAE	MAE Gain (%)
Expanding	AR(1)	0.8174	18.26	0.8293	17.07
Rolling	AR(1)	1.0013	-0.13	1.0579	-5.79

Notes. Relative RMSE /MAE is defined as the ratio of the forecasting error of Ridge to that of the benchmark model. Values below 1 indicate superior Ridge performance. RMSE Gain (%) and MAE Gain (%) are computed as 1 − Relative value, representing the percentage reduction in forecast error.

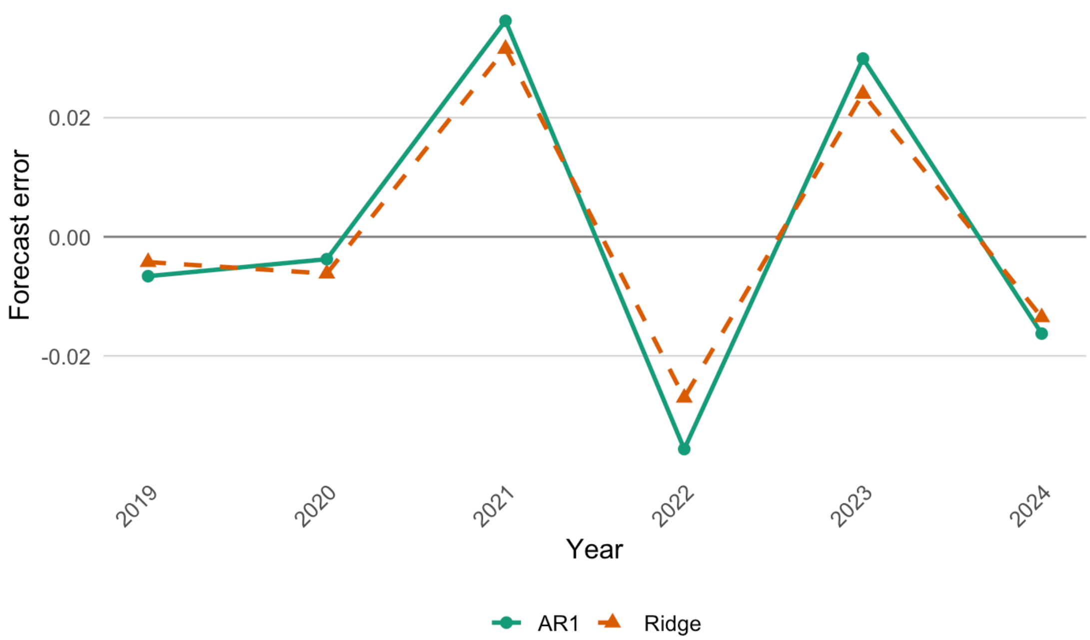

Figure 1 - Forecast Errors under Expanding Window: Ridge vs AR(1)

Note. Forecast error = actual - predicted. Positive values indicate underprediction, while negative values indicate overprediction

This study compares the best-performing model, Ridge, with the AR(1) benchmark model (see Table 4). Under the expanding-window setting, Ridge achieves lower forecasting errors than the AR(1) benchmark, with Relative RMSE and Relative MAE both below 1. Specifically, Ridge reduces RMSE and MAE by approximately 18% and 17%, respectively. However, under the rolling-window setting, the Relative RMSE and Relative MAE are both close to 1, suggesting that Ridge and the AR(1) model exhibit very similar forecasting performance. Figure 1 reports the forecasting errors of Ridge and the AR(1) benchmark under the expanding-window setting. The forecasting errors of both models fluctuate around zero and follow broadly similar patterns over time, although Ridge exhibits slightly smaller deviations in several periods.

The results indicate that a simple autoregressive specification already explains much of the variation in energy consumption under small-sample forecasting conditions. At the same time, regularized linear models such as Ridge appear to perform more steadily when additional exogenous variables are included in the rolling forecasting framework.

4.2.2 Robustness to Adding Lagged Dependent Variables

Table 5 - Performance difference between models with and without lagged dependent variables

Model	RMSE (X only, Exp)	RMSE (X+Y, Exp)	ΔRMSE (Exp)	RMSE (X only, Roll)	RMSE (X+Y, Roll)	ΔRMSE (Roll)
Ridge	0.020582	0.020565	-0.000017	0.020295	0.020286	-0.000009
RF	0.024582	0.023176	-0.001405	0.020743	0.022035	0.001292
Lasso	0.025913	0.023993	-0.001920	0.020851	0.020784	-0.000067
SVR	0.025134	0.024755	-0.000379	0.022039	0.021533	-0.000506
XGBoost	0.022227	0.026893	0.004666	0.021656	0.025744	0.004088
KNN	0.028660	0.029935	0.001275	0.021087	0.022792	0.001706

Notes. $\Delta RMSE$ is defined as the difference between the $RMSE$ of the model with both X-lags and Y-lags and that with X-lags only. Positive values indicate a deterioration in forecasting performance after including lagged dependent variables, while negative values indicate an improvement.

To further examine the robustness of the forecasting setting, this study compares models with and without lagged dependent variables. One specification includes only lagged explanatory variables (X-lags), while the other additionally incorporates lagged dependent variables (X-lags + Y-lags). Model performance under both expanding and rolling window settings is evaluated using RMSE and MAE.

The results reported in Table 5 indicate that including lagged dependent variables leads to only minor changes in forecasting accuracy. For most models, including Ridge, Lasso, SVR, RF, and KNN, the differences in RMSE remain relatively small, and some models even perform slightly worse after adding lagged dependent variables. This suggests that the explanatory variables already contain much of the information relevant for forecasting energy consumption, whereas lagged dependent variables contribute relatively little additional predictive information.

4.2.3 Robustness to Feature Ordering

Given the small-sample setting and the large number of lagged explanatory variables used in this study, substantial correlations among predictors may arise. In addition, feature pre-filtering is conducted using the mutual information method before model estimation. Under such conditions, the feature selection process and model estimation may potentially be affected by variations in predictor ordering, particularly in small-sample forecasting environments.

To examine the robustness of the results, an additional robustness analysis is conducted by randomly permuting the ordering of explanatory variables while keeping all model specifications and tuning parameters unchanged. The Ridge model is then re-estimated under both the expanding-window and rolling-window pseudo-out-of-sample forecasting frameworks.

Table 6- Performance difference between models with and without lagged dependent variables

Predictor Ordering	Expanding RMSE	Expanding MAE	Rolling RMSE	Rolling MAE
Original	0.020582	0.017757	0.020295	0.016254
Random 1	0.020544	0.017581	0.020988	0.017130
Random 2	0.020452	0.017505	0.021499	0.017461

Notes. “Original” refers to the original predictor ordering used in the baseline specification. “Random 1” and “Random 2” represent alternative randomized predictor orderings

Table 6 reports the forecasting performance obtained under different randomized predictor orderings. Overall, the RMSE and MAE values remain highly similar across different permutations, indicating strong stability in the forecasting results. Under the expanding-window framework, the forecasting errors are nearly identical across different random orderings. Slightly larger fluctuations are observed under the rolling-window framework, which is likely related to the smaller effective training sample size under rolling estimation. Overall, the results suggest that the Ridge model's forecasting performance is not materially affected by randomized predictor ordering. This indicates that the main findings of this study are relatively robust and not driven by unstable feature selection or arbitrary predictor ordering.

4.3 Model interpretation and economic insights

4.3.1 Global Feature Importance

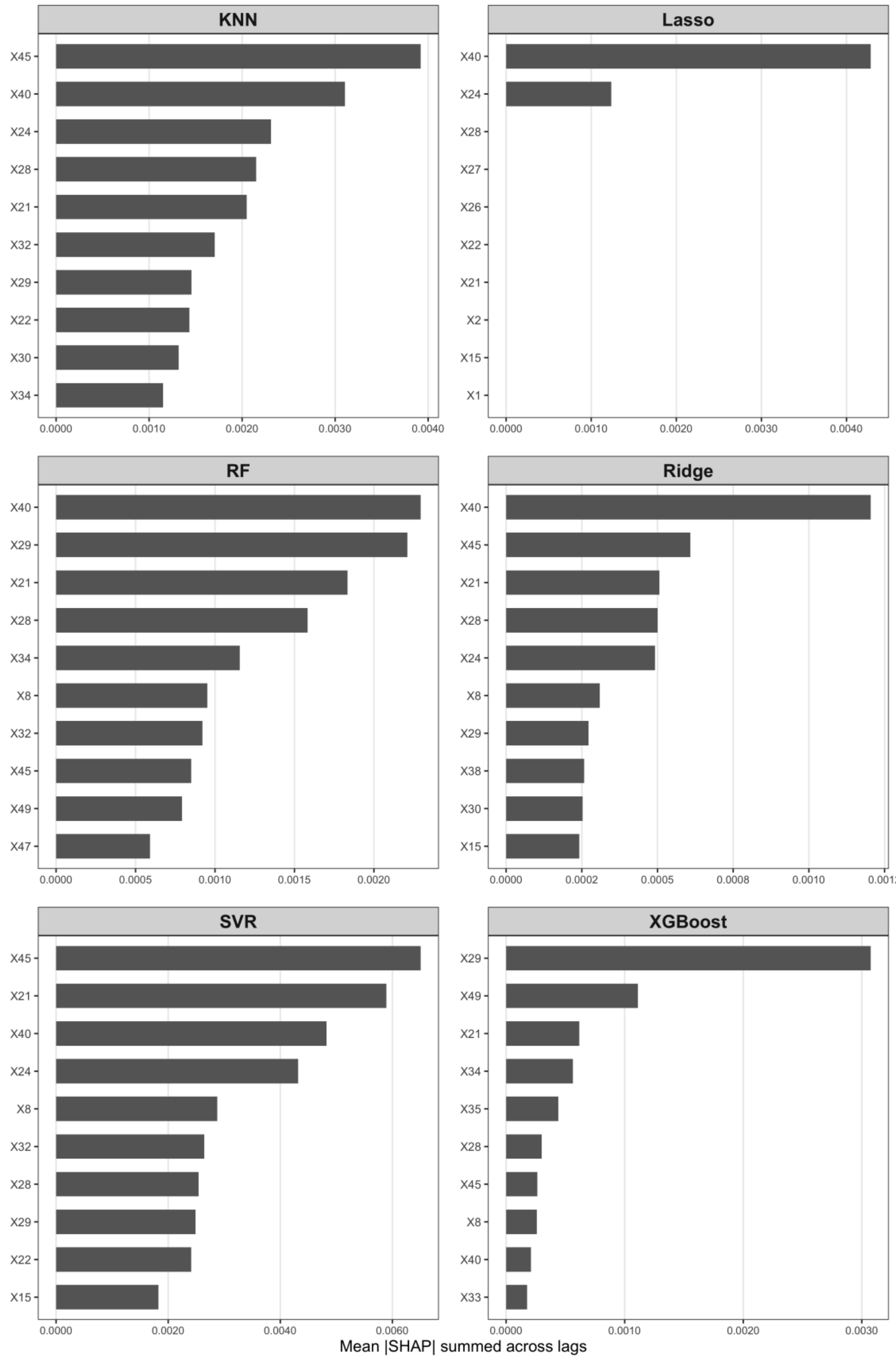

Figure 2 - Global SHAP Top 10 Feature Importance across Models

To further understand the prediction mechanisms of different models, this study uses SHAP values to examine global feature importance. From Figure 2, the different models exhibit some overlap in the distributions of important variables.

In the SHAP results for the Lasso model, only Foreign Direct Investment *(X40)* and Thermal Power Generation *(X24)* exhibit noticeable importance, whereas the SHAP values of most other variables remain close to zero. This pattern is consistent with the sparsity structure typically observed in Lasso estimation. Because the model applies L1 regularization, many coefficients are shrunk to zero, leaving only a small number of selected variables in the final specification. As a result, the SHAP distribution is concentrated on a limited set of features.

To further examine this result, the Lasso model's regression coefficients are also reported. As shown in Table 7, most coefficients are zero, consistent with the sparse feature importance pattern observed in the SHAP analysis.

Table 7 - Sparsity Summary of the Final Lasso Model

Metric	Value
Total selected predictors before Lasso shrinkage	42
Non-zero coefficients after Lasso shrinkage	2
Zero coefficients after Lasso shrinkage	40
Share of zero coefficients (%)	95.24

In contrast, the SHAP values of the Ridge model are generally smaller, and the importance distribution is smoother. This does not mean that Ridge has weaker explanatory power. Instead, because Ridge uses L_2 regularization, it spreads the influence across more variables, so each variable's contribution is relatively small.

Therefore, the SHAP results of Ridge show a “distributed” prediction mechanism. The model does not rely on a few individual variables. Instead, it uses the combined small effects of many variables to make predictions. This is also consistent with its earlier stable performance.

It should be noted that the absolute size of SHAP values cannot be directly compared across different models. This is because the scale of SHAP values depends on the range of model outputs, the function form, and how contributions are assigned to variables. Therefore, to better describe how different models depend on variables, the cumulative SHAP importance for the top 5, 10, and 20 features is calculated.

Table 8 - Concentration of Global SHAP Feature Importance Across Models

Model	Top5	Top10	Top20
KNN	0.446	0.679	0.946
Lasso	1.000	1.000	1.000
RF	0.483	0.702	0.950
Ridge	0.516	0.724	0.944
SVR	0.506	0.752	0.963
XGBoost	0.742	0.896	1.000

*Note.*Top5, Top10 and Top20 denote the cumulative share of total SHAP importance explained by the top 5, 10, and 20 features, respectively.

Figure 3 - Robust Features Across Models: Frequency in Top-20 Global SHAP Importance Lists

Table 8 shows that, in most models, the top 20 variables account for roughly 94% to 100% of total feature importance. This suggests that model predictions are mainly driven by a relatively small group of variables. Among the models, Lasso achieves 100% with only the top 5 variables, reflecting its strong sparsity. XGBoost also assigns a relatively large share of importance to the top-ranked variables, likely because the boosting procedure repeatedly concentrates on variables with stronger predictive contributions. By comparison, Ridge, SVR, and RF display a more even distribution of importance across variables, implying that these models rely on a broader set of information. KNN shows a lower concentration among the top 5 variables, suggesting that its predictions depend more on the overall distance structure across many variables than on a few dominant predictors.

The analysis further examines feature consistency across models. Based on the concentration results, the top 20 variables from each model are collected and their frequencies across models are summarized (see Figure 3).

Variables with the highest frequency (frequency = 6) include Foreign Direct Investment *(X40),* Total Energy Production *(X35),* Grain Output *(X34),* Manufactured Goods Exports *(X29),* Fuel Exports *(X28),* Crude Oil Production *(X27),* Coke Production *(X26),* Thermal Power Generation *(X24),* and Total Fixed Asset Investment *(X21).* These variables appear among the top 20 features in all models, suggesting that their importance is relatively stable across different model specifications and may reflect common factors associated with energy consumption dynamics.

It should also be noted that the dependent variable in this study is the second difference of energy consumption. The models capture changes in the rate of change of energy consumption rather than its level or growth rate. In this context, the key variables identified by the models can be interpreted as factors associated with the dynamics of energy consumption adjustment.

From an interpretive perspective, these variables can be grouped into three related mechanisms: the energy-supply-constraint mechanism, the open-economy transmission mechanism, and the real economic activity mechanism. However, under the second difference framework, these mechanisms more fundamentally reflect the dynamic adjustment responses of economic agents, who adjust their energy use paths in response to different constraints and incentives.

First, on the supply side, Total Energy Production *(X35),* Thermal Power Generation *(X24),* Coke Production *(X26),* and Crude Oil *Production (X27)* form the basic supply capacity of the energy system. The importance of these variables lies not only in the direct constraint of energy supply on consumption but also in a dynamic mechanism in which changes in supply capacity affect the “decision space” of economic agents.

Specifically, changes in energy supply not only set a limit on energy use but also influence price expectations, input availability, and production possibilities. This affects optimal production decisions of firms and sectors(Stern, 2011). For example, higher thermal power generation generally indicates fewer constraints on electricity supply, which may support the expansion of energy-intensive industries in the short run. Changes in coke and crude oil supply can also influence sectors such as steel production and transportation, leading to short-term adjustments in energy use. Under the second-difference specification, these supply-side variables are therefore more closely related to changes in the adjustment speed of energy consumption rather than its overall level. In this sense, supply conditions may affect how quickly energy consumption responds to changes in production activity(Acemoglu et al., 2012).

Second, from the perspective of an open economy, Foreign Direct Investment *(X40),* Manufactured Goods Exports *(X29),* and Fuel Exports *(X28)* reflect the influence of external demand and international resource allocation on domestic economic activity. In the environmental economics literature, these variables are commonly associated with the scale, composition, and technique effects (Grossman & Krueger, 1995). Under the second-difference setting, these effects are better interpreted as short-term responses to external shocks and changes in economic activity. For example, FDI inflows may expand production and introduce new technologies through spillover effects, both of which can influence energy use in the short run (Borensztein et al., 1998). Based on this, firms adjust their input structure, thereby affecting the optimal allocation of energy inputs. Similarly, changes in manufactured goods exports reflect external demand shocks. When external demand increases, firms often adjust their production capacity and energy input in advance to meet future orders. This leads to an acceleration in the rate of change in energy consumption. Fuel exports reflect the reallocation of energy between domestic and international markets. Changes in fuel exports affect domestic energy supply expectations and price mechanisms, which then influence energy use decisions.

Therefore, the main role of open economy variables is not simply to change the scale of energy use, but to affect how economic agents respond to uncertainty and expectations about future market conditions. This then alters their production and technology decisions, leading to dynamic adjustments in energy-use paths. This mechanism is consistent with findings in the literature on trade policy uncertainty(Y. Xi et al., 2024).

Finally, at the level of real economic activity, Total Fixed Asset Investment *(X21)* and Grain Output *(X34)* reflect the activity levels of the industrial and agricultural sectors, respectively. The importance of these variables indicates that the acceleration of energy consumption mainly stems from changes in the real economy's business cycle. Specifically, fixed asset investment is an important part of overall demand, and its decisions depend strongly on expectations about future returns and the economic environment. Therefore, it has a forward-looking feature (Bernanke et al., 1999). When investment increases, infrastructure construction and industrial expansion will increase energy demand in advance, leading to faster growth in energy consumption. When investment slows, energy demand adjusts quickly, leading to slower growth in energy consumption. In comparison, grain output usually reflects the scale and technology level of agricultural production. Its changes not only reflect agricultural activity but also mechanization and irrigation inputs. Therefore, the importance of agricultural variables shows that energy consumption is not only driven by industry but also affected by adjustments across different sectors of the economy. Under the second-difference framework, these variables together describe how real economic activity affects the path of change in energy consumption through “cyclical adjustment”.

In conclusion, within the framework of energy consumption acceleration, supply constraints, open-economy factors, and real economic activity influence expectations, constraints, and optimal decisions of economic agents, driving the dynamic adjustment of energy consumption. This shows that changes in energy consumption are essentially a “behavioral response process” influenced by multiple economic factors.

4.3.2 Temporal Importance by Lag Order

This study examines how models depend on lagged information over time. Unlike the previous analysis based on global feature importance, temporal importance does not focus on “which variables are important,” but rather on “how information works through different time lags.” It should be noted that all models use MI for feature pre-selection before training. The temporal importance results should therefore be interpreted as the relative importance of the lagged information selected by each model, rather than a direct comparison across all original lagged variables. Under the second-difference specification, the results mainly show how models rely on information from different lag periods when forecasting changes in energy consumption.

Overall, Figure 4 shows noticeable differences in lag patterns across models. Most models, including Ridge, SVR, KNN, and RF, place greater importance on the first lag, suggesting that recent information is more useful for forecasting changes in energy consumption. However, while short-term information is dominant, models show different preferences for medium- and long-term lags. Except for RF, most models treat the third lag as the second most important source of information, while the second lag has a smaller role. RF shows a more balanced use of different lags, which reflects its ability to combine information across multiple time scales.

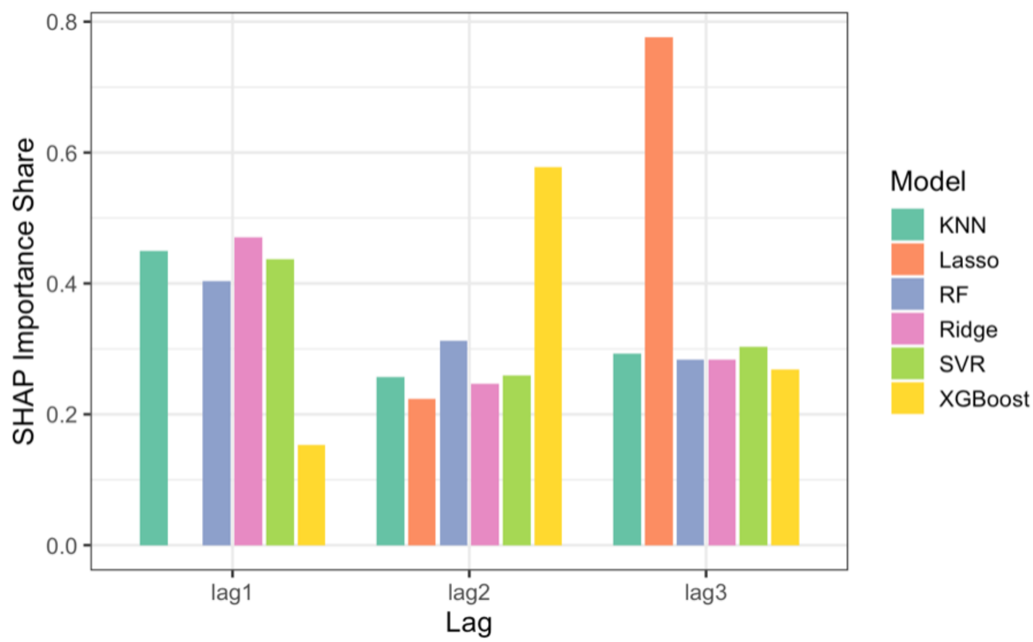

Figure 4 -Temporal SHAP Importance Across Different Models

The differences between models mainly come from their feature selection and learning methods. Lasso relies almost completely on the third lag. This strong long-term preference can be explained by the combination of MI pre-selection and L1 regularization, which keeps only the most representative variables among highly related lagged variables. In contrast, XGBoost depends more on the second lag, which is related to its step-by-step learning process. In this process, the model continues to use variables that help reduce prediction errors. In this dataset, lag2 provides useful signals, so it is used more often during training.

Overall, temporal importance shows that short-term information is the most important for most models, while medium- and long-term information plays a supporting but different role across models.

5. Conclusion

This study examines energy consumption forecasting in China using a machine-learning-based forecasting framework with macroeconomic variables. In addition to comparing the predictive performance of different forecasting models, this study also applies explainable machine learning methods to analyze how the models utilize information in the forecasting process.

The empirical results provide several important findings. First, there are clear differences between in-sample and out-of-sample forecasting performance. Complex machine learning models such as XGBoost perform well during rolling cross-validation tuning, but their forecasting performance deteriorates in the pseudo-out-of-sample evaluation. In contrast, Ridge consistently achieves the most stable forecasting performance across different forecasting settings. This finding is broadly consistent with the existing literature on small-sample forecasting, which suggests that highly flexible machine learning models are more likely to suffer from overfitting when the available sample size is limited (Makridakis et al., 2018). By comparison, regularized linear models may provide a better balance between model flexibility and forecasting stability.

Second, compared with the benchmark models, Ridge performs slightly better than the AR(1) model under the expanding-window setting, whereas the two models produce broadly similar results under the rolling-window framework. This suggests that autoregressive information alone already explains much of the variation in energy consumption in a small-sample forecasting environment. At the same time, including additional macroeconomic variables through regularized regression may still help improve forecasting performance in some settings. The robustness analysis further shows that adding lagged dependent variables leads to only small changes in forecasting accuracy. This result suggests that the lagged explanatory variables already contain most of the information relevant for forecasting energy consumption.

Third, the SHAP results provide additional evidence on how different models generate predictions. Although the models differ in structure, several variables repeatedly appear among the most important predictors across models, including foreign direct investment, energy production, exports, and fixed asset investment. These variables are closely related to energy supply conditions, external economic activity, and domestic investment. The SHAP results also reveal noticeable differences across models. Lasso concentrates importance on a relatively small number of variables, whereas Ridge and tree-based models distribute importance more broadly across lagged predictors. In terms of lag structure, most models assign greater importance to recent lagged information, indicating that short-term economic conditions are more relevant for forecasting changes in energy consumption. At the same time, some differences remain across models. For example, Lasso places relatively more importance on longer lags, while XGBoost assigns a larger share of importance to medium-term lag periods.

The robustness tests further confirm the main results. Additional experiments using randomized predictor orderings show that Ridge maintains relatively stable forecasting performance under different input sequences. This suggests that the results are not sensitive to variable ordering or changes in the feature-selection process.

Overall, the findings suggest that, in annual small-sample forecasting settings, more complex machine learning models do not necessarily produce better forecasting results. Regularized linear models, especially Ridge, tend to perform more steadily while remaining relatively easy to interpret. The SHAP analysis further shows how different models assign importance to macroeconomic variables and lagged information in the forecasting process.

From an empirical perspective, the results indicate that changes in energy consumption are closely related to macroeconomic activity, external openness, and energy supply conditions. These factors may therefore need to be considered jointly in long-term energy and economic policy planning.

This study also has several limitations. First, the analysis is based on annual macroeconomic data with a relatively limited sample size, which may not fully capture short-term fluctuations in energy consumption. Second, the explanatory variables are mainly constructed from macro-level indicators, while micro-level or high-frequency information is not included. Third, although SHAP improves model interpretability, it remains an explanatory tool rather than a causal inference framework. Finally, the interpretability analysis is conducted on static fitted models and may not fully reflect the dynamic updating process of rolling forecasting models.

Future research may extend this study in several directions. First, higher-frequency or regional-level data could be incorporated to improve forecasting precision and better capture short-term dynamics. Second, future studies may include firm-level or sector-level information to enrich the explanatory variable system. Third, explainable machine learning methods could be combined with causal inference frameworks to better distinguish predictive relationships from causal mechanisms. Finally, future research may further develop dynamic interpretability approaches that are more closely aligned with real-time forecasting environments.

Reference

Acemoglu, D., Aghion, P., Bursztyn, L., & Hemous, D. (2012). The environment and directed technical change. *American Economic Review*, *102*(1), 131–166. https://doi.org/10.1257/aer.102.1.131

Adom, P. K., Bekoe, W., & Akoena, S. K. K. (2012). Modelling aggregate domestic electricity demand in ghana: An autoregressive distributed lag bounds cointegration approach. *Energy Policy*, *42*, 530–537. https://doi.org/10.1016/j.enpol.2011.12.019

Akintande, O. J., Olubusoye, O. E., Adenikinju, A. F., & Olanrewaju, B. T. (2020). Modeling the determinants of renewable energy consumption: Evidence from the five most populous nations in africa. *Energy*, *206*, 117992. https://doi.org/10.1016/j.energy.2020.117992

Akyürek, Z. (2020). LMDI decomposition analysis of energy consumption of Turkish manufacturing industry: 2005–2014. *Energy Efficiency*, *13*(4), 649–663. https://doi.org/10.1007/s12053-020-09846-8

Bagherzadeh, F., Nouri, A. S., Mehrani, M.-J., & Thennadil, S. (2021). Prediction of energy consumption and evaluation of affecting factors in a full-scale WWTP using a machine learning approach. *Process Safety and Environmental Protection*, *154*, 458–466. https://doi.org/10.1016/j.psep.2021.08.040

Bernanke, B. S., Gertler, M., & Gilchrist, S. (1999). *Handbook of Macroeconomics*. Elsevier.

Bhavsar, N., & Khushbu, D. (2025). *Empirical analysis of linear interpolation for missing data recovery in soft computing applications* (SSRN Scholarly Paper No. 5197761). Social Science Research Network. https://doi.org/10.2139/ssrn.5197761

Bloch, H., Rafiq, S., & Salim, R. (2015). Economic growth with coal, oil and renewable energy consumption in China: Prospects for fuel substitution. *Economic Modelling*, *44*, 104–115. https://doi.org/10.1016/j.econmod.2014.09.017

Bontempi, G., Ben Taieb, S., & Le Borgne, Y.-A. (2013). *Machine learning strategies for time series forecasting* (Vol. 138). https://doi.org/10.1007/978-3-642-36318-4_3

Borensztein, E., De Gregorio, J., & Lee, J.-W. (1998). How does foreign direct investment affect economic growth? *Journal of International Economics*, *45*(1), 115–135. https://doi.org/10.1016/S0022-1996(97)00033-0

Campbell, J. Y., & Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies*, *21*(4), 1509–1531. https://doi.org/10.1093/rfs/hhm055

Canh, N. P., Schinckus, C., Thanh, S. D., & Chong, F. H. L. (2021). The determinants of the energy consumption: A shadow economy-based perspective. *Energy*, *225*, 120210. https://doi.org/10.1016/j.energy.2021.120210

Clements, A. E., Hurn, A. S., & Li, Z. (2016). Forecasting day-ahead electricity load using a multiple equation time series approach. *European Journal of Operational Research*, *251*(2), 522–530. https://doi.org/10.1016/j.ejor.2015.12.030

de Oliveira, E. M., & Cyrino Oliveira, F. L. (2018). Forecasting mid-long term electric energy consumption through bagging ARIMA and exponential smoothing methods. *Energy*, *144*, 776–788. https://doi.org/10.1016/j.energy.2017.12.049

Dokas, I., Panagiotidis, M., Papadamou, S., & Spyromitros, E. (2022). The Determinants of Energy and Electricity Consumption in Developed and Developing Countries: International Evidence. *Energies*, *15*(7), 2558. https://doi.org/10.3390/en15072558

Esseghir, A., & Haouaoui Khouni, L. (2014). Economic growth, energy consumption and sustainable development: The case of the union for the mediterranean countries. *Energy*, *71*, 218–225. https://doi.org/10.1016/j.energy.2014.04.050

Esso, L. J. (2010). Threshold cointegration and causality relationship between energy use and growth in seven african countries. *Energy Economics*, *32*(6), 1383–1391. https://doi.org/10.1016/j.eneco.2010.08.003

Fang, D., Shi, S., & Yang, J. (2017). A study on forecasting and early warning of China's energy demand under the new normal. *Resource Development & Market*. https://doi.org/10.%203969%20/%20j.%20issn.%201005%20-%208141.%202017.%2001.%20002

Grossman, G. M., & Krueger, A. B. (1995). Economic growth and the environment. *The Quarterly Journal of Economics*, *110*(2), 353–377. https://doi.org/10.2307/2118443

Huang, N., Hu, Z., Cai, G., & Yang, D. (2016). Short term electrical load forecasting using mutual information based feature selection with generalized minimum-redundancy and maximum-relevance criteria. *Entropy*, *18*(9), 330. https://doi.org/10.3390/e18090330

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice*. OTexts. https://books.google.com/books?hl=en&lr=&id=_bBhDwAAQBAJ&oi=fnd&pg=PA7&dq=info:QamsSnE0lWMJ:scholar.google.com&ots=Tjl-snUJJM&sig=8zbWbuR_DK27ysXr0k3ItwRjl7c

IEA. (2022, October 27). *World Energy Outlook 2022 – Analysis*. IEA. https://www.iea.org/reports/world-energy-outlook-2022

Inoue, A., Jin, L., & Rossi, B. (2017). Rolling window selection for out-of-sample forecasting with time-varying parameters. *Journal of Econometrics*, *196*(1), 55–67. https://doi.org/10.1016/j.jeconom.2016.03.006

Jevons, W. S., & Flux, A. W. (with New York Public Library). (1965). *The coal question; an inquiry concerning the progress of the Nation, and the probable exhaustion of our coal-mines*. New York, A. M. Kelley. http://archive.org/details/coalquestionani00jevogoog

Karatasou, S., Santamouris, M., & Geros, V. (2006). Modeling and predicting building's energy use with artificial neural networks: Methods and results. *Energy and Buildings*, *38*(8), 949–958. https://doi.org/10.1016/j.enbuild.2005.11.005

Karmaker, S. (2025). *Feature engineering in time series forecasting*. https://doi.org/10.5281/zenodo.17573456

Kouakou, A. K. (2011). Economic growth and electricity consumption in cote d'ivoire: Evidence from time series analysis. *Energy Policy*, *39*(6), 3638–3644. https://doi.org/10.1016/j.enpol.2011.03.069

Lee, C.-C., & Chien, M.-S. (2010). Dynamic modelling of energy consumption, capital stock, and real income in G-7 countries. *Energy Economics*, *32*(3), 564–581. https://doi.org/10.1016/j.eneco.2009.08.022

Leung, K. K., Rooke, C., Smith, J., Zuberi, S., & Volkovs, M. (2023). *Temporal Dependencies in Feature Importance for Time Series Predictions* (arXiv:2107.14317). arXiv. https://doi.org/10.48550/arXiv.2107.14317

Li, J., Zhang, X., Ali, S., & Khan, Z. (2020). Eco-innovation and energy productivity: New determinants of renewable energy consumption. *Journal of Environmental Management*, *271*, 111028. https://doi.org/10.1016/j.jenvman.2020.111028

Liu, C., & Qian, X. (2023). Forecast of carbon emissions from energy consumption in China under the "dual carbon" targets (in Chinese). *Resources Science*, *45*(10), 1931–1946.

Liu, N., Tang, Q., Zhang, J., Fan, W., & Liu, J. (2014). A hybrid forecasting model with parameter optimization for short-term load forecasting of micro-grids. *Applied Energy*, *129*, 336–345. https://doi.org/10.1016/j.apenergy.2014.05.023

Liu, X., Moreno, B., & García, A. S. (2016). A grey neural network and input-output combined forecasting model. Primary energy consumption forecasts in Spanish economic sectors. *Energy*, *115*, 1042–1054. https://doi.org/10.1016/j.energy.2016.09.017

Ma, B. (2019). Calculation method of China's energy consumption based on least squares support vector machine (in Chinese). *Industrial Technology Economics, 139–144*, *38*(6).

Mahalik, M. K., & Mallick, H. (2014). Energy Consumption, Economic Growth and Financial Development: Exploring the Empirical Linkages for India. *The Journal of Developing Areas*, *48*(4), 139–159.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLOS ONE*, *13*((3)), e0194889. https://doi.org/https://doi.org/10.1371/journal.pone.0194889

Malik, A. A., Tao, L., Allias, N., & Hamzah, I. H. (2025). Analysing feature selection: Impacts towards forecasting electricity power consumption. *International Journal of Reconfigurable and Embedded Systems (Ijres)*, *14*(1), 265. https://doi.org/10.11591/ijres.v14.i1.pp265-272

michael.lecuivre@reacfin.com, samuel mahy@reacfin com. (2022). *Machine Learning Interpretability*.

Moghram, I., & Rahman, S. (1989). Analysis and evaluation of five short-term load forecasting techniques. *IEEE Transactions on Power Systems*, *4*(4), 1484–1491. https://doi.org/10.1109/59.41700

National Bureau of Statistics of China. (2025). *China statistical yearbook 2025*. China Statistics Press. https://www.stats.gov.cn/sj/ndsj/2025/indexeh.htm

Neubauer, A., Brandt, S., & Kriegel, M. (2025). Explainable multi-step heating load forecasting: Using SHAP values and temporal attention mechanisms for enhanced interpretability. *Energy and AI*, *20*, 100480. https://doi.org/10.1016/j.egyai.2025.100480

Nowicka-Zagrajek, J., & Weron, R. (2002). Modeling electricity loads in California: ARMA models with hyperbolic noise. *Signal Processing*, *82*(12), 1903–1915. https://doi.org/10.1016/S0165-1684(02)00318-3

Omri, A., & Kahouli, B. (2014). Causal relationships between energy consumption, foreign direct investment and economic growth: Fresh evidence from dynamic simultaneous-equations models. *Energy Policy*, *67*, 913–922. https://doi.org/10.1016/j.enpol.2013.11.067

Pao, H. T. (2009). Forecasting energy consumption in Taiwan using hybrid nonlinear models. *Energy, 11th Conference on Process Integration, Modelling and Optimisation for Energy Saving and Pollution Reduction*, *34*(10), 1438–1446. https://doi.org/10.1016/j.energy.2009.04.026

Rahman, A., Srikumar, V., & Smith, A. D. (2018). Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. *Applied Energy*, *212*, 372–385. https://doi.org/10.1016/j.apenergy.2017.12.051

Sadorsky, P. (2010). The impact of financial development on energy consumption in emerging economies. *Energy Policy*, *38*(5), 2528–2535. https://doi.org/10.1016/j.enpol.2009.12.048

Sen, P., Roy, M., & Pal, P. (2016). Application of ARIMA for forecasting energy consumption and GHG emission: A case study of an Indian pig iron manufacturing organization. *Energy*, *116*, 1031–1038. https://doi.org/10.1016/j.energy.2016.10.068

Shirzadi, N., Nizami, A., Khazen, M., & Nik-Bakht, M. (2021). Medium-Term Regional Electricity Load Forecasting through Machine Learning and Deep Learning. *Designs*, *5*(2), 27. https://doi.org/10.3390/designs5020027

Steenhof, P. A. (2006). Decomposition of electricity demand in China’s industrial sector. *Energy Economics*, *28*(3), 370–384. https://doi.org/10.1016/j.eneco.2006.02.006

Stern, D. I. (2011). The role of energy in economic growth. *Annals of The New York Academy Of Sciences*, *1219*(1), 26–51. https://doi.org/10.1111/j.1749-6632.2010.05921.x

Štrumbelj, E., & Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems*, *41*(3), 647–665. https://doi.org/10.1007/s10115-013-0679-x

Szoplik, J. (2015). Forecasting of natural gas consumption with artificial neural networks. *Energy*, *85*, 208–220. https://doi.org/10.1016/j.energy.2015.03.084

Thomas, S., & Rosenow, J. (2020). Drivers of increasing energy consumption in Europe and policy implications. *Energy Policy*, *137*, 111108. https://doi.org/10.1016/j.enpol.2019.111108

Wang Y., Hong Z., & Fu B. (2020). Research on characteristics of highway passenger flow based on multivariate data—Taking ningbo city as an example. *Open Journal of Transportation Technologies*, *09*(01), 32–42. https://doi.org/10.12677/OJTT.2020.91005

Wang, Y., & Li, Y. (2010). Applying LS-SVM to predict primary energy consumption. *2010 International Conference on E-Product E-Service and E-Entertainment*, 1–4. https://doi.org/10.1109/ICEEE.2010.5660473

Wang, Y., Wang, L., & Zhang, Q. (2019). Decomposition of manufacturing-related electricity consumption intensity in China using the LMDI method: 1990–2015. *Energy Efficiency*, *12*(7), 1837–1855. https://doi.org/10.1007/s12053-019-09794-y

World Bank Group. (2022). *Global Economic Prospects* [Text/HTML]. https://www.worldbank.org/en/publication/global-economic-prospects

Xi, J. (2020). *Full text of Xi's statement at the General Debate of the 75th Session of the United Nations General Assembly | english.scio.gov.cn*. Xinhua. Full Text of Xi's Statement at the General Debate of the 75th Session of the United Nations General Assembly. http://english.scio.gov.cn/topnews/2020-09/23/content_76731466.htm

Xi, Y., Zhang, L., Zhou, Y., & Wu, L. (2024). Trade policy uncertainty and firm-level pollution emissions: Evidence from China's accession to the WTO. *Finance Research Letters*, *59*, 104713. https://doi.org/10.1016/j.frl.2023.104713

Yao, Y., Ivanovski, K., Inekwe, J., & Smyth, R. (2019). Human capital and energy consumption: Evidence from OECD countries. *Energy Economics*, *84*, 104534. https://doi.org/10.1016/j.eneco.2019.104534

Zaman, K., Khan, M. M., Ahmad, M., & Rustam, R. (2012). Determinants of electricity consumption function in pakistan: Old wine in a new bottle. *Energy Policy, Special Section: Past and Prospective Energy Transitions - Insights from History*, *50*, 623–634. https://doi.org/10.1016/j.enpol.2012.08.003

Zhao, Z., & Liu, Q. (2019). Research on China's energy strategy planning: Forecast based on energy consumption, production, and structure (in Chinese). *Resources and Industries*, *21*, 1–8. https://doi.org/10.13776/j.cnki.resourcesindustries.20191206.007

Zhong, W., Song, J., Yang, W., Fang, K., & Liu, X. (2021). Evolving household consumption-driven industrial energy consumption under urbanization: A dynamic input-output

analysis. *Journal of Cleaner Production*, *289*, 125732.
https://doi.org/10.1016/j.jclepro.2020.125732

Appendix

Appendix 1 Explanation and Justification for the Retention of Outliers

China’s highway passenger traffic *(47)* reached a stage-specific peak during 2010–2012 and has declined steadily since then. This change should not be regarded as an abnormal fluctuation, but rather as a result of structural adjustments in the transportation system. The 14th Five-Year Plan for the Development of a Modern Comprehensive Transportation System points out that, with the continuous improvement of the integrated transport system and the ongoing optimization of the transport structure, the shares of railway and civil aviation will increase, while the proportion of medium- and long-distance highway passenger transport will gradually decline. This indicates that the decrease in highway passenger traffic is a long-term trend. Wang et al.(2020), using Ningbo as a case study, found that highway passenger volume has decreased at an average annual rate of about 7% since 2012, and that the opening of high-speed rail is a major factor contributing to this decline. Therefore, the decline in highway passenger traffic around 2012 should be understood as a structural change driven by competition and substitution among multiple modes of transport.

In the early 1990s, the large fluctuations in the Producer Price Index *(X37),* Retail Price Index *(X38),* and Foreign Direct Investment *(X40)* had a common cause. In 1992, Deng Xiaoping’s Southern Tour confirmed the direction of accelerating Reform and Opening up. After that, the 14th National Congress officially proposed building a socialist market economy, which pushed Reform and Opening up to a deeper stage. Under this background, investment and demand increased quickly, and foreign capital flowed in large amounts. These became important institutional factors in the early 1990s, driving macroeconomic fluctuations. With stronger macroeconomic control and the impact of the Asian financial crisis, demand declined, prices fell, and foreign investment stabilized. Therefore, the fluctuations in this period are the result of macroeconomic cycles and policy changes, rather than abnormal values.

Energy Production Elasticity *(X45)* was significantly elevated in 2022, mainly due to the combined effects of strengthened energy supply security policies and a slowdown in economic growth. Against the backdrop of power shortages in 2021, the government intensified efforts in 2022 to increase coal production and ensure energy supply. In addition, the global energy tightness triggered by the Russia–Ukraine conflict further accelerated the rapid growth in energy production(IEA, 2022). Meanwhile, affected by the pandemic, economic growth remained relatively subdued, resulting in energy production growth significantly outpacing economic growth, and thus leading to a sharp rise in the elasticity coefficient(World Bank Group, 2022). Therefore, this high value reflects a temporary fluctuation due to a supply–demand mismatch during a specific period, rather than an abnormal data point.

Appendix 2 Supplementary Tables and Figures

Table A-Definition and Description of Variable

Code	Category	Variable Name	Unit	Data Source
X1	National Accounts	Per Capita GDP	yuan	China Statistical Yearbook 2025
X2	National Accounts	Gross Domestic Product	100 million yuan	China Statistical Yearbook 2025
X3	National Accounts	Value Added of Primary Industry	100 million yuan	China Statistical Yearbook 2025
X4	National Accounts	Value Added of Secondary Industry	100 million yuan	China Statistical Yearbook 2025
X5	Industry	Industrial Value Added	—	China Statistical Yearbook 2025
X6	National Accounts	Value Added of Tertiary Industry	100 million yuan	China Statistical Yearbook 2025
X7	Population	Total Population	10,000 persons	China Statistical Yearbook 2025
X8	Population	Urban Population	10,000 persons	China Statistical Yearbook 2025
X9	Population	Rural Population	10,000 persons	China Statistical Yearbook 2025
X10	Population	Natural Population Growth Rate	%	China Statistical Yearbook 2025
X11	Transportation, Postal Services and Software Industry	Railway Passenger Traffic	10,000 persons	China Statistical Yearbook 2025

Code	Category	Variable Name	Unit	Data Source
X12	Transportation, Postal Services and Software Industry	Highway Passenger Traffic	10,000 persons	China Statistical Yearbook 2025
X13	Transportation, Postal Services and Software Industry	Waterway Passenger Traffic	10,000 persons	China Statistical Yearbook 2025
X14	Transportation, Postal Services and Software Industry	Civil Aviation Passenger Traffic	10,000 persons	China Statistical Yearbook 2025
X15	Transportation, Postal Services and Software Industry	Total Freight Volume	10,000 tons	China Statistical Yearbook 2025
X16	Transportation, Postal Services and Software Industry	Railway Freight Volume	10,000 tons	China Statistical Yearbook 2025
X17	Transportation, Postal Services and Software Industry	Highway Freight Volume	10,000 tons	China Statistical Yearbook 2025
X18	Transportation, Postal Services and Software Industry	Waterway Freight Volume	10,000 tons	China Statistical Yearbook 2025
X19	Transportation, Postal Services and Software Industry	Civil Aviation Freight Volume	10,000 tons	China Statistical Yearbook 2025
X20	Retail Trade	Total Retail Sales of Consumer Goods	100 million yuan	National Bureau of Statistic Database

Code	Category	Variable Name	Unit	Data Source
X21	Fixed Asset Investment	Total Fixed Asset Investment	100 million yuan	National Bureau of Statistic Database
X22	Energy	Electricity Generation	100 million kWh	National Bureau of Statistic Database
X23	Energy	Hydropower Generation	100 million kWh	National Bureau of Statistic Database
X24	Energy	Thermal Power Generation	100 million kWh	National Bureau of Statistic Database
X25	Energy	Natural Gas Production	100 million m^3	National Bureau of Statistic Database
X26	Energy	Coke Production	10,000 tons	National Bureau of Statistic Database
X27	Energy	Crude Oil Production	10,000 tons	National Bureau of Statistic Database
X28	International Trade	Fuel Exports	million USD	National Bureau of Statistic Database
X29	International Trade	Manufactured Goods Exports	million USD	National Bureau of Statistic Database

Code	Category	Variable Name	Unit	Data Source
X30	International Trade	Fuel Imports	million USD	National Bureau of Statistic Database
X31	International Trade	Manufactured Goods Imports	million USD	National Bureau of Statistic Database
X32	Agriculture	Total Sown Area of Crops	1,000 hectares	National Bureau of Statistic Database
X33	Agriculture	Agricultural Machinery Power	10,000 kW	National Bureau of Statistic Database
X34	Agriculture	Grain Output	10,000 tons	National Bureau of Statistic Database
X35	Energy	Total Energy Production	10,000 tons standard coal	China Statistical Yearbook 2025
X36	Prices	Consumer Price Index	(last year=100)	National Bureau of Statistic Database
X37	Prices	Producer Price Index	(last year=100)	National Bureau of Statistic Database
X38	Prices	Retail Price Index	(last year=100)	National Bureau of Statistic Database
X39	International Trade	Total Imports and Exports	100 million yuan	National Bureau of Statistic Database

Code	Category	Variable Name	Unit	Data Source
X40	International Trade	Foreign Direct Investment	100 million USD	National Bureau of Statistic Database
X41	Energy	Per Capita Raw Coal Production	tons	National Bureau of Statistic Database
X42	Energy	Per Capita Crude Oil Production	kg	National Bureau of Statistic Database
X43	Energy	Per Capita Electricity Generation	kWh	National Bureau of Statistic Database
X44	Energy	Electricity Production Elasticity	—	National Bureau of Statistic Database
X45	Energy	Energy Production Elasticity	—	National Bureau of Statistic Database
X46	Energy	Rural Electricity Consumption	100 million kWh	National Bureau of Statistic Database
X47	Transportation, Postal Services and Software Industry	Railway Operating Mileage	10,000 km	National Bureau of Statistic Database
X48	Transportation, Postal Services and Software Industry	Highway Mileage	10,000 km	National Bureau of Statistic Database

Code	Category	Variable Name	Unit	Data Source
X49	Transportation, Postal Services and Software Industry	Railway Petroleum Freight	10,000 tons	National Bureau of Statistic Database
X50	Transportation, Postal Services and Software Industry	Railway Major Freight Volume	10,000 tons	National Bureau of Statistic Database

Table B - ADF Test Results and Stationarity Assessment

Variable	Differencing Order	ADF Test Statistic	10% Critical Value	Stationary (10%)
Y	2	-3.467296973	-3.18	TRUE
X1	2	-5.016250239	-3.18	TRUE
X2	2	-5.022467429	-3.18	TRUE
X3	0	-3.273890405	-3.18	TRUE
X4	2	-5.889446412	-3.18	TRUE
X5	0	-3.632053266	-3.18	TRUE
X6	2	-5.004677779	-3.18	TRUE
X7	2	-3.04410116	-3.18	FALSE
X8	2	-4.591375266	-3.18	TRUE
X9	0	-3.862244511	-3.18	TRUE
X10	1	-3.359797913	-3.18	TRUE
X11	2	-2.060919815	-3.18	FALSE
X12	2	-3.82750552	-3.18	TRUE
X13	2	-2.875006607	-3.18	FALSE
X14	2	-1.539364684	-3.18	FALSE
X15	2	-6.295522338	-3.18	TRUE
X16	1	-3.434026588	-3.18	TRUE
X17	1	-3.26079635	-3.18	TRUE
X18	2	-5.696368504	-3.18	TRUE
X19	1	-3.999026732	-3.18	TRUE

Variable	Differencing Order	ADF Test Statistic	10% Critical Value	Stationary (10%)
X20	2	-4.040926432	-3.18	TRUE
X21	1	-3.342860897	-3.18	TRUE
X22	2	-6.231816443	-3.18	TRUE
X23	1	-3.877121707	-3.18	TRUE
X24	1	-3.188493085	-3.18	TRUE
X25	2	-6.532888825	-3.18	TRUE
X26	2	-4.316171398	-3.18	TRUE
X27	2	-4.801052064	-3.18	TRUE
X28	1	-6.391668235	-3.18	TRUE
X29	1	-3.830038004	-3.18	TRUE
X30	1	-5.910107673	-3.18	TRUE
X31	1	-4.173171419	-3.18	TRUE
X32	1	-3.845185497	-3.18	TRUE
X33	1	-3.704740912	-3.18	TRUE
X34	1	-3.511703356	-3.18	TRUE
X35	2	-5.028072908	-3.18	TRUE
X36	0	-4.655004286	-3.18	TRUE
X37	0	-3.331620447	-3.18	TRUE
X38	0	-3.316615248	-3.18	TRUE
X39	1	-4.216208084	-3.18	TRUE
X40	0	-4.879819253	-3.18	TRUE
X41	2	-3.914324072	-3.18	TRUE
X42	2	-4.772431389	-3.18	TRUE
X43	2	-6.20410766	-3.18	TRUE
X44	1	-6.036346729	-3.18	TRUE
X45	1	-4.294877835	-3.18	TRUE
X46	1	-3.98704302	-3.18	TRUE
X47	2	-3.968830455	-3.18	TRUE
X48	1	-4.071079201	-3.18	TRUE
X49	1	-5.789399564	-3.18	TRUE

Variable	Differencing Order	ADF Test Statistic	10% Critical Value	Stationary (10%)
X50	1	-3.474639141	-3.18	TRUE

Table C - Hyperparameter Tuning and Cross-Validation Performance

Model	Hyperparameter Grid	Selected Parameters	RMSE	MAE
XGBoost	nrounds ∈ {50,100}; max_depth ∈ {2,3,4}; η ∈ {0.01,0.05,0.1}; subsample = 0.8; colsample = 0.8	nrounds = 50; max_depth = 2; η = 0.01; subsample = 0.8; colsample = 0.8	0.0231	0.0189
RF	ntree ∈ {200,500}; mtry ∈ {√p, p/3, p/2}	ntree = 200; mtry = 6	0.0253	0.0219
Ridge	λ ∈ log-linear sequence (glmnet default)	λ = 0.5730	0.0258	0.0205
SVR	C ∈ {0.1,1,10}; γ ∈ {0.01,0.05,0.1}; ε ∈ {0.01,0.1}	C = 10; γ = 0.01; ε = 0.01	0.0260	0.0199
KNN	k ∈ {3,5,7,9}	k = 3	0.0281	0.0227
Lasso	λ ∈ log-linear sequence (glmnet default)	λ = 0.0080	0.0322	0.0259

Appendix 3 AI Use Disclosure

AI tool (ChatGPT) was used only for limited technical and language assistance outside the core intellectual contribution of this thesis. All generated outputs were reviewed, edited, and verified by the author before inclusion.

Table D - AI Use Disclosure

Activity	Scope of Assistance	Author Review
Language polishing and grammar correction	Assistance with improving academic phrasing, grammar, sentence clarity, and readability throughout the thesis.	All edits were reviewed and approved by the author before inclusion.
R code support	Assistance with troubleshooting and refining R scripts for data analysis and visualization.	All code was inspected, adapted, and tested by the author prior to use.
LaTeX formatting support	Assistance with LaTeX formatting, table layout, and document structure adjustments.	The author verified all formatting and compiled the final document manually.

Resümee

HIINA KOGUENERGIATARBE PROGNOOSIMINE MASINÕPPE MUDELITE ABIL

Xinran Zhang

Globaalse energiasiirde ja Hiina „kahe süsiniku“ eesmärgi kontekstis muutub energiatarbimise täpne prognoosimine ning selle mõjutegurite mõistmine üha olulisemaks. Käesolevas uuringus töötatakse välja välja prognoosiraamistik Hiina aastaste makromajanduslike andmete analüüsimiseks, ühendades masinõppemudelid selgitatavate meetoditega. Kaasatud muutujad pärinevad mitmest valdkonnast, sealhulgas majandusest, tööstusest, energiast ja rahvusvahelisest kaubandusest, ning ajalise sõltuvuse arvestamiseks moodustatakse viitajaga tunnused. Tunnuste valikuks kasutatakse vastastikust informatsiooni. Töös rakendatakse mitut masinõppemudelit, sealhulgas Ridge'i, Lasso't, tugivektorregressiooni (SVR), juhumetsa (RF), XGBoosti ja k-lähima naabri meetodit (KNN), ning võrreldakse neid esimese järgu autoregressiivse võrdlusmudeliga AR(1). Mudelite prognoosivõimet hinnatakse veereva ristvalideerimise ja pseudo-valimivälise prognoosimise raamistikus.

Tulemused näitavad, et väikese valimiga makromajanduslike aegridade puhul on regulariseeritud lineaarsed mudelid, eriti Ridge'i mudel, prognoosimisel stabiilsemad ja usaldusväärsemad kui paindlikumad masinõppemudelid. Ridge'i mudel annab AR(1) võrdlusmudeliga võrreldavaid prognoositulemusi, kuid säilitab eri prognoosistsenaariumides suurema stabiilsuse. Tõlgendatavuse parandamiseks kasutatakse SHAP-meetodit (SHapley Additive exPlanations), et analüüsida mudelite käitumist nii muutujate kui ka ajamõõtme lõikes. Tulemused viitavad sellele, et energiavarustuse tingimused, majanduse avatust kirjeldavad näitajad ja reaalmajanduslik aktiivsus on tihedalt seotud energiatarbimise dünaamikaga. Lisaks tugineb enamik mudeleid eelkõige lühiajalisele viitajaga teabele, mis rõhutab hiljutise teabe olulisust prognoosimisel. Prognoosimise ja selgitatava masinõppe meetodite ühendamise kaudu suurendab uuring masinõppe prognoosimudelite läbipaistvust ning pakub kasulikke sisendeid Hiina energiapoliitika ja makromajandusliku reguleerimise jaoks.

Märksõnad: energiatarbimine; masinõpe; prognoosimine

JEL: C41, C45, C53

Non-exclusive licence to reproduce the thesis and make the thesis public

I, Xinran Zhang ,

1. grant the University of Tartu a free permit (non-exclusive licence) to

   reproduce, for the purpose of preservation, including for adding to the digital archives of the University of Tartu until the expiry of the term of copyright, my thesis

   Forecasting China's Total Energy Consumption Using Machine Learning Models ,

   supervised by Luca Alfieri, PhD ;

2. grant the University of Tartu a permit to make the thesis specified in point 1 available to the public via the web environment of the University of Tartu, including via the digital archives, under the Creative Commons licence CC BY NC ND 4.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright;

3. am aware of the fact that the author retains the rights specified in points 1 and 2;

4. confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Xinran Zhang
19/05/2026