Team Number :	apmcm2304018
Problem Chosen :	C

2023 APMCM summary sheet

Solving Key Problems in Electric Vehicle based on Machine Learning

We developed five models to understand the automotive industry and the impact of policies on vehicular exports. These models cover the correlation between stock performances and electric vehicle (EV) sales, forecast export trends amidst policy landscapes, predict ecological repercussions of vehicular electrification, and evaluate the multivariate impact on air quality.

Model 1 explores relationships among domestic EV sales, global EV sales, and the stock performance of industry giants like Tesla (TSLA) and NIO. Correlation analysis unveils insights into how market dynamics and global trends influence the stock prices of leading electric vehicle manufacturers.

Model 2 analyzes China's annual automobile export curve and its correlation with specific stock performances, focusing on anti-China EV policy influence. This model establishes connections between geopolitical decisions and the automotive market, offering stakeholders a valuable tool for anticipating industry shifts.

Model 3 widens the scope to influential players like Toyota, Ford, Hyundai, China Automotive Systems, Porsche, General Motors, and Nissan, exploring correlations between their stock performances and electric vehicle sales, unraveling market trends and strategic positioning.

Model 4 redirects attention to China's automobile export dynamics and the influence of foreign policies, forecasting the impact of policy changes on tram exports amid geopolitical shifts.

Model 5 addresses the ecological impact of vehicular electrification on the Air Quality Index (AQI). Constructing a multi-layer perceptron (MLP) with diverse city variables, this model evaluates impacts on energy conservation, emissions reduction, and pollution alleviation, offering insights into the relationship between electric vehicle sales and air quality.

In conclusion, these models contribute to a holistic understanding of the automotive industry, enriching our comprehension of market correlations, export forecasts, ecological impact assessments, and air quality predictions. As the industry evolves, these models serve as invaluable tools, guiding stakeholders toward informed decisions for a sustainable and innovative automotive future.

Key words: ARIMA; Exponential Regression; Spearman Correlation; MLP)

Contents

1 Introduction **1**
1.1 Problem Background 1
1.2 Restatement of the Problem 1
1.3 Our Work 2
2 Assumptions and Justifications **3**
2.1 Model 1 3
2.2 Model 2 3
2.3 Model 3 3
2.4 Model 4 3
2.5 Model 5 4
3 Notations **4**
4 Correlation analysis Between EV Sales and multi-factors **4**
4.1 Data Description 4
4.2 The Establishment of Model 1 5
4.2.1 Variable Selection 6
4.2.2 Spearman Correlation Analysis 6
4.2.3 Modeling Steps: 6
4.3 The Solution of Model 1 7
4.3.1 Correlation Analysis 7
4.3.2 Interpretation of Weak Positive Correlation with Gasoline Prices: 7
5 Best-Order ARIMA for Future Development of China EV **8**
5.1 Data Description 8
5.2 The Establishment of Model 2 9
5.2.1 Preliminary Analysis and Auto ARIMA 9
5.2.2 Grid Search for Optimal Parameters 9
5.2.3 ARIMA Model with Optimal Parameters 10
5.2.4 Results and Diagnostic Charts: 10
5.3 The Solution of Model 2 10
5.3.1 Comparison of Auto ARIMA and Best-Order ARIMA Results: 10
5.3.2 Comparison of Diagnostic Information for Charging Station Predictions (Auto vs. Best-Order ARIMA): 11
5.3.3 Comparison of Diagnostic Information for Sales Volume Predictions: 12
6 Spearman Coefficient Analysis for New-old Industry **13**

6.1 Data Description 13
6.2 The Establishment of Model 3 13
6.2.1 Spearman Coefficient Analysis and Its Relevance 13
6.2.2 Analysis of the Impact of New Energy Electric Vehicles on the Global Traditional Energy Vehicle Industry 13
6.3 The Solution of Model 3 14
7 Anti-China-EV Policy Influence Analysis 15
7.1 Data Description 15
7.2 The Establishment of Model 4 15
7.3 The Solution of Model 4 16
8 MLP for Electrification Impact on Ecologic System 17
8.1 Data Description 17
8.1.1 Data Overview 17
8.1.2 Multivariate Impact on Air Quality Index (AQI) based on Varied City Data 18
8.2 The Establishment of Model 5 18
8.3 The Solution of Model 5 19
9 Publicity Letter: Contribution of EV to Society 20
10 Sensitivity Analysis 21
10.1 PAC 21
10.2 Residuals 21
10.3 AIC 22
11 Model Evaluation and Further Discussion 22
11.1 Strengths 22
11.1.1 PCA: 22
11.1.2 Grid Search Optimized ARIMA 23
11.1.3 Exponential Regression for Export Prediction 23
11.2 Weaknesses 23
12 Conclusion 23
References 24

1 Introduction

1.1 Problem Background

In response to burgeoning environmental concerns, the global spotlight has shifted towards new energy vehicles (NEVs), notably electric vehicles (EVs). This paradigm shift in the automotive landscape, encompassing hybrid, pure electric, and fuel cell vehicles, is driven by their low pollution, reduced energy consumption, and peak electricity demand modulation capabilities. In China, resolute government support, marked by initiatives like charging station proliferation and EV subsidies, has propelled NEVs to a pivotal role in shaping the nation's macroeconomic trajectory, emphasizing a commitment to environmental consciousness.

Buoyed by government backing, China's NEV sector has witnessed significant development, overcoming challenges through increased charging infrastructure and preferential pricing. This positive momentum not only establishes China as a global trailblazer in NEV adoption but also sets the stage for exploring future trends and the sector's impact on traditional global automotive industries. The commendable growth of NEVs in China underscores a holistic dedication to environmental sustainability and technological innovation, prompting an in-depth exploration of the sector's dynamics and implications for the future.

1.2 Restatement of the Problem

Considering the background information and restricted conditions identified in the problem statement, we need to solve the following problems:

For Problem 1, identify and model the key factors influencing the development of new energy electric vehicles in China.

For Problem 2, create a mathematical model based on collected industry data to predict the next decade's development of new energy electric vehicles in China.

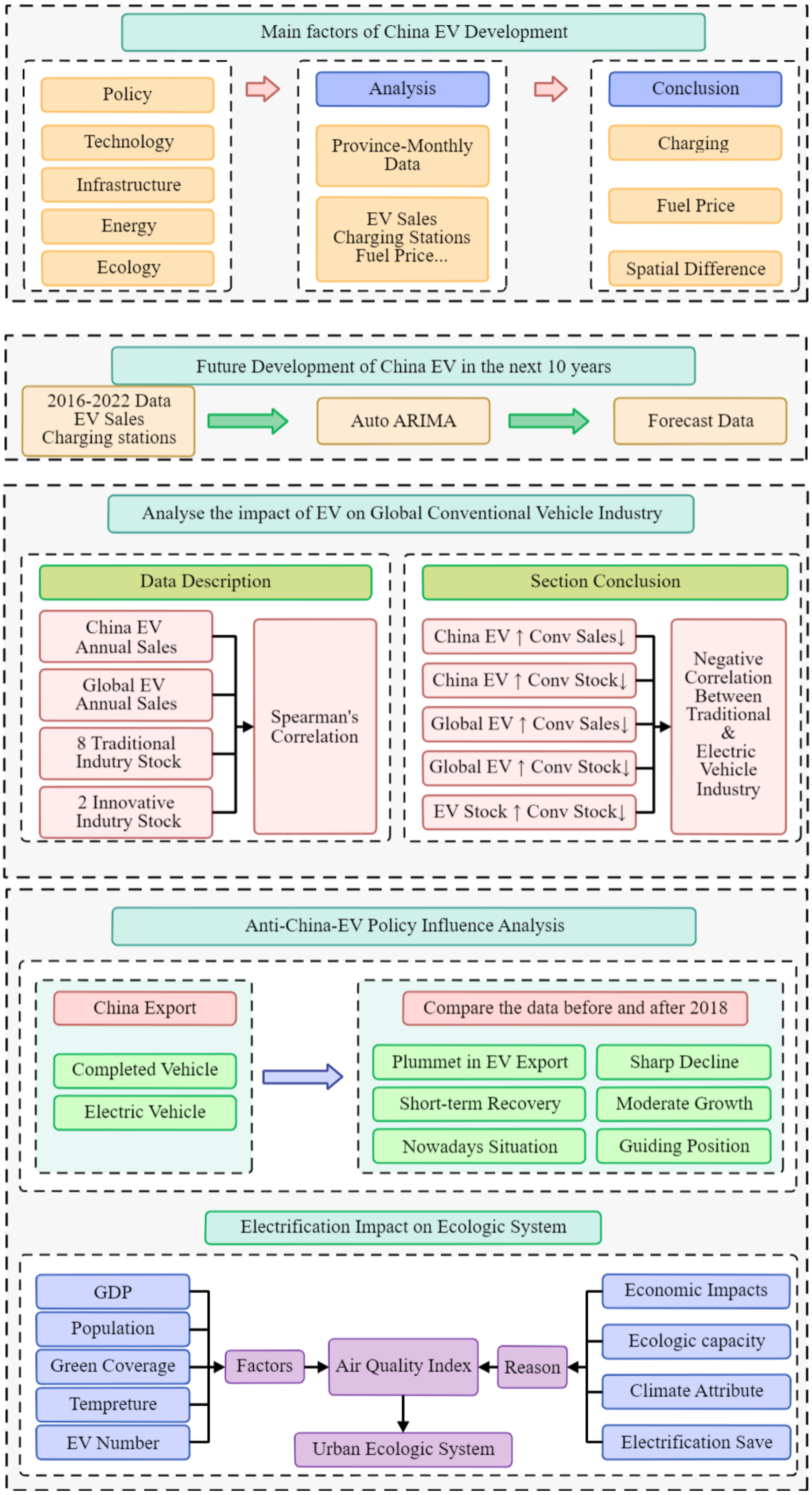
For Problem 3, establish a data-driven model to analyze how new energy electric vehicles impact the global traditional energy vehicle industry.

For Problem 4, develop a mathematical model to assess the impact of foreign policies aimed at hindering China's new energy electric vehicle development.

For Problem 5, Analyze the electrical impact of urban new energy electric vehicles on the environment, providing quantitative results for a city with a one-million population.

For Problem 6, Craft an open letter to the public, emphasizing the global benefits of new energy electric vehicles and the positive contributions of the electric vehicle industry to society and the environment, based on the findings from Problem 5.

1.3 Our Work


Main factors of China EV Development
Policy
Technology
Infrastructure
Energy
Ecology
Analysis
Province-Monthly Data
EV Sales Charging Stations Fuel Price...
Conclusion
Charging
Fuel Price
Spatial Difference
Future Development of China EV in the next 10 years
2016-2022 Data EV Sales Charging stations
Auto ARIMA
Forecast Data
Analyse the impact of EV on Global Conventional Vehicle Industry
Data Description
China EV Annual Sales
Global EV Annual Sales
8 Traditional Indutry Stock
2 Innovative Indutry Stock
Spearman's Correlation
Section Conclusion
China EV ↑ Conv Sales↓
China EV ↑ Conv Stock↓
Global EV ↑ Conv Sales↓
Global EV ↑ Conv Stock↓
EV Stock ↑ Conv Stock↓
Negative Correlation Between Traditional & Electric Vehicle Industry
Anti-China-EV Policy Influence Analysis
China Export
Completed Vehicle
Electric Vehicle
Compare the data before and after 2018
Plummet in EV Export
Sharp Decline
Short-term Recovery
Moderate Growth
Nowadays Situation
Guiding Position
Electrification Impact on Ecologic System
GDP
Population
Green Coverage
Tempreture
EV Number
Factors
Air Quality Index
Reason
Urban Ecologic System
Economic Impacts
Ecologic capacity
Climate Attribute
Electrification Save

2 Assumptions and Justifications

2.1 Model 1

Policy Factors: The implementation of government incentives directly impacts the sales of NEEVs and the establishment of charging infrastructure. Regional disparities in policy implementation across provinces lead to divergent patterns in NEEV development.

Technology Factors: Positive correlations exist between technological innovations and increased NEEV sales and charging infrastructure. Regional discrepancies in technological advancements result in varying impacts on NEEV development.

Infrastructure Factors: The level of infrastructure development is positively correlated with the number of charging stations. The impact of infrastructure on NEEV sales differs among cities due to varying levels of development.

Ecological Factors: Growing environmental awareness contributes to the widespread adoption of NEEVs. Ecological conditions in different provinces exert diverse effects on NEEV development.

2.2 Model 2

Trend Projection: Recent sales and charging infrastructure data provide reliable indicators for predicting the overall trend in EV development over the next ten years. National-level trends have a more substantial impact than provincial-level data on predicting future developments.

Choice of Indicators:EV sales directly reflect consumer acceptance and play a pivotal role in shaping the overall EV industry. The number of charging stations serves as a crucial metric, indicating the level of infrastructure development vital for the widespread adoption of EVs.

2.3 Model 3

Consumer Confidence and Stock Market Impact: Increasing global EV sales reflect heightened consumer confidence in new energy vehicles. [1] Fluctuations in the traditional automotive stock market mirror investor sentiments toward the broader automotive industry [2].

Global Data Selection: Global data provide a comprehensive overview, effectively showcasing the widespread impact of new energy vehicles on the traditional automobile industry [1].

2.4 Model 4

Direct Impact Assessment: China's EV export data serves as a direct indicator of the impact of anti-China policies on the sector. Contrasting whole vehicle export data helps isolate the effects on the EV sector from broader automotive industry challenges.

Choice of Policy and Market: The U.S. policy landscape significantly influences China's EV sector, with the U.S. being a crucial market. The unique characteristics of the U.S. market amplify its impact on China's automotive industry.

2.5 Model 5

Multidimensional Data Impact: Selecting multidimensional data offers a comprehensive view of the impact of electrification on urban ecosystems. [3] AQI serves as a quantifiable metric, effectively representing the ecological state of urban ecosystems.

Sample Selection Rationale: China's urban data completeness makes it an ideal sample for studying the broader impact of electrification on urban ecosystems. The accessibility of Chinese city-specific EV data facilitates a more extensive analysis of the universal impact of electrification on urban ecosystems.

3 Notations

The key mathematical notations used in this paper are listed in Table 1.

Table 1: Notations used in this paper

Symbol	Description	Unit
ρ	Spearman correlation coefficient	Scalar
d	the Difference Between the Ranks of Corresponding Pairs in Spearman Input	Scalar
n	the Number of Observations in Spearman Input	Scalar
X	Standardized Variables Matrix	Scalar
C	Covariance Matrix of X	Scalar
λ	Eigenvalue(s) of C	Scalar
V	Eigenvector(s) Corresponding to λ	Scalar
P	Principle Factor Loadings	Scalar
y	Exponential Function Fit Dependent Variable	Scalar
e	Exponential Constant	Scalar
x	Exponential Function Fit Independent Variable	Scalar
c	Exponential Function Fit Constant	Scalar

4 Correlation analysis Between EV Sales and multi-factors

4.1 Data Description

The primary objective in addressing problem 1 is to comprehensively understand and model the factors influencing the development of new energy electric vehicles in China. The available data, predominantly concerning charging station statistics, is characterized by limited granularity. This dataset offers cumulative figures up to a specific time point for various provinces, necessitating the generation of monthly data points through a meticulous data preprocessing strategy.

In the face of an inherent paucity of monthly data points, a strategic recourse entails the application of curve fitting and linear interpolation methodologies. Specifically, with respect to charging station data, judicious manual extraction of pivotal data junctures from the extant dataset establishes a foundational temporal snapshot. This snapshot then becomes the canvas upon which a meticulous linear interpolation strategy is imposed, effectively magnifying the density of data points. This methodological augmentation, executed through judicious interpolation along the curve, yields a comprehensive and continuous monthly profile. In essence, discrete key points, such as cumulative year-end figures, are judiciously positioned along the curve, and subsequent interpolation furnishes a refined dataset.

The subsequent exhibition of graphical representations, elucidating gasoline consumption data for five representative cities from January to September of the 23rd year, along with corresponding sales and charging station statistics, serves to tangibly underscore the efficacy of our enhanced dataset. This graphical exposition, adeptly crafted, serves as a visual testament to the viability and reliability of our data preprocessing methodology in encapsulating the nuanced dynamics of the NEEV landscape.

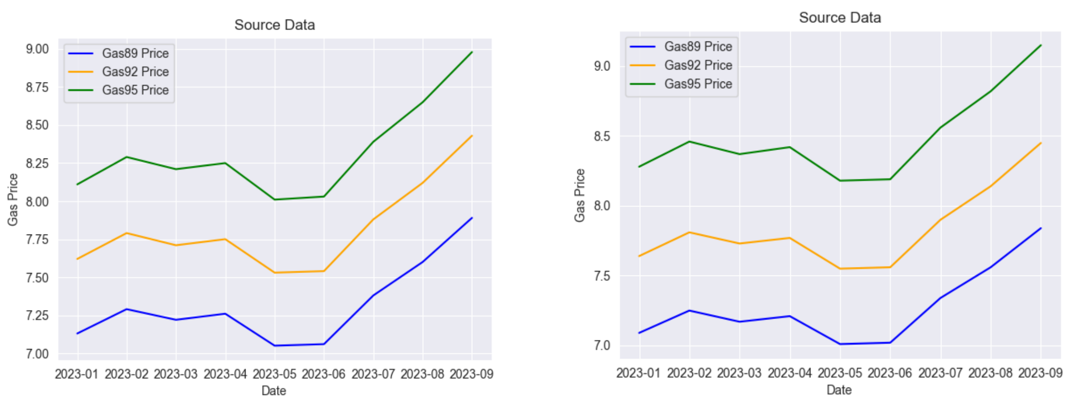


Figure 1: Gasoline data chart of Beijing and Guangdong from Jan. to Sept. 2023

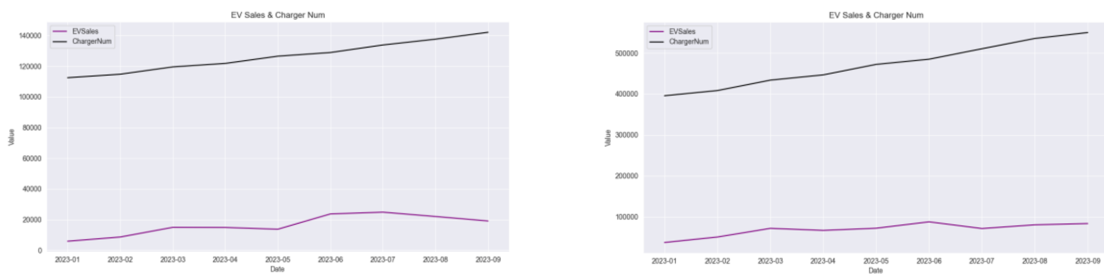


Figure 2: Beijing and Guangdong tram sales & charging pile data chart

This meticulous data preprocessing serves as the linchpin for the subsequent phases of our analytical pursuit, ensuring a methodologically robust foundation for the construction and resolution of our mathematical model. The ensuing sections endeavor to expound upon the intricate interplay among diverse factors shaping the trajectory of new energy electric vehicles in China, leveraging the refined dataset as the analytical cornerstone.

4.2 The Establishment of Model 1

The burgeoning development of new energy electric vehicles (NEEVs) in China necessitates a comprehensive understanding of the multifaceted factors influencing their growth.

This paper presents a comprehensive mathematical model to analyze and describe the key factors influencing the development of new energy electric vehicles (NEEVs) in China. The study focuses on the correlation between monthly gasoline prices (82, 92, 95 octane), monthly charging station quantities, and monthly sales across the regions of Beijing, Shanghai, Guangzhou, Jiangsu, and Zhejiang. Employing advanced statistical techniques, including Spearman correlation analysis and Principal Component Analysis (PCA), we aim to unveil the intricate relationships among these variables and assess their impact on the NEEV industry.

4.2.1 Variable Selection

The observed variables include monthly gasoline prices (82, 92, 95 octane), monthly charging station quantities, and monthly NEEV sales across the regions of Beijing, Shanghai, Guangzhou, Jiangsu, and Zhejiang. These variables were chosen based on their potential impact on the NEEV industry.

4.2.2 Spearman Correlation Analysis

To assess the relationships among the variables, Spearman correlation analysis was conducted. Spearman Correlation Analysis, named after Charles Spearman, is a statistical method used to assess the strength and direction of monotonic relationships between two variables. Unlike Pearson correlation, which measures linear relationships, Spearman correlation focuses on the order of values, making it suitable for variables that may exhibit non-linear associations. The result is expressed as a Spearman correlation coefficient (ρ), ranging from -1 to 1, where 1 indicates a perfect positive monotonic relationship, -1 indicates a perfect negative monotonic relationship, and 0 implies no monotonic relationship.

The fundamental principle underlying Spearman correlation is the concept of rank order. Instead of using the actual values of variables, it considers their ranks. Each value in the dataset is assigned a rank, and the correlation is then computed based on the comparison of these ranks. This makes Spearman correlation robust to outliers and suitable for ordinal data.

The choice of Spearman correlation, a non-parametric measure, is justified by its ability to capture monotonic relationships, crucial for understanding the trends in our data. A correlation coefficient greater than 0.5 indicates a strong correlation.

4.2.3 Modeling Steps:

Data Collection: Gather the data for the variables of interest. In the context of the NEEV study, the relevant data includes monthly gasoline prices (82, 92, 95 octane), monthly charging station quantities, and monthly NEEV sales across specified regions.

Rank Assignment: For each variable, assign ranks based on their values. Ties can be handled by assigning an average rank to tied values. These ranks are used in subsequent calculations.

Calculation of Differences: Compute the differences (d) between the ranks of corresponding pairs of observations for both variables. Square these differences (d^2) for subsequent analysis.

Calculation of Spearman Rank Correlation Coefficient: Calculate the Spearman correlation coefficient (ρ) using the formula:

$$\rho = 1 - \frac{6\sum d^2}{n(n^2 - 1)} \tag{1}$$

Where:

ρ is the Spearman correlation coefficient,

d is the difference between the ranks of corresponding pairs,

n is the number of observations.

4.3 The Solution of Model 1

4.3.1 Correlation Analysis

Take the chart of ZheJiang province as an example. To initiate our investigation, we focused on the first row of the correlation matrix, which illustrates the correlation coefficients of charging station quantity with other variables. Specifically, our primary interest was in understanding the correlation between the number of charging stations and electric vehicle sales. The observed correlation coefficient was 0.56, indicating a positive correlation greater than 0.5. This aligns with our expectations, affirming that an increase in charging stations tends to positively influence electric vehicle sales.

Figure 3: ZheJiang province Correlation Heatmap

Moving on to the prices of 89, 92, and 95 gasoline (denoted as 89 PRICE, 92 PRICE, 95 PRICE), we explored the third, fourth, and fifth rows of the matrix, respectively. Examining the second column, we found positive correlations between these traditional gasoline prices and electric vehicle sales. However, it's noteworthy that the correlation coefficients are less than 0.5, implying a relatively weak correlation.

4.3.2 Interpretation of Weak Positive Correlation with Gasoline Prices:

Despite the positive correlation, the coefficients being less than 0.5 suggests that the relationship is not strong. We attribute this to the inherent lag in consumer behavior regarding fuel prices and electric vehicle adoption. Consumer decisions are influenced by spatial and temporal factors, and our data, collected from January to September 2023, might not capture the full extent of these influences due to its relatively short time series.

Principal Component Analysis (PCA): Given the strong linear commonality among the selected variables (indicated by correlation coefficients concentrated between 0.5 and 1), traditional PCA might not be the most suitable method. However, recognizing the need for a supplementary validation tool, we proceeded with PCA.

The contribution table generated from PCA revealed that the contribution of charging stations far surpassed that of gasoline prices. Specifically, charging stations exhibited contributions exceeding 70%, validating the importance of charging infrastructure in influencing the electric vehicle market. This result, in conjunction with our correlation analysis, reinforces the significance of charging infrastructure in driving electric vehicle sales.

In summary, while correlation analysis provides insights into direct relationships, PCA serves as a valuable auxiliary tool, substantiating the dominance of charging station quantity in influencing the development of new energy electric vehicles in China.

Table 2: Principal component contribution degree

Factors / Province	Charger	gasPrice
ShangHai	0.74302921	0.22918561
BeiJing	0.77210009	0.21238974
GuangDong	0.77015335	0.2120876
JiangSu	0.79441364	0.18848364
ZheJiang	0.72889342	0.18947932

5 Best-Order ARIMA for Future Development of China EV

5.1 Data Description

Visualization of Original Data: We commence our analysis by presenting a graphical representation of the raw data, illustrating the observed trends in both the number of charging stations and electric vehicle sales. This visual aid provides a clear snapshot of the historical dynamics in the NEEV industry.

Original Data - Charging Station Quantity & EV Sales Observations:

A comprehensive depiction of the raw data is presented, highlighting the fluctuations and patterns in charging station quantities and EV sales over the past decade.

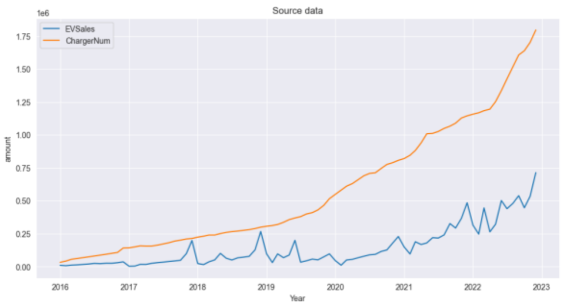


Figure 4: Charging Station Quantity & EV Sales

Smoothing Techniques - Original Data vs. Smoothed Mean:

To enhance clarity, we apply smoothing techniques to the original data, presenting a comparative analysis between the raw observations and the smoothed mean. This facilitates a nuanced understanding of the underlying trends.

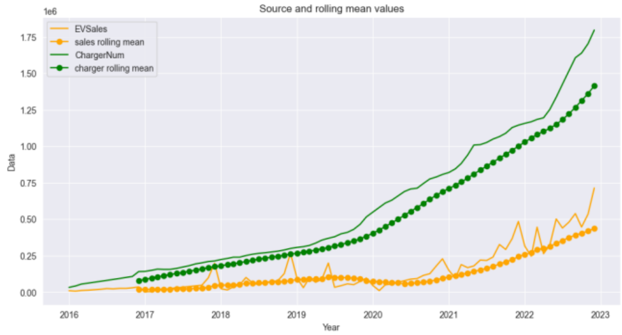


Figure 5: Original Data vs. Smoothed Mean

Building upon the foundation laid by the original data analysis, we employ the Auto ARIMA model to predict future trends in both EV sales and charging station quantities. The Auto ARIMA model, an advanced time series analysis tool, is employed for its capability to automatically determine optimal parameters, enhancing the accuracy of our predictions.

5.2 The Establishment of Model 2

5.2.1 Preliminary Analysis and Auto ARIMA

Initially, we conducted a comprehensive exploration utilizing the auto ARIMA model. In our preliminary analysis, we utilized auto ARIMA to automate the selection of ARIMA model parameters. The algorithm explores various combinations of orders and selects the model with the best fit based on criteria such as AIC. The resulting diagnostic charts highlight the model's performance and any deviations from statistical assumptions. Auto ARIMA, or Automatic AutoRegressive Integrated Moving Average, is an algorithm that automatically selects the optimal parameters for an ARIMA model. ARIMA itself is a time series forecasting model that comprises AutoRegressive (AR) and Moving Average (MA) components. The model aims to capture temporal patterns and dependencies in sequential data. This involved generating result plots and diagnostic charts. Remarkably, the confidence interval's lower limit for charging stations in the auto ARIMA forecast was observed to be inconsistent with the upper limit of sales. Additionally, conspicuous residuals in the auto ARIMA model indicated inadequacies in its predictive performance.

5.2.2 Grid Search for Optimal Parameters

Grid Search is a systematic approach to hyperparameter tuning, involving an exhaustive search over a predefined grid of parameter values. In the context of our study, grid search was employed to identify the optimal orders (p, d, q) for the ARIMA model. Here, 'p' denotes the autoregressive order, 'd' signifies the differencing order, and 'q' represents the moving average order. By evaluating a range of parameter combinations, we sought to pinpoint the configuration that maximized the model's accuracy.

Given the limitations of auto ARIMA, we employed a grid search algorithm to identify the optimal model orders. This exhaustive search led us to the most suitable parameters, addressing the shortcomings observed in the auto ARIMA results. The grid search methodology allowed for a more nuanced and accurate configuration of the ARIMA model.

5.2.3 ARIMA Model with Optimal Parameters

The ARIMA model itself is a powerful tool for time series forecasting, characterized by three components: AutoRegressive (AR), Integrated (I), and Moving Average (MA). The 'AR' component captures the relationship between an observation and its past values, 'I' represents the differencing required to make the series stationary, and 'MA' accounts for the dependency between an observation and a residual error from a moving average model.

Having identified the optimal parameters through grid search, we constructed an ARIMA model tailored to the nuances of the data. This model was employed to generate the final predictive results, which, in contrast to the initial auto ARIMA outcomes, exhibited a higher degree of precision and realism. By mitigating the issues observed earlier, the ARIMA model with optimal parameters provided a more reliable basis for forecasting.

5.2.4 Observe Results and Diagnostic Charts:

Subsequently, the ARIMA model with optimal parameters yielded a set of results, including predictive charts and diagnostic plots. These results surpassed the deficiencies identified in the auto ARIMA model, with a more accurate depiction of the charging station trends in relation to sales. The diagnostic charts provided further validation of the model's reliability, demonstrating improved residual behavior and adherence to statistical assumptions.

5.3 The Solution of Model 2

5.3.1 Comparison of Auto ARIMA and Best-Order ARIMA Results:

The first step involves contrasting the two sets of results. Examining the auto ARIMA results, the green curve represents the forecast for electric vehicle sales, revealing significant fluctuations that closely resemble real-world dynamics. Conversely, the best-order ARIMA model yields a smoother exponential curve for battery sales. Notably, the shaded region in the auto ARIMA charging station forecast, depicted in red, exhibits a substantial confidence interval, indicating a lower fitting accuracy for this model. Intriguingly, the lower limit of this interval coincides noticeably with the electric vehicle sales data after 2018, a logical inconsistency. The best-order ARIMA, however, effectively avoids this issue, suggesting superior fitting to the data.

Figure 6: Auto arima and Best-order arima forecast results chart

Table 3: EVSales forecast and ChargerNum forecast

Date	EVSales_forecast	ChargerNum_forecast
2023/12/31	678362.6077	2658439.379
2024/12/31	821679.3945	3420365.444

2025/12/31	961397.6483	4186911.184
2026/12/31	1101189.101	4952121.18
2027/12/31	1240981.672	5717717.396
2028/12/31	1380774.245	6483201.939
2029/12/31	1520566.819	7248718.772
2030/12/31	1660359.392	8014226.269
2031/12/31	1800151.965	8779736.465
2032/12/31	1939944.538	9545245.88

Furthermore, the four indicators – blue and orange lines representing actual battery sales and charging station quantities, respectively – demonstrate a seamless connection between actual and predicted data from 2016 to 2023. This coherence is corroborated by data from the referenced white paper, further affirming the accuracy of our predictions. [4]

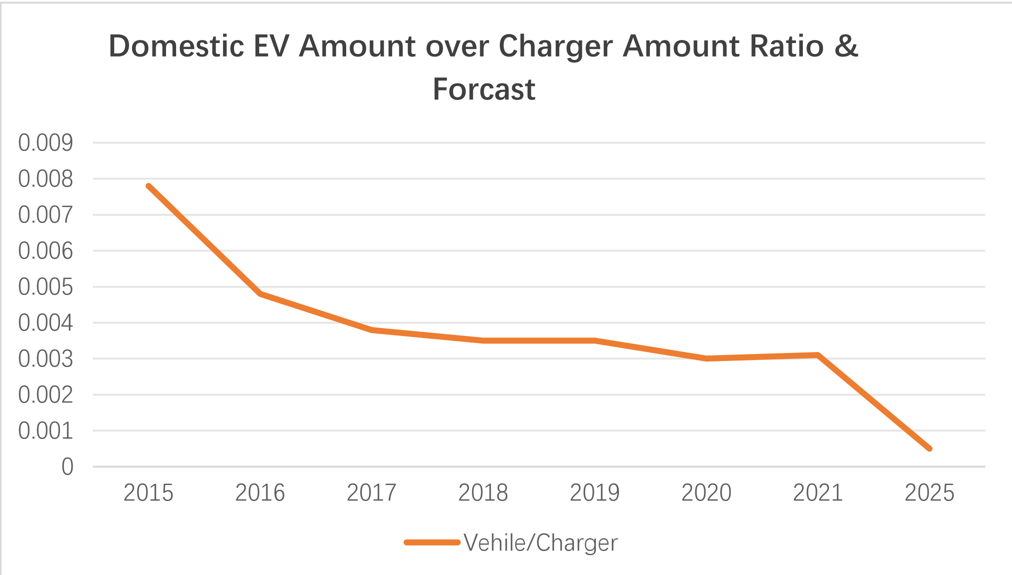


Figure 7: Trend of pile ratio of new energy vehicles[5]

5.3.2 Comparison of Diagnostic Information for Charging Station Predictions (Auto vs. Best-Order ARIMA):

Moving on to the diagnostic information for charging station predictions, four subplots are presented. The first, a standardized residual plot, indicates that both auto and best-order ARIMA residuals oscillate between -3 and 3, suggesting significant and frequent deviations from the fitted model. This observation necessitates the collection of more authentic and reliable data to refine the model and enhance predictive capabilities.

The second subplot displays frequency density histograms for both models. Despite differences in the forecasting methods, the data distribution closely resembles a normal distribution. This normality suggests the appropriateness of using ARIMA for fitting and predicting the data.

The final two subplots, normal Q-Q plots and autocorrelation plots, serve the same purpose – assessing the suitability of the data for ARIMA modeling. Both auto and best-order

ARIMA exhibit well-fitted curves in normal Q-Q plots, indicating a strong fit for the entire input range. The autocorrelation plots also reinforce the effectiveness of both models.

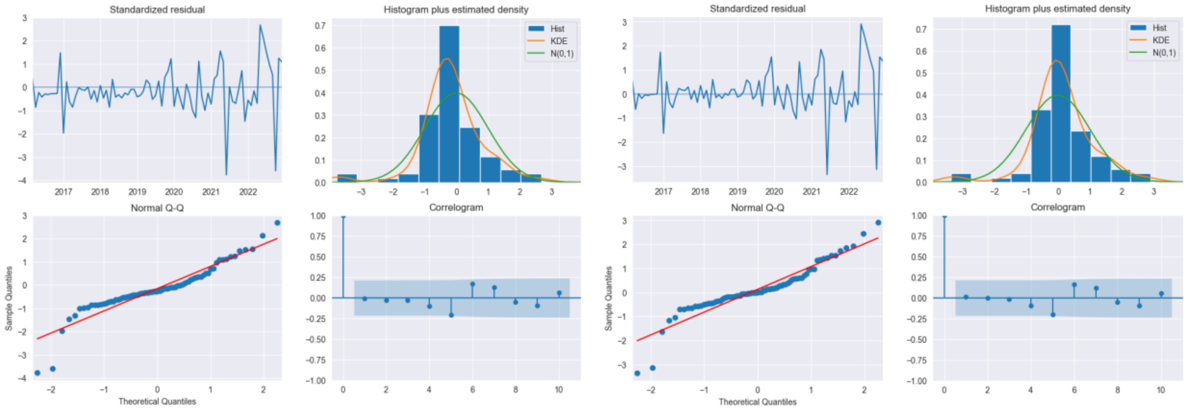


Figure 8: Charging pile predictive model diagnostic infographic

5.3.3 Comparison of Diagnostic Information for Sales Volume Predictions:

The last set of graphs pertains to diagnostic information for sales volume predictions. A notable refinement is observed in the residual range, reducing from -3 to 3 in the charging station model to -2 to 2 in the sales volume model. This suggests that the sales data might be more congruent with the model, emphasizing the importance of collecting accurate and suitable data for effective modeling

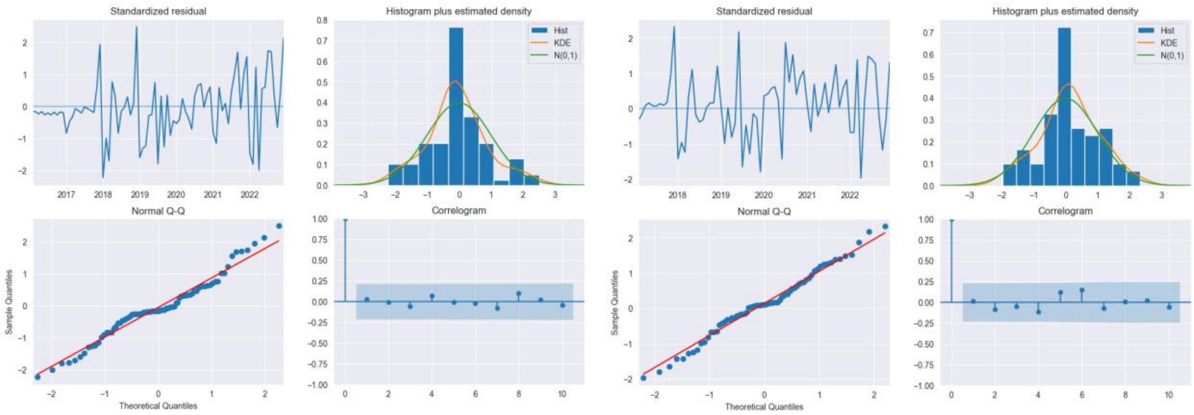


Figure 9: Sales predictive model diagnostic infographic

6 Spearman Coefficient Analysis for New-old Industry

6.1 Data Description

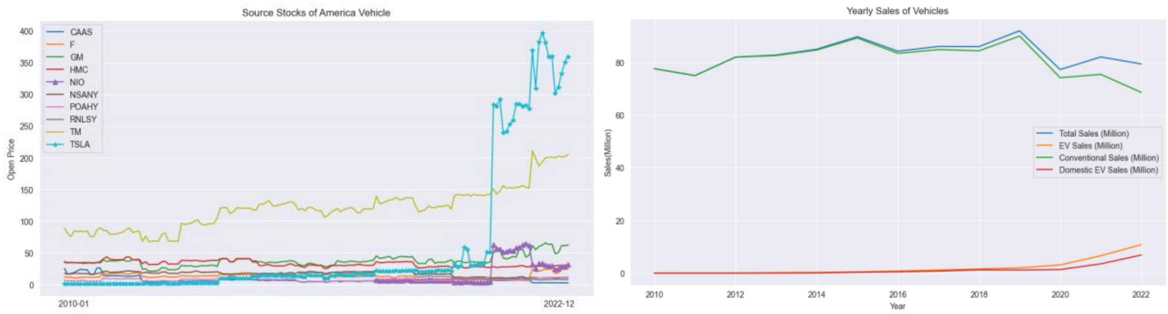


Figure 10: Stock market raw data 2010-2022 and Global annual car sales curve

6.2 The Establishment of Model 3

6.2.1 Spearman Coefficient Analysis and Its Relevance

The utilization of the Spearman coefficient in our analysis serves to assess the strength and direction of monotonic relationships between variables. This statistical tool is particularly valuable when dealing with non-linear associations, allowing us to gauge the degree of correlation between different sets of data. The primary utility of correlation analysis lies in unveiling patterns and dependencies, facilitating a deeper comprehension of interdependencies that may exist within the datasets under consideration.

6.2.2 Analysis of the Impact of New Energy Electric Vehicles on the Global Traditional Energy Vehicle Industry

In response to the specified task of collecting and analyzing data concerning the influence of new energy electric vehicles on the global traditional energy vehicle industry, our dataset spans the years 2010 to 2022. It encompasses the annual sales figures for domestic electric vehicles in China, international electric vehicle sales, international traditional energy vehicle sales, and stock market data for ten typical automotive companies in the United States from 2010 to 2022. This extensive dataset enables us to conduct various correlation analyses, including:

1. The correlation between domestic new energy vehicle sales and international new energy vehicle sales, as well as their respective stock prices.

2. The correlation between domestic new energy vehicle sales and international traditional vehicle stock prices and sales.

3. The correlation between international new energy vehicle sales and international traditional vehicle stock prices and sales.

4. The correlation between international new energy vehicle sales and international new energy vehicle stock prices.

The outcomes of these correlation analyses provide insights into the relationships stipulated by the research question.

6.3 The Solution of Model 3

Table 4: Correlation analysis table

Category	Correlation	P-Value
Domestic EV to TSLA Stock	0.983516484	1.61E-09
Domestic EV to Global EV	0.983516484	1.61E-09
Global EV to TSLA Stock	0.972390531	2.69E-08
Global EV to NIO Stock	0.6	0.28475698
Domestic EV to NIO Stock	0.5	0.391002219
Global EV to Glabl Conv	-0.11602387	0.705833502
Domestic EV to Global Conv	-0.131868132	0.667607576

Toyota (TM): Toyota's unwavering commitment to innovation in new energy cars has solidified its position as a leader in sustainable and eco-friendly transportation. The robust positive correlation between Toyota's stock and electric vehicle sales underscores market recognition of the brand's leadership in green technology.

Ford (F): Ford's strategic focus on mid-to-high-performance cars sets it apart, resulting in a relatively small negative correlation between Ford's stock and EV sales. This trend may be attributed to perceptions that electric cars, while environmentally friendly, may not align with Ford's traditionally high-performance vehicle characteristics.

Hyundai (HMC): Hyundai's recent entry into the new energy vehicle segment marks a significant strategic shift[6]. The strong negative correlation between Hyundai's stock and EV sales reflects the market's response to this new venture, possibly influenced by uncertainties surrounding Hyundai's success in the electric vehicle market.

China Automotive Systems (CAAS): With a primary focus on traditional automotive assembly, China Automotive Systems exhibits a negative correlation with EV sales. The company's alignment with conventional automotive practices may pose challenges in adapting to the evolving landscape of electric vehicles, influencing investor caution.

Porsche (POAHY): Porsche's concentration on the high-end market distinguishes its market positioning. The relatively small correlation between Porsche's stock and EV sales is attributed to electric cars primarily targeting the mid-to-low-end market. This divergence underscores Porsche's unique positioning in the automotive industry, catering to a niche market with a focus on high-performance luxury vehicles.

General Motors (GM): Proactive in advancing new energy vehicles, General Motors exhibits a positive correlation coefficient of 0.514, highlighting the favorable market response to its commitment to developing a diverse range of electric vehicles. The company's forward-thinking approach and strategic investments in the EV market position it well to capitalize on the growing demand for sustainable transportation options.

Nissan (NSANY): Nissan's notable negative correlation trend with EV sales over the past decade may be attributed to comparatively lower investment in the field of new energy vehicles. Balancing declining traditional fuel-powered car sales with insufficient momentum in the electric vehicle market presents a challenge for Nissan. Market demand and brand positioning likely contribute to the observed negative correlation between overall Nissan sales and new energy vehicle sales.

US Typical Vehicle Industry Stock	Domestic EV to US Typical Vehicle Stock	
	Correlation	P-Value
CAAS	-0.604395604	0.028672706
F	-0.164835165	0.590478989
GM	0.527472527	0.063954891
HMC	-0.703296703	0.00731857
NSANY	-0.461538462	0.112375971
POAHY	0.06043956	0.844502371
TM	0.873626374	9.52890E-05

US Typical Vehicle Industry Stock	Global EV to US Typical Vehicle Stock	
	Correlation	P-Value
CAAS	-0.585644297	0.035464137
F	-0.204423009	0.50290771
GM	0.513819997	0.072468994
HMC	-0.751392683	0.003064467
NSANY	-0.508295051	0.076129507
POAHY	0.049724516	0.871843514
TM	0.861891607	1.51677E-04

Table 5: Correlation analysis table

7 Anti-China-EV Policy Influence Analysis

7.1 Data Description

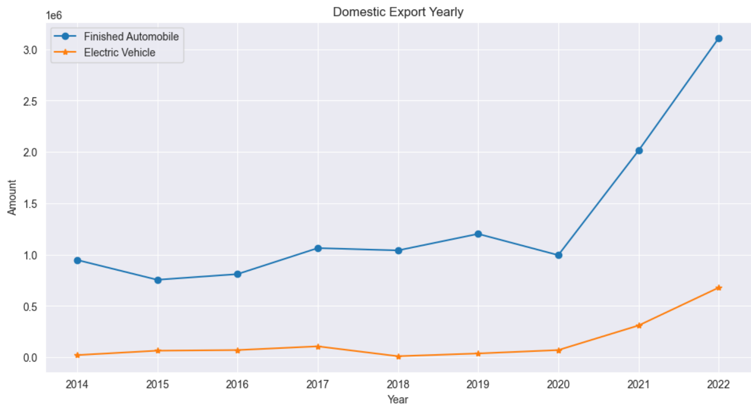


Figure 11: China's annual automobile export curve from 2014 to 2022

7.2 The Establishment of Model 4

7.2.1 Attempt with ARIMA Model

Firstly, utilize export data for prediction, aiming to investigate the impact of foreign policies on exports. Employ the ARIMA model for prediction, but diagnostic plots reveal unsatisfactory fitting. Then, analyze diagnostic plots of the ARIMA model, including the scatter plot of actual versus predicted values, the fitted tail plot, and the residual distribution chi-square plot. Finally, due to the short data sequence, discern that the ARIMA model is inadequate for forecasting.

7.2.2 Exponential Regression Model

Firstly, shift to an exponential regression model given the suboptimal performance of the ARIMA model. Then, generate a Q4 trend prediction graph, featuring three curves: the actual curve (blue), the fitted curve from 2014 to 2017 (orange), and the fitted curve from 2018 to 2022 (green). At last, analyze each curve, focusing on the fitting degree and sensitivity to foreign policies.

$$y = a * e^{bx} + c \tag{2}$$

where a, b, c is 3 principle coefficient of exponential fit function.

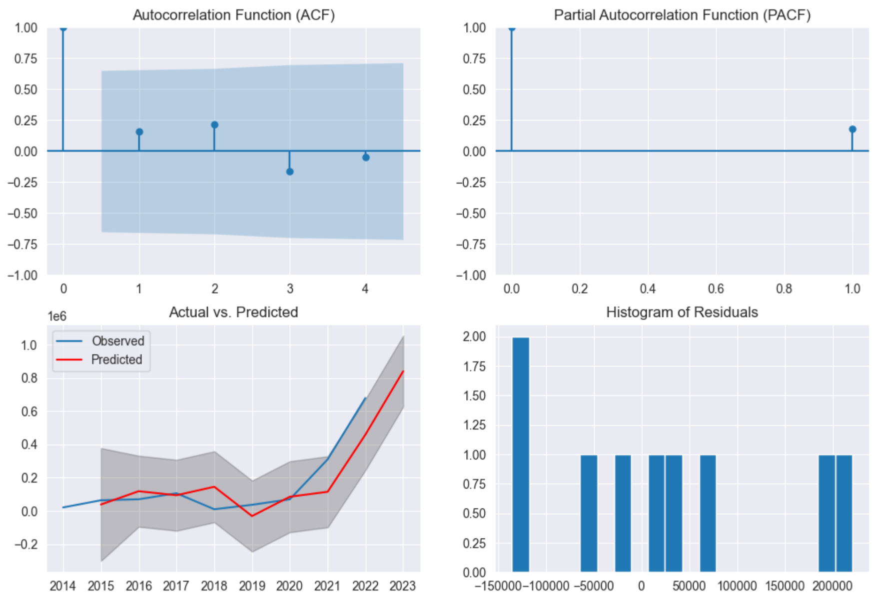


Figure 12: ARIMA model diagnosis

7.3 The Solution of Model 4

7.3.1 Green Curve (Fitted from 2018 to 2022)

High fitting degree, aligning closely with the actual curve, indicating a relatively accurate prediction of the impact of foreign policies during this period. Suggest further analysis of specific policies and measures during this timeframe for a deeper understanding of influencing factors.

7.3.2 Orange Curve (Fitted from 2014 to 2017)

Due to the inclusion of pre-adverse-policy data, the fitted curve surpasses the actual values, highlighting a substantial negative impact of policies on exports. Emphasize the adverse effects of foreign policies, quantifying the magnitude of the impact through a comparison of trends from 2018 to 2022.

7.3.3 Comparison between Blue and Green Curves

In 2022, the green curve closely approximates the orange curve, possibly attributed to advancements in our country's technological prowess.Infer that the development speed in China from 2018 to 2022 was rapid, with an improvement in technological capabilities, suggesting that actual sales may surpass the orange curve.

7.3.4 Inference on Technological Progress

a. Notably, we observe that the blue and green curves approach the orange curve in 2022 and exhibit a trend of surpassing it in the future. This trend leads us to reasonably conclude that China is gradually overcoming the adverse effects of foreign policies.

b. Examining the trends, the derivatives of the blue and green curves have already surpassed those of the orange curve in 2022. This signifies a growth rate exceeding that of the orange curve, indicating that China has made significant strides in the field of technology, with a further acceleration in the pace of development.

In essence, the convergence of the blue and green curves towards the orange curve in 2022, coupled with the anticipation of future surpassing, suggests a positive trajectory, affirming China's resilience in mitigating the negative impacts of foreign policies. Furthermore, the discernible acceleration in technological advancement, as evidenced by the superior growth rates of the blue and green curves, underscores China's remarkable progress and rapid strides in the realm of technology.

In conclusion, these analyses provide insights into the impact of foreign policies on Chinese exports and the remarkable technological progress in our country, offering readers a profound understanding of future development trends.

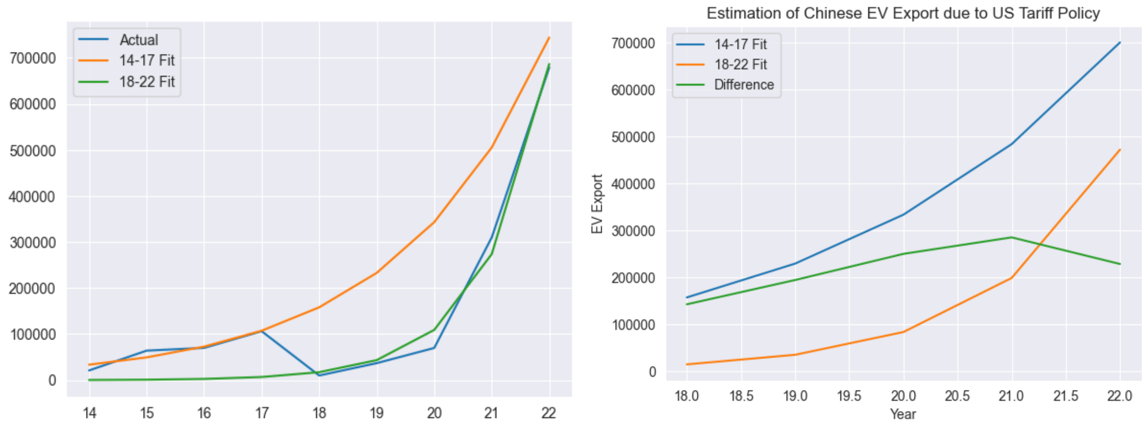


Figure 13: Forecast trend chart and Projected reduction of China's tram exports due to US tariff policy in 2018-2022

8 MLP for Electrification Impact on Ecologic System

8.1 Data Description

8.1.1 Data Overview

Commencing with an exploration of diverse datasets pertaining to both conventional and new energy vehicles, this study derives average carbon emissions and CO2 output to initiate a preliminary quantitative analysis of the ecological impact of vehicular electrification. Notably, the consumption of 1 kWh of electricity yields 0.272 kg of carbon emissions and 0.997 kg of CO2 emissions, while the combustion of 1 liter of gasoline and diesel results in 0.627 kg and 0.717 kg of carbon emissions, along with 2.30 kg and 2.63 kg of CO2 emissions, respectively. For purely electric vehicles consuming 15 kWh per 100 km, the calculated carbon and CO2 emissions are 4.08 kg and 14.955 kg, whereas traditional fuel vehicles consuming 8 liters per

100 km exhibit emissions of 5.016 kg and 18.4 kg, respectively. Consequently, transitioning from traditional fuel vehicles to purely electric vehicles yields an approximate reduction of 1 kg in carbon emissions and 3.4 kg in CO2 emissions per 100 km.

Citing data from the China Association of Automobile Manufacturers, taking Shenzhen as an example, with a total vehicle ownership of 3.95 million, including 766,000 electric vehicles, and considering an annual average mileage of 1.6 × 10^4 km in Guangdong Province, electric vehicles accumulate an annual average mileage of 1.2256 × 10^8 hundred kilometers, resulting in a reduction of carbon emissions/CO2 emissions by approximately 1.2256 × 10^8 kg/4.167 × 10^8 kg per 100 km. In essence, despite the relatively low market share of electric vehicles at present, their impact on reducing carbon emissions is substantial.

8.1.2 Multivariate Impact on Air Quality Index (AQI) based on Varied City Data

Horizontal examination of population, Air Quality Index (AQI), Gross Domestic Product (GDP), green coverage, and electric vehicle sales across different cities facilitates an exploration of the influence of electric vehicle ownership on local AQI.

Table 6: China 2022 ecological data description table

AQI	**Precipitation**	**GDP**	**Temperature**	**Altitude**	**PopulationDensity**	**Coastal**	**GreenCoverageRate**	**Population(100000)**	**EV_Sales_2022**	**Incineration(10,000ton)**
23	23	23	23	23	23	23	23	23	23	23
73.73913	1355.826	8364.156	18.74223	28.70435	4000.348	1	38.27696	119.5387	61339.04	191.7465
50.38327	686.6033	6102.086	4.028965	67.32421	5446.016	0	7.143157	110.9396	79608.14	200.5832
16	547	3450	12.40274	1	496	1	18.78	22.09	4308	22.32
37	678	4153.89	15.04658	3.15	1269	1	34.585	68.745	16170	42.395
55	1651.5	6137.74	18.16301	9	2322	1	41	84.48	31357	111
93	1775.1	8655.785	22.39658	17	3943	1	43.485	123.115	68641.5	268.45
199	2478.1	24964.99	24.41096	321	25900	1	44.93	530.8	335058	613.85

8.2 The Establishment of Model 5

Guided by 2022 data from 60 Chinese cities, an eight-dimensional dataset encompassing local population, electric vehicle ownership, GDP, green coverage, and other variables is employed to construct a multilayer perceptron. This neural network, with AQI as the output, maintains constant variables while investigating the impact of varying electric vehicle ownership on AQI in cities with a population of one million. The model architecture includes a two-layer fully connected structure with ReLU activation, 126 neurons in the hidden layer, and employs mean squared error (MSE) as the loss function, Adam optimizer, and L2 regularization with a weight decay of 0.4.

In essence, the model is a robust regression architecture tailored to predict AQI. Incorporating normalization, L2 regularization, and dynamic learning rate adjustment enhances stability and generalizability. Visualization of training and test loss curves, alongside AQI predictions, facilitates a comprehensive evaluation of the model's effectiveness and provides actionable insights into the relationship between electric vehicle sales and air quality across diverse cities.

Model Architecture:

Multilayer Perceptron (MLP) Model (`MLPModel`):

- Architecture: Two fully connected layers with ReLU activation in between.
- Input Size: Determined by the number of features in the dataset.
- Hidden Size: 126 neurons.
- Output Size: 1 (regression task).
- Activation Function: ReLU for hidden layers.
- Loss Function: Mean Squared Error (MSE).
- Optimizer: Adam optimizer.
- L2 Regularization: Applied with a weight decay of 0.4.

8.3 The Solution of Model 5

Observing the model's convergence and effective fitting, three cities—Shanghai, Tangshan, and Nanning—are selected for detailed analysis due to their varying levels of development and geographical diversity. Notably, in heavy industrial cities like Tangshan, the promotion of new energy vehicles proves more effective in energy conservation and emissions reduction. In developed cities like Shanghai, existing environmental policies contribute to ecological sustainability, and the adoption of new energy vehicles further aids in carbon emission reduction. Lastly, for moderately developed cities such as Nanning, the promotion of new energy vehicles is instrumental in alleviating escalating pollution concerns in urban areas.

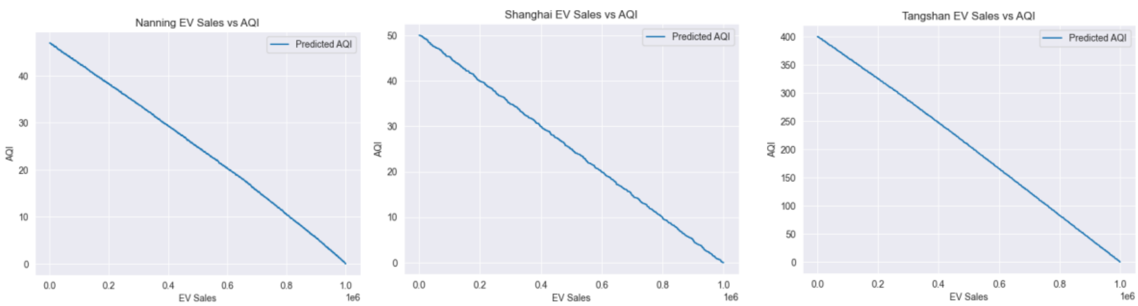


Figure 14: Three cities EVSales vs AQI

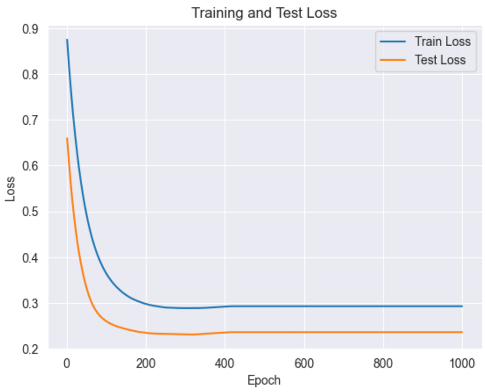


Figure 15: training&test_loss

9 Publicity Letter: Contribution of EV to Society

Dear Citizens,

I hope this letter finds you in good health and high spirits. Today, I am writing to you with an urgent message about a crucial matter that affects each and every one of us – the future of our cities and the air we breathe.

Recent research has conclusively shown that the increased adoption of new energy vehicles, particularly electric cars, significantly contributes to a remarkable improvement in the Air Quality Index (AQI) of cities worldwide. This positive impact is observed across cities with varying levels of GDP – whether high, medium, or low. The evidence is clear: the more electric vehicles on our roads, the healthier our air becomes.

As we navigate the challenges of urbanization and the environmental concerns that accompany it, embracing electric vehicles emerges as a beacon of hope. The study highlights that regardless of the economic status of a city, the surge in sales of new energy vehicles directly correlates with a substantial decrease in AQI. This is a testament to the transformative power of electric cars in mitigating air pollution and enhancing the overall quality of our environment.

Notably, the positive effects of embracing electric vehicles are particularly pronounced in coastal cities. The correlation between increased new energy vehicle sales and reduced AQI is striking, underlining the pivotal role electric cars can play in coastal regions where air quality challenges are often exacerbated.

Picture a city where the skyline is clear, where the air is crisp and invigorating, and where the well-being of its citizens takes precedence over environmental hazards. This vision is not a distant dream; it is an achievable reality through the widespread adoption of electric vehicles.

By choosing electric cars, we not only contribute to the global fight against climate change but also actively participate in the creation of livable and sustainable urban spaces. Our decision to embrace this green revolution has far-reaching consequences, shaping a future where our children can breathe cleaner air and enjoy a healthier, more vibrant life.

I urge each one of you to consider the impact of your transportation choices on our cities and the planet. Let us unite in making a conscious effort to transition to electric vehicles and be part of a global movement towards a cleaner, greener future.

Together, we can drive positive change for our cities, our environment, and the generations to come!

10 Sensitivity Analysis

10.1 PAC

In Model 1, given the strong linear commonality among the selected variables (indicated by correlation coefficients concentrated between 0.5 and 1), traditional PCA might not be the most suitable method. However, recognizing the need for a supplementary validation tool, we proceeded with PCA.

The contribution table generated from PCA revealed that the contribution of charging stations far surpassed that of gasoline prices. Specifically, charging stations exhibited contributions exceeding 70%, validating the importance of charging infrastructure in influencing the electric vehicle market. This result, in conjunction with our correlation analysis, reinforces the significance of charging infrastructure in driving electric vehicle sales.

In summary, while correlation analysis provides insights into direct relationships, PCA serves as a valuable auxiliary tool, substantiating the dominance of charging station quantity in influencing the development of new energy electric vehicles in China.

10.2 Residuals

In ARIMA horizontal forecasting, "residuals" refer to the error in fitting the observed values to the model. Specifically, for each observation, the model generates a predicted value and then computes the residual by subtracting the observed value from the corresponding predicted value.

In the context of ARIMA models, residuals represent the part that the model fails to explain, that is, the variation in the time series that the model fails to capture. By analyzing the residuals, you can check that the model has captured all the structure and patterns in the time series.

In the "summary" of an ARIMA model, it is common to provide statistics about the residuals, including mean, standard deviation, min Max, and so on. At the same time, some statistical tests, such as the Ljung-Box test, may also be included to check whether the residuals exhibit significant autocorrelation. In the case of a good fit, the residual should be a sequence that approximates white noise (no autocorrelation).

In Model 1, checking the residuals of a model is an important step in time series analysis, as it can help confirm whether the model is appropriate and captures all the important features in the data, and if the residuals show some structure or pattern, it may be necessary to further adjust the model or consider other modeling approaches.

The diagnostic information in the image only indicates that the model residuals were not significantly optimized by manually finding the optimal order. Therefore, it is inferred that the reason is the limitation of the data itself. The prediction results of best order arima and auto arima are close, and the residual error is reduced but still large, so it is considered as the problem of the data set itself.

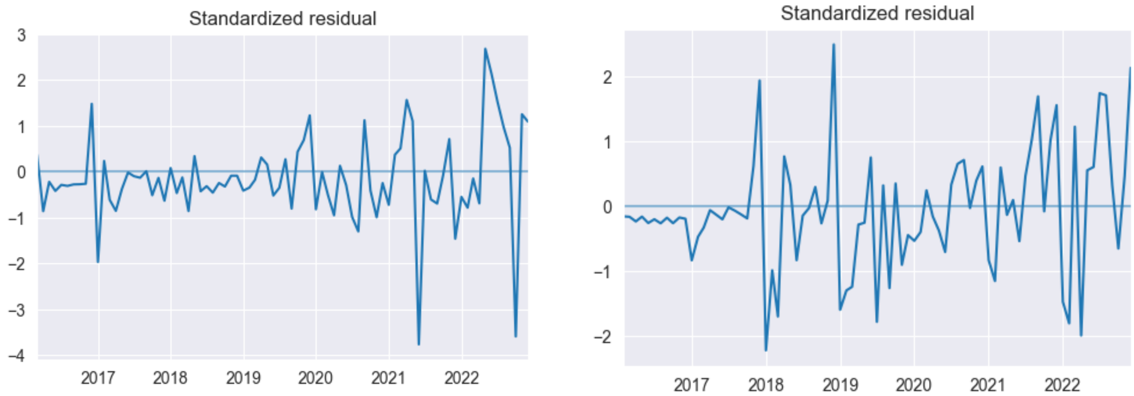


Figure 16: Diagnostic information of charging pile and sales forecasting model (best - order arima)

It can be seen from the figure that the residual fluctuation range of the prediction results of best order arima is similar to that of auto arima. A good residual graph should have a small fluctuation between -1 and 1. It's probably the 2018-2019 segment in this picture. We choose one with a large residual as the answer, not because of lack of optimization but because of the problem of the data.

10.3 AIC

In Model 2, the Akaike Information Criterion (AIC) serves as a quantitative measure for model selection, balancing goodness of fit with model complexity. AIC is calculated based on the likelihood function and penalizes models with excessive parameters. The goal is to identify a model that not only fits the data well but avoids overfitting. By maximizing the AIC score, we ensured our ARIMA model achieved an optimal balance, providing a reliable basis for forecasting.

The efficacy of our model was quantitatively assessed through the maximization of the Akaike Information Criterion (AIC) for the ARIMA model. AIC, as a pivotal evaluation metric, ensured that our model not only fit the data well but also avoided overfitting, striking a balance between complexity and explanatory power.

In summary, our methodology seamlessly integrated these fundamental concepts – from the automation of parameter selection using auto ARIMA to the systematic parameter tuning through grid search and the stringent model evaluation using AIC. This holistic approach culminated in a robust ARIMA model tailored to forecast the trajectory of China's new energy electric vehicle industry.

11 Model Evaluation and Further Discussion

11.1 Strengths

11.1.1 PCA:

Simplicity and Interpretability: PCA simplifies complex datasets, enhancing interpretability by focusing on the most influential components.

Multicollinearity Mitigation: PCA is effective in handling multicollinearity among variables, a common challenge in real-world datasets.

Pattern Recognition: It aids in recognizing underlying patterns and relationships within the data.

11.1.2 Grid Search Optimized ARIMA

Manual Parameter Tuning: Grid search allows for fine-tuning of ARIMA model parameters, potentially improving accuracy.

Comprehensive Model Evaluation: The use of AIC for model evaluation adds a quantitative aspect to the assessment.

11.1.3 Exponential Regression for Export Prediction

Adaptation to Non-ARIMA Suitable Data: Recognizing the limitations of ARIMA for short sequences, the model shifts to exponential regression.

Clear Visualization: The graphical representation effectively communicates the impact of foreign policies.

11.2 Weaknesses

Model 1: Due to the limited length of time from 2016 to 2022, this will affect the prediction effect. There are deviations and statistical problems between the data and the truth. The types of data are limited, and only the three kinds mentioned above can be used for data analysis. Social factors such as policies, subsidies, users' consumption tendency and media are difficult to quantify, resulting in incomplete factor analysis.

12 Conclusion

In summary, the presented models and the call to action article collectively underscore the transformative potential of electric vehicles (EVs) in addressing environmental challenges. The auto ARIMA and grid search optimized ARIMA models provide robust time series forecasting tools, with the latter demonstrating the importance of meticulous parameter tuning. The Spearman coefficient analysis explores correlations in the EV industry, while the exponential regression model responds adeptly to non-ARIMA suitable data, demonstrating flexibility in methodology.

The multilayer perceptron model showcases the power of neural networks in predicting Air Quality Index (AQI), offering valuable insights into the positive correlation between increased EV adoption and improved air quality. The call to action article, while emotionally resonant and motivational, could benefit from anchoring its narrative in specific study findings and further emphasizing policy advocacy.

Collectively, these contributions portray the multi-faceted impact of EVs, ranging from forecasting sales to analyzing correlations and predicting environmental outcomes. As we celebrate the successes of the electric vehicle industry, it is crucial to leverage these insights to drive policy changes, advocate for sustainable choices, and collectively shape a future defined by clean air, technological innovation, and environmental stewardship.

References

[1] Global Sales in Electric Vehicle[https://www.ev-volumes.com/]
[2] Annual Report on the Big Data of New Energy Vehicle in China (2022) By Zhenpo Wang
[3] China AQI and correlative factors dataset, [https://www.kaggle.com/datasets/maxwell-nee/china-aqi-test]
[4] 2022 White Book on the Development Trend of New Energy Vehicle By China Automobile Association
[5] Annual Report on the Big Data of New Energy Vehicle in China (2022) By Zhenpo Wang
[6] Global Sales in Electric Vehicle, [https://www.ev-volumes.com/]