Team Number :	apmcm2303831
Problem Chosen :	C

2023 APMCM summary sheet

Dynamic evaluation model of China's new energy electric vehicle

This research delves into the multifaceted analysis and solutions pertaining to the growth of new energy electric vehicles (NEEVs) in China. we employ a multivariable linear regression model for model one comprehensively analyze the primary factors influencing NEEV development. This analytical approach explores governmental policies, technological advancements, infrastructure development, and economic considerations. Utilizing the multivariable linear regression model and TOPSIS, we aim to quantify and elucidate the complex relationships among these factors.

For Problem Two, involving the forecast of NEEV development over the next decade, we focus on data collection related to NEEV sales. Employing the grey model (1,1), we establish mathematical models to describe and predict the current state and future trajectory of China's NEEV industry. Additionally, the efficacy of the grey model and ARIMA[i] for forecasting will be compared to discern which provides more accurate predictions.

Problem Three examines the transformative impact of new energy vehicles on the global traditional energy vehicle industry. ARIMA and curvilinear regression models are utilized to analyze sales data, assess trends, and predict future sales of traditional vehicles. The findings offer valuable insights for stakeholders and policymakers navigating the evolving automotive landscape.

Problem Four concentrates on evaluating the impact of policies resisting NEEV growth in China. Through curvilinear regression modeling, we analyze the effects of these policies on NEEV development. Results will contribute to a comprehensive understanding of the effectiveness of current policies in shaping the trajectory of NEEVs in China.

Problem Five assesses the impact of NEEV electrification on the ecological environment in urban areas. Calculations, assuming an urban population of 1 million, reveal the positive effects of the transition to electric vehicles on environmental sustainability. These results to play a pivotal role in fostering a more sustainable and environmentally friendly urban transport system.

Keywords: Multivariable Linear Regression, TOPSIS, Grey Model (1,1), ARIMA, Curvilinear Regression, NEEV, Ecological Environment, Sustainability.

Contents

1 Introduction

1.1 Problem Background

Hawaii wildfires, unrelenting rain, and extreme weather around the world serve as an urgent reminder that we still have a long way to go before curbing climate change. With great power comes great responsibility. As we witness the way China's GDP has boosted, we must also take responsibility for the environment. The introduction of a new energy vehicle has been a net positive for the city, reducing air pollution and relying less on fossil fuels. China, particularly since 2011, has played a pivotal role in propelling the development of NEEVs. The government has proactively implemented a series of preferential policies to stimulate growth within this sector. New energy electric vehicles have garnered significant attention and experienced rapid growth, attributable to their inherent qualities of low pollution, reduced energy consumption, and the capacity to regulate peak electricity demand. This category encompasses electric buses and family electric cars with fewer than seven seats, gaining widespread popularity among both consumers and governments globally. This result is remarkable, positioning the new energy electric vehicle industry as a symbol of Chinese innovation and progress, akin to the renowned 'China High-speed railway ‘.

1.2 Restatement of the Problem

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

- Problem 1 What are the pivotal factors shaping the development of new energy electric vehicles (NEEVs) in China?
- Problem 2 How do NEEVs impact the market, and what predictive model can depict their evolution over the next decade?
- Problem 3 To what extent do NEEVs pose a challenge to the global traditional energy vehicle industry?
- Problem 4 In the face of resistance from tech industry isolation, how does it impede China's long-term NEEV development?
- Problem 5 What environmental ramifications arise from the emergence of NEEVs, particularly in urban areas with a population of 1 million?
- Problem 6 Compose an open letter leveraging insights from Problem 5 to articulate the benefits of NEEVs and underscore their global contributions.

2 Problem analysis

2.1 Analysis of Problem one

In order to analyze the primary factor that affect influencing the growth of new energy electric vehicles in China, we initiate a comprehensive analysis. we employed a multivariable linear regression model. This analytical approach involves a detailed exploration of various aspects, encompassing governmental policies, technological advancements, infrastructure development, and economic considerations. Utilizing this **multivariable linear regression model,** and **TOPSIS** we aim to quantify and elucidate the complex relationships among these factors.

2.2 Analysis of Problem two

To address question 2 that requires us to forecast the new energy electric vehicle development over the next decade. We initiate the data collection process focus on the sales of new energy electric vehicle as indicator. With the collected data we apply the **grey model (1,1)** to established mathematical models. These models will serve to describe the current state of China's NEEV trending. The grey model and Arima both renowned for their forecasting capabilities, will be employed to compare their efficacy in predicting the future trajectory of China's NEEV industry. This comparative analysis aims to discern which model provides more accurate and reliable predictions.

2.3 Analysis of Problem Three

This analysis investigates the transformative impact of new energy vehicles on the global traditional energy vehicle industry. Utilizing **ARIMA and curvilinear regression models**, we gathered and analyzed sales data, assessed trends, and predicted future sales of traditional vehicles. The ARIMA model helped identify temporal patterns, while the regression model examined the nonlinear relationship between various factors and traditional vehicle sales, particularly the influence of NEEVs. The findings offer valuable insights for stakeholders and policymakers to navigate the evolving automotive landscape and make informed decisions in response to the rise of electric vehicles.

2.4 Analysis of Problem four

This study focuses on evaluating the impact of policies aimed at impeding the growth of new energy electric vehicles (NEEVs) in China. Through the application of **curvilinear regression modeling,** we aim to analyze the effects of these policies on the development of NEEVs in the Chinese market. By collecting relevant data on policy implementations and NEEV sales, the regression model will help uncover nonlinear relationships between policy variables and the adoption of electric vehicles. The results of this analysis will contribute to a comprehensive understanding of the effectiveness of current policies in shaping the trajectory of NEEVs in China, offering insights for policymakers and industry stakeholders.

2.5 Analysis of Problem Five

The analysis of the impact of new energy electric vehicles' electrification, including electric buses, on the ecological environment in urban areas involves assessing the comprehensive energy consumption. Assuming an urban population of 1 million, our calculations reveal that the transition to electric vehicles significantly contributes to environmental sustainability. By utilizing general rules for the calculation of comprehensive energy consumption, we compare and demonstrate the positive effects of this transition on reducing ecological harm. The results highlight the potential of electrification to play a pivotal role in fostering a more sustainable and environmentally friendly urban transport system.

3 Assumptions and Justification

To simplify the problem, we make the following basic assumptions, each of which is properly justified.
Assumption 1: Assuming that the data source is reliable and accurate
Assumption 2: Assume no influenced by policies of resistance.
Assumption 3: Assume influenced and profoundly affected by favorable government policies.
Assumption 4: Assume there are no significant events affecting the development of electric vehicles in the next ten years.
Assumption 5: Assume that, apart from energy consumption, only basic parameters are considered, and more complex calculation methods are not taken into account.

4 Model Preparation

4.1 Notations

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

Table 1: Notations used in this paper

Symbol	Description	Unit
E	The total electric energy of vehicle Category	kwh/km
V	The total volume fuel of vehicle Category	L/km
SCEelectric	Conversion coefficients for electricity to standard coal	kgce/kwh

SCEfuel	Conversion coefficients for fuel to standard coal	kgce/km
E	The total electric energy of vehicle Category	kwh/km

4.2 The Data

4.2.1 Data Collection

We conduct extensive quantitative data collection to train our model and the following are some quantitative data variables we employed: Charging station, number of patents in clean energy, GDP per capita etc.

We obtain the data from multiple reputable sources to ensure accuracy and reliability. The following platforms were utilized: National Bureau of Statistics, IEA, Wind, Patsnap, State Grid Corporation of China, General Administration of Customs of the People's Republic of China and China Statistical Yearbook. We adhere to strict standard of data integrity and transparency, you will find the detailed references to specific sources for each dataset, allowing for traceability and verification.

5 Model for Question I : TOPSIS and Multilinear Regression

5.1 Data Description

For this question, we have collected three key indicators spanning the years 2011 to 2022, which are suitable as dependent variables. These variables include the sales volume of new energy vehicles, the ownership of new energy vehicles, and the temporal coverage rate of new energy vehicles. In assessing the main factors influencing the development of new energy vehicles in China, considering that development encompasses various aspects, we intend to evaluate the data based on the years using the Topsis model to quantify a new indicator called 'Development Score'.

Year	China's NEV Sales (10,000 units)	China's NEV Ownership (10,000 units)	China's NEV Market Share (%)
2011	0.82	1850	0.044
2012	1.28	1931	0.066
2013	1.7642	2199	0.080
2014	7.48	2350	0.318
2015	33.11	2460	1.346
2016	50.7	2803	1.809
2017	77.7	2888	2.690
2018	125.6	2808	4.473
2019	120.6	2576	4.682
2020	136.7	2531	5.401
2021	352.1	2627	13.403
2022	606.7	2686	22.587

5.2 The Establishment of TOPSIS

TOPSIS [1][ii]is a highly effective method in multi-objective decision analysis. It first normalizes the data matrix through forward scaling and normalization (dimensionless transformation). Subsequently, it identifies the optimal and worst-case objectives among multiple targets. Distances between each evaluation objective and the optimal or worst-case objectives are then calculated. The closeness of each objective to the ideal solution is obtained, and the objectives are ranked based on the magnitude of their closeness. This serves as the basis for evaluating the superiority or inferiority of the objectives. Closeness values range from 0 to 1, with a value closer to 1 indicating that the evaluation objective is closer to the optimal level, while a value closer to 0 indicates proximity to the worst level.

5.2.1 Data Normalization

As the data here are already positive indicators, there is no need for further normalization.

5.2.2 Data standardization

The purpose of standardization is to eliminate the influence of different dimensions or units. Assuming there are n objects to be evaluated, and a matrix consisting of m indicators (already normalized) is as follows:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix}$$

The matrix after standardization is denoted as Z, where the elements in Z are:

$z_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{n} x_{ij}^2}}$, (each element / The sum of the squares of the elements in its column)

The standardized data is shown in the following table.

Year	China's NEV Sales (10,000 units)	China's NEV Ownership (10,000 units)	China's NEV Market Share (%)
2011	0.0011	0.0004	0.0016
2012	0.0017	0.0007	0.0024
2013	0.0024	0.0018	0.0029
2014	0.0101	0.0071	0.0114
2015	0.0446	0.0250	0.0484
2016	0.0683	0.0544	0.0650
2017	0.1047	0.0914	0.0967
2018	0.1692	0.1553	0.1608
2019	0.1625	0.2268	0.1683

2020	0.1842	0.2930	0.1942
2021	0.4744	0.4669	0.4820
2022	0.8175	0.7801	0.8122

Table x. standardized data

5.2.3 Entropy weight method

To determine the weights for the three indicators, we use the entropy weight method for calculation. For the j-th indicator, the calculation formula for information entropy is:

$$e_j = -\frac{1}{\ln n}\sum_{i=1}^{n} p_{ij}\ln p_{ij} \quad (j = 1,2,\dots,m)$$

Where

$$p_{ij} = \frac{z_{ij}}{\sum_{i=1}^{n} z_{ij}}$$

After calculation, it is found that the three weights are the same.

5.2.4 Calculate scores

Define the maximum Z^+ and minimum Z^- as

$$Z^+ = (Z_1^+, Z_2^+, .., Z_m^+)$$
$$= (\max\{z_{11}, z_{21}, \dots, z_{n1}\}, \max\{z_{12}, z_{22}, \dots, z_{n2}\}, \max\{z_{1m}, z_{2m}, \dots, z_{nm}\})$$
$$Z^- = (Z_1^-, Z_2^-, .., Z_m^-)$$
$$= (\min\{z_{11}, z_{21}, \dots, z_{n1}\}, \min\{z_{12}, z_{22}, \dots, z_{n2}\}, \min\{z_{1m}, z_{2m}, \dots, z_{nm}\})$$

Define the distance between the i-th (i=1,2,...,n) evaluation object and Z^+, Z^- as:

$$D_i^+ = \sqrt{\sum_{j=1}^{m}(Z_j^+ - z_{ij})^2}$$

$$D_i^- = \sqrt{\sum_{j=1}^{m}(Z_j^- - z_{ij})^2}$$

So, we can calculate that the normalized score for the i-th (i=1,2,...,n) evaluation object is:

$$S_i = \frac{D_i^-}{D_i^+ + D_i^-}$$

5.2.5 Result of Topsis

The final development score indicators for each year are shown in the following table.

Year	Development scores
2011	0.000
2012	0.000

2013	0.001
2014	0.004
2015	0.019
2016	0.030
2017	0.047
2018	0.078
2019	0.091
2020	0.110
2021	0.230
2022	0.390

Table x. development scores

5.3 The Establishment of Multilinear Regression

Multiple linear regression analysis is a method used to assess the relationship between a dependent variable and multiple independent variables. In this context, the independent variables are represented by x, signifying variables that can independently and freely vary. On the other hand, the dependent variable is denoted as y, indicating a variable influenced by other variables and not independent.

After extensive data research and literature review, the following indicators have been selected as the primary factors influencing the development of new energy vehicles (NEVs) in China. These indicators include Year, Charging Piles (10,000s), Cost per km (RMB), New Energy Patents, NEV Production (10,000s), Per Capita Disposable Income (RMB), Positive Government Policies, NEV Enterprise Registrations (10,000s), and Fuel Cell Sales (10,000s).

（1） Policy Support: Policies introduced by the government, such as subsidies or preferential policies, significantly impact the continuous improvement of new energy development.

（2） Infrastructure: Charging Piles (10,000s): The quantity of infrastructure directly influences overall development.

（3） Technological Development:
- New Energy Patents: Indicative of advancements in technology.
- NEV Enterprise Registrations (10,000s): Reflects the number of enterprises contributing to technological innovation.

（4） Economic Factors: GDP Per Capita (RMB): The disposable income per capita has a substantial effect on the development of electric vehicles.

（5） Production Volume: NEEV Production (10,000s): The production volume of new energy vehicles directly influences their development. Stagnation may occur without sufficient production.

（6） Fuel Cell Sales (10,000s): Believing that peripheral sales influence overall development. This comprehensive set of indicators ensures a thorough examination of factors influencing the development of new energy vehicles in China, encompassing policy support, infrastructure, technological advancements, economic conditions, production volume, and peripheral sales considerations.

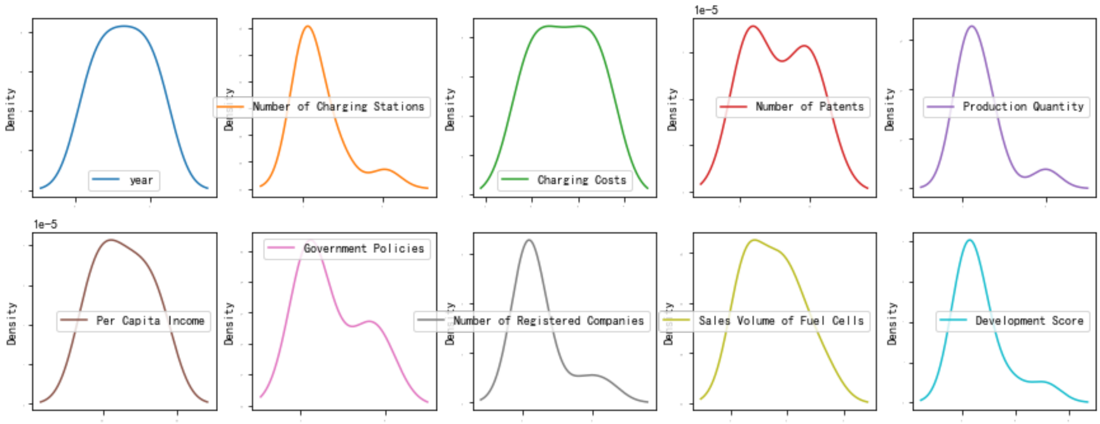


Fig x. The density graphs

Before conducting linear regression analysis, it is essential to assess the Pearson correlation between the key factors and the development scores derived from the Topsis model. Initially, the data is subjected to standardization, resulting in the following dataset, which is then concatenated with the development scores obtained from the Topsis model. After standardization, we can observe the variations of each factor over time.

Year	Charging Piles (10,000s)	Cost per km (rmb)	New Energy Patents	NEV Production	Per Capita Disposable Income (rmb)	Positive Government Policies	NEV Enterprise Registrations (10,000s)	Fuel Cell Sales (10,000s)
2011	0.0076	1.3	883	0.84	14551	3	0.12	435
2012	0.8281	1.3	1344	1.26	16510	5	0.21	1770
2013	1.9	1.2	2024	1.7533	18311	7	0.51	1516
2014	2.8	1.1	3111	7.85	20167	10	0.92	601
2015	4.9	1	4566	34.05	21966	13	1.28	425
2016	14.12	0.9	8205	51.7	23821	15	1.92	2045
2017	44	0.8	12262	79.4	25974	18	2.51	801
2018	77.7	0.7	17647	127	28228	40	3.85	2314
2019	121.9	0.6	18498	124.2	30733	70	4.1	2715
2020	168.1	0.5	19739	136.6	32189	81	7.52	500
2021	261.7	0.4	21300	216.6	35128	92	17.06	1556
2022	520.9	0.4	24700	500	36883	101	23.94	3368

Table x. Data used for multilinear regression

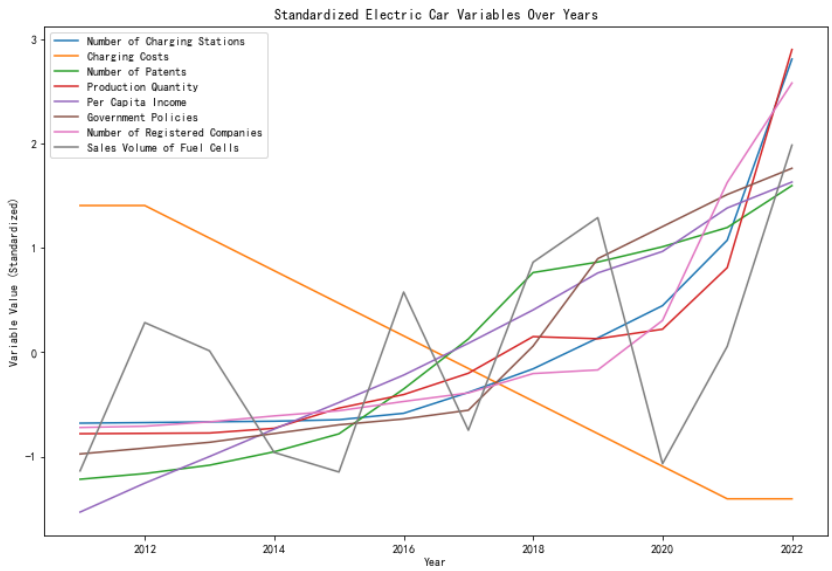


Fig x. The relationship chart of all factors with respect to the years.

Year	Charging Piles (10,000s)	Cost per km (rmb)	New Energy Patents	NEV Production	Per Capita Disposable Income (rmb)	Positive Government Policies	NEV Enterprise Registrations (10,000s)
2011	0.0000	0.4132	0.0182	0.0014	0.1595	0.0166	0.0039
2012	0.0013	0.4132	0.0277	0.0021	0.1810	0.0277	0.0068
2013	0.0030	0.3814	0.0416	0.0029	0.2007	0.0388	0.0164
2014	0.0045	0.3496	0.0640	0.0131	0.2211	0.0554	0.0296
2015	0.0078	0.3178	0.0940	0.0570	0.2408	0.0720	0.0412
2016	0.0226	0.2860	0.1688	0.0865	0.2611	0.0830	0.0618
2017	0.0704	0.2543	0.2523	0.1328	0.2847	0.0997	0.0808
2018	0.1242	0.2225	0.3631	0.2125	0.3094	0.2214	0.1239
2019	0.1949	0.1907	0.3806	0.2078	0.3369	0.3875	0.1320
2020	0.2688	0.1589	0.4062	0.2285	0.3528	0.4484	0.2421
2021	0.4184	0.1271	0.4383	0.3623	0.3851	0.5093	0.5491
2022	0.8329	0.1271	0.5082	0.8364	0.4043	0.5592	0.7706

Table x. Data used for multilinear regression after Standardization(Partially)

By calculating the Pearson correlation coefficient, we obtained a value ranging between -1 and 1. The formula for the Pearson correlation coefficient is as follows. Before conducting multiple linear regression analysis, it is necessary to perform feature selection.

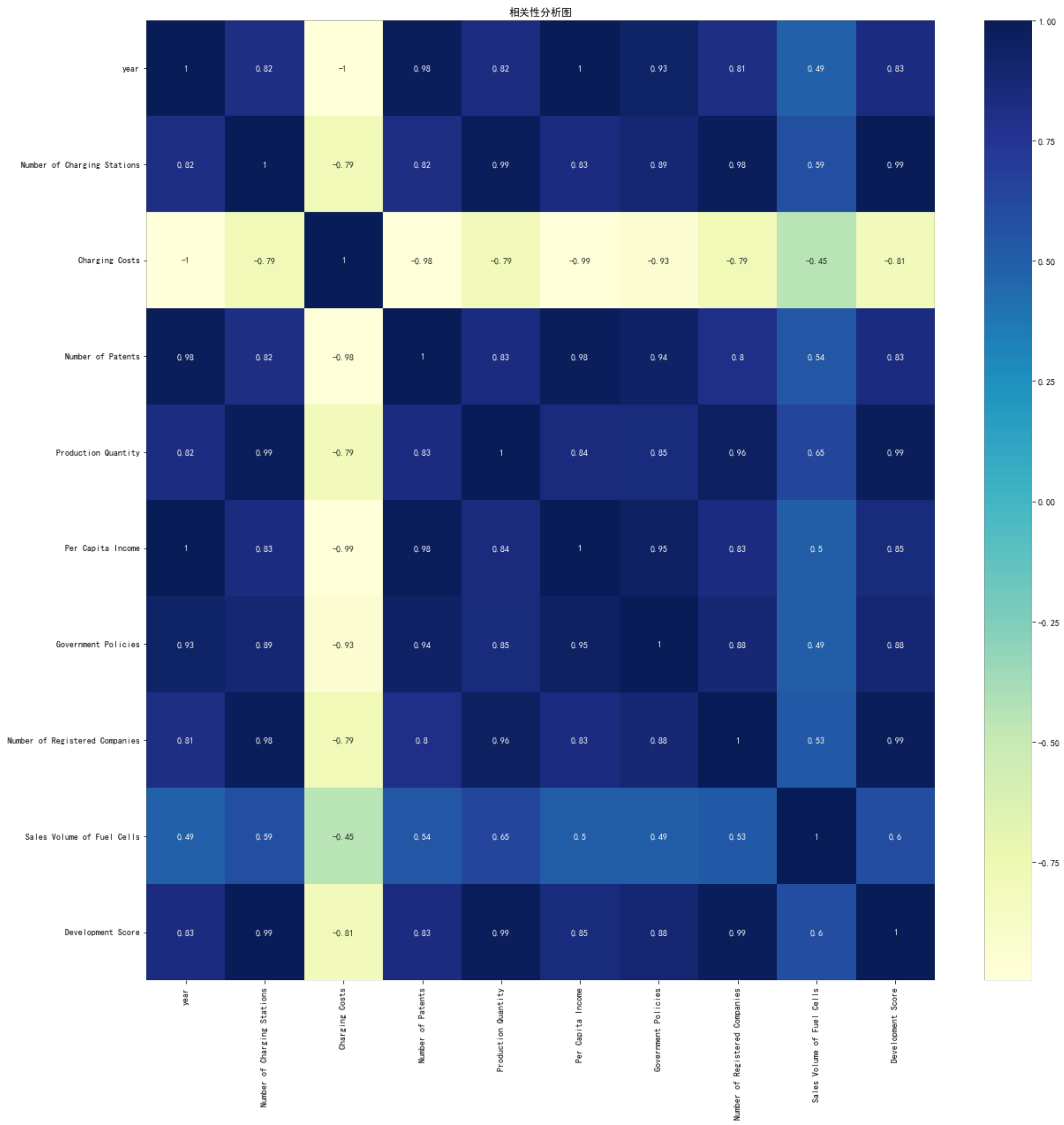


Fig x. Pearson correlation coefficient

The final results indicate that, from the graph, we observe a correlation coefficient of 0.6 between Fuel Cell Sales (in units of 10,000s) and Development Score. Since this value is not close to -1 or 1, we have decided to exclude this feature in the final analysis.

The final analysis results indicate that, from the graph, we observe a correlation coefficient of 0.6 between Fuel Cell Sales (in units of 10,000s) and Development Score. Since this value is not close to -1 or 1, we have ultimately decided to exclude this feature.

Year	Charging Piles (10,000s)	Cost per km (rmb)	New Energy Patents	NEV Production	Per Capita Disposable Income (rmb)	Positive Government Policies
2011	0.0000	0.4132	0.0182	0.0014	0.1595	0.0166
2012	0.0013	0.4132	0.0277	0.0021	0.1810	0.0277
2013	0.0030	0.3814	0.0416	0.0029	0.2007	0.0388

2014	0.0045	0.3496	0.0640	0.0131	0.2211	0.0554
2015	0.0078	0.3178	0.0940	0.0570	0.2408	0.0720
2016	0.0226	0.2860	0.1688	0.0865	0.2611	0.0830
2017	0.0704	0.2543	0.2523	0.1328	0.2847	0.0997
2018	0.1242	0.2225	0.3631	0.2125	0.3094	0.2214
2019	0.1949	0.1907	0.3806	0.2078	0.3369	0.3875
2020	0.2688	0.1589	0.4062	0.2285	0.3528	0.4484
2021	0.4184	0.1271	0.4383	0.3623	0.3851	0.5093
2022	0.8329	0.1271	0.5082	0.8364	0.4043	0.5592

Table x. Data after processed used for multilinear regression(Partially)

5.4 Results

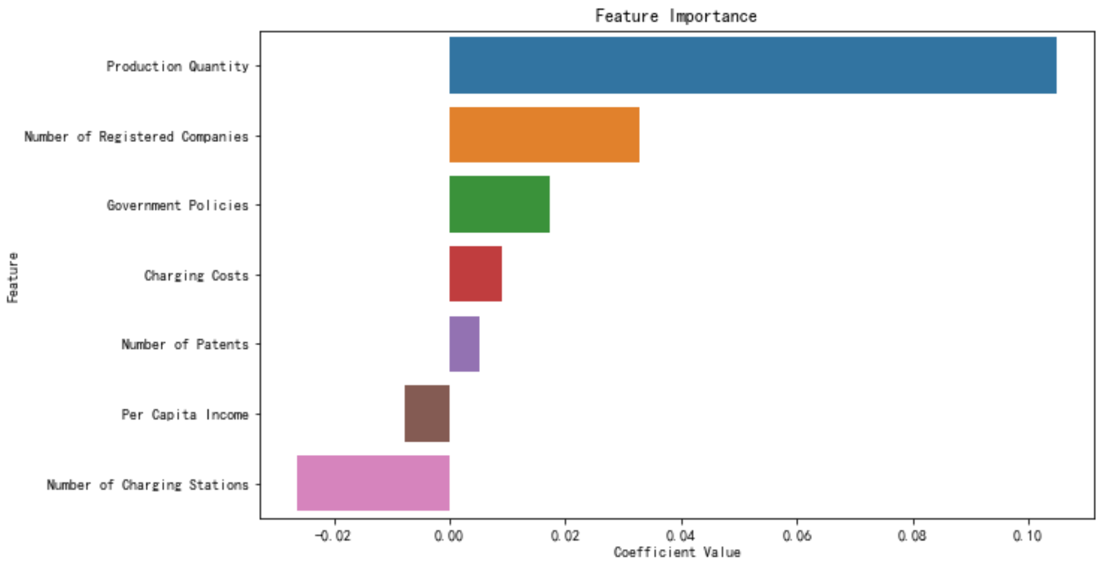


Fig x. The Importance Features.

Finally, the following weights can be obtained. Multiple linear regression was performed using Python, and the error calculations yielded:

MSE: 0.0004511817851570348
MAE: 0.01658187789086477

The corresponding weights, aligned with the independent variables from left to right according to the table, are as follows:

Weights: [-0.02624563 0.00915644 0.00527758 0.10484835 -0.00777705
0.01738091 0.03289783]

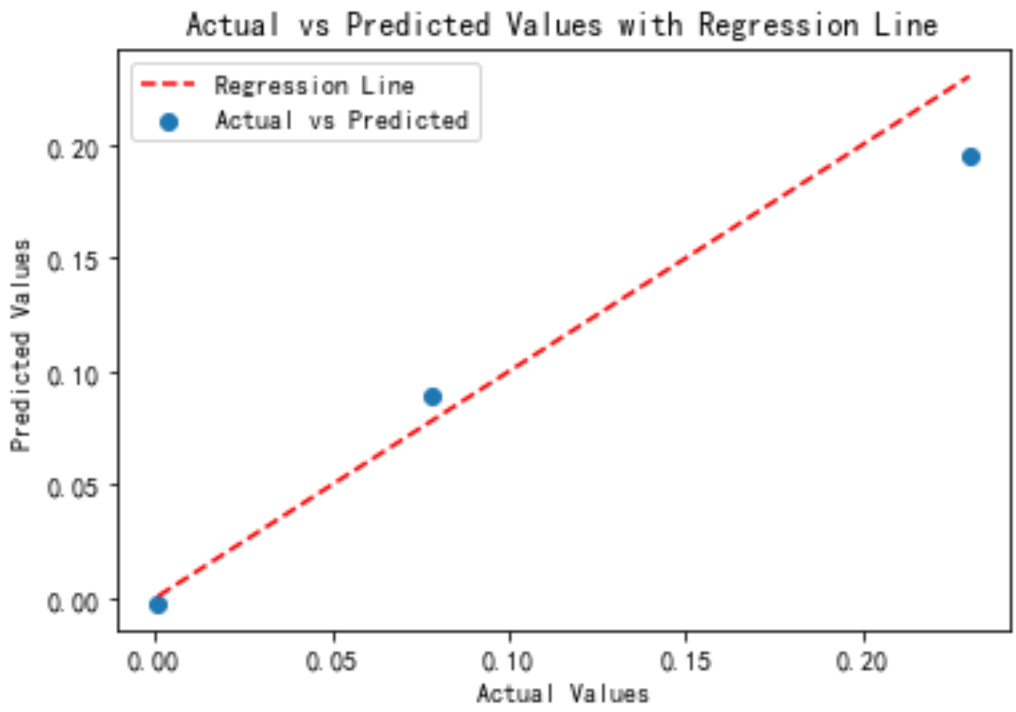


Fig x. The regression graph

The regression fit can be observed to be excellent from the graph below, as shown in the figure. From the analysis results, it is evident that the development of new energy vehicles is predominantly influenced by the production volume of new energy vehicles and the number of registered companies. Therefore, we should strongly encourage the development of related industries.。

6 Model for Question II : Grey Model(1,1)

6.1 Data Description

For this data we use sales as our indicator.

Year	2016	2017	2018	2019	2020	2021	2022
Sales	50.7	77.7	125.6	120.6	136.7	352.1	606.7

Table x. NEEV sales in China (unit:10000)

6.2 The Establishment of GM(1, 1) iii

6.2.1 Examination of data and data preprocessing

Firstly, to ensure the feasibility of the modeling approach, it is necessary to conduct essential checks and processing on the known data columns. Calculate the Grading test for the reference series, aiming to evaluate its stability. This step is crucial for selecting an appropriate modeling method.

$$\lambda(\mathrm{k}) = \frac{x^0(k-1)}{x^0(k)}, \qquad k = 2,3,\dots,n \tag{1}$$

If the results λ(k) falls in the acceptable range $\Theta = \left(e^{-\frac{2}{n+1}}, e^{\frac{2}{n+1}}\right)$, then the sequence $x^{(0)}$ can be used as data for the GM (1, 1), otherwise it is necessary to perform essential transformation on the sequence $x^{(0)}$to bring it within the acceptable range. This involves aggregate a constant c for a shift transformation.

$$y^{(0)}(k) = x^0(k) + c, \qquad k = 1,2,\dots,n \tag{2}$$

Making the sequence $y^0 = \left(y^0(1), y^0(2), \dots, y^0(n)\right)$ meets the requirement.

$$\lambda(k) = \frac{x^0(k-1)}{x^0(k)} \in \Theta, \qquad k = 2,3,\dots,n \tag{3}$$

6.2.2 Mathematical modelling

Later on, establishing Differential Equation Model.

$$\frac{dx^1(t)}{dt} + ax^1(t) = b, \tag{4}$$

The model is a first-order, single-variable differential equation, denoted as GM(1,1). In order to identify model parameters a and b, in the interval k-1 < t ≤ k, let..

$$x^{(1)}(t) = z^{(1)}(k) = \frac{1}{2}\left[x^{(1)}(k-1) + x^{(1)}(k)\right] \tag{5}$$

$$\frac{dx^1(t)}{dt} = x^{(1)}(k) - x^{(1)}(k-1) = x^{(0)}(k) \tag{6}$$

Then, equation (4) transformed into a discrete model,

$$y^{(0)}(k) = az^{(1)}(k) = b, \qquad k = 2,3,\dots,n \tag{3}$$

Equation (15.4) is referred to as the grey differential equation, while Equation (15.3) is termed the corresponding whitening equation. Record as:

$$u = [a,b]^T, Y = \left[x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n)\right]^T, B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix}$$

The least square method is employed to obtain the estimated value of u that minimizes $J(u) = (Y - Bu)^T (Y - Bu)$

$$\hat{u} = [\hat{a}, \hat{b}]^T = \left[B^T B\right]^{-1} B^T Y \tag{4}$$

Solve the equation (6) , we get

$$\hat{x}^{-1}(t) = \left(x^{(0)}(1) - \frac{\hat{b}}{\hat{a}}\right) e^{-\hat{a}t} - \frac{\hat{b}}{\hat{a}}, \tag{5}$$

And predicted value,

$$\hat{x}^{(1)}(k+1) = \left(x^{(0)}(1) - \frac{\hat{b}}{\hat{a}}\right) e^{-\hat{a}k} + \frac{\hat{b}}{\hat{a}}, \qquad k = 0,1,2, \dots \tag{6}$$

Moreover, $\hat{x}^{(0)}(1) = \hat{x}^{(1)}(1), \hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) \qquad k = 0,1,2, \dots$

6.2.3 Residual examination

There are two ways to verify, the method we utilize is relative error method , to calculate the residual we have:

$$\delta(k) = \frac{|x^0(k) - \hat{x}^{(0)}(k)|}{\hat{x}^{(0)}(k)}, \quad k = 1,2,3, \dots, n \tag{7}$$

In this situation $\hat{x}^{(0)}(1)$=$x^{(0)}(1)$, If $\delta(k) < 0.2$, it can be considered as meet general requirements, if $\delta(k) < 0.1$, it is considered to meet higher requirements.

6.2.4 Prediction and Forecasting

Obtain the predicted values at specified points from the GM (1,1) model, and provide corresponding predictions and forecasts according to the needs of the actual problem.

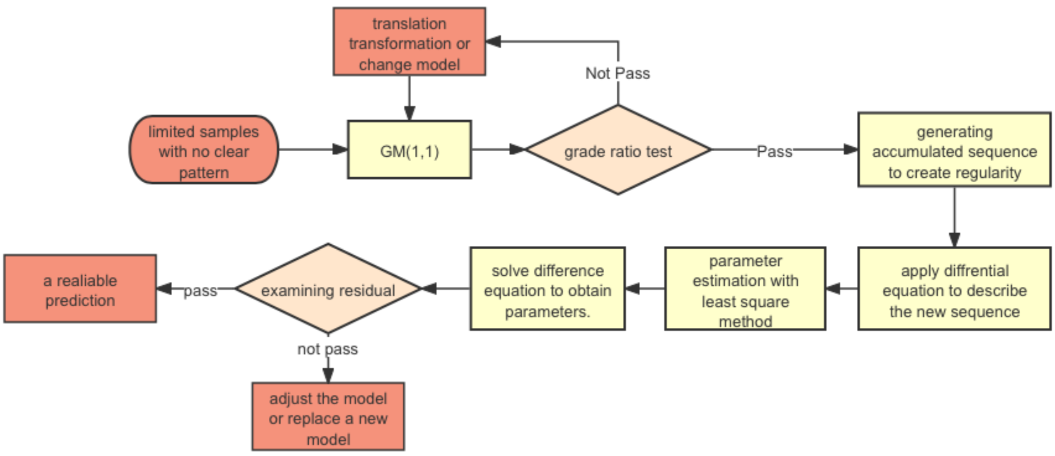
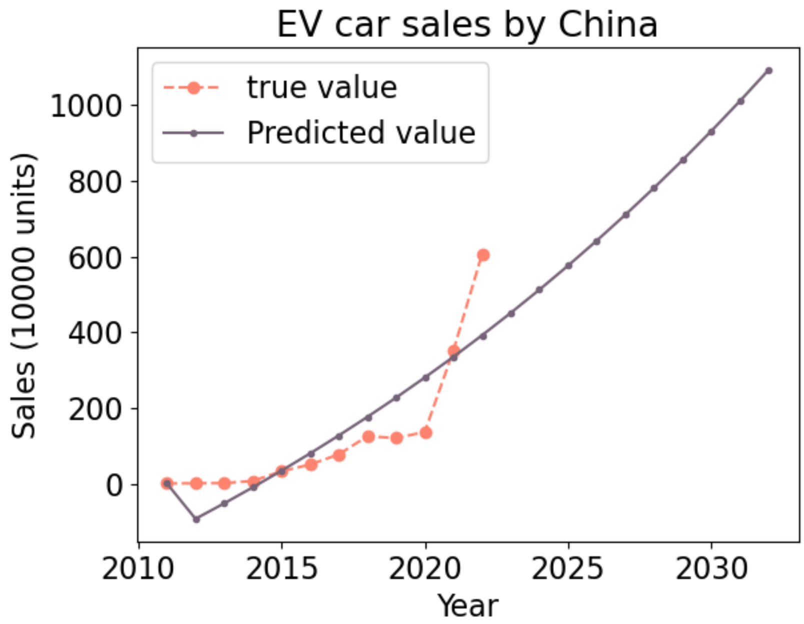


Fig x. The proposed modified GM (1,1) model.

6.3 Results



From the graph, it can be observed that the future development of China's new energy electric vehicles will continue to trend upward.

7 Model for Question 3: Curvilinear Regression and ARIMA

7.1 Data Description

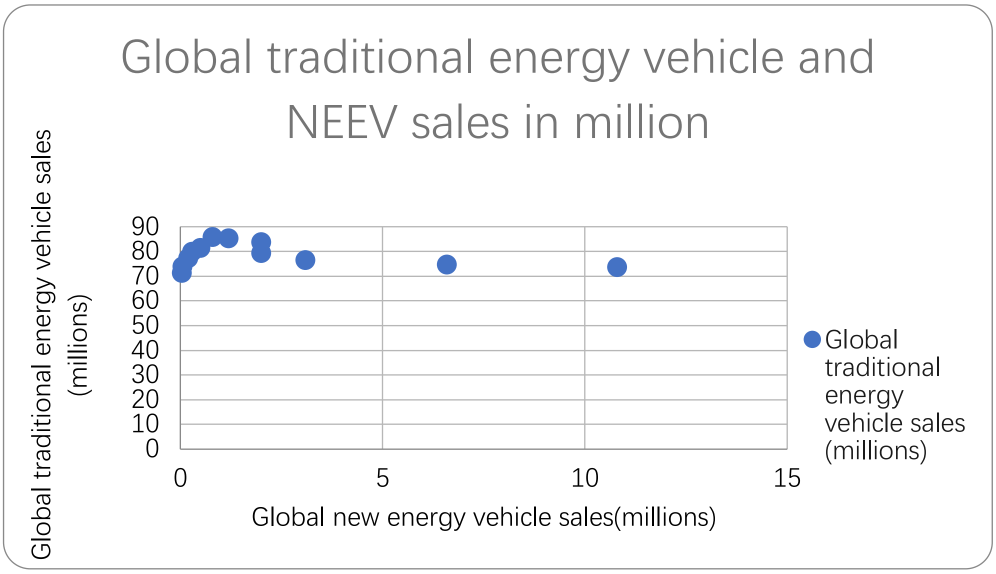
The dataset comprises annual global automotive sales data, categorizing the information into the following key columns: Year, Global Total Vehicle Sales, Global Traditional Energy Vehicle Sales, Global New Energy Vehicle Sales, and " Global New Energy Vehicle Market Share.

Year	Global vehicle sales (millions)	Global traditional energy vehicle sales (millions)	Global new energy vehicle sales(millions)	Global new energy vehicle market share (%)
2011	71.239	71.2	0.039	0.05
2012	73.758	73.7	0.058	0.08
2013	77.5	77.3	0.2	0.26
2014	80	79.7	0.3	0.38
2015	81.8	81.3	0.5	0.61
2016	86.4	85.6	0.8	0.93
2017	86.3	85.1	1.2	1.39
2018	85.6	83.6	2	2.34

2019	81.3	79.3	2	2.46
2020	71.8	68.7	3.1	4.32
2021	74.9	68.3	6.6	8.81
2022	74.8	64	10.8	14.44

7.2 Relation of conventional vehicle and NEEV

To analyze the relationship between traditional and new energy vehicles, a scatter plot is created with global new energy vehicle sales on the x-axis and global traditional energy vehicle sales on the y-axis. However, due to significant disruptions in data post-2020 caused by the pandemic, it's proposed to treat post-2020 traditional energy vehicle sales as outliers. Utilizing the ARIMA model, we can predict new data to replace these outliers and provide a more accurate representation of the relationship between the two sectors.



Year	Global new energy vehicle sales(millions)	Global traditional energy vehicle sales (millions)
2011	0.039	71.2
2012	0.058	73.7
2013	0.2	77.3
2014	0.3	79.7
2015	0.5	81.3
2016	0.8	85.6
2017	1.2	85.1
2018	2	83.6
2019	2	79.3

2020	3.1	76.5
2021	6.6	74.67
2022	10.8	73.48

By observing the plot, it is evident that there is no linear relationship between the two. Here, we consider using a curve regression method to fit the nonlinear relationship between them.

7.3 The Establishment of Model III

Curvilinear regression [iv] is the model that attempts to fit a curve as opposed to a straight line. To established such a model, here is the steps:

1. Plot the scatter diagram.
2. Determine the type of relationship curve based on the scatter plot.
3. Estimate the unknown parameters.

Based on the observed pattern in the scatter plot, we can assume the relationship curve to be

$$y = axe^{-bx}$$

The equation represents a decreasing exponential function. In this equation:

- y is the dependent variable,
- x is the independent variable,
- a is a scale factor that determines the overall height of the curve,
- b is a rate constant that influences the decay, and
- e is the mathematical constant approximately equal to 2.71828 (the base of the natural logarithm).

This type of equation is commonly used to model processes or phenomena that exhibit exponential decay.
The above relationship curve passes through the origin, but our data has an intercept of 70. Therefore, the relationship curve needs to be adjusted.

$$y = axe^{-bx} + 70$$

7.3.1 Curve Transformation

The relationship curve is given by:

$$y = axe^{-bx}$$

Dividing both sides by x, and taking the natural logarithm (ln),

we obtain $ln\frac{y}{x} = -bx + lna$

Let $y' = ln\frac{y}{x}$, $b' = -b, a' = lna$,

we get $y' = b'x + a'$
The data after this transformation is shown partially in the table below.

x	0.3	0.5	0.8	1.2	2	2	3.1	6.6	10.8
y'	3.48	3.12	2.97	2.53	1.92	1.54	0.74	-0.35	-1.13

7.4 ARIMA

An Autoregressive Integrated Moving Average (ARIMA) model is an extension of the Autoregressive Moving Average (ARMA) model. These models are utilized in time series analysis for understanding data patterns or predicting future points in the series (forecasting). ARIMA models are particularly useful when the data exhibit non-stationarity, requiring one or more differencing steps (represented by the 'integrated' part of the model) to eliminate non-stationarity. The ARIMA model is denoted as ARIMA(p, d, q), where

- 'p' is the autoregressive (AR) parameter,
- 'd' is the differential parameter,
- 'q' is the moving average (MA) parameter.

The forecasting equation is constructed as below. Let Y denote the d^{th} difference of Y, which means:
If $d = 0, y_t = Y_t$
If $d = 1, y_t = Y_t - Y_{t-1}$
If $d = 2, y_t = Y_t - 2Y_{t-1} + Y_{t-2}$

The note clarifies that in ARIMA models, the second difference (d=2) doesn't represent the difference from two periods ago. Instead, it captures the discrete analog of a second derivative, signifying the local acceleration of the series rather than its local trend. The note emphasizes this distinction in the context of time series analysis.

$$y_t = \mu + \phi y_{t-1} + \cdots + \phi_p y_{t-p} - \theta_1 e_{t-1} \ldots - \theta_q e_{t-q}$$

In a stationary and autocorrelated series, the forecast can be expressed as a multiple of its own previous value plus a constant. The forecasting equation, in this case, is $\hat{Y}_t = \mu + \phi_1 Y_{t-1}$, where Yt is regressed on its lagged value by one period. This corresponds to an "ARIMA(1,0,0) + constant" model. If the mean of Y is zero, the constant term would be omitted.

7.5 Results

For After the transformation, the relationship curve becomes linearized, allowing for linear regression calculations. By utilizing linear regression, we obtain $a' = 3.27$ and $b' = 0.479$, which implies $a = 26.31$ and $b = 0.479$. Simultaneously, the coefficient of determination R^2 is found to be 0.874, indicating a good fit.

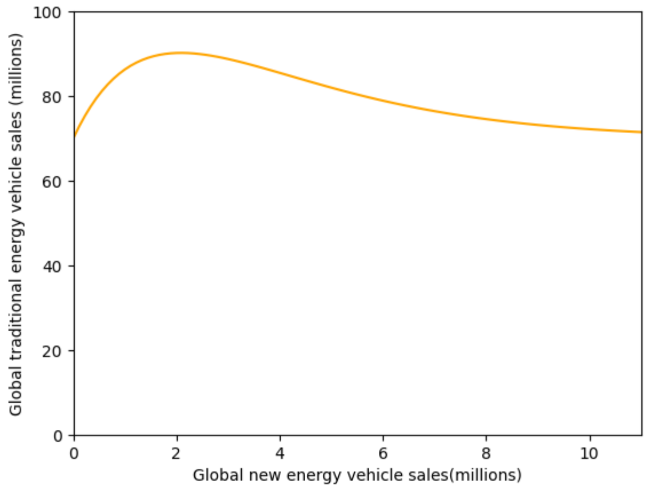


Fig x. Global NEEV sales in millions

For The model ARIMA result is $y_t = 77.32 + 0.83y_{t-1} + \epsilon_t$, which y_t is dependent variable at the time t , μ is the constant term, $\phi$is the coefficient of the autoregressive term, $y_{t-1}$is the previous time step, ϵ_t is the error term.

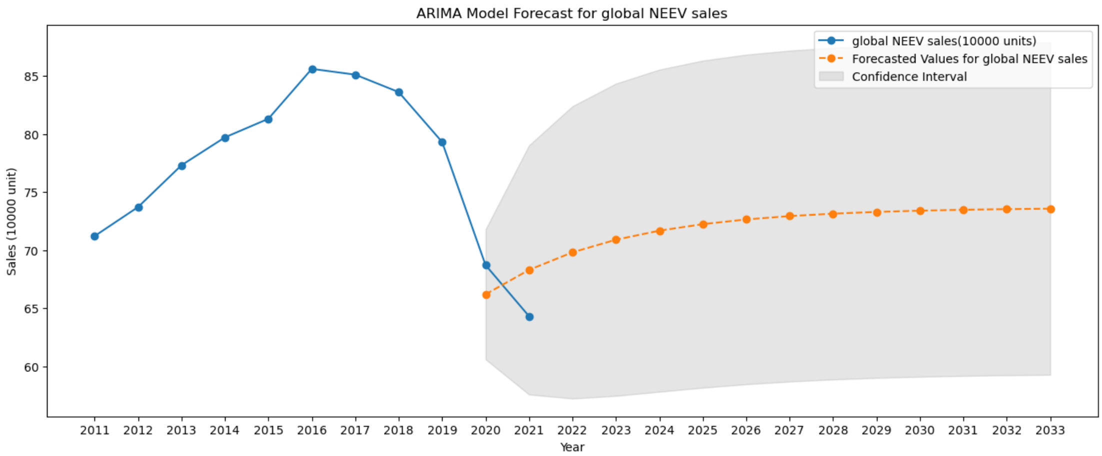


Fig x. predicted global NEEV sales using ARIMA

With the development of new energy electric vehicles, we predict that the market share of traditional energy vehicles will be gradually overtaken by new energy vehicles. However, some individuals may still opt for traditional fuel vehicles until new energy vehicle technology advances enough to fully replace traditional one.

8 Model for Question IV: Curvilinear Regression

8.1 Background Review

Some policies that may adversely affect the export of new energy vehicles include the IRA (International Revenue Acceleration), disruptions in the US battery supply chain, and anti-subsidy investigations in Europe. The IRA, as outlined by the Center on

Global Energy Policy at Columbia University SIPA, along with other policy measures and investigations, can have a significant impact on the international trade of new energy vehicles. It is essential to consider these policies and investigations when analyzing the export dynamics of China's new energy electric vehicles, as they may contribute to fluctuations in export volumes and market conditions, particularly in relation to subsidies.

8.2 Data Description

The data set we are using here is China's exported NEEV data from 2011 to 2023 and show partially as below.

year	2016	2017	2018	2019	2020	2021	2022	2023
NEEV Exported by China in 10000 units	7.6	10.6	14.7	25.4	22.4	59.0	122.0	220

Based on the data, a scatter plot is created to observe the variation pattern in export volumes.

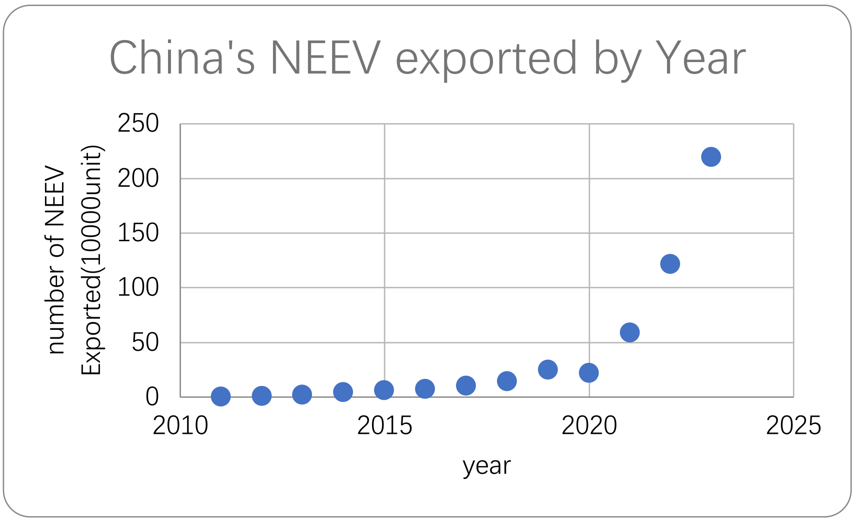


Fig x China's NEEV exported by year.

Observing the scatter plot reveals that the export volume in 2020 did not increase but declined. This decline can be attributed to the impact of the COVID-19 pandemic and a certain country implementing a tech blockade, leading to insufficient production materials.

The foreign resistance policy (IRA) was enacted in 2023. Therefore, we use data from 2011 to 2019 for nonlinear regression fitting to obtain the export volume data for 2020-2023 may be considered as unaffected by resistance policies. We then compare this with the actual export volume data to understand the development of China's new energy electric vehicles under the influence of relevant policies.

8.3 Established model

The model used for this question is the same as the curve regression model used in question three.

Based on the observed scatter plot patterns, here we opt for a power function curve as the relationship curve, which can be expressed as $y = ae^{bx}$。 Taking the natural logarithm (ln) on both sides of the relationship curve, we obtain $\ln y = bx + lna$. Let $y' = \ln y$, $a' = \ln a$, then we have $y' = bx + a'$。
Taking the year 2010 as the reference point (0) and setting its export volume to 0, the transformed data for new energy electric vehicle exports is presented partially at the table below.

x	1	2	3	4	5	6
y‘	-0.693	0.182	0.875	1.568	1.856	2.028

Table x. Transformed Data for New Energy Electric Vehicle Exports.

8.4 Result

Utilizing linear regression, we calculate $a' = -0.655, b = 0.444$, and $a = 0.5194$ thus obtaining a = 0.5194. Additionally, R^2 = 0.95, indicating a good fit.

Substituting the values of a and b into the relationship curve, we derive the curve regression equation as follows:

$$y = 0.5194e^{0.444x}$$

Using the above formula, we predict the export volumes for the years 2020 to 2023, obtaining the following values: 44.032, 68.643, 107.01, and 166.821 million units, respectively, we get the result below.

Year	Number predicted of NEEV Exported in 10000 units	NEEV Exported in 10000 units
2011	0.810	0.5
2012	1.262	1.2
2013	1.968	2.4
2014	3.068	4.8
2015	4.782	6.4
2016	7.455	7.6
2017	11.622	10.6
2018	18.118	14.7
2019	28.245	25.4
2020	44.032	22.4
2021	68.643	59.0
2022	107.010	122.0
2023	166.821	220

Table x. Predicted and Actual Data for
China's New Energy Electric Vehicle Exports

9 Model V :

9.1 data description:

Due to the requirements of the task, which involve limiting the study to cities with a population of one million, and the need to consider the ecological impact of public buses and other modes of transportation, we are considering energy consumption as a metric for evaluating the environmental impact of electric and fuel-powered vehicles. If electric vehicles prove to be more energy-efficient than gasoline cars, it will demonstrate a positive environmental impact of electric transportation.[1]v

The data used below is based on the 2020 population of Beijing, which is 21.893 million, and the total number of various types of vehicles in Beijing in 2020. The energy consumption of various electric transportation modes in that year is standardized based on the consumption of kilowatt-hours per 100 kilometers, while fuel-powered vehicles are standardized based on the consumption of liters of gasoline per 100 kilometers. +---

Category	NEEV (kWh/100km)	Fuel Vehicles （L/100km）	number of vehicles
Private car	17.4	5.61	526915
Taxi	17.4	5.61	17972
Bus	73.6	30	27800
Tourist Bus	73.6	30	6699
Sanitation Bus	24	18	2670
Logistic Vehicle	33.8	18	143445

9.2 Established Model

This model is primarily based on the comprehensive energy consumption calculation guidelines of the People's Republic of China's national standards, establishing a mathematical model for comparing new energy vehicles with conventional fuel vehicles. Before applying the comprehensive energy consumption calculation guidelines, it is necessary to convert the energy consumption of new energy vehicles (electricity consumption) and conventional fuel vehicles (fuel consumption) into values that can be multiplied by SCE parameters and the respective vehicle quantities to obtain the total energy consumption.

The unit for SCE parameters of new energy is kgce/kwh, and the energy consumption unit for new energy vehicles is kwh/km. It can be observed that without any conversion, the total energy consumption for new energy vehicles used by one million people can be obtained.

$$Electric\ Vehicle = \frac{E}{S} * SCE_{electric}$$

The unit for SCE parameters of fuel is kgce/kg. To convert the fuel consumption unit (L) to kg, we need to multiply it by the density of fuel. This result can then be multiplied by SCE parameters to obtain the total energy consumption for fuel vehicles.

$$Fuel\,Vehicle = \frac{V}{S} * P * SCE_{fuel}$$

Finally, the calculation of Relative Energy Efficiency (REE) is based on different categories of vehicles. The numerator is the total energy consumption of electric vehicles, and the denominator is the total energy consumption of fuel vehicles. If R < 1, it indicates that electric vehicles are more energy-efficient than fuel vehicles; conversely, if R > 1, it indicates that fuel vehicles are more energy-efficient.

$$Relative\,Energy\,Efficiency = \frac{Electric\,Vehicle}{Fuel\,Vehicles}$$

9.3 Result

According to our mathematical model above, we can conclude that as long as R is less than 1, electric vehicles are more energy-efficient than fuel-powered vehicles, significantly contributing to environmental sustainability. The result shown below.

Types	Average Electric(kWh/100km)|	AverageFuel (L/100km)	REE
Car	2,353,263.0	758,724.45	0.35733
Taxi	80248.8	25873.32	0.35733
Bus	525,136	214050	0.28264
Tourist Bus	126518.4	51570	0.28264
Sanitation Car	16440	12330	0.15361
Logistic vehicle	1244448.4	662724	0.21634

Table x. result of REE

10 Question 6

Dear Citizens,
We hope this letter finds you in good health and high spirits. Today, we wish to shed a light on a composing to highlight the pivotal role matter of paramount importance - the revolutionary impact of New Energy Electric Vehicles (NEEV) on our environment. Our extensive research has unequivocally demonstrated that various types of NEEVs, including cars, taxis, buses, coaches, sanitation vehicles, and logistics vehicles, play a crucial role in reducing carbon dioxide emissions and promoting energy efficiency. These vehicles represent a collecting effort towards building a sustainable and eco-friendly environment. We contribute to achieving carbon peaks by 2030 and carbon neutrality by 2060, aligning with global sustainability goals. Let's drive change together for a cleaner, healthier planet.

i [2]Yamak, Peter T., Yujian Li, and Pius K. Gadosey. 2020. A Comparison between ARIMA, LSTM, and GRU for Time Series Forecasting. In \textit{Proceedings of the 2019 2nd International Conference on Algorithms, Computing and Artificial Intelligence (ACAI '19)}, edited by Association for Computing Machinery. New York, NY, USA. Pages 49–55. ISBN 10.1145/3377713.3377722.
ii [1] Riahi, S., Fathianpour, N., Tabatabaei, S. H., "Improving the Accuracy of Detecting and Ranking Favorable Porphyry Copper Prospects in the East of Sarcheshmeh Copper Mine Region Using a Two-step Sequential Fuzzy - Fuzzy TOPSIS Integration Approach," Journal of Asian Earth Sciences: X, 2023, 100166, ISSN 2590-0560
iii [1] Liu, X., Peng, H., Bai, Y., Zhu, Y., Liao, L., "Tourism Flows Prediction based on an Improved Grey GM(1,1) Model," Procedia - Social and Behavioral Sciences, Volume 138, 2014, Pages 767-775, ISSN 1877-0428
iv [1] Naeem, M., Rauf, A., Akhtar, M. S., Iqbal, Z., "QSPR Modeling with Curvilinear Regression on the Reverse Entropy Indices for the Prediction of Physicochemical Properties of Benzene Derivatives," Polycyclic Aromatic Compounds, 2023, ISSN 1040-6638
[1]v [1] Wang, N., Cai, G., Yan, P., et al. (2015). "Analysis and Countermeasures of Automobile Popularization and Energy Issues." Automotive Industry Research, 2015(08), 17-20.