

Research On the Development Trend of New Energy Vehicles Based on Multi-Model Analysis

Yimeng Zhao¹, Xin Wu¹, Hao Luo^{1, 2, 3, *}

¹ School of Information Science and Engineering of Gansu University of Chinese Medicine, Gansu, China

² School of Electronic and Information Engineering, Lanzhou Jiaotong University, Gansu, China

³ Key Laboratory of Media convergence Technology and Communication, Gansu Province, Gansu, China

* Corresponding Author Email: luoh804@gszy.edu.cn

Abstract. This study aims to explore the dynamic factors and future trend of the development of the new energy automobile industry and its impact on the traditional automobile industry. First of all, through the gray correlation analysis, the research points out that the new energy vehicle industry is affected by many factors, especially the popularity of charging piles and the implementation of government subsidy policies are the key elements to promote the development of the industry. Then, the research uses the ARIMA model to predict the market trend of China's new energy electric vehicles in the next decade, showing an optimistic development blueprint. Further, using multiple linear regression analysis to comparing the differences between new energy and traditional fuel vehicle industry, it is found that new energy vehicles are gradually subverting the pattern of the traditional automobile market and gradually occupy the dominant position. This study not only provides important industry insights, but also proposes coping strategies and adjustment directions for policy makers and automobile manufacturers.

Keywords: New Energy Vehicles, Gray Correlation Analysis, ARIMA, Multiple Linear Regression.

1. Introduction

The background of this study is based on the global energy structure and the rapid transformation of the automotive industry. With the increasingly severe problem of global climate change, the dependence of countries on traditional fossil fuels has attracted widespread attention. Reducing carbon dioxide emissions and achieving green and low-carbon development has become an urgent task in today's world [1]. As an important part of urban transportation and human activities, the energy consumption and emission problems of automobiles are particularly prominent. Therefore, it is the inevitable trend of new energy vehicles to use the use of clean and renewable energy.

As the world's largest automobile market, China's new energy vehicle industry development is highly valued by the government. Through a series of policy support and market-driven mechanism, China's new energy vehicle industry has achieved leap-forward development. Various types of new energy vehicles, such as electric vehicles, plug-in hybrid electric vehicles and fuel cell vehicles, have been rapidly promoted. However, the development of the new energy vehicle industry also faces challenges in technology, cost, market acceptance and related infrastructure construction [2].

At the same time, the rise of the new energy vehicle industry has had a profound impact on the traditional energy vehicle industry. From the oil industry to the automobile manufacturing industry, from the supply chain management to the end-consumption market, the development of new energy vehicles is forcing these industries and markets to carry out structural restructuring and transformation. Understanding and analysis of these impacts are of great value for promoting the healthy development of new energy vehicles, alleviating the transformation pressure of the traditional energy vehicle industry and promoting the sustainable development of the whole industry.

In this context, this study will use the grey association analysis model to deeply analyze the key driving factors of the development of the new energy vehicle industry [3]. By studying the relationship between multiple variables, the model reveals which factors have the most significant

impact on the new energy vehicle industry. Subsidies, environmental awareness, battery prices and the number of charging piles will all be taken into account.

Further, this study will use the ARIMA (autoregressive integral sliding average) model [4] based on the results of existing data and grey association analysis to predict the development trend of new energy electric vehicles in China in the next decade. With its unique time-series analysis ability, this model helps us to build the future development trend of the new energy electric vehicle industry based on historical data.

In addition to analyzing and predicting the development of the new energy vehicle industry, this study will also explore the profound impact of the rise of the new energy vehicle industry on the traditional fuel vehicle industry. With the expanding proportion of new energy vehicles in the market share, the traditional energy vehicle industry will undoubtedly encounter unprecedented challenges and shocks. The study will involve the comparison of the energy consumption of a variety of new energy vehicles and traditional energy vehicles and reveal the long-term impact of the new energy vehicle industry on the global energy pattern.

The in-depth investigation of this study is of great significance to ensure the healthy and sustainable development of China's new energy vehicle industry. Predicting the future development trend not only provides support for enterprise layout and policy making, but also provides a framework for researchers to deeply understand the complex dynamics of industrial development.

2. Research method

2.1. Data acquisition and preprocessing

(1) As for the analysis of the factors affecting the development of the new energy vehicle industry, the data set relied on by the Institute covers the monthly related data of new energy vehicles from September 2019 to July 2024, with a total of 35 data points. The data set contains seven key characteristic indicators: the output of new energy vehicles, the configuration of public charging piles, subsidy policies for new energy vehicles, electricity price, cost price of new energy batteries, carbon emissions and public awareness of environmental protection.

(2) In order to predict the development trend of China's new energy vehicle industry in the next ten years, this study comprehensively uses the data of public charging piles released by China Charging Alliance between 2016 and 2022, including the number of DC charging piles, AC charging piles and AC-DC integrated charging piles [5]. In addition, this study also records the sales and production data of China's new energy vehicles from 2010 to 2022 released by China Association of Automobile Manufacturers and the Ministry of Industry and Information Technology. It aims to reflect the effect of government subsidy policies through the characteristics of production and sales, and accordingly predict and analyze the development trajectory of China's new energy electric vehicles in the following ten years.

(3) In discussing the interaction influence of the new energy vehicle industry and the traditional energy vehicle industry, two data sets were introduced in this study. The first dataset records information on vehicle production in China, the US, Japan, Germany, South Korea and India between 2010 and 2021. The second set of data involves the inventory of new energy vehicles in China, the European Union, the United States and other countries between 2013 and 2021^[6]. The combined analysis of these two sets of data is conducive to revealing the impact and challenges of the rise of the new energy vehicle industry on the traditional energy vehicle industry.

2.2. Method introduction

(1) Gray association analysis: Gray association analysis is a statistical method used to identify the degree of influence among various factors, especially suitable for those cases with incomplete information and small amount of data. This method relies on establishing correlation coefficients with target factors to characterize the relative importance among factors. It is not strict with data

requirements, is relatively simple in model construction, and can quantify the uncertain information, and is suitable for use in the condition of lack of data.

(2) ARIMA: The ARIMA model is a classic model in the field of time series analysis, which is applicable to the data with sequence stability or regular existence. It stabilizes the non-stationary data by difference, and then builds an autoregressive model and a moving average model on this basis. Its modeling is more flexible, and it can be finely adjusted through model identification and parameter estimation, and the prediction accuracy of time series is relatively high.

(3) Multiple linear regression: Multiple linear regression [7] is a basic method in statistical analysis, which is used to analyze the linear relationship between two or more independent variables and dependent variables. This method is widely used in cases where the relationship between variables is nearly linear, and the data are relatively sufficient. Its model is concise, easy to understand and interpret, can give a quantitative evaluation of the relationship between variables, and can give the credibility of the model through statistical tests.

2.3. Model evaluation index

This study evaluated the effectiveness of the model using RMSE (Root Mean Square Error), MSE (Mean Square Error), and R^2 (Coefficient Of Determination).

(1) MSE

MSE is a common indicator to measure the deviation between the model and the actual observed value. The calculation is calculated by summing the sum of the difference between the predicted value and the true value for all data points and then dividing by the number of data points. The calculation formula is as follows:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (1)$$

Among, n is the total number of data points, Y_i is the true value of the i observation point, while \hat{Y}_i is the predicted value of the i observation point.

(2) RMSE

RMSE is the square root of MSE, which provides a metric in the same unit as the observed value, which more intuitively reflects the deviation of the model predicted value. The smaller the RMSE is, the closer the predicted value of the model is to the true value. The formula is expressed as follows:

$$RMSE = \sqrt{MSE} \quad (2)$$

(3) R^2

R^2 is a statistical measure of the goodness of fit of the model to the data, which has values between 0 and 1. The closer R^2 is to 1, the closer the variation explained by the model is to the observed value, the better the effect on the goodness of model fit [8]. The formula is:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (3)$$

Among them, SS_{res} is the sum of squared residuals, i.e

$$SS_{res} = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (4)$$

SS_{tot} is the total sum of squares, i.e

$$SS_{tot} = \sum_{i=1}^n (Y_i - \bar{Y}_i)^2 \quad (5)$$

\bar{Y} is the average value of Y . The coefficient of determination reflects the percentage of variation of the dependent variable in the data, which can be explained by the regression relationship predicted by the independent variable.

When evaluating the model, the performance of these three indicators is considered by the combination of the model. MSE and RMSE were used to visually represent the magnitude of the error, and R2 was used to demonstrate the ability of the model to fit the data. The combination of multiple indicators can provide a more comprehensive assessment rather than simply relying on a single index.

3. Model building and solution

3.1. Exploration on the influencing factors of new energy electric vehicles

3.1.1 Data preprocessing

First of all, the monthly data set related to new energy vehicles collected from September 2019 to July 2024 was preprocessed, and the inspection found that no missing data and duplicate data existed. Subsequently, the seven characteristics of new energy vehicle output, the configuration of public charging pile, new energy vehicle subsidy policy, electricity price, new energy battery cost price, carbon emissions and public environmental protection awareness are drawn to see whether there are outlier values. The boxplot is shown in Figure 1. It can be seen that only the three characteristics of subsidy policy, electricity price and public environmental protection awareness have corresponding outlier points. Based on the specific actual analysis, considering that the above three characteristics have corresponding fluctuations under actual circumstances, the outlier points in these three characteristics are not considered as outliers.

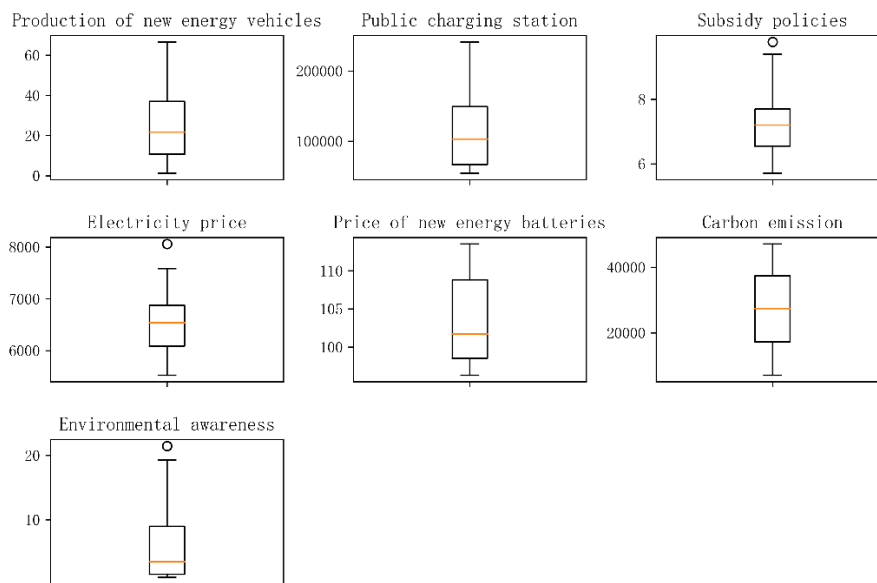


Figure 1 Boxplots of each feature

At the same time, it can be found that there is a large order of magnitude difference between the configuration of public charging piles and the new energy vehicle subsidy policy and the public's environmental protection awareness. Considering that the magnitude and magnitude difference of different characteristics may affect the accuracy of the model, the maximum and minimum value of the data is normalized. The formula is as follows:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (6)$$

Among them, X is the original data value, X_{min} is the minimum value in the dataset, and X_{max} is the maximum value in the dataset. X_{norm} is the normalized new value.

The maximum and minimum normalization method preserves the relative relationship between the original values in the data^[9], and linearly maps the values of the data points to the [0,1] interval, which ensures that all the data points maintain the order between each other before and after scaling.

3.1.2 Establishment and solution of the grey association analysis model

Unable to quantify the new energy vehicle industry index, the output of new energy vehicles is used to describe the development of the new energy vehicle industry. And in order to explore the configuration of public charging pile, new energy vehicle subsidy policy, electricity price, new energy battery cost, carbon emissions and public environmental protection consciousness six indicators and the influence of the relationship between new energy vehicle production, and considering the data is small, not easy to explore the statistical law between them, so the method of grey correlation analysis, reveals the dynamic correlation between the index characteristics and degree.

(1) Child and mother sequences were determined

The mother sequence, also known as the reference sequence, can reflect the behavior characteristics of the system. The output of new energy vehicles is taken as the mother sequence and recorded as:

$$x_0 = \{x_0(k) | k = 1, 2, \dots, n\} \quad (7)$$

The subsequence, also known as the comparative sequence, is composed of factors affecting the system behavior. Six indicators include the configuration of public charging piles, the subsidy policy for new energy vehicles, electricity price, the cost price of new energy batteries, carbon emissions and public environmental protection awareness as subsequences and are recorded as:

$$x_i = \{x_i(k) | k = 1, 2, \dots, n\} \quad (8)$$

Among them, $i = 1, 2, \dots, m$.

(2) Gray correlation degree was calculated

The formula for calculating the grey correlation is as follows:

$$\gamma(x_0(k), x_i(k)) = \frac{a + \rho b}{|x_0(k) - x_i(k)| + \rho b} \quad (9)$$

Among them, $\gamma(x_0(k), x_i(k))$ represents the correlation degree between the k-th data in the i-th subsequence and the mother sequence. a is the minimum difference between the two levels, and b is the maximum difference between the two levels, which are the maximum and minimum values of $|x_0(k) - x_i(k)|$, respectively; ρ is the resolution coefficient, taken as 0.5 here.

Since each indicator of each piece of data has a correlation degree, the information appears too scattered to compare, so we define:

$$\gamma(x_0, x_i) = \frac{1}{n} \sum_{k=1}^n \gamma(x_0(k), x_i(k)) \quad (10)$$

The grey correlation degree between sequence x_i and reference sequence x_0 .

Draw gray, gray correlation plot as shown in Figure 2:

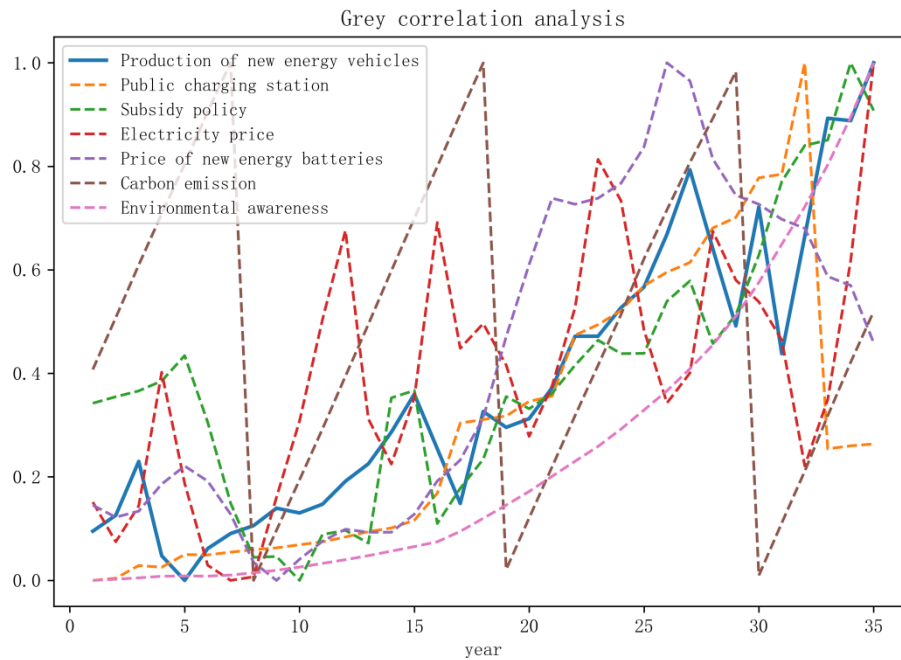


Figure 2 Gray correlation line plot

3.1.3 Factors of new energy electric vehicles

Through the above model building steps, the grey correlation between the configuration of public charging piles, new energy vehicle subsidy policy, electricity price, new energy battery cost price, carbon emission and public environmental protection awareness and the output of new energy vehicles is solved, as shown in Table 1:

Table 1 Gray correlation between the output of new energy vehicles

designation	Number of public charging piles	subsidy policy	electrovalency
Grey correlation degree	0.81	0.80	0.76
designation	Battery price	carbon emission	environmental awareness
Grey correlation degree	0.75	0.62	0.77

According to the data in Table 1, it can be seen that there is a significant positive relationship between the output of new energy vehicles and the number of public charging piles and the government subsidy policies. Specifically, the increase in the number of charging piles may reduce consumers' concerns about the convenience of daily charging, which will help improve the attractiveness of new energy vehicles, which in turn stimulate demand and production. At the same time, the government's subsidy policy, providing additional incentives for consumers by reducing the cost of cars, which may also have a positive impact on the sales and production of new energy vehicles. These factors suggest that it is critical to continue to invest in charging infrastructure and implement subsidies to attract consumers to promote the development of the neV market.

3.2. Development forecast of China's new energy electric vehicles

3.2.1 data preprocessing

In the above study, it can be learned that the number of public charging piles and subsidy policies are the main factors affecting the new energy electric vehicle industry. For the research on the development of new energy electric vehicles in China, the public charging pile data released from 2016 to 2022 and the subsidy policy data from 2010 to 2022 are used, and the development of new

energy electric vehicles in China in the next decade is predicted. Similarly, the data was cleaned first, and no missing values, duplicate values, and outliers were found in the data.

Further, the relevant data is visualized. There are three different charging piles in the public charging piles, and the line chart is used to observe the changing trend, as shown in Figure 3. It can be seen that both DC charging piles and AC charging piles show an increasing trend, while AC and DC integrated charging piles show a downward trend, and the number of AC charging piles is greater than the number of DC charging piles.

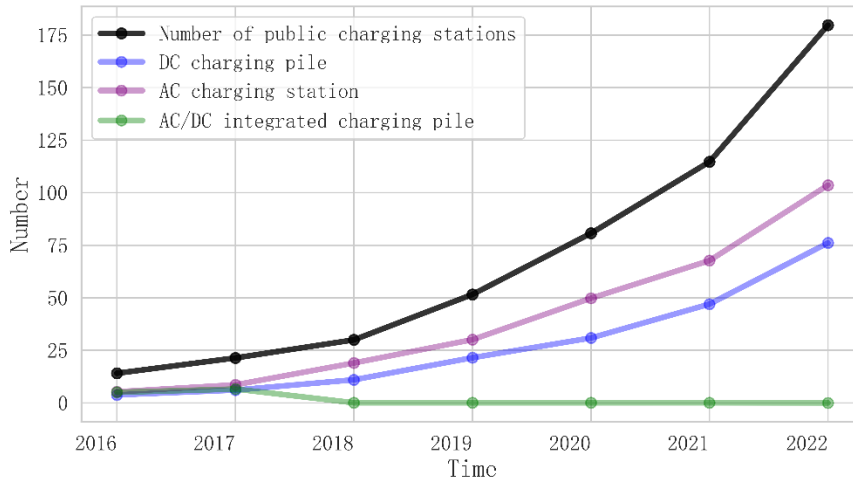


Figure 3 Changes in the number of public charging piles

Continue to visualize the subsidy policy. Since the subsidy policy is not easy to quantify, the sales volume of new energy vehicles is selected to indirectly reflect the impact of the subsidy policy on the new energy electric vehicles, as shown in Figure 4. It can be seen that the production and sales of new energy vehicles have shown an upward trend, and in recent years, due to the impact of policies, the sales of new energy vehicles have undergone great changes.

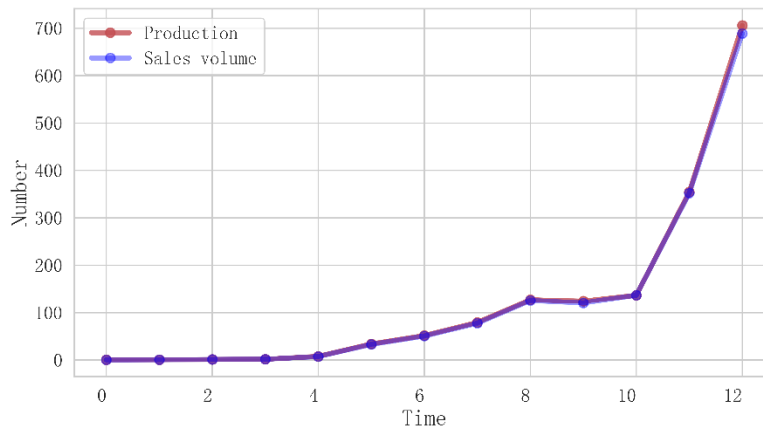


Figure 4 Changes in the output and sales volume of new energy electric vehicles

3.2.2 Establishment and solution of the ARIMA model

(1) Smooth test

Due to the small amount of data, the ARIMA model was used for the prediction. First, the ADF unit root stationarity test was used to test whether the data are stable^[10], which is the premise of ARIMA prediction. First, perform hypothesis testing:

H0 is an alternative hypothesis: The sequence is unstable, i. e. there exists a unit root.

H1 is an alternative hypothesis: the sequence is stable and there is no unit root

The test statistic, the ADF, was subsequently constructed. The test statistic was defined as the unit root test statistic of the autoregressive model with a lag term. In general, the ADF tests will consider choosing different lag terms to explore the best model fit.

The ADF test statistics were then calculated. Where the cut-off value was determined based on the selected sample size and confidence level. This can be determined by viewing the cut-off table for the ADF test statistic. Finally, the calculated ADF test statistic is compared to the cutoff value at the corresponding confidence level. Discovering that the test statistic is greater than the critical value, rejecting the null hypothesis H_0 that the time series is stable. Similarly, to eliminate the effect of dimension on model accuracy, the data were normalized according to Equation (1.1).

(2) model building

The mathematical expressions for the AR and MA models are as follows:

$$AR: Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + \xi_t \tag{11}$$

$$MA: Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \tag{12}$$

When $d = 0$, the ARIMA model can be seen as a direct combination of the AR model and the MA model. Generally speaking, the formula of the ARIMA model can be expressed as:

$$Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \tag{13}$$

Among them, Y_t represents time series data. φ_1 to φ_p are parameters of the AR model, used to describe the relationship between the current value and past time points. θ_1 to θ_q are parameters of the MA model, used to describe the relationship between the current value and the error at past q time points. ε_t is the error term at time t .

The residual analysis of the established ARIMA model was conducted by using the autocorrelation graph (ACF) and partial autocorrelation graph (PACF), taking the public charging pile as an example, as shown in Figure 4. It can be seen that the fitting degree of the model is relatively good, and then the model is used to predict the number of public charging piles and subsidy policies for new energy electric vehicles in China in the next 10 years.

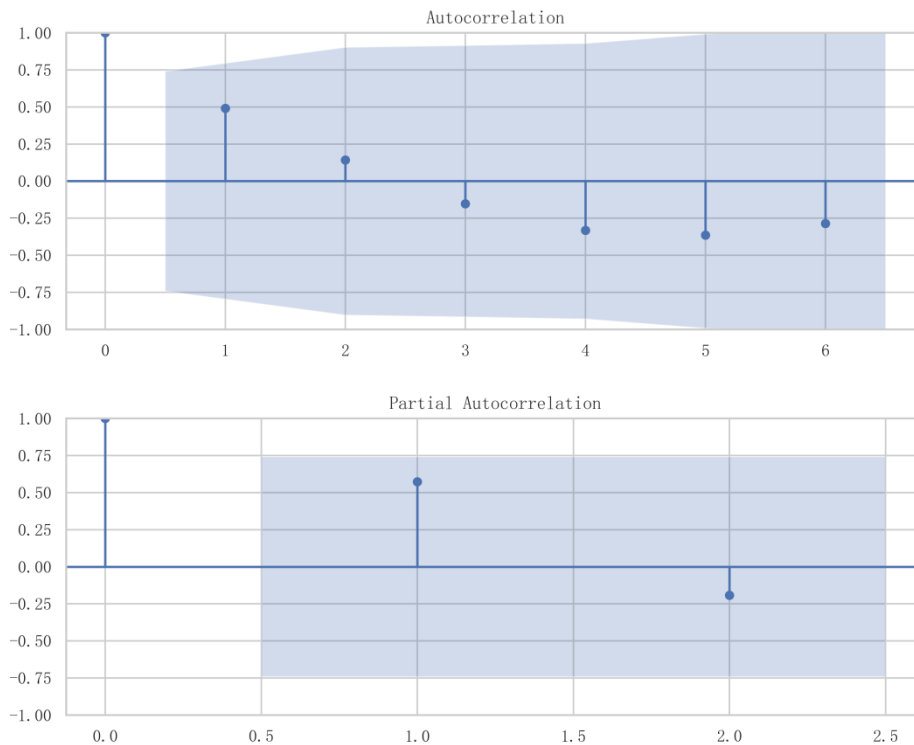


Figure 4 Residual diagram of the public charging pile

(3) China's new energy electric vehicles development forecast in the next decade

Taking the number of public charging piles for new energy electric vehicles in China in the next 10 years as an example, the forecast trend is shown in Figure 5:

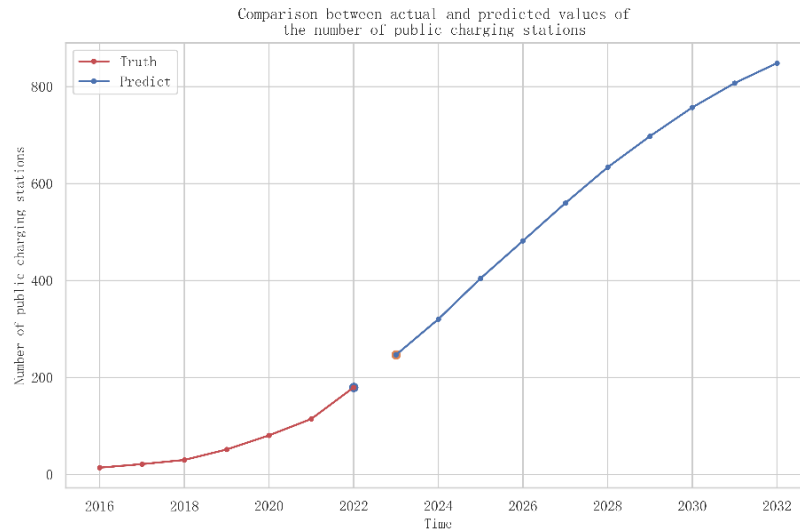


Figure 5. Trend prediction of public charging piles

The relevant prediction results are shown in Table 2. It can be seen that the number of public charging piles is increasing year by year, which represents the gradual development of China's new energy vehicle industry.

Table 2. Number of public charging piles in the next ten years

time	Number of public charging piles	time	Number of public charging piles
2023	246.80	2028	633.79
2024	320.46	2029	697.95
2025	404.82	2030	757.35
2026	482.06	2031	807.49
2027	559.92	2032	848.76

The models were evaluated, as shown in Table Table 3. It can be seen that the prediction effect of this model is relatively good, but the large error is mainly due to the small amount of data collected.

Table 3. ARIMA model evaluation

type	RMSE	MSE	R2
Public charging pile	420	176725	-56
sales volume	1577	2489279	-7194

3.2.3 Development forecast of China's new energy vehicles in the next ten years

In general, for the development of new energy vehicles in the next decade, the development environment of new energy vehicles is gradually showing a good development trend, its demand is increasing, sales are growing rapidly, and the development prospect is bright.

3.3. Impact of new energy vehicle industry on traditional energy vehicle industry

3.3.1 Data visualization

In the study on the impact of the new energy vehicle industry on the traditional energy vehicle industry, the vehicle production data of China, the United States, Japan, Germany, South Korea and India from 2010 to 2021 and the new energy vehicle stock data of China, the European Union and the United States from 2013 to 2021 are used. First, the vehicle production and new energy vehicle inventory in China, USA, Japan, Germany, Korea, and India were visualized in 2010-2021, as shown in Figure 6 and Figure 7. It can be seen that China's automobile production is far higher than that of other five countries, and after 2016, China's new energy vehicles have developed rapidly, far more than other countries in the world. The global production of new energy vehicles has increased year by year.

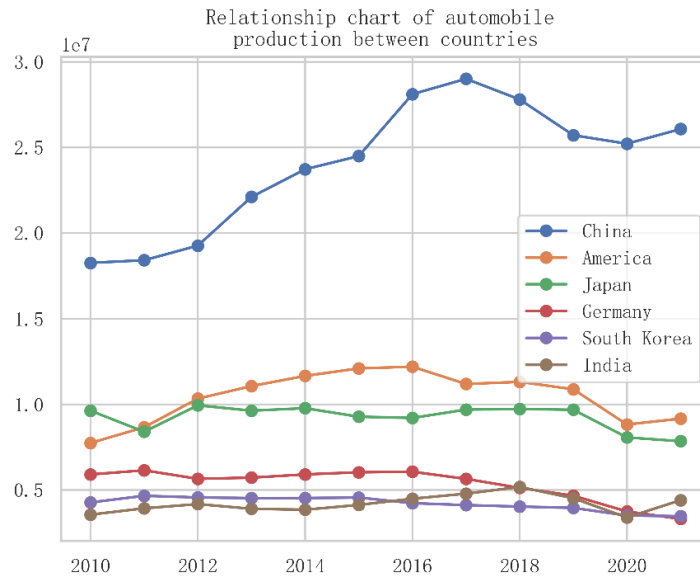


Figure 6. Production output of new energy vehicles in various countries

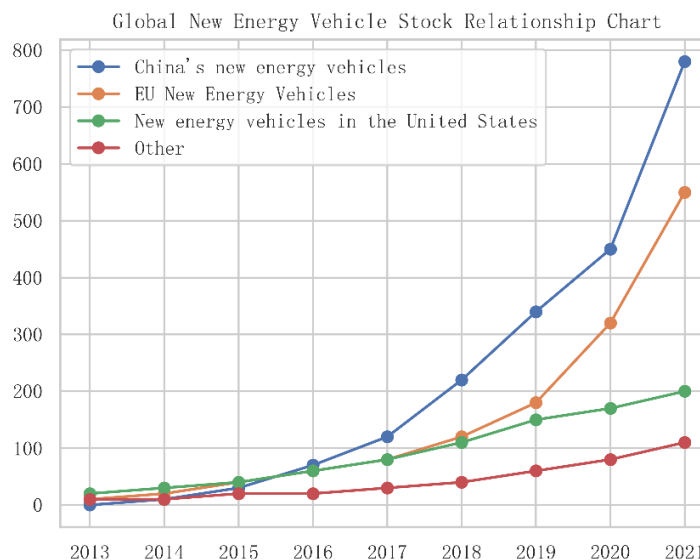


Figure 7. Line chart of new energy vehicle stock in various countries

3.3.2 Establishment and solution of the multiple linear regression model

According to the collected data, car sales in China and the United States can well represent the changing trends in global car sales. Therefore, representing China and the United States, we will analyze the impact of new energy vehicles on the global traditional energy vehicle industry. Since the production data of new energy vehicles is from 2013 to 2021, according to the formula:

$$\text{Traditional energy vehicles sales} = \text{carsales} - \text{new energy vehicles sales}$$

Calculate the annual sales volume of traditional energy vehicles in 2013-2021. Here, it can be approximate that the sales of new energy vehicles are equal to the output of new energy vehicles. Draw a line chart to see the sales changes of new energy vehicles and traditional energy vehicles, as shown in Figure 8:

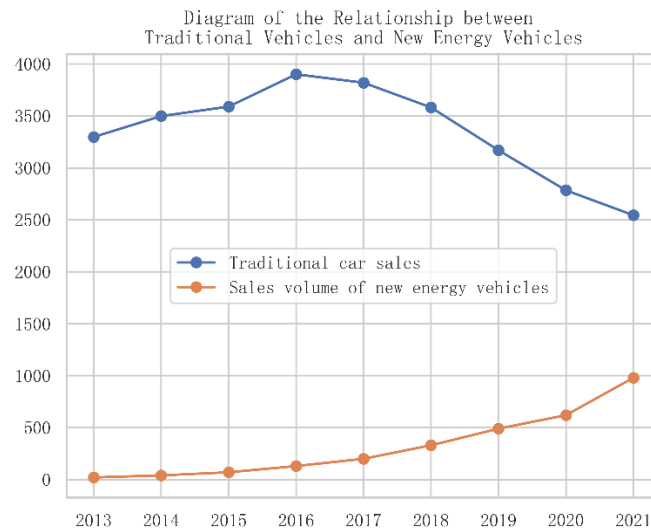


Figure 8. Changes in the sales volume of new energy vehicles and traditional energy vehicles

As can be seen from Figure 8, since 2017, the output of traditional energy vehicles has decreased year by year, and the output of new energy vehicles has increased year by year. In order to quantify the impact of new energy electric vehicles on the global traditional energy vehicle industry, the multiple linear regression model is selected to solve it, and the establishment process is as follows:

Let x_1 and x_2 be the independent variables of new energy vehicle sales and years, and y be the dependent variable of traditional energy vehicle sales. The linear relationship is as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \mu_i \tag{14}$$

$\beta_1 \beta_2 \beta_3$ is the regression coefficient, and μ_i is observed and satisfies certain conditions to obtain the predicted value:

The multiple linear regression model established by solving the above process shows that β_1 is -2.6505, β_2 is 187.1596, and the significance levels of x_1 and x_2 are both less than 0.05. It can be seen that there is a negative correlation between the sales volume of traditional energy vehicles and the sales volume of new energy vehicles, and there is a positive correlation between the sales volume of traditional energy vehicles and the year, which strongly indicates that the sales volume of traditional energy vehicles is affected by the sales volume of new energy vehicles, which can be summarized as "loss to profit". It is inevitable that the traditional energy automobile industry will be impacted, which is also in line with the changing law of alternative commodities in economics.

4. Conclusions

The rise of the new energy vehicle industry is affecting the traditional energy vehicle market with an obvious negative correlation. The data show that the new charging facilities and government subsidies have significantly promoted the production and sales growth of new energy vehicles, while the negative correlation with the sales of traditional vehicles, revealing that the market demand is changing. As consumers become more concerned about environmental protection and operating costs, this trend indicates that the new energy sector will continue to contain huge growth potential, posing a challenge to traditional automakers. In the future, long-term plans must be made to expand the electric vehicle charging infrastructure and maintain incentive policies to promote the transformation of green transportation.

Acknowledgments

This work was supported by the Natural Science Foundation for Young Scientists of Gansu **Province** (Grant No. 22JR5RA595), the Youth Project of Social Science Planning of Gansu **Province** (Grant No. 2021QN027), and the Scientific Research and Innovation Fund Project of Gansu University of Chinese Medicine (Grant No. 2021KCYB-10).

References

- [1] Che Shuai, Shi Yuzheng the current situation and suggestions for the development of China's new energy vehicle industry [J] *Automotive Electrical Appliances*, 2022 (10): 16-19,24
- [2] Pang Bo Research on the Agglomeration Development of China's New Energy Vehicle Industry [D] Yunnan University, 2020
- [3] Liu Lanling. A Support Vector Machine Combination Model Based on Grey Relational Analysis [D]. Liaoning Normal University, 2020
- [4] Xiong Jingjing Research on the Application of ARIMA Model and AR-GARCH Model [D] Central China Normal University, 2018
- [5] Ren Zeping, Lian Yishi, Guo Shuangtao Charging Station: New Infrastructure, Moving towards the Era of New Energy Vehicles Development Research, 2020, (10): 22-30.
- [6] Li Yongxing The construction of charging networks for electric vehicles in Europe and America is accelerating *Automotive Horizontals*, 2021, (06): 96-98.
- [7] He Xiuli Multiple linear models and ridge regression analysis [D] Huazhong University of Science and Technology, 2005
- [8] Gelman A, Goodrich B, Gabry J, et al. R-squared for Bayesian Regression Models [J]. *The American Statistician*, 2019, 73 (3): 307-309.
- [9] Yang Hanyu, Zhao Xiaoyong, Wang Lei Overview of Data Normalization Methods [J] *Computer Engineering and Applications*, 2023, 59 (03): 13-22
- [10] Li K, Li G, Zhang Y. Study on Rainfall Distribution in Yunnan Province Based on ARIMA Model by Big Data Computation [J]. *Academic Journal of Computing & Information Science*, 2023, 6 (1):