New Energy Vehicle Development Influencing Factors and Trend Forecast Exploration

Kelong Hu [1,*], Baoli Li [2]

[1] School of Computer Science, Guiyang University, Guiyang, China, 550005

[2] School of Culture and Media, Guiyang University, Guiyang, China, 550005

* Corresponding Author Email: hklzyhs@163.com

Abstract. The rapid development of new energy electric vehicles (EVs) stems from the urgent need to solve the problems of environmental pollution and energy consumption of traditional fuel vehicles. In this study, four main factors affecting the development of new energy electric vehicles in China are selected from the official open source website: the number of charging piles, energy density, vehicle sales and market share. Through multiple linear regression analysis, Pearson's correlation coefficient and scatter plot analysis, it is found that the number of charging piles and energy density are the main influencing factors. In addition, the ARIMA algorithm is applied to predict the development trend of new energy electric vehicles in China in the next ten years. Finally, the negative correlation between new energy electric vehicles and traditional energy vehicles is revealed by comparing the development data of global new energy and traditional energy vehicles from 2011 to 2023, combined with scatter plot analysis.

Keywords: New Energy Electric Vehicles, Multiple Linear Regression Analysis, Pearson Correlation Coefficient, Time Series Forecasting ARIMA Algorithm.

1. Introduction

Since 2011, China's new energy vehicles have become another symbol of "China's high speed rail" due to their low pollution, low energy consumption, significant effect in regulating peak electricity consumption and relying on the government's active promotion and numerous preferential policies to make the new energy vehicle industry develop at a high speed, and gradually become the main driving force of the vehicles, which are driven by the new energy technologies such as batteries, fuel cells and hybrids. The use of new energy technologies, such as batteries, fuel cells, hybrids, etc., as its main driving force of the car. New energy vehicles are promoted to solve the environmental pollution and energy consumption problems of traditional fuel vehicles [1-3]. There are four main types: hybrid, pure electric, fuel cell and other new energy vehicles.

Wei [4] suggests that the factors affecting the development of China's new energy vehicle industry are national policy orientation, key technology, infrastructure facilities, and comprehensive cost. Thus put forward to reduce the new energy vehicle production and use of cost, improve the supporting industry construction and after-sales protection mechanism is conducive to the development of China's new energy automobile industry countermeasures and recommendations; Tian Gen [5] and other use of cloud computing technology service model fits the network environment of convenience, high efficiency guidelines, in the realization of a set of high modularity, high intelligence, high security in the integration of the whole scene of the power battery manufacturing from the multi-dimensional summary Intelligence to the new energy vehicle power battery manufacturing benefits.

Scholars for the development of China's new energy vehicle industry influencing factors from the multi-dimensional, substantive theoretical summary, for the development of China's new energy vehicle industry is of great significance, based on this this paper from the official open source website to select a comprehensive and representative of the four main factors to study the development of new energy electric vehicles in China, reflecting the authority of the source of the research data and scientific, so that the study is more credible. Multiple linear regression analysis, Pearson's correlation coefficient and scatterplot are used to determine the charging pile and energy density may be the main influencing factors, and this multi-method approach can more accurately and comprehensively reveal the relationship between the factors and the degree of influence. Using the time series forecasting model ARIMA algorithm to forecast the development trend in the next ten years, it provides a forward-looking reference for the development of the industry, which is of great significance as a guide. By analysing global data over a long period of time (2011-2023) to explore the interactions between new energy electric vehicles and traditional energy vehicles, and presenting the negative correlation between the two clearly and intuitively through scatter plots, the conclusions are more convincing and visualised.

2. Model construction

2.1. Data sources

The data in this paper comes from the following open source web site, see Table 1 for details.

Table 1. Presentation of data sources

Data	Web address
Data1	https://www.stats.gov.cn/sj/
Data2	http://www.caam.org.cn/ddzn

2.2. Multivariate linear regression analysis and Pearson's model

Firstly, the assumption is carried out that there is some kind of linear relationship between new energy vehicles and several main influencing factors, and each factor influences and constrains each other, in order to verify the truth of this assumption, this paper establishes a multivariate linear regression model [6].

Assuming that the level of development of new energy electric vehicles in China is Z, and each of the factors such as state-owned policies, economy, new energy vehicle technology, social factors, and basic security facilities are $X_1, X_2, X_3, X_4, X_5, \beta, \alpha$ respectively, there are:

$$Z = \partial_0 + \partial_1 X_1 + \partial_2 X_2 + \partial_3 X_3 + \partial_4 X_4 + \partial_5 X_5 + \beta \tag{1}$$

Since 2011, the active introduction of state-owned policies, the continuous development of the economy, the continuous iteration and updating of new energy vehicle technology, the social demand for low-energy, low-emission vehicles, and the strengthening of the relevant security can have a positive impact on the development of China's new energy vehicles, and we can assume that the regression coefficients of Z are all positive.

A scatter plot of the correlation of open source data between the development of new energy vehicles in China and a number of major influencing factors is plotted, as shown in Figure 1.

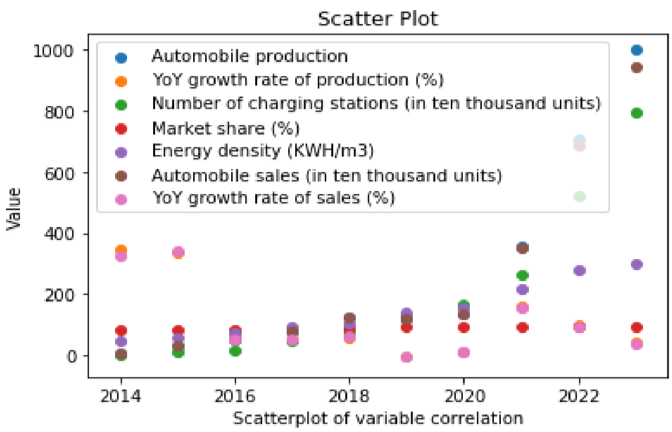

Figure 1. Data scatter plot

A data table of correlation coefficients between the development of new energy vehicles in China and a number of major influencing factors is plotted, as shown in Figure 2.

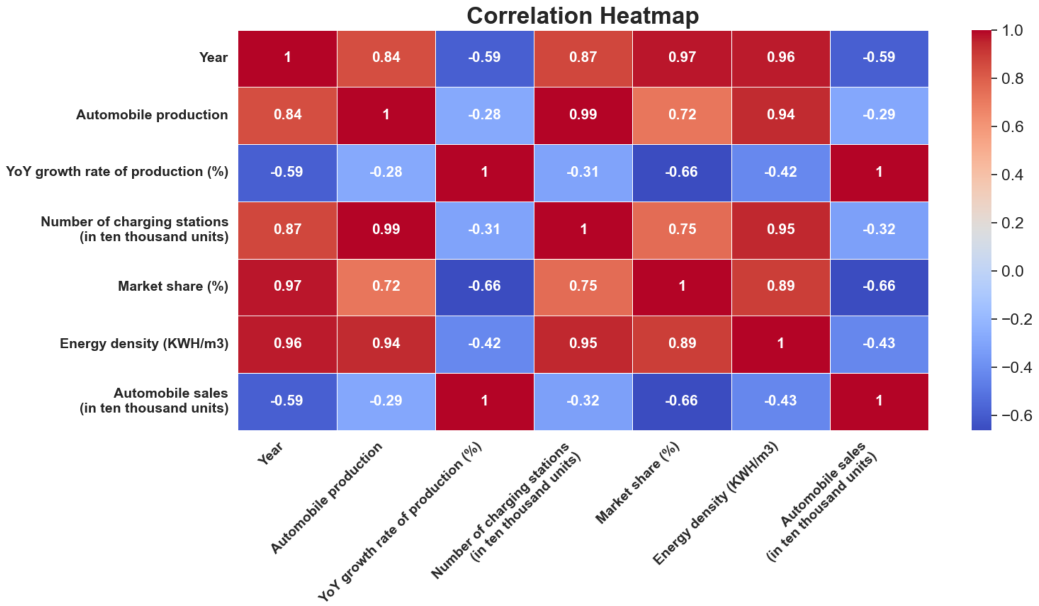

Figure 2. Data table of correlation coefficients of factors

Based on the correlation coefficient data table in Fig. 2, a correlation heat map is generated, as shown in Fig. 3. The next step of Pearson's analysis is then carried out to obtain the corresponding plot, as shown in Figure 4 [8].

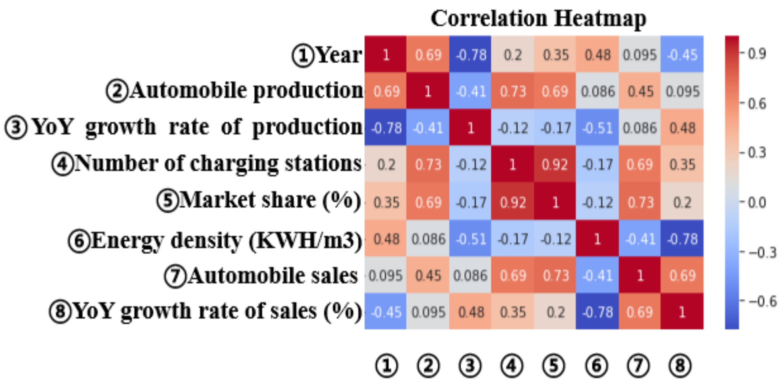

Figure 3. Heat map of Pearson's coefficient

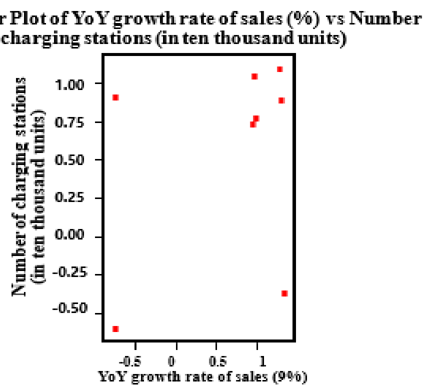

Figure 4. Correlation data Pearson analysis main factors chart

By analysing Figures 3 and 4, four of the indicators with stronger correlation were extracted using Pearson analysis, and the conclusion can be drawn from this figure. Charging piles (10,000), energy density (%), production (10,000), sales (10,000) and other factors that mainly affect the development of China's new energy vehicle industry [7].

Through the official website open source data of China's new energy electric vehicle industry development data, due to the limited nature of the data, so consider using the classic ARIMA algorithm of the time series prediction model, Pearson analysis of these data after the correlation of the three indicators are stronger through the corresponding code prediction, prediction assessment for evaluation after [8]. The corresponding score is given to reflect the general trend of the development of the next 10 years with the data integration of the previous twelve years. After the data integration, but also with the sales of new energy vehicles, ownership, energy density of the three indicators of the impact of new energy vehicles to produce relevant line graphs, as shown in Figure 5, Figure 6.

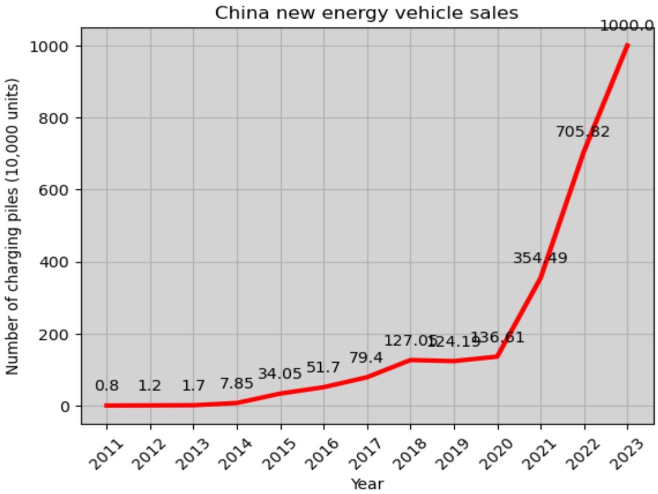

Figure 5. China's new energy vehicle sales

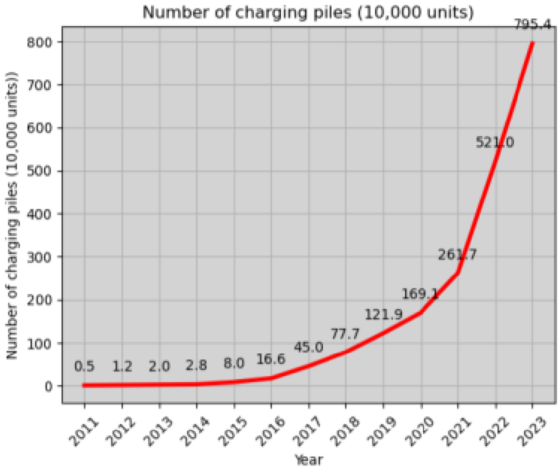

Figure 6. Number of charging piles (10,000 units)

Time series prediction model ARIMA algorithm:

As one of the classical algorithms for studying changes over time, in classical time series models, data correlation is generally strong, and time series analysis helps this paper to go for a matching model based on such correlation. It will also be due to the conversion of unstable time series to stable time series. Model setting, data testing and its diagnosis to complete the analysis of the model. The common time series models are divided into the following four major types: autoregressive (AR) model, moving average (MA) model, autoregressive moving average (ARMA) model and differential autoregressive moving average (ARIMA) [9-10]. Through the official website open-source data showing the global new energy electric vehicles and traditional energy vehicles development data from 2011-2023, the sales of new energy vehicles from 2011-2023 can be plotted as Figure 7.

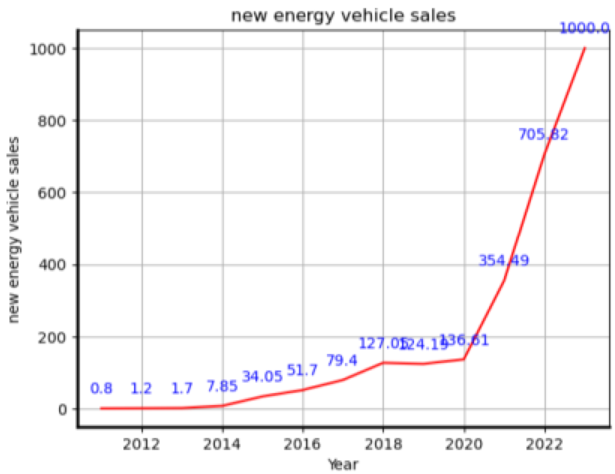

Figure 7. New energy vehicle sales, 2011-2023

ARIMA algorithm time series smoothness analysis:

In order to meet the ARIMA modelling conditions, it is necessary to preprocess the data of the corresponding time series of China's new energy vehicles from 2011 to 2023 to verify the smoothness of the series, which can be clearly seen in Figures 8, 9 and 10, which show that the sales volume of China's new energy vehicles shows a relatively obvious trend of incremental changes in each year.

ARIMA algorithm model is called autoregressive moving average model, expressed as a kind of time series analysis, the main idea is: the forecast object with the advancement of time after the formation of the data series can be regarded as a random sequence, with a specific model to fit, through the historical value of the time series to judge the future trend. The ARIMA autoregressive moving average model has three covariates: the autoregressive analysis order p, the number of differences d and the moving average component order q. The previous convention is to use ARIMA (p, d, q). Its expression is satisfied:

$$\omega_t = \lambda_1 Z_{t-1} + \lambda_2 Z_{t-2} + \lambda_3 Z_{t-3} + \cdots + \lambda_i Z_{t-i} \tag{2}$$

Where λ_t is the regression parameter, t=1, 2, 3, ..., p is the number of lagged variables, and ω_t is the white noise process, then the observation of linear data is the p-order autoregressive model can be expressed as $AR_{(p)}$. The white noise test is expressed in terms of the lag operator as:

$$\varphi_t = \mathrm{A}(L)Z_t = (1 - \lambda_1 L - \lambda_2 L^2 - \cdots - \lambda_P L^{\mathrm{p}})\omega_t \tag{3}$$

Where A(L) is the autoregressive operator. The autoregressive operator expression is:

$$\mathrm{A}(L) = (1 - G_1^{-1}L)(1 - G_2^{-1}L)\cdots(1 - G_{\mathrm{p}}^{-1}L) \tag{4}$$

$G_1^{-1}, G_2^{-1}, \cdots, G_P^{-1}$ is each characteristic root of the autoregressive equation, and the AR model is smooth at stage p when it satisfies A(L)=0. If the observation ω_t satisfies:

$$\omega_t = \varphi_t + \mu_t \varphi_{t-1} + \mu_2 \varphi_{t-2} + \mu_t \varphi_{t-3} + \cdots + \mu_q \varphi_{t-q} \tag{5}$$

Here, $\mu_1, \mu_2, \mu_3, \cdots, \mu_q$ is the parameter of the equation and φ_{t-q} is the corresponding white noise at time t-q. The observation ω_t is the moving average model of order q. R is denoted as MA(q), and the autoregressive arithmetic is deformed to give:

$$\omega_t = \mu(L)\varphi_t = (1 + \mu_1 L + \mu_2 L^2 + \cdots + \mu_p L^p)\varphi_t \tag{6}$$

$\mu(L)$ is the moving average operator and its characteristic equation can be expressed as:

$$\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q \tag{7}$$

The characteristic expression for the moving average operator is:

$$\theta(L) = (1 - M_1^{-1} L)(1 - M_2^{-1} L) \cdots (1 - M_p^{-1} L) \tag{8}$$

$M_1^{-1}, M_2^{-1}, \cdots, M_P^{-1}$ is the characteristic root of the moving average characteristic equation, and from here, the data of the observations under study:

$$\omega_t = \theta(L)^{-1} \varphi_t = \left(\frac{R_1}{1 - H_1 L} + \frac{R_2}{1 - H_2 L} + \cdots + \frac{R_q}{1 - H_q L} \right) \varphi_t \tag{9}$$

$R_1, R_2, \cdots, R_q$ are constants in Eq.

The MA model is invertible to order q if the characteristic equations satisfy $\theta(L) = 0$. It is possible to combine the ARMA and MA models into a structure, after which the expression is expressed as:

$$\omega_t = \lambda_1 Z_{t-1} + \lambda_2 Z_{t-2} + \lambda_3 Z_{t-3} + \cdots + \lambda_i Z_{t-i} + \varphi_t + \mu_1 \varphi_{t-1} + \mu_2 \varphi_{t-2} + \mu_3 \varphi_{t-3} + \cdots + \mu_q \varphi_{t-q} \tag{10}$$

An additional deformation of the ARMA algorithm is obtained synthetically as:

$$A(L)\omega_t = A(L)\varphi_t \tag{11}$$

If the obtained time series does not have smoothness, it is necessary to deal with differential anomalies of the non-smooth model, in this way, then the forecasting model can be roughly considered as an ARIMA model.

The ARIMA forecasting of the time series of China's new energy vehicle sales (10,000 units) will be taken as an example: Figures 8 and 9.

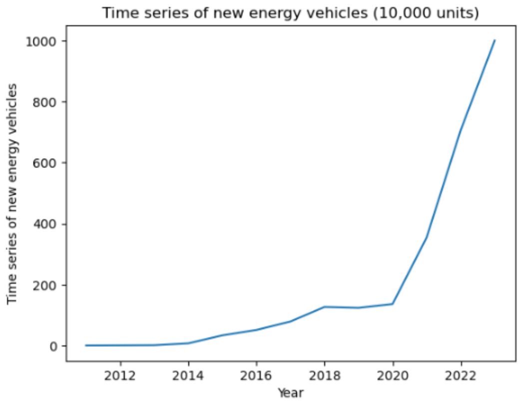

Figure 8. Time series of new energy vehicles (10,000 units)

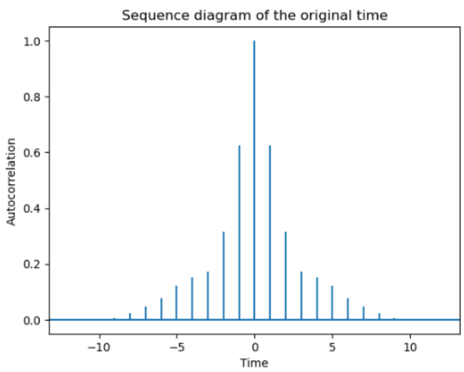

Figure 9. Sequence diagram of the original time

From the above graph, it can be learnt that the data is not very smooth and hence a differencing needs to be done to facilitate the execution of the later steps.

Test of p,q values of ARIMA autoregressive moving average model:

In this paper, we will directly determine its related BIC and AIC, using the method of traversal search to get the minimum value to fit the model parameters under the influence of different conditional factors, while the results obtained are shown in Fig. 10, Fig. 11:

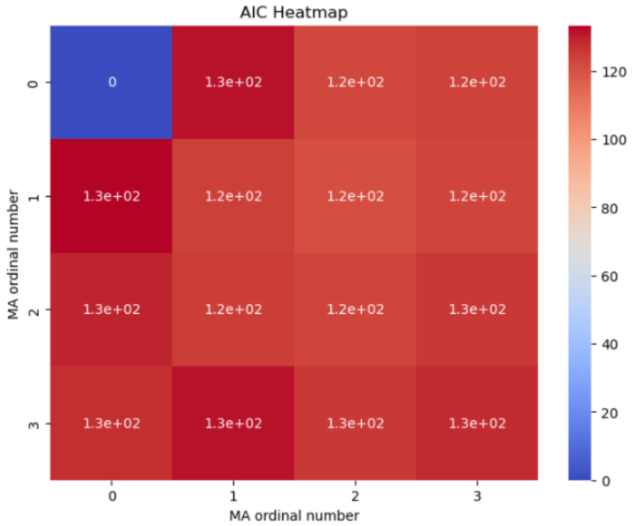

Figure 10. Time series of new energy vehicles (10,000 units)

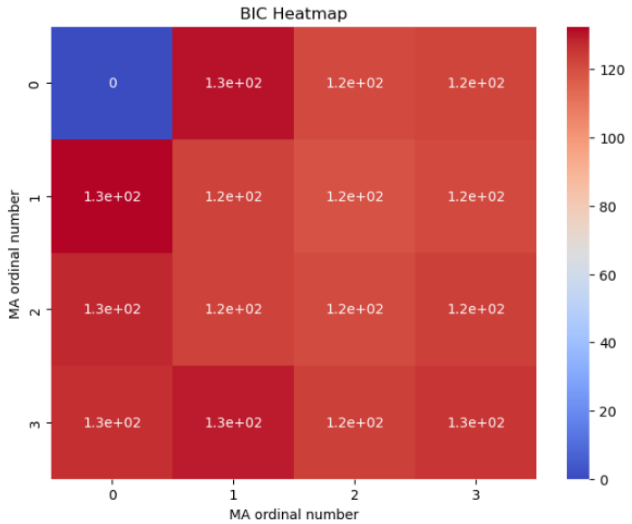

Figure 11. Time series of new energy vehicles (10,000 units)

It is clear from the graph that q=1, p=2, ARIMA (2, 1, 1) model is chosen to go for prediction. The prediction results are obtained as shown in Table 2.

Table 2. BIC quasi-thermal map ARIMA Autoregression Model Predictions

serial number	Development of new energy vehicles in the next 10 years	Future New Energy Vehicle Sales Forecast
1	2024	1124.280000
…	…	…
10	2033	1379.260000

This results in a projection for the next ten years as shown in Figure 12 and plots the general direction of its future trends:

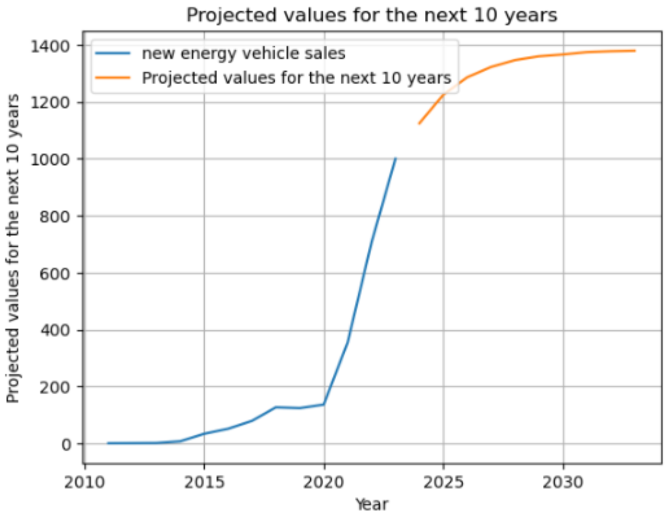

Figure 12. Projected values for the next 10 years

Analogous to the methodology above, two other indicators, predicted according to the same methodology, yielded the results shown in Table 3:

Table 3. New Energy Vehicle Sales Forecast for the Next 10 Years

Serial number	Year	New Energy Vehicle Sales Forecast	Projected car ownership	Projected maket share
1	2024	1124.280000	1676.200000	20.650000
…	…	…	…	…
3	2033	1379.260000	4368.950000	23.890000

As can be seen from the above chart, on the influence of factors larger new energy vehicle sales, car ownership are having a gradual increase, but the new energy vehicle market share in a certain range of level fluctuations.

3. Impact of new energy electric vehicles on the global traditional energy vehicle industry

3.1. Data sources

The data in this paper comes from the following open source web site, see Table 1 for details.

The global development data of new energy electric vehicles and traditional energy vehicles were collected through the open source data on the official website. The data are visualised and pre-processed, and the development trend of the data is analysed, and it can be seen that their growth rates are both showing an upward trend. Therefore, this paper adopts the study of their development growth rate to judge the impact of new energy electric vehicles on traditional energy vehicles. In addition to the collection of domestic relevant data, foreign data about the relevant indicators of traditional new energy vehicles are also collected, as shown in Table 4.

Table 4. Global Conventional and New Energy Vehicle Data Sheet

Serial number	Year	Global Automobile Sales (in 10,000 units)	Global Conventional Energy Vehicle Sales (in 10,000 units)	Global new Energy Vehicle Sales (in 10,000 units)	Global new Energy Vehicle Share (%)
1	2011	6348	6354	6	0.000945
…	…	…	…	…	…
3	2023	8707	9941	1234	0.141725

By visualising the relevant data in the above table, it is concluded that the development trends of new energy electric vehicles and traditional energy vehicles are non-linearly correlated with each other, as shown in Figure 13.

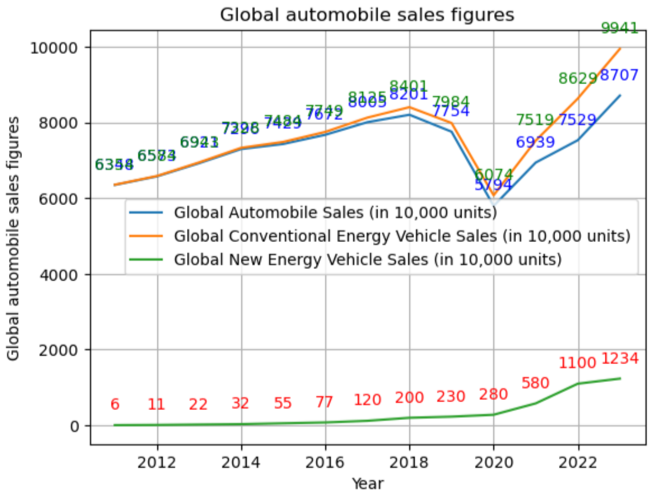

Figure 13. New Energy Vehicle Sales Comparison Chart

The global sales of new energy vehicles have been rising steadily, while the growth of traditional energy vehicles in recent years has not increased significantly, as can be seen from the table is in 2019-2020 affected by the epidemic when the sales of both clearly show a downward trend, followed by a clear upward trend. This paper treats the data after 2019 as an outlier [11]. In this paper, the ARIMA model is used to predict the data for the next four years after 2019, as shown in Table 5.

Table 5. Global Conventional and New Energy Vehicle Data Sheet

Serial number	Year	Global Conventional Energy Vehicle Sales (in 10,000 units)	Global new Energy Vehicle Sales (in 10,000 units)
1	2011	6354	6
…	…	…	…
13	2023	9700	1349

Plotting the scatterplot through the data in Table 5 yields the scatterplot shown in Figure 14:

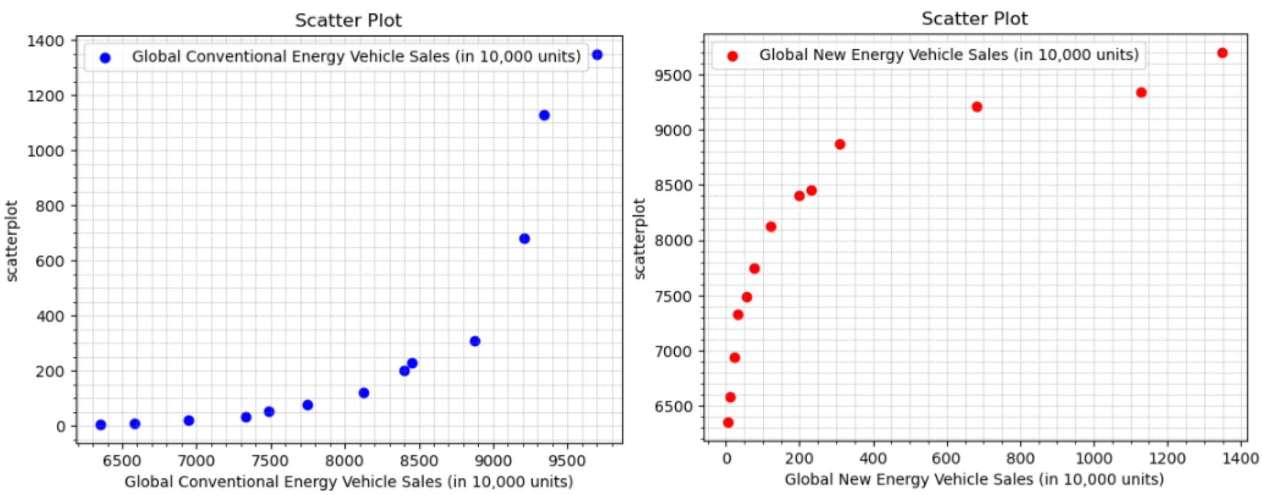

Figure 14. Global Conventional Energy Vehicle Sales (in 10,000 units)

Judging by the scatterplot that they conform to the characteristics of a hyperbola, assume that it exists with.

$$\hat{y} = \frac{x}{a + bx} \tag{12}$$

Taking the reciprocal $y' = \frac{x}{\hat{y}}$ on both sides yields the data shown in Fig. 15 when brought into the formula $y' = ax + b$ and calculated:

	y'	x
0	0.134000	4.500000
1	0.145000	5.810000
2	0.073000	10.800000
3	0.075000	45.100000
4	0.083000	100.300000
5	0.087000	209.300000
6	0.086000	256.900000
7	0.091000	320.800000
8	0.123000	589.800000
9	0.234000	673.200000
10	0.456000	679.300000
11	0.578000	761.400000
12	0.798000	1000.200000

Figure 15. Data on y and x

Linear regression calculations were performed to obtain a=0.0534, b=0.007 respectively, which were obtained by substituting the respective values of a, b into the hyperbolic formula:

$$\hat{y} = \frac{x}{0.0534 + 0.007x} \tag{13}$$

It is concluded that new energy electric vehicles show negative correlation with global traditional energy vehicles. Secondly, to analyse the link between them, the sales data of new energy electric vehicles as the data on the horizontal axis of the coordinates, and the sales data of traditional energy vehicles as the data on the vertical axis [11].

Through the visual analysis of the data to get the link between them, as shown in Figure 16, the analysis concluded that the sales of traditional energy vehicles with the growth of new energy vehicle sales growth of its development growth rate showed a flat trend. It shows that new energy electric vehicles have a certain impact on the development of traditional energy vehicles.

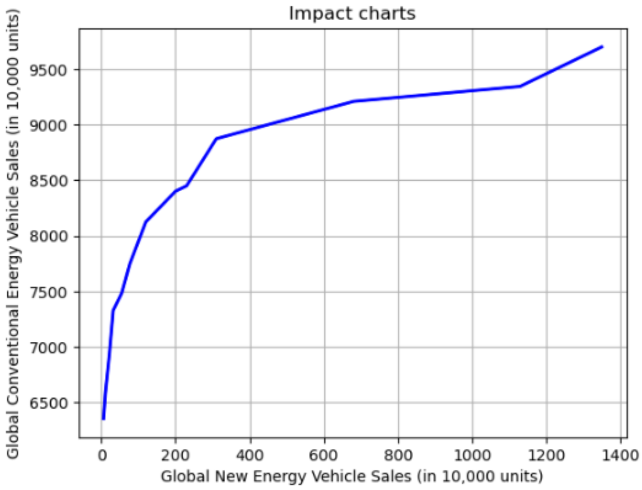

Figure 16. Impact of New Energy Electric Vehicles on Conventional Energy Vehicles

4. Conclusion

Based on the authoritative data from the official open source website, this study uses multiple linear regression analysis, Pearson correlation coefficient (the correlation coefficient is roughly close to 1) and scatterplot analysis to clarify the number of charging piles and energy density as the key influencing factors for the development of new energy electric vehicles in China. At the same time, a time series forecasting model was constructed through the ARIMA algorithm to predict the development trend of China's new energy EV industry in the next ten years, and the negative correlation between new energy EVs and traditional energy vehicles was revealed through linear regression analysis model fitting and scatter plot analysis.

When constructing the time series forecasting model, the ARIMA algorithm has a high sensitivity to outliers and missing values, so data cleaning and preprocessing are particularly important. In order to pursue more accurate prediction, the next step is to use the LSTM algorithm, which can not only deal with long series data effectively, but also circumvent the problem of disappearing gradient, and then provide more scientific and rigorous decision support for the rapid development of the new energy electric vehicle market.

References

[1] JIANG Jia-Chen, MEI Yuxin, PAN Yao-Yao, et al. Development assessment and sales forecast of new energy vehicles [J]. Journal of Taizhou College, 2024, 46 (03): 9-15.

[2] Wang Pin. Development Status and Countermeasures of New Energy Vehicles in China [J]. Automobile Practical Technology, 2024, 49 (08): 187-191.

[3] FU Ruoqi, XUE Jingyi, GUO Yuqi, et al. Development status and suggestions of charging pile facilities for new energy vehicles in China [J]. Times Automotive, 2024, (08): 130-132.

[4] WEI Guo, JIANG Hongmei, WEI Kedin, et al. Influencing factors and countermeasure suggestions for the development of China's new energy vehicle industry [J]. Equipment Manufacturing Technology, 2023, (12): 136-139.

[5] Tian Gen. Application and development of cloud computing technology in intelligent manufacturing of new energy vehicles [J]. Energy Storage Science and Technology, 2024, 13 (05): 1748-1750.

[6] ZHANG Mingming, ZHANG Mingdong. Multivariate linear model based on grey load prediction for predicting settlement of high-rise buildings [J]. Surveying and Spatial Geographic Information, 2024, 47 (05): 195-197+201.

[7] ZHOU Kun, XU Yunfei, QI Haowei. Short-term dynamic dispatch model of new energy generation based on improved ARIMA [J]. Computer and Information Technology, 2024, 32 (01): 56-61.

[8] DONG Lipeng, NIE Qinghao, SUN Xiaokun, et al. Analysis of the influence of shield tunneling parameters on surface settlement based on Pearson's correlation coefficient method [J]. Construction Technology (in Chinese and English), 2024, 53 (01): 116-123.

[9] CHEN Zhaomeng, LIU Yue. Threshold exploration of early warning signals in time series - based on variable point autoregressive model [C] // Chinese Psychological Association. Abstracts Collection of the 25th National Psychology Academic Conference - Symposium. Institute of Brain and Psychological Sciences, Sichuan Normal University; 2023: 3.

[10] Sun QT, Han YN, Liu Y, et al. Application of an autoregressive moving average (ARIMA) model to predict rat density trends in Shandong Province [J]. Chinese Journal of Vector Biology and Control, 2021, 32 (06): 744-748.

[11] Xia Zhibin, Jing Shi. Miao Wei: The trend of new energy vehicles replacing traditional fuel vehicles has been formed [N]. China Business News, 2023-12-11 (C08).