Research on Sales of New Energy Vehicles Based on ARIMA Model

Qiming Liu#, Qiyue Zhang*, #, Dashun Ma#

School Of Electrical Engineering, China University Of Mining And Technology, Xuzhou, China, 221116

* Corresponding Author Email: 18852181827@163.com

#These authors contributed equally.

Abstract. Firstly, this article identifies seven indicators based on electric vehicles: battery endurance, number of public charging facilities, government subsidies, gasoline prices, electricity prices, carbon emissions, and number of power battery companies. Spearman correlation analysis is conducted between indicator data and sales of new energy electric vehicles. The correlation coefficients of the first five indicators are all greater than 0.6, which is the main influencing factor. Then, the sales of new energy vehicles are predicted using a yearly forecasting method. Due to the close correlation between the sales of new energy vehicles and the five main indicators, the ARIMA time series model is first used to predict the annual quantity of the five indicators. Then, using these five indicators as inputs and sales as outputs, a BP neural network model is established to predict the annual sales of new energy vehicles in China to be 46051748 units in ten years. Finally, quantifying the score of ecological environment assessment using four evaluation indicators: sulfur dioxide, smoke emissions, ammonia nitrogen, and carbon dioxide. The entropy weight method was used to determine the weights of each indicator, and the results were 37.25%, 30.96%, 21.24%, and 10.55%, respectively. Then apply it to the TOPSIS model to obtain ecological environment scores for different years, and analyze the correlation between the degree of electrification and ecological environment scores. Finally, it can be concluded that the larger the market share of new energy vehicles, the better the ecological environment.

Keywords: Spearman Correlation Analysis, ARIMA, BP Neural Network, Entropy Weight TOPSIS Evaluation.

1. Introduction

New energy vehicles refer to the use of new on-board fuels as power sources or the use of conventional gasoline or diesel to build new energy systems, and integrate advanced technologies in power control and driving to produce new vehicle models[1][2]. The current types of new energy vehicles mainly include hybrid electric vehicles (such as oil and gas, oil electric hybrid, etc.), pure electric vehicles, and fuel cell electric vehicle[3]. At present, hybrid electric vehicles and pure electric vehicles account for the majority of the new energy vehicle market. As one of the new energy vehicles developed in recent years, fuel cell electric vehicles have not been widely used due to outdated production technology and difficult storage[4].

With the progress of society, China's requirements for automobile exhaust emissions are constantly increasing. New energy vehicles, as a new, environmentally friendly, and efficient industry, conform to the concept of green environmental protection and can effectively reduce pollution to the environment. In recent years, the technological level of new energy vehicles has been continuously improving, and the overall performance of vehicles has also been greatly improved. At the same time, with the full promotion of the country, the new energy vehicle industry has experienced rapid development and the market size continues to expand.(Data sources: National Bureau of Statistics, Ministry of Industry and Information Technology of China, China National Knowledge Infrastructure, China Association of Automobile Manufacturers, International Energy Agency, Google Scholar, Zhihu).

2. Factors affecting the development of new energy vehicles

This article analyzes relevant statistical data to determine the main factors or indicators that affect the development of new energy vehicles in China. Through extensive data search, seven indicators were ultimately determined, namely battery endurance, number of public charging facilities, government subsidy amount, price of power batteries, electricity price, and number of power battery companiesp[5]. Next, a correlation analysis will be conducted between these seven indicators and the sales of new energy vehicles. Analyze the degree of correlation by the size of the correlation coefficients, and then integrate them into the ranking indicator model. The main indicators that affect the development of new energy vehicles in China will be selected based on the size of the obtained values.

2.1. Data preprocessing

By collecting statistical data on new energy vehicle sales, battery life, number of public charging stations, subsidy amounts, electricity prices, gasoline retail prices, carbon emissions, etc. in the past three years, it was found that there are the following problems:

(1) The statistical data release cycle for each indicator is different, with some taking months as the cycle and others taking years or quarters as the cycle;

(2) The statistical data for individual indicators is incomplete.

To address the above issues, the following measures will be taken:

(1) Considering the timeliness of the impact of indicators, it is proposed to use the sales volume of new energy vehicles and statistical data of various indicators from September 2019 to July 2022

(2) Select statistical data for each indicator on a monthly basis;

(3) Interpolate incomplete indicator data;

2.2. Spearman correlation analysis model

Spearman correlation analysis is a hierarchical transformation of two variables X and Y, represented by the levels RX and RY[6]; Then calculate the correlation between RX and RY using Pearson correlation analysis. If there are no duplicate values in the data, and when two variables are completely monotonically correlated, the Spearman correlation coefficient is+1 or -1.

The correlation coefficient is an indicator that describes the degree of correlation between variables, represented by γ [-1, 1]. The larger the value γ, the smaller the error Q, and the higher the degree of linear correlation between variables; The closer the value γ is to 0, the greater Q, and the lower the linear correlation between variables[7].

2.3. Spearman correlation analysis method to analyze the degree of influence of various indicators

The process of calculating Spearman correlation coefficient is as follows:

First, rank the variables X and Y in ascending order, and represent them with rank orders RX and RY. When sorting, data equality occurs, resulting in the phenomenon of causing the same rank is called tie, in which the average rank is taken as the rank of each data. Spearman phase the formula for calculating the number of relationships is:

$$\gamma_{correlationcoefficient} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})}} \tag{1}$$

2.4. Spearman correlation analysis results display

Here we have a total of 7 indicators, namely battery endurance, number of public charging stations, subsidy amount, electricity price, retail price of gasoline, carbon emissions, and number of power battery companies. Analyze these seven indicators separately, and the specific results are shown below Fig 1:

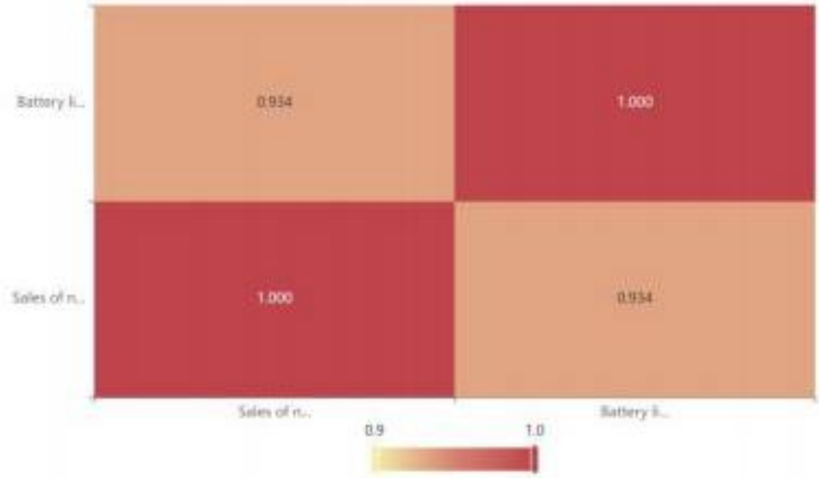

Fig. 1 Heat map of the correlation coefficient between sales volume of new energy vehicles and battery endurance

Fig. 2 Sales volume of new energy vehicles and total heat map of various indicators

2.5. Result analysis

From Figs 1 to 2, the paper can observe the magnitude and ranking of the correlation between different indicators and the sales of new energy vehicles. It can be found that the number of public charging stations and battery endurance have the highest correlation with the sales of new energy vehicles, with a value of 0.934. The correlation between carbon emissions and the number of power battery enterprises with the sales of new energy vehicles is the smallest, with a value of less than 0.2. The other three electricity prices the price and subsidy amount of gasoline also have a significant impact on the sales of new energy vehicles[8]. Therefore, we can consider that the five indicators of the number of public charging stations, subsidy amount, electricity price, retail price of gasoline, and battery range capacity have the greatest impact on the sales of new energy electric vehicles, and they are the main determining factors for the development of new energy vehicles.

3. Analysis based on ARIMA time series model and BP neural network

Because it is necessary to predict the sales of new energy electric vehicles in the next ten years, but the collected data is the sales of cars in the past three years in months, if the direct prediction of the sales of new energy electric vehicles in the next ten years will be relatively large, so we use the method of year-by-year forecasting. That is, according to the data from 2019.9 to 2022.7, the data from 2022.8 to 2023.8 is predicted, and then the data of this year is integrated into the existing data, and the data from 2019.9 to 2023.8 is used to further predict the data of the next year, and finally the data forecast of the decade is completed[9]. At the same time, according to the results of the first question, we found that the sales volume of new energy electric vehicles is highly correlated with the number of public charging piles, subsidy prices, electricity prices, gasoline retail prices, and battery life, so if only the number of new energy electric vehicles is predicted in time series, the error may be relatively large, so we can first use the time series model to predict the number of these five indicators in the next ten years by predicting them year by year, and then predict the sales of new energy electric vehicles year by year according to the BP neural network.

3.1. Predictive analysis based on ARIMA time series model

A time series, also known as a dynamic series, refers to a numerical series that arranges the indicator values of a certain phenomenon in chronological order. Time series analysis can be roughly divided into three parts, which are describing the past, analyzing patterns, and predicting the future. Time series is the numerical manifestation of the long-term change of the value of a certain indicator, so the regularity of the numerical transformation must be contained behind the numerical change of the time series value, which is the entry point of the time series analysis.

In general, there are four types of numerical variations in time series,as shown in Fig 3:

Fig. 3 The numerical variation of the time series

A time series is often the result of the superposition of the above four components.

The ARIMA (Autoregressive Integrated Moving Average model) is an autoregressive model with integrated moving average, which is one of the time series prediction and analysis methods. ARIMA (p, d, q) is a combination of an autoregressive model (AR), a moving average model (MA), and a difference method. In ARIMA(p,d,q), AR is "autoregressive", and p is the number of autoregressive terms; MA is the "moving average", q is the number of terms of the moving average, and d is the number of differences (order) that has been done to make it a stationary series[10]. Here, the paper will use the ARIMA time series model to predict the number of changes in each indicator over the next decade.

3.2. Data preprocessing

Firstly, the paper export the time series data of each indicator from 2019.9 to 2022.8 to obtain the time series of the number of each indicator, so as to establish a time series model of the number of each indicator. Here, we take the battery life of the index battery as an example, and establish a time series model to predict the number of changes in the battery life in the next ten years through the year-by-year prediction method. First of all, we make a time series graph of battery life from 2019.9 to 2022.8, as shown in the Fig 4, to prepare for the establishment of the ARIMA time series model with spss.

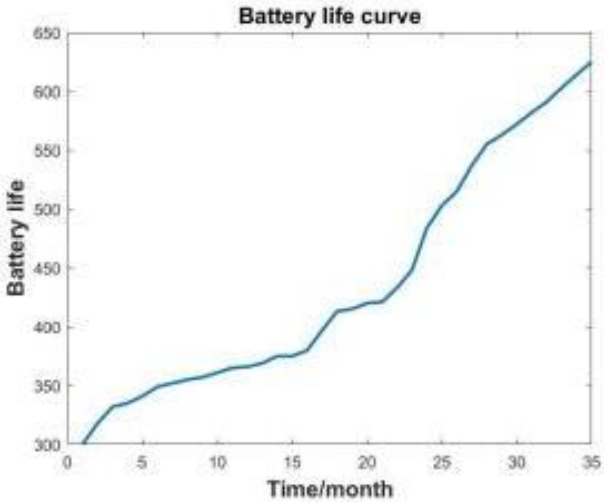

Fig. 4 Diagram of the time series model

3.3. ADF test

The ARIMA time series model requires that the time series satisfies the stationarity test, and if P<0.05, the series is stationary. If the original time series does not satisfy the stationarity, it is divided into differences and seasonal differences until the stationarity is satisfied. Therefore, we first use the spss software to perform the ADF test on the original time series, as shown in Table 1, to determine whether the series satisfies stationarity.

Table 1. ADF Test Form

variable	Differential order	ADF Test Form		AIC	Threshold		
		t	P		1%	5%	10%
Battery life	0	1.202	0.996	167.025	-3.646	-2.954	-2.616
	1	-3.215	0.019**	158.48	-3.646	-2.954	-2.616
	2	-5.881	0.000***	159.322	-3.661	-2.961	-2.619

Note: ***, **, and * represent the significance levels of 1%, 5%, and 10%, respectively

Through the ADF test, it can be found that the original time series based on the battery life data has a significance P value of 0.000** when the difference is divided into order 1, which is horizontally significant, rejecting the null hypothesis, the series is a stationary time series, and we can determine that d is 1.

3.4. Residual ACF and residual PACF

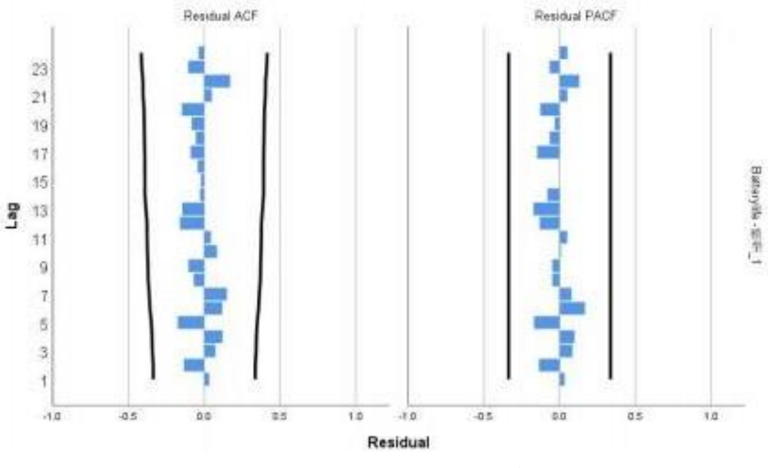

Fig. 5 ACF and PACF plots of residuals

As shown in Fig 5 the ACF and PACF graphs of the residuals, there is no significant difference between the autocorrelation coefficients and partial autocorrelation coefficients of all lag orders and 0. Therefore, the established ARIMA model can basically fit the sequence data, and then we can analyze and obtain that the value of ARIMA(p, d, q)p is 1, and the value of q is 0, so that we can establish the ARIMA(1, 1,0) model.

3.5. Model parameters and model fitting evaluation

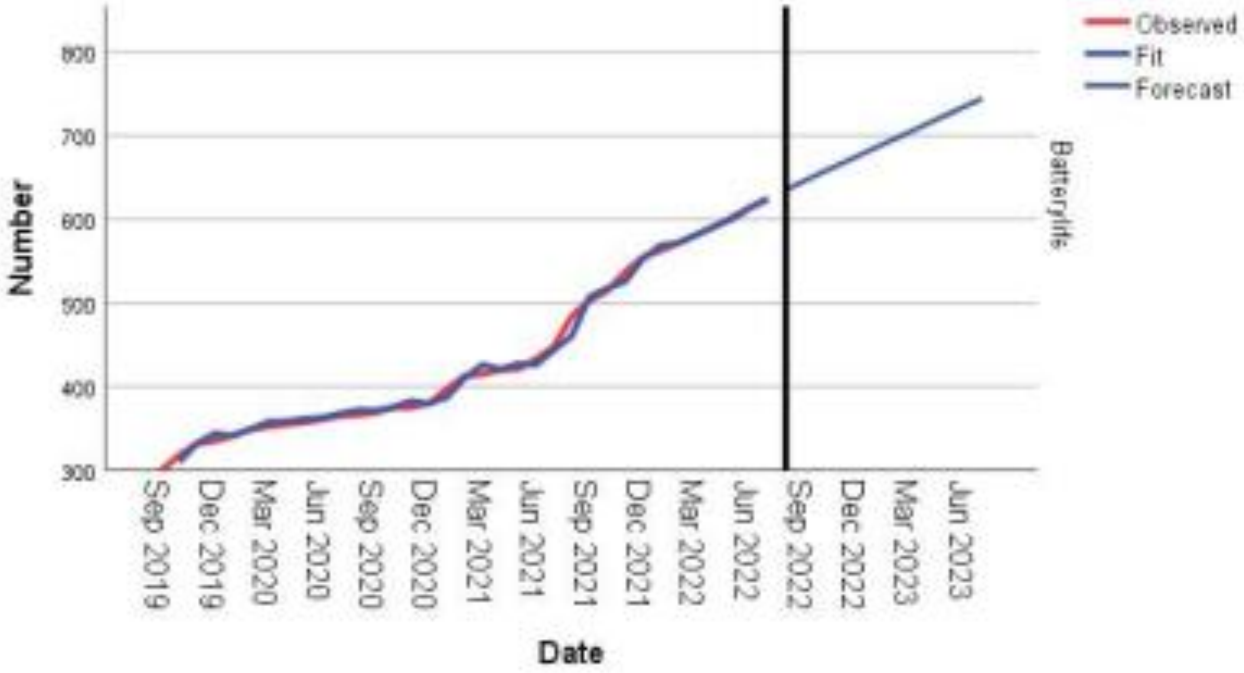

Fig. 6 Comparison of model-fitted time series plots with original time series plots

As shown in Fig 6,based on the battery life data, the spss system automatically finds the optimal parameters based on the AIC information criterion, and finally the model result we establish is the ARIMA model (1, 1,0), and the model formula is as follows: $y(t)=4.689+0.524*y(t-1)$. It can be seen from the model fit that the stationary R side is 0.996, which is close to 1, so the model fit is very high. By comparing the time series diagram and the original time series diagram fitted by the model, it can be found that the data of the original time series and the fitting are basically the same, so the model we established has a good fitting effect and basically meets the requirements.

Table 2. Predicted value(Battery life)

Predicted value	
Order (Time)	Battery life
1	635
2	646
3	656
4	667
5	676
6	685
7	695
8	705
9	715
10	724
11	734
12	745

As shown in Table 2, the paper predict that the battery range will reach 745 km by July 2023. Using the same method, we can predict the changes of the remaining four indicators in the first year, and the analysis method is the same as the battery life, without going into details, here directly give the prediction results of the first year of the five indicators.

Table 3. Predicted value(All)

Predicted value					
Order(Time)	Battery life	Number of charging piles public	Subsidy cost	Charge electricity of	Gasoline price
1	635	1659947	9.468040401	7633.1	9.468040401
2	646	1701786	9.535623814	6405.4	9.535623814
3	656	1764902	9.603207226	6464	9.603207226
4	667	1803723	9.670790639	6522.5	9.670790639
5	676	1854550	9.738374051	6581.1	9.738374051
6	685	1891668	9.805957463	6639.6	9.805957463
7	695	1935561	9.873540876	6698.2	9.873540876
8	705	1971718	9.941124288	6756.7	9.941124288
9	715	2011699	10.0087077	6815.3	10.0087077
10	724	2047313	10.07629111	6873.8	10.07629111
11	734	2085086	10.14387453	6932.4	10.14387453
12	745	2120395	10.21145794	6990.9	10.21145794

According to Table 3,the above are the predicted data for the five indicators in the first year. We then integrated these data with the data from September 2019 to July 2022 to predict the data for the next year. We also established an ARIMA model, which will not be repeated here. Finally, the paper obtained the predicted data for the five indicators in the next ten years, which is relatively large and is included in the appendix.

4. Conclusions

This article concludes that the electrification of new energy electric vehicles (including electric buses) in cities can effectively improve the ecological environment of Chinese cities. Combine the degree of car electrification in our city with the indicators related to the urban ecological environment we have collected, such as carbon dioxide emissions, ammonia nitrogen compound emissions, and sulfur oxide compound emissions, and use entropy weight method for evaluation and analysis. The final results indicate that improving the electrification level of urban new energy electric vehicles (including electric buses) can effectively reduce the emissions of these indicators, greatly improve the ecological environment of our city, and greatly improve the air quality, greenhouse effect, urban heat island effect, and soil conditions of the city.

The use of new energy electric vehicles can not only save daily expenses, but also help improve the living environment. At present, countries around the world are actively promoting the development of new energy electric vehicles, not only reducing taxes for the new energy electric vehicle industry, but also providing subsidies for residents who purchase new energy electric vehicles. Especially with China as a representative, countries have increased their investment in the research and development of new energy electric vehicles, continuously improving their range, economic benefits, and comfort. They have made significant breakthroughs in lithium battery research and fast charging, making important contributions to the development of new energy electric vehicles worldwide.

References

[1] Wang Li; Cao Shiyi; Song Kaifa; Huang Jigui Application of ARIMA model in predicting hand, foot, and mouth disease epidemic in Jingzhou City [J]. Chinese Journal of Social Medicine,2023,40(05):623-626.

[2] Zhang Yong. Evaluation of water resource carrying capacity of Xingfu River Lake in Anhui Province based on entropy weight method and TOPSIS model [J]. Shaanxi Water Resources,2023,(09):45-47.

[3] Fan Cairui; Zheng Hui; Zhang Wenying; Ren Xinmei. Research on the Health Evaluation Index System of Daihai Lake Based on AHP Entropy Weight Method [J]. Inner Mongolia Water Conservancy,2023,(07):9- 11.

[4] He Xiaolong. Research on Comprehensive Data Quality Evaluation of Power Dispatching Based on TOPSIS Method [J]. Electrical Application,2023,42(10):55-61.

[5] Lian Qiang. Spearman rank correlation coefficient of comprehensive interval numbers and its application [J]. Journal of Chongqing University of Technology (Natural Science Edition),2020,37(06):71-75.

[6] Zhou Li; Sun Xiaohan. Evaluation of Water Resources Carrying Capacity in Zhejiang Province Based on BP Neural Network [J]. Journal of Wenzhou University (Natural Science Edition),2023,44(04):25-31.

[7] Xing Yuqi; Zhou Jia; Huang Weili; Wang Junjuan; Liu Meiyan; Hu Xiaoping. A prediction model for wheat scab in rice wheat rotation areas based on BP neural network [J]. Northwest Agricultural Journal,2023,32(11):1842- 1848.

[8] Teng is steady and steady; Zhang et al. Research on Risk Control of Logistics Supply Chain Finance Based on Entropy Weight Method [J]. Times Economics and Trade, 2023,20 (10): 60-64.

[9] Zheng S , Huang J H .New Energy Vehicles Sales Prediction Method and Empirical Research Under the Environment of Big Data[J]. 2018.

[10] Da-Hai Z , Shi-Fang J , Yan-Qiu B I ,et al.Study of power system load forecast based on wavelet neural networks[J].Electric Power Automation Equipment, 2003.