Forecasting New Energy Vehicle Sales in China based on GM (1,1) Model

Mingshang Ma *
Dundee International Institute, Central South University, Changsha, 410006, China
* Corresponding author: 7805220202@csu.edu.cn

Abstract. With the intensification of the conflict between economic development and green energy, the prediction of sales figures of new energy vehicles, especially electric vehicles, has become one of the key concerns in society. This study aims to help businesses understand market trends and adjust marketing strategies and planning in a timely manner through sales volume forecasting. The research data in this paper is sourced from the monthly sales volume data of new energy vehicles in China from January to June 2023, obtained from the global automotive industry platform Mark Lines. Taking into account various factors and adopting a time series approach, the GM(1,1) model is utilized to forecast the sales of the research subject. for the future time period in 2023 on a monthly basis. The results indicate that this method can accurately predict the research subjects in this study, the new energy vehicle industry is expected to enter a faster development stage. This provides an important reference for promoting the vigorous development of China in the new energy vehicle industry.

Keywords: time series, new energy vehicle, grey prediction, GM (1, 1) model.

1. Introduction

Against the backdrop of expeditious socioeconomic advancement in China and the global energy and environmental crisis, new energy vehicles, with electric cars as the main component, have become one of China's emerging strategic industries. The Chinese government has been strengthening the strategic layout of the new energy vehicle industry, promoting its vibrant market development. Against this backdrop, it is crucial to scientifically plan and strategically invest in businesses, as well as implement macro policies that align with the market. Predicting sales can provide a clear market fluctuation trend, and it can offer fundamental guidance for policy adjustments [1]. This paper aims to use quantitative forecasting methods to predict monthly sales. This will help manufacturers and businesses understand the trends in the new energy vehicle market, allocate resources effectively, and formulate sales strategies.

Sales forecasting is a research topic focused on regional markets, and the automotive industry, in particular, has high market value and attention. At the current stage, there are not many sales forecasts specifically for the automotive market. However, there is a mature foundation of research and related literature in the field of sales forecasting both domestically and internationally [2]. Cai applied grey system theory to dynamically forecast the demand of the Chinese automotive market in the coming years, which proved to be practical and feasible, but the focus of the study was on predicting the private car ownership in China [3]. Zhu and Zhang employed the GM (1,1) model to obtain the annual sales of cars in China for the next five years with good accuracy, but this prediction falls under medium-term forecasting [4]. Chen and Miu used historical sales data from the Chinese automotive market to establish an ARIMA model, considering the seasonal factors of historical sales data. The final results demonstrate that the model had good predictive performance [5, 6]. Gong constructed a comprehensive forecasting model based on grey theory and exponential smoothing, which achieved good predictive results, but it overlooked multiple factors that affect sales [7]. Wu and Chen combined principal component analysis and general regression neural networks to build a feasible prediction model. However, the model had limitations due to limited data and the lack of consideration for factors such as political and economic influences [8].

In reality, the relationship between sales volume and influencing factors is quite complex, and ordinary linear models are easily affected by the irregularity of data, which affects accuracy [9]. In this regard, Zhao summarized the main factors influencing the sales volume of cars in China, such as GDP, transportation, and policies. They used Granger causality tests to select factors causally related to sales volume and conducted grey correlation analysis. Finally, they established a regression prediction model using principal component analysis [10]. Ma et al. analyzed the factors affecting the purchase of new energy vehicles, such as cost, reliability, image, and convenience, using analytic hierarchy process to analyze their weights. They then used a logit model to regress the market share of new energy vehicles against consumer utility, considering market shares and consumer utility in other countries. They constructed predictive models for the market share of new energy vehicles under three government scenarios, achieving good results in analyzing the correlation between sales volume and influencing factors [11]. Pan and Han used a combination forecasting theory, considering multiple economic factors affecting the market. They established single prediction models for future demand in the Chinese automotive market using grey system, multivariate regression, and triple exponential smoothing methods, overcoming the limitations of individual models [12]. Yang established a multiple linear regression and BP neural network model to predict the passenger car market by analyzing the relevant factors of car sales. The results indicated that the comprehensive model had a high degree of fit [13]. Marco and Detlef analyzed historical sales data of cars on an annual, quarterly, and monthly basis. They built a sales forecasting model for the automotive market, based on time series analysis and classical data mining algorithms. They standardized economic data in the model, obtained relevant parameters, introduced the concept of tree decision-making in the prediction process, and improved accuracy through model correction [14].

Currently, there are various models and methods for such forecasting problems, mainly divided into time series models, machine learning models, and hybrid models based on these two methods [2]. However, different methods need to be used for different types of data, and there is no universally applicable model.

In conclusion, considering the requirements of short-term forecasting, the limited amount of data used, and the relatively weak expression of seasonal trends, this paper will adopt the grey prediction and forecast the data based on the GM (1, 1) model.

2. Method

2.1. Data Sources

The data for this study is sourced from the global automotive industry platform, Mark Lines database. The data was retrieved by filtering the country as China and the vehicle types as EV/PHV/FCV. The collected data includes the latest new energy vehicle sales figures for June 2023, which were published on July 11th, 2023.

2.2. Variable Selection

The data used in this study includes statistical results from January to June 2023. The vehicle types included are EV/PHV/FCV. The data was filtered based on the production and sales figures published by Mark Lines. The following table presents the collected data.

Table 1. Statistical table (Unit:10,000 vehicles)

Months	1	2	3	4	5	6
HV	28.7	37.6	49	47.1	52.2	59.3
PHV	12.1	14.9	16.3	16.5	19.4	23.2
FCV	0.02	0.004	0.05	0.03	0.04	0.1
Total	40.8	52.5	65.3	63.6	71.7	80.6

Table 1 lists the individual sales and total sales for three different car models, with data accurate to one decimal place. It can be observed that there is an overall upward trend in sales on a monthly basis. Additionally, the HV vehicle type has the highest sales figure, while the FCV vehicle type has significantly lower sales compared to HV and PHV.

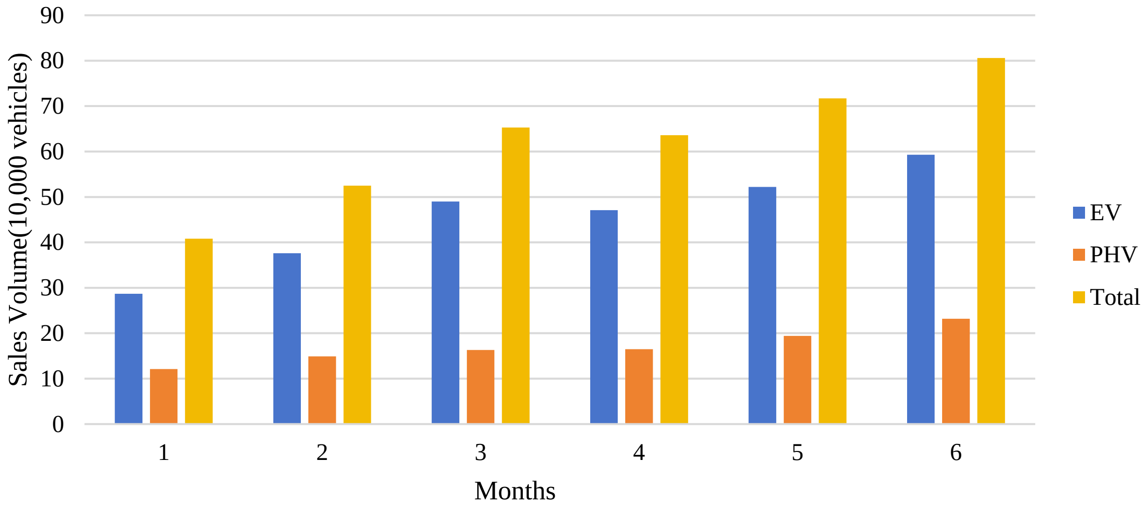

Figure 1. Monthly display of sales in China, Jan-June 2023

Figure 1 displays the individual sales and total sales of HV/PHV vehicle types, while the sales of FCV vehicles are not shown in the figure due to their relatively small numbers. On a monthly basis, the sales trend of electric vehicle is similar to that of conventional vehicles, showing a nonlinear growth trend over time. However, due to the limited data range, the seasonal characteristics of sales are not reflected.

2.3. Research Protocol

2.3.1. Model introduction

For this forecasting problem, all kinds of factors are associated with the outcome, such as national policies, GDP, population size, and other macro factors, as well as micro factors like changes in car performance and prices. Moreover, there are often inherent connections among these factors. Time series models are precisely designed to capture these inherent connections and make accurate predictions, thereby avoiding the need to discuss the impact of various complex factors on the results. The grey system theory is a significant method for handling small-sample discrete datasets, emphasizing the utilization and development of limited information from the original data. The basic idea of this theory is to perform Accumulated Generating Operation (AGO) on the original data. The GM (1,1) model used in this paper represents a first-order differential equation form of a prediction model with a unique variable. It can reduce the complexity of the problem and analyze and predict the changing patterns of the research object, while disregarding other influencing factors and having a small sample size. It is applicable to addressing problems with a small number of samples, limited data information, or incomplete information, and serves as a method for predicting systems with uncertain factors [4, 15].

2.3.2. Protocol introduction

In this study, the GM (1,1) model is utilized to transform the original data into a univariate time series dataset. The dataset is then accumulated to generate a new set of data with distinct trend characteristics. Based on this, a model is established, and SPSSAU is used to obtain numerical results as well as assess the fitting and prediction performance.

2.4. Model Principle

2.4.1. Data preprocessing

Letting the original data sequence is denoted as with n being the number of data points.

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \ldots\ldots, x^{(0)}(n)) \tag{1}$$

Let $x^{(1)}$ represent the 1-AGO sequence of $x^{(0)}$.

$$x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i)\,, k = 0,1,2,\ldots \tag{2}$$

The ratio test is conducted on the original sequence to assess the suitability of constructing a model for the data sequence.

2.4.2. Modelling

Establishing a first-order linear differential equation with respect to k:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b \tag{3}$$

In equation (3), a and b are undetermined coefficients. The sequence$x^{(1)}$(k)is then modified into a mean-generated sequence.

$$z^{(1)}(k) = \frac{1}{2}x^{(1)}(k) + \frac{1}{2}x^{(1)}(k-1), k = 2, \ldots \ldots n \tag{4}$$

$$x^{(0)}(k) = b - az^{(1)}(k) \tag{5}$$

The matrix equation can be expressed as Y = BU, where Y represents the vector of values, B is the coefficient matrix, and U is the vector of unknowns or coefficients.

$$\begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(N) \end{bmatrix} = \begin{bmatrix} -\frac{1}{2}\left[x^{(1)}(2) + x^{(1)}(1)\right] & 1 \\ -\frac{1}{2}\left[x^{(1)}(3) + x^{(1)}(2)\right] & 1 \\ \vdots & \vdots \\ -\frac{1}{2}\left[x^{(1)}(N) + x^{(1)}(N-1)\right] & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \tag{6}$$

By means of the method of least squares, find the value of U that minimizes $(Y - BU)^T(Y - BU)$.The estimated value of U is $\hat{U} = [\hat{a} \quad \hat{b}]^T = (BB^T)^{-1}B^T$, and determine $\hat{a}$ and $\hat{b}$ based on $\hat{U} = \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix}$.

By substituting the sum into (3), the solution is obtained as follows:

$$\hat{x}^{(1)}(k+1) = \left(x^{(0)}(1) - \frac{\hat{b}}{\hat{a}}\right)e^{-\hat{a}k} + \frac{\hat{b}}{\hat{a}}, k = 0,1,2,\ldots \tag{7}$$

The simulated value of $\hat{x}^{(0)}$ is:

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) \tag{8}$$

2.4.3. Error check

Calculation of residuals:

$$E(k) = x^{(0)}(k) - \hat{x}(k), k = 2,3,\ldots,n \tag{9}$$

Calculation of relative residuals:

$$e(k) = \frac{x^{(0)}(k) - \hat{x}^{(0)}(k)}{x^{(0)}(k)}, k = 2,3, \ldots, n \tag{10}$$

Calculation of the mean squared deviation of $x^{(0)}$:

$$S_1 = \sqrt{\frac{1}{N}\sum_{k=1}^{n}[x^{(0)}(k) - \bar{X}]^2} \tag{11}$$

Residual mean:

$$\bar{E} = \frac{1}{N-1}\sum_{k=2}^{n} E(k) \tag{12}$$

Residual variance:

$$S_2 = \sqrt{\frac{1}{N-1}\sum_{k=2}^{n}[E(k) - \bar{E}]^2} \tag{13}$$

Posterior odds ratio:

$$C = \frac{S_2}{S_1} \tag{14}$$

Probability of small error:

$$P = P\{|E(k) - \bar{E}| < 0.6745 S_1\} \tag{15}$$

2.5. Model Testing

After model construction, the obtained development coefficient a and grey action quantity b were used as model parameters. However, their practical significance is relatively small. A posterior odds ratio "C" less than or equal to 0.35 indicates a very good level of model accuracy. Subsequently, the residual values were examined, where a smaller relative error indicates better performance, numerical results below 0.2 indicate that the model meets the basic requirements, while results below 0.1 indicate that the model achieves a higher level of accuracy. Similarly, a smaller deviation in the ratio indicates better performance, with a value below 0.2 meeting the requirements and less than 0.1 indicating a higher level of achievement.

3. Results and Discussion

The table below presents the ratio deviation values after conducting the ratio deviation test.

Table 2. Ratio Deviation for GM (1,1) Model

Index	Original Value	Ratio Deviation λ	Original/Transformed Value (shift=0)	Transformed Ratio Deviation λ
1	40.800	-	40.800	-
2	52.500	0.777	52.500	0.777
3	65.300	0.804	65.300	0.804
4	63.600	1.027	63.600	1.027
5	71.700	0.887	71.700	0.887
6	80.600	0.890	80.600	0.890

Based on Table 2, it is evident that all the ratio values, before performing the shifting transformation, are within the interval of the standard range [0.751, 1.331]. This implies that the data is applicable to constructing a GM (1,1) model.

The following table displays the values of four parameters during the model construction process.

Table 3. The results of model construction

Development Coefficient a	Grey action quantity b	C	*p*
-0.0939	48.4250	0.0396	1.000

Table 3 presents the numerical values of the four parameters in the model establishment process. The value of C is 0.0396, which is below 0.4, indicating that the model has extremely high accuracy. Additionally, the value of *p* is equal to 1 also signifies that the model accuracy is excellent.

Table 4. The table of predicted values by the model

Index	Original value	The predicted value
1	40.800	40.800
2	52.500	54.785
3	65.300	60.177
4	63.600	66.098
5	71.700	72.603
6	80.600	79.747
Next 1 periods	-	87.595
Next 2 periods	-	96.215
Next 3 periods	-	105.683
Next 4 periods	-	116.083
Next 5 periods	-	127.506
Next 6 periods	-	140.054
Next 7 periods	-	153.836
Next 8 periods	-	168.975
Next 9 periods	-	185.603
Next 10 periods	-	203.867
Next 11 periods	-	223.929
Next 12 periods	-	245.965
Note: RMSE=2.802		

GM (1,1) model prediction table displaying fitted values and predictions for the next 12 periods. The fitted values for the current period can be used for further calculations such as residual, correlation error, and ratio bias. The RMSE value represents the root mean square error, with lower values indicating better performance and can be utilized to compare the effectiveness of different models. The table 4 provides numerical results for the predictions.

Table 5. The GM (1,1) model verification

Index	Original value	Predicted value	Residual	Relative error value	Ratio bias
1	40.800	40.800	0.000	0.000%	-
2	52.500	54.785	-2.285	4.353%	0.146
3	65.300	60.177	5.123	7.846%	0.117
4	63.600	66.098	-2.498	3.928%	-0.128
5	71.700	72.603	-0.903	1.259%	0.026
6	80.600	79.747	0.853	1.058%	0.023

The fitted values obtained from Table 5 were used to conduct residual analysis, relative error testing, and ratio bias testing for the model. For the relative error value, smaller values are better. The value corresponding to a well-performing model should be less than 0.2, while the value corresponding to high accuracy should be less than 0.1. Similarly, smaller values are desirable for the ratio bias. The definition range for a good and excellent model is also defined as less than 0.2 or 0.1.

By analyzing the data, it can also find that the largest relative error value is 0.078, which is less than 0.1. The maximum ratio bias value of the model is 0.146, which is less than 0.2. This indicates that the model's fitting performance meets higher accuracy requirements.

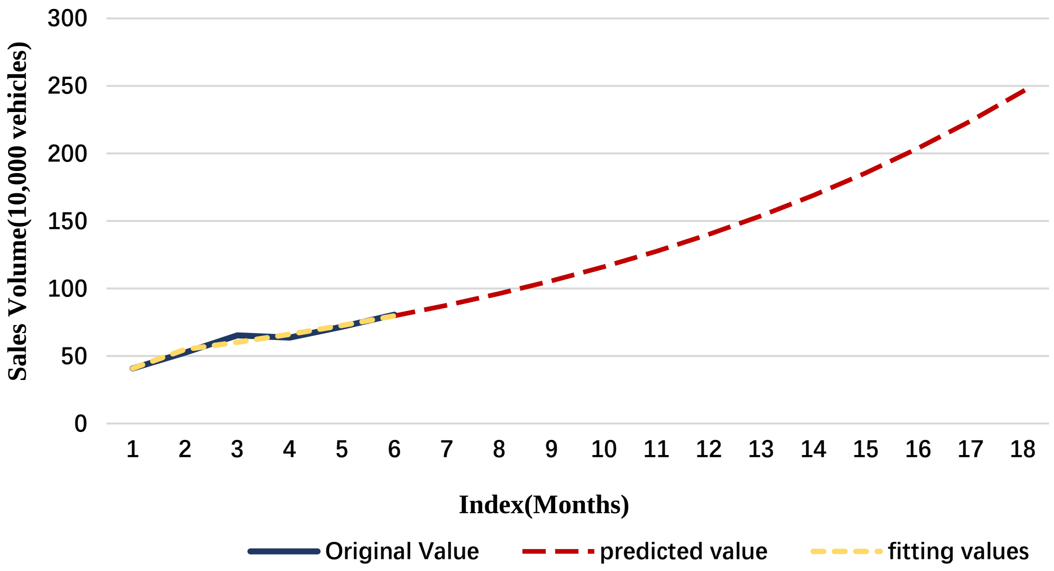

Figure 2. Model Fitting and Prediction Curves

From Figure 2, it can be visually observed that the predicted curve obtained by the model has a high degree of fit with the original curve. According to the model, the monthly sales data for electric powered vehicle in China from July to December 2023 are 87.595, 96.215, 105.683, 116.083, 127.506, and 140.054, respectively. It is evident that the predicted curve shows a continuous upward trend in the later period, which is a characteristic of the GM (1,1). It is only applicable too short to medium-term forecasting, and therefore, data beyond the forecast target will lose value. It is important to exercise caution when dealing with data that exceeds the scope of the forecast.

4. Conclusion

This study utilized a time series approach and employed the GM (1,1) model to forecast vehicle sales in China's new energy sector for the year 2023. The conclusion is that the sales statistics of new energy vehicles (HV/PHV/FCV) in China for the months of July to December 2023 are projected to be 87.595, 96.215, 105.683, 116.083, 127.506, and 140.054 thousand units, respectively. The forecasting results demonstrate a high level of accuracy. It indicates that the market sales quotes in China will continue to exhibit an upward trend for the remaining period of 2023, with the highest sales volume exceeding 1.4 million units.

It is undeniable that the grey theory is not applicable to predicting time series data with seasonality. Therefore, its value is mainly reflected in short-term forecasting. As the span of the original time series data used increases, the seasonality of new energy vehicle sales becomes more prominent, making this method no longer suitable for prediction. Additionally, this model did not analyze the relationship between various factors and the sales of new energy vehicles, nor did it provide intuitive conclusions regarding the correlation between sales and other factors. This is a limitation of this study. However, this research still has many values and advantages. Firstly, it has a strong foundation and innovation. On one hand, the grey system theory, with GM (1,1) as the main model, is widely used in sales forecasting. In this study, the classic model was applied to predict new energy vehicle sales, yielding accurate results. This enriches the research on short-term forecasting for this market. On the other hand, the GM (1,1) model can be combined with other methods to tackle more complex and demanding forecasting problems. For example, combining data analysis methods with the GM (1,1) model can incorporate the seasonal characteristics of research target, enabling better utilization of the grey system theory and obtaining more accurate results for more complex predictions. Apart from that, sales forecasts directly impact a company's market assessment. The results of this study will assist companies in formulating reasonable short-term plans and sales strategies, thereby promoting the development of the new energy vehicle industry.

References

[1] Ling Tuo. Research on Sales Forecast of Electric Vehicles in China. Chongqing University of Technology, 2018.

[2] Xu Yuenin. Car Sales Forecast under Combination Model. Northwest University, 2022.

[3] Cai Jiaming. Application of Grey System Model in Automotive Market Demand Forecasting. Journal of Shanghai University of Engineering Science, 2003, (01): 72 - 74+80.

[4] Zhu Manyu, Zhang Zheng. Research on Sales Forecast of New Energy Vehicles Based on GM (1, 1) Model. Logistics Engineering and Management, 2023, 45 (02): 138 - 141.

[5] Chen De. Chinese automobile demand prediction based on ARIMA model. 2011 4th International Conference on Biomedical Engineering and Informatics (BMEI), Shanghai, China, 2011, 2197 - 2201.

[6] Miao Hui, Tang Chentian, Luo Lulu. Sales Forecast of New Energy Vehicles Based on ARIMA Model. Enterprise Science and Technology Development, 2020, (10): 97 - 98.

[7] Gong Wenwei, Huang Jing. Integrated Model for Demand Forecasting Based on Grey Theory and Exponential Smoothing Method. Statistics and Decision Making, 2017, (01): 72 - 76.

[8] Wu Minfeng, Chen Wen. Forecast of Electric Vehicle Sales in the World and China Based on PCA-GRNN. Sustainability, 2022, 14 (4): 2206.

[9] Huang Hongyun, Liu Weixiao, Ding Zuohua. Sales Forecast Based on Multivariate Grey Model and Neural Network. Journal of Software, 2019, 30 (04): 1031 - 1044.

[10] Zhao Ying. Research on Forecasting Model of Automobile Sales in China Based on Regression Analysis. Huazhong Normal University, 2014.

[11] Ma Jun, Wang Ning, Kong Deyang. Forecasting Model of New Energy Vehicle Market Based on AHP and Logit Regression. Journal of Tongji University (Natural Science Edition), 2009, 37 (08): 1079 - 1084.

[12] Yang Bingrong. Forecasting Model of Passenger Car Market Based on Multivariate Linear Regression and BP Neural Network. Huazhong University of Science and Technology, 2017.

[13] Pan Zhigang, Han Ying. Application of Combination Forecasting Method in Demand Forecasting of China's Automobile Market. Business Research, 2006, (20): 126 - 129.

[14] Hulsmann M, et al. General sales forecast models for automobile markets and their analysis. Transactions on Machine Learning and Data Mining, 2012, 5 (2): 65 - 86.

[15] Liu Yuan. Long-term and Short-term Forecasting of New Energy Vehicle Sales. Jilin University, 2020.