Forecasting the Development of New Energy Electric Vehicles in China Based on ARIMA Time Series Models

Zhenzi Jin[1,a], Mengyun Chen[1,b], Tongxuan Zhao[1,c], Yutong Pan[2,d], Ye Tao[1,e], Yizhe Deng[1,f], Xiaoling Qin[1,*]

{1589395265@qq.com[a], 358094365@qq.com[b], 2076768024@qq.com[c], 837202642@qq.com[d], 3110998767@qq.com[e], 483831973@qq.com[f], *Corresponding author: 474495883@qq.com}

Wuhan Donghu University, Wuhan, Hubei, 430212, China[1]
Jinan University, Guangzhou, Guangdong, 510632, China[2]

Abstract. With the Chinese government's strong support and promotion of the new energy electric vehicle industry since 2011, this field has made remarkable progress and become another national landmark industry after the ‘Chinese high-speed railway’. This study utilizes the ARIMA time-series forecasting model for quantitative analysis in order to accurately forecast the future development of new energy electric vehicles in China and to evaluate the effect of foreign policies on the country's industry. By identifying the key influencing factors on the sales of new energy electric vehicles, we focus on the core indicators such as average battery energy density, market share and number of charging piles, and normalise them. Based on these processed data, this study constructs a prediction model and applies ARIMA with nonlinear fitting method [1] for effective solution, with a view to providing decision support for policy makers and promoting the sustainable and healthy development of the industry.

Keywords: new energy electric vehicle, ARIMA, nonlinear fitting, development trend

1 Introduction

New energy electric vehicles cover a wide range of types, including hybrid vehicles, pure electric vehicles and fuel cell vehicles. In recent years, new energy electric vehicles have achieved remarkable leaping development by virtue of their low-pollution and low-energy consumption characteristics as well as their excellent power peak regulation ability, winning the favour of consumers and the government. An in-depth study of the development trend of new energy electric vehicles is of great significance for promoting sustainable economic development and building an ecological civilised society. In view of this, it has become an indispensable research topic to accurately predict the development trend of China's new energy electric vehicles with the help of advanced artificial intelligence calculation methods.

2 Research status:

Scholars have used different methods to forecast the development trend of new energy electric vehicles in China: for example, Liao, F. (2017) [2] Offering advice for further exploration into

MSEA 2024, May 24-26, Jinan, People's Republic of China

DOI 10.4108/eai.24-5-2024.2350126

the growth of the new energy vehicle sector, market surveys and investigations into consumer behavior, policies and rules, and technological advancement are conducted; Zhou Jinlong and his team members (2020)[3] use grey forecasting to forecast new energy vehicle The model has high accuracy and provides reference value for the sustainable development of China's new energy vehicle industry; Domarchi, (2023) predicts the future development trend by analyzing the sales data[4], technological advancement, and market trends of new energy electric vehicles over the past few years; Shiqi Ou and his team members (2020) combine the charging infrastructure quantification with a vehicle market analysis tool - NEOCC model [5], and concluded that more family parking spaces can stimulate more sales of personal plug-in hybrid electric vehicles; Chunli He (2020) established a binomial logit model [6], empirically demonstrated the demand for new energy vehicles based on consumption preferences, proposed corresponding policy guidance for the demand of new energy vehicle market at different levels, and provided an insight into the demand for new energy vehicle industry chain. policy orientation and conducted research on the collaborative innovation of the new energy vehicle industry chain. The literature used in the above literature shows great variability in methodology, data and background, and cannot provide accurate forecasts. Therefore, this paper mainly adopts the ARIMA time series forecasting method to obtain a data selection model with greater relevance regarding the development of new energy electric vehicles in China.

3 Data pre-processing

We conducted an examination of pertinent literature [7] and data to ascertain the state of new energy electric vehicles in China in recent times, thus examining their development trend in Table 1:

Table 1. Status of New Energy Electric Vehicles in China

Year	China's new energy vehicle efficiency (10,000 units)	Total automobile sales in China (10,000 units)	Retention (10,000 units)	China's new energy vehicle market share (%)
2011	0.8	1852	1.2	0.04
2012	1.2	1931.6	2.7	0.06
2013	1.7	2198.3	5.9	0.08
2014	7.5	2340.6	14.8	0.32
2015	33.1	2471.5	33.1	1.34
2016	50.7	2811.9	65.1	1.8
2017	77.7	2890.5	150.6	2.69
2018	125.6	2808.1	253.4	4.47
2019	120.6	2571.6	377.5	4.69
2020	136	2514.6	484.6	5.41
2021	351	2530	784.1	13.87
2022	688.7	2686.4	1310	25.6

We selected seven data indicators based on the open-source dataset for five factors: sample policy, economy, technology, society and infrastructure. In this paper we perform data integration, data cleaning, principal component analysis data dimensionality reduction and data standardization on the raw data. The results after pre-processing the raw data are shown in Table 2:

Table 2. Tables for data standardization

year	Production of new energy vehicles in China	Industry synergy rate	Number of charging stations	Subsidy estimates	Motor efficiency	market share	The average energy density of the battery	Sales of new energy vehicles
2016	0.0198969	0.083333	0.008269	0.15625	0.114441	0.018259282	0.055026	0.020236
2017	0.0305573	0.095238	0.022417	0.21875	0.11921	0.027388923	0.069699	0.031013
2018	0.0492226	0.107143	0.038707	0.1875	0.122616	0.045648205	0.077036	0.050611
2019	0.048068	0.119048	0.060725	0.125	0.126703	0.047474133	0.102715	0.048136
2020	0.0561499	0.130952	0.08374	0.109375	0.127384	0.054777845	0.112252	0.056239
2021	0.1415486	0.142857	0.130368	0.09375	0.128747	0.238385068	0.15774	0.140856
2022	0.2697044	0.154762	0.25954	0.0625	0.130109	0.259687563	0.205429	0.275365
2023	0.3848522	0.166667	0.396234	0.046875	0.13079	0.308378982	0.220103	0.377545

4 ARIMA time series modelling [8]

The Autoregressive Differential Moving Average Model (ARIMA), composed of Autoregressive Model (AR), Differential Process (I) and Moving Average Model (MA), is a widely recognized model. This model between time points is expressed in this way: the labelled value of a given point in time can be represented by a linear combination of all the labelled values of a certain time period in its past. This linear combination can be understood as a weighted summation of past information, with each past time point being represented by a corresponding weight, as reflected in the formula below:

$$AR: Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \ldots + \varphi_p Y_{t-p} + \xi_t \quad (1)$$

ξ_t is the error term at time point t, also known as the white noise term, expressed in terms of the lag operator as:

$$\zeta_t = (1 - \varphi_1 L - \varphi_2 L^2 - \ldots - \varphi_p L^p) Y_t \quad (2)$$

The MA model describes the relationship between the data at the current point in time and the past noise, where the data at each point in time are independent and follow the same distribution with constant mean and variance. Given a white noise sequence ς_t, the MA model is defined as:

$$MA: Y_t = \mu + \varsigma_t + \theta_1 \varsigma_{t-1} + \theta_2 \varsigma_{t-2} + \ldots + \theta_q \varsigma_{t-q} \tag{3}$$

Since ς_t are white noise sequences, they are unpredictable. However, once we have the data Y_t, we can solve for ς_t using the following equation:

$$\varsigma_t = Y_t - c - \theta \varsigma_{t-1} \tag{4}$$

We can then use this newly solved ς_t to compute Y_{t+1} at the next moment:

$$Y_{t+1} = c + \varsigma_{t+1} + \theta \varsigma_t \tag{5}$$

This is a recursive process and we can make predictions by continually solving for the above values.

Assuming that the difference term is 0, the ARIMA model can be viewed as a direct combination of the AR model and the MA model, which is given by:

$$Y(t) = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \ldots + \varphi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-w} + \varepsilon_t \tag{6}$$

The above model is therefore called model $ARIMA(p,d,q)$. The model's lagged values, denoted by 'P', are the 'autoregressive part', and the order of 'p' in the equation dictates the number of observations the model refers to. The ‘moving average part', denoted by 'q', is the lagged values of the errors used in the model. The equation's order q, denoting the ‘moving average part' of the model's error terms' lagged values, determines the amount of white noise referred to; d, the difference order, converts a non-stationary series into a stationary one. Figure 1 below shows the ARIMA flowchart:

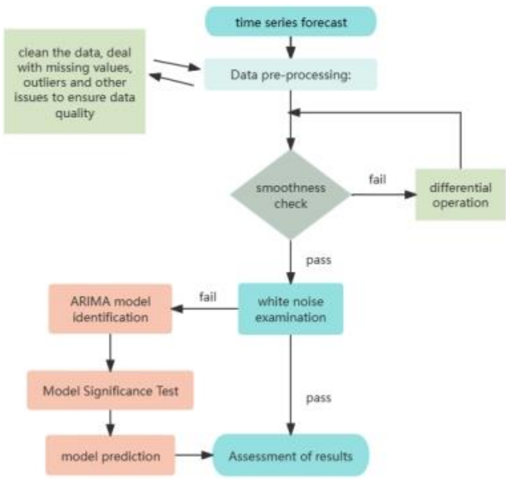

Fig. 1 ARIMA flow chart

- Determination of p and q values for ARIMA

Autocorrelation and partial autocorrelation coefficients were calculated to test whether the preprocessed data met the requirements of ARIMA modelling.

The ACF for any lag $(lag)k_t$ autocorrelation function $\rho(k)$ can be expressed as:

$$\mathrm{ACF_k} = \rho_k = \frac{Cov(Y_t, Y_{t-k})}{Var(Y_t)} \tag{7}$$

The ACF's range is from -1 to 1; when it is close to 1, the two observations have a strong positive correlation; when it is near -1, the two observations are strongly negative; and when it is near 0, the correlation between the two observations is feeble.

The PACF, a measure of the direct correlation between observations at the present moment and those that have been lagged after other, prior lagged observations have been taken into account, is utilized.

The PACF autocorrelation function $\varphi(k)$ for any lag $(lag)k_t$ can be expressed as:

$$\varphi(k) = \frac{Cov(Y_t - E[Y_t \mid X_{t-1}, \ldots, Y_{t-k+1}], Y_{t-k} - E[Y_t \mid Y_{t-k}, \ldots, Y_{t-1}])}{Var(Y_t)} \tag{8}$$

The value of PACF also ranges from -1 to 1. When PACF is near 1, it implies a strong positive correlation between the two observations; when it is near -1, it implies a strong negative correlation; and when it is close to 0, it suggests a feeble direct correlation between the observations at the two time points.

The visualization of the plotted autocorrelation and partial autocorrelation is shown in Figure 2 as follows [9].

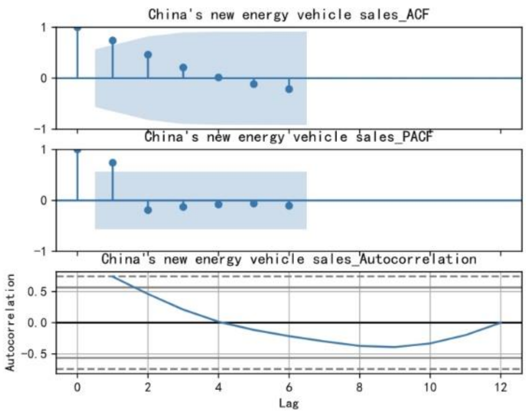

Fig. 2 Autocorrelation and partial autocorrelation

Based on the truncation of the partial correlation function we initially choose AMIMA (4, 0, 1), and combined with the AIC criterion to order the model, we choose p=1, q=2, and select the AMIMA (1, 1, 2) model for prediction.

We then perform a white noise test by finding out whether the autocorrelation coefficient tends to zero. If the autocorrelation coefficient tends to 0, then it can be considered as a white noise series and the ARIMA model applies. We end up with p=-0.142, which tends to 0. The ARIMA model applies.

We diagnose the estimated ARIMA model predictions using both absolute and relative error methods. Then we plot the data model approximation curve as shown in Figure 3:

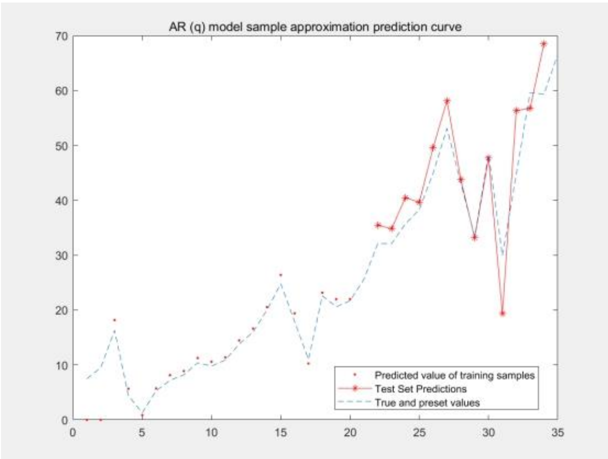

Fig. 3 Data model approximation curve

From the above figure we can intuitively see that the test set predicted values fit well with the true values and preset values. We consult the relevant literature and conclude that the error of single-step prediction is lower and the accuracy is higher. So we choose single-step prediction [10] to predict the development of new energy electric vehicles in China. The prediction results are shown in Figure 4 below [11]:

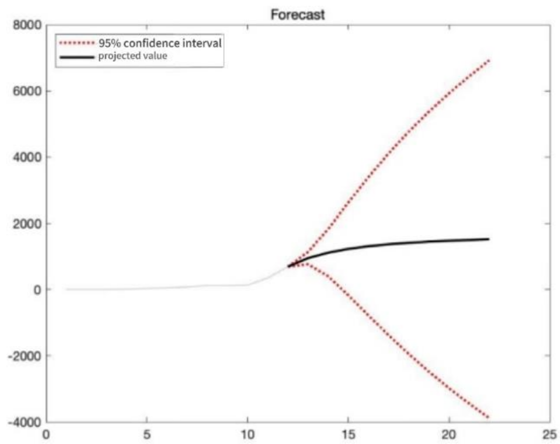

Fig. 4 Forecast Results

The above graph shows that the predicted values are within the confidence interval and the predicted values are true and valid.

China's new energy vehicle ownership and sales in the next ten years are shown in the table 3 below:

Table 3. Development forecast values

Year	2023	2024	2025	2026	2027
Sales volume	950.2	1113.9	1227.8	1308.6	1367.4
Year	2028	2029	2030	2031	2032
Sales volume	1411.5	1445.8	1473.7	1497.3	1518

As can be seen from the figure, the real value fits well with the predicted value. In the next ten years, with the support of China's policies, the improvement of people's environmental awareness, and the gradual improvement of new energy electric vehicle infrastructure, the development trend of new energy electric vehicles in China is gradually rising [12].

5 Conclusion

This paper constructs a forecasting framework based on the ARIMA time series model, aiming to accurately predict the development trend of China's new energy electric vehicles in the next ten years. The results of the model analysis show that China's new energy electric vehicles will show a solid upward trend in the next decade. This positive development trend is attributed to the support of industrial policies, enhanced capital injection, promotion of technological innovation, expansion of market demand, and optimisation of the supply chain. However, in the face of increasingly fierce market competition, China's new energy vehicle companies must continue to maintain their core competitiveness in order to ensure that they remain invincible in the long run.

References

[1] B. Guo, B. Wang, X. Liang, et al. Nonlinear least squares method for fitting fault surface parameters and its MatLab implementation[J]. Sichuan Earthquake,2016,(03):29-33.DOI:10.13716/j.cnki.1001-8115.2016.03.006.

[2] Liao, F., Molin, E., & van Wee, B. (2017). Consumer preferences for electric vehicles: a literature review. Transport Reviews, 37(3), 252–275.

[3] Domarchi, Cristian, and Elisabetta Cherchi. "Electric Vehicle Forecasts: A Review of Models and Methods Including Diffusion and Substitution Effects." Transport Reviews 43, no. 6 (2023): 1118–43. doi:10.1080/01441647.2023.2195687.

[4] ZHOU Jinlong, ZHANG Tianhong, HU Peng, et al. Forecasting the development of new energy vehicles based on grey forecasting[J]. Electronic World,2020,(03):17-18.DOI:10.19353/j.cnki.dzsj.2020.03.007.

[5] Ou, S., Lin, Z., He, X., Przesmitzki, S., & Bouchard, J. (2020). Modelling charging infrastructure impact on the electric vehicle market in China. Transportation Research Part D-transport and Environment, 81,. 102248.
[6] He Chunli. Research on new energy vehicle market demand and policy orientation [D]. Southwest University of Finance and Economics, 2020.DOI: 10.27412/d.cnki.gxncu.2020.000326.
[7] Wang Na. Identification and analysis of key influencing factors for the development of new energy vehicles in China[J]. Journal of Nanjing University of Technology (Social Science Edition),2017,16(04):20-27.
[8] Yasmin S ,Moniruzzaman M .Forecasting of area, production, and yield of jute in Bangladesh using Box-Jenkins ARIMA model[J].Journal of Agriculture and Food Research,2024,16101203-.
[9] Yanya X ,Dongxi L ,Chenlong L .Time series prediction via elastic net regularization integrating partial autocorrelation[J].Applied Soft Computing Journal,2022,129
[10] Kudinov A A ,Nousiainen A ,Koskinen H , et al.Single-step genomic prediction for body weight and maturity age in Finnish rainbow trout (Oncorhynchus mykiss)[J].Aquaculture,2024,585740677-.
[11] A Jiang Youyi,Hu Ruijuan. Development and forecast of new energy vehicle market[J]. Automobile and New Power,2020,3(06): 36-40.DOI:10.16776/j.cnki.1000-3797.2020.06.003.
[12] Xu Xiaoqin,Ren Chaoyang. Analysis of the development status and trend of new energy vehicles[J]. Hubei Agricultural Mechanisation,2020,(02):31.