Trajectory Of Development in China's New Energy Vehicle Industry Through Data Analysis and Expectation

Kai Zhou [1], Zihao Li [2], Zhaofeng Liu [2, *]

[1] International College of CQUPT, Chongqing University of Posts and Telecommunications, Chongqing, China, 400065

[2] College of Science, Hebei Agricultural University, Baoding, China, 071001

* Corresponding Author Email: liu2811680636@outlook. com

Abstract. Amidst a global shift towards sustainable transportation, this study conducts an in-depth analysis of data about China's new energy vehicle (NEV) sector from 2013 to 2022. This research primarily evaluates the standards and key factors influencing the NEV industry's evolution. Spearman correlation and decision tree models indicate that the average price of new energy vehicles and the government subsidies for them have the most significant impact on the development of the new energy industry in China. Expanding on these insights, a robust LASSO linear regression model was developed to further explore these dynamics. Additionally, an ARIMA time series model was employed, leveraging historical data to forecast the factors likely to influence the NEV industry in the coming decade. Integrating these forecasts into the initial evaluation model, the study anticipates a positive growth trajectory for China's NEV development, especially between 2023 and 2025. This research not only sheds light on the current state of the NEV industry in China but also provides valuable predictions for its future direction, contributing to the broader understanding of sustainable vehicle evolution in the global context.

Keywords: ARIMA time series; Regression model; New energy vehicle; Sustainability.

1. Introduction

As global awareness of climate change and environmental protection continues to rise, the New Energy Vehicle (NEV) industry is encountering unprecedented opportunities for growth. This is particularly evident in China, one of the world's largest automotive markets, where the development of NEVs not only reflects technological innovation but also mirrors shifts in national policy directives and market demands. This study aims to delve deeply into the development data of China's NEV industry from 2013 to 2022, assessing the standards and factors influencing the industry's growth, and forecasting future trends.

To achieve the research objectives mentioned above, this study utilizes Spearman correlation analysis, decision trees, and the ARIMA model. Currently, these three models have been successfully applied in multiple fields. Yu et al successfully employed Spearman correlation analysis to explore the relationship between the average annual outage losses in China's electric grid and the system's inertia constant[1]. Chen et al applied the decision tree algorithm in the field related to energy storage devices at distributed integrated energy stations[2]. The ARIMA model has shown promising applications in forecasting, and it has already been utilized in areas such as energy load[3], wind power generation[4], and other related fields. Although these algorithms have not yet been applied in the field of new energy vehicles, their successful applications in various other domains have already demonstrated their potential.

In this study, relevant data on new energy vehicles were first collected from various sources, and several potential key influencing factors were identified based on related literature. A Spearman correlation analysis was then conducted on these factors to explore their interrelationships. To delve deeper into the contributions of these factors, a decision tree algorithm was applied for key factor identification. Following ARIMA diagnostics performed on the data, ARIMA forecasting was conducted. Subsequently, the forecasted data was incorporated into a LASSO linear regression model for predictions, ultimately providing insights into the future development trends of China's new energy vehicles.

2. Analysis of Influential Factors in China's New Energy Vehicle Industry

Initially, gather data from the National Bureau of Statistics of China, the China Association of Automobile Manufacturers, etc. Encompassing factors influencing the progress of new energy vehicles and the developmental status of the new energy vehicle sector in China. According to the relative literature, the factors impacting the advancement of new energy vehicles are categorized into political, economic, technological, pricing, environmental protection, and other dimensions[5]. Specifically, politically pertinent factors include China Government Subsidies for New Energy Vehicles (CGSNEV) and the number of new energy vehicle charging piles in China (NEVCP, supporting infrastructure construction). Economically relevant factors comprise the Disposable Income per Capita for Chinese Residents (DIPCCR) and Per Capita expenditure on transportation and communication for Chinese residents (PCETCCR)[6]. Technology-related factors include the Number of Patent Applications for New Energy Vehicles (PANEV), the energy efficiency of new energy vehicles (EENEV), and the maximum range of pure electric new energy vehicles (MRPENEV). Pricing-related factors consist of China electric vehicle charging cost (CEVCC), China fuel prices (CFP), average prices of China electric vehicles (APCEV), and average prices of China gasoline vehicles (APCGV). Environmental protection-related factors involve China’s carbon dioxide emissions (CDE).

$$\rho=\frac{\sum_{i}\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right)}{\sqrt{\sum_{i}\left(x_i-\overline{x}\right)^2\left(y_i-\overline{y}\right)^2}} \tag{1}$$

Herein, establish an assessment framework for the development indicators of China's new energy vehicle industry, utilizing China's new energy vehicle production, sales, market size, and ownership figures from 2013 to 2022 as evaluation criteria. The production of new energy vehicles directly mirrors the industry's manufacturing capabilities and production vigor, while sales and market size are pivotal indicators of market demand and product popularity. Additionally, the quantity of new energy vehicles reflects their long-term penetration rate in the overall vehicle market, serving as a crucial aspect of the industry's sustainable development. Through vigilant monitoring and analysis of these indicators, we can deepen our comprehension of the dynamic evolution of China's new energy vehicle industry, offering robust insights for future decision-making and strategic planning.

To address the aforementioned objectives, this study conducts a Spearman correlation analysis of the factors influencing the development of the new energy vehicle industry. For a sample with a sample size of n, where n raw data points are transformed into ranked data, the correlation coefficient ρ is given by Eq1, the results are shown in figure 1. The examination reveals that barring China’s fuel prices, all variables exhibit a robust correlation with the overall development score. Perform a significance test on China’s fuel prices: The P value is 0.68, exceeding the 0.05 threshold, indicating a lack of statistical significance. Consequently, there is no discernible correlation between China’s fuel prices and the comprehensive development score.

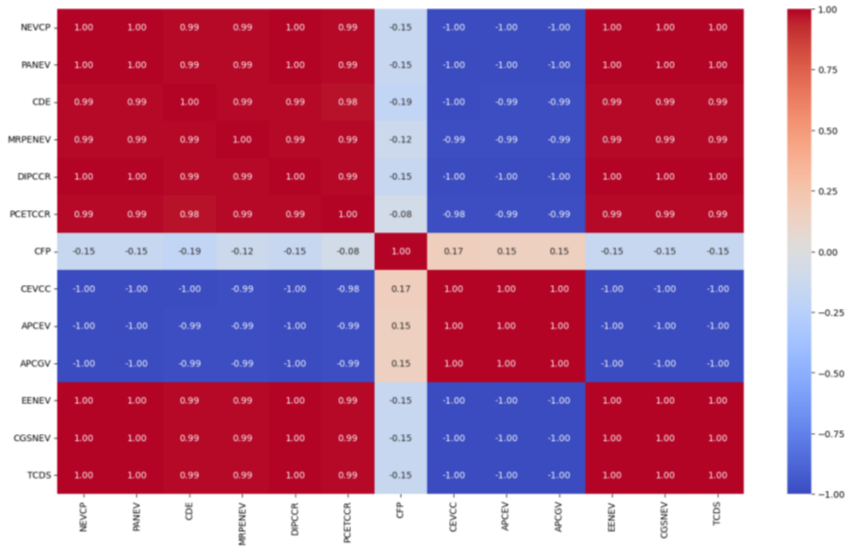

Figure. 1 Results of Spearman coefficient correlation analysis

Furthermore, to gain deeper insights into the contributions of the aforementioned factors[7], this study employs a decision tree model for analyzing the weights of influencing factors. The resulting ranking of factor importance is depicted in Figure 2.

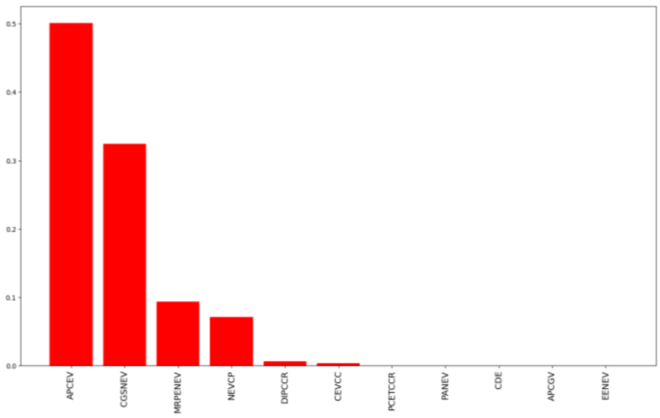

Figure. 2 The weight of influencing factors

The factors influencing the development of new energy vehicles in China ranked in descending order of impact, are as follows: APCEV, CGSNEV, MRPENEV, NEVCP, DIPCCR, CEVCC, and PCETCCR. Notably, the average price of China's new energy vehicles exerts the most significant impact, with government subsidies for new energy vehicles and the maximum cruising range of pure electric new energy vehicles also holding considerable importance. The remaining factors exhibit comparatively lower levels of impact.

3. Forecasting Future Trends of China's New Energy Vehicles

Derived from the findings of the previous part, the seven factors influencing the progress of China's new energy electric vehicle industry are identified: APCEV, CGSNEV, MRPENEV, NEVCP, DIPCCR, CEVCC, and PCETCCR. Building upon the historical data of these seven factors spanning from 2010 to 2022, an ARIMA time series model was established.

The ARIMA model, also known as the differential integrated moving average autoregressive model, also known as the integrated moving average autoregressive model (moving can also be called sliding), is one of the time series forecast analysis methods. In ARIMA (p, d, q), AR is autoregression, p is the number of autoregressive terms; MA is the moving average, q is the number of moving average terms, and d is the number of differences (order) to make it a stationary sequence[8, 9].

The ARIMA (p, d, q) model is an extension of the ARMA (p, q) model. The ARIMA (p, d, q) model can be expressed as:

$$\left(1-\sum_{i=1}^{p}\phi_i L^i\right)(1-L)^d X_t=\left(1+\sum_{i=1}^{q}\theta_i L^i\right)\varepsilon_t \tag{2}$$

Where L is the lag operator，d∈Z, d>0.

Using Python to draw residual plots, density estimation histograms, standard Q-Q plots, and autocorrelation plots and analyze them. As shown in figure 3.

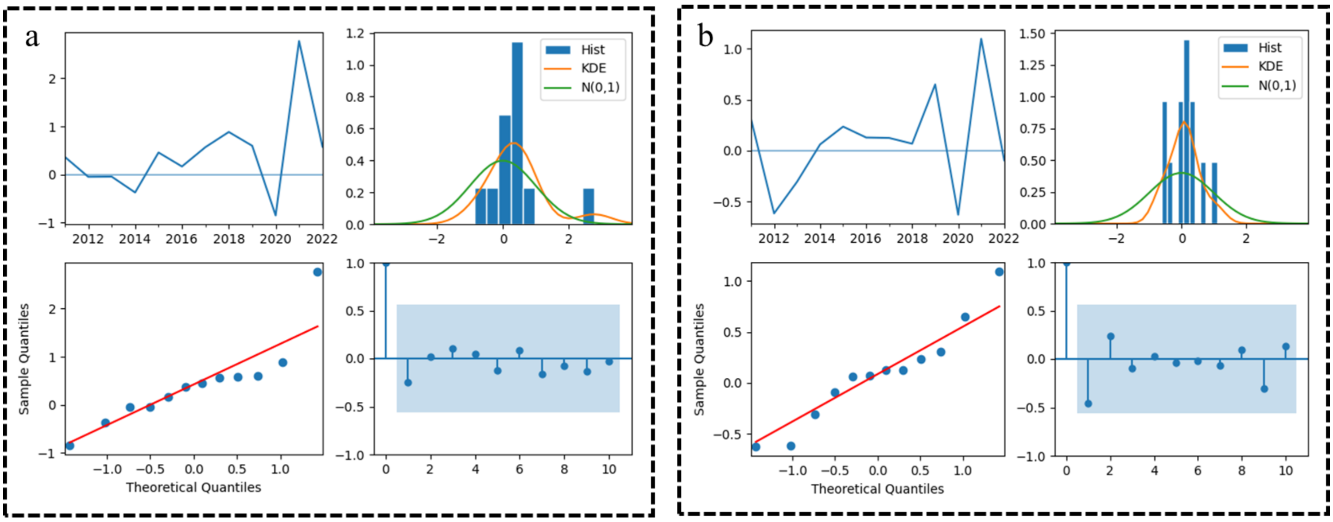

Figure. 3 Schematic graph of ARIMA diagnostics. (a) Maximum cruising range (km) of new electric energy vehicles (b) China’s per capita disposable income (￥)

In the residual plot, the residuals fluctuate around the zero axis, indicating that the fitting effect is good. The density estimation histogram shows that the data conforms well to the original distribution. In the standard Q-Q diagram, the trend of the scatter points coincides with the straight line, proving that the data sequence satisfies the probability distribution. It can be seen from the autocorrelation diagram that the autocorrelation coefficient of the series has always been relatively small and is almost always controlled within the range of twice the standard deviation. The sequence constantly fluctuates near the zero axis, and the line can be considered a random—stationary, solid time series.

Predict the values of seven factors for the development of China's new energy electric vehicle industry in the next ten years. And use Python for visualization, as shown in Figure 4. The figure clearly illustrates that the predicted values closely align with the actual values, indicating an outstanding model-fitting performance.

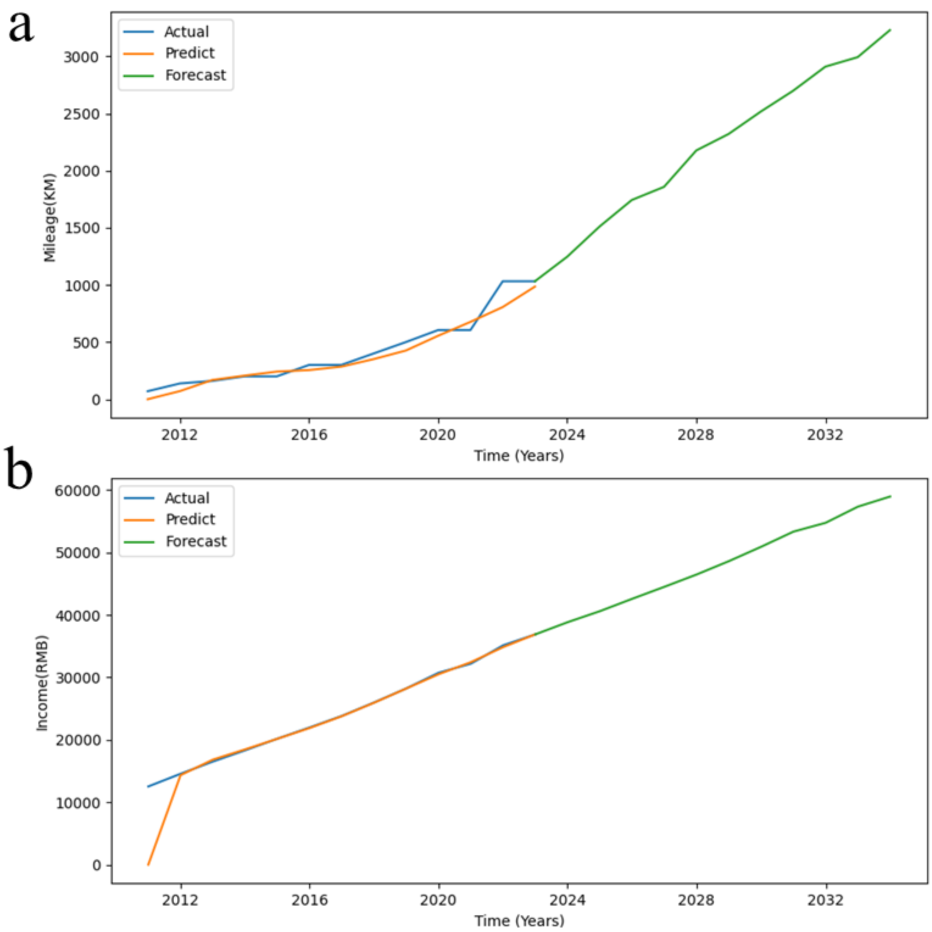

Figure. 4 Schematic graph of ARIMA prediction results. (a) Maximum cruising range (km) of new electric energy vehicles (b) China's per capita disposable income (￥)

The Spearman correlation analysis revealed strong correlations among variables, excluding domestic fuel prices in China. Hence, we employed LASSO regression to mitigate the impact of multicollinearity among independent variables on the regression fit. Using Python, we conducted LASSO regression on four indicators, yielding the corresponding LASSO regression equations. Observations from cross-validation plots and analysis of the regression fit's mean absolute error (MAE) values were made. The MAE values for regression fit with Chinese new energy vehicle (NEV) stock, production, sales, and market size as dependent variables were 2.47, 2.52, 2.58, and 2.70, respectively. These findings indicate a good fit for the regression equations.

The prediction results of the ARIMA model are substituted into the LASSO regression equation into the four indicators respectively, and the predicted values of China's new energy vehicle development indicators are obtained[10]. Use Python for visualization, and the effect is shown in Figure 5. As depicted in the preceding figure, the margin between the fitted and actual values is very small, underscoring the exceptional effectiveness of the model fitting.

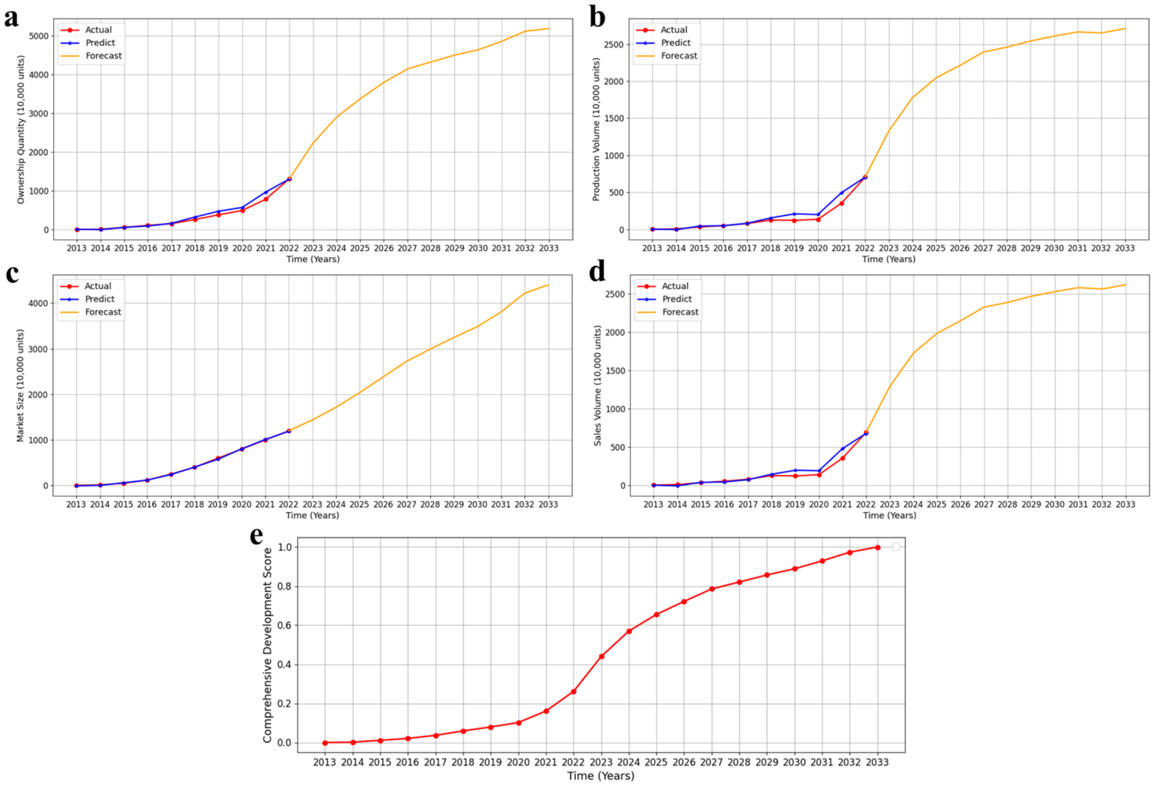

Figure. 5 Schematic diagram of the prediction results. (a) China's New Energy Vehicle Ownership. (b) China's new energy vehicle production. (c) China's new energy vehicle market size. (d) China's new energy vehicle sales. (e) Development status of new energy vehicles in China

As seen from the above figure, from 2023 to 2033, the development situation of China's new energy vehicles will improve, and the comprehensive development index will gradually increase. Among them, 2023 to 2025 will be the year of significant development of China's new energy vehicles, and the extensive development will advance by leaps and bounds and then enter the next eight years' Steady growth stage.

4. Conclusions

This study delves into the analysis of China's new energy development data spanning from 2013 to 2022, specifically focusing on evaluating the standards and influencing factors shaping the new energy vehicle industry. Spearman correlation coefficient analysis identified the average price of China’s new energy vehicles as the most significant factor, with government subsidies and the cruising range of pure electric vehicles also playing crucial roles. Furthermore, we established a robust LASSO linear regression model that performed well. An ARIMA time series model was developed using historical data to predict the factors influencing the development of China's new energy vehicles over the next decade. Substituting these results into the previous model, our predictions suggest a positive trajectory for new energy vehicle development in China, particularly from 2023 to 2025.

References

[1] YU Q, HUO X, HE J, et al. Trend Prediction of Power Blackout Accidents in Chinese Power Grid Based on Spearman's Correlation Coefficient and System Inertia [J]. Proceedings of the Chinese Society of Electrical Engineering, 2023,43(14): 5372-5380.

[2] CHEN H, WU G L, ZHANG L Y, et al. Data Collection Method for Energy Storage Device of Distributed Integrated Energy Station based on Double Decision Tree [J]. INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS, 2022,13(7): 527-534.

[3] MPAWENIMANA I, PEGATOQUET A, ROY V, et al. A comparative study of LSTM and ARIMA for energy load prediction with enhanced data preprocessing: 2020 IEEE SENSORS APPLICATIONS SYMPOSIUM (SAS 2020) [Z]. 15th IEEE Sensors Applications Symposium (SAS): 2020.

[4] BISWAS A K, AHMED S I, BANKEFA T, et al. Performance Analysis of Short and Mid-Term Wind Power Prediction using ARIMA and Hybrid Models: 2021 IEEE POWER AND ENERGY CONFERENCE AT ILLINOIS (PECI)[Z]. IEEE Power and Energy Conference at Illinois (PECI): 2021.

[5] HUI HE L J. HOW CHINA PUT NEARLY 5 MILLION NEW ENERGY VEHICLES ON THE ROAD IN ONE DECADE[EB/OL]. https://theicct.org/how-china-put-nearly-5-million-new-energy-vehicles-on-the-road-in-one-decade/.

[6] HANDWIKI. New Energy Vehicles in China [EB/OL]. https://handwiki.org/wiki/Engineering: New_energy_vehicles_in_China.

[7] FU M X, ZHANG C, HU C F, et al. Achieving Verifiable Decision Tree Prediction on Hybrid Blockchains [J]. ENTROPY, 2023,25(7).

[8] LU S, LI Y, TIAN Y, et al. Prediction of epidemic trends of Ebola virus during regional armed conflicts based on the ARIMA model [J]. Military Medical Sciences, 2023,47(6): 417-421, 426.

[9] CHEN X X, JIANG Z S, CHENG H, et al. A novel global average temperature prediction model--based on GM-ARIMA combination model [J]. EARTH SCIENCE INFORMATICS, 2023.

[10] WU X, ZHANG Y, GAO Y, et al. Relationships between chemical components and sensory indexes of flue-cured tobacco based on Lasso regression model [J]. Tobacco Science & Technology, 2023,56(12): 49-57.