

Research on the Development of New Energy Electric Vehicles in China Based on Correlation Analysis and Ridge Regression Models

Bowei Cui^{*, #}, Xuzeng Cao[#], Zunming Shen[#]

Nanjing University of Information Science and Technology, Nanjing, China, 210044

* Corresponding Author Email: Bowei_Cui@126.com

[#]These authors contributed equally.

Abstract. This paper analyses the main factors affecting China's new energy electric vehicles and their influencing factors. Firstly, it is necessary to collect data, construct indicators, analyse the correlation between indicators, then construct a ridge regression model to obtain the correlation coefficient of each indicator, and finally compare the size of the absolute value of the normalised correlation coefficient to analyse the degree of influence of each indicator. It is found that the development of new energy electric vehicles in China is influenced by four areas: policy, technology, economy and market. Eleven factors were selected as sub-indicators, including policy subsidies, number of policies, battery life, energy efficiency ratio, number of patents, GDP per capita, petrol price, average price of new energy electric vehicles, charging cost, industry chain size and number of charging piles. The sales volume as the dependent variable reflects the impact of the development of new energy electric vehicles, the higher the sales volume, the better the development of new energy electric vehicles. Data analysis and modelling were carried out through Spearman correlation analysis, ridge regression model construction and other methods to derive the degree of influence of each indicator on sales volume. The results show that the number of charging piles has the greatest impact on the development of new energy electric vehicles in China, while the energy efficiency ratio has the least impact. The ridge regression model predicts sales better, but performs poorly under special circumstances.

Keywords: Correlation Analysis, Regression Modelling, New Energy Vehicles.

1. Introduction

In the rapidly evolving landscape of transportation, the adoption and impact of new energy electric vehicles (NEEVs) in China have become pivotal areas of investigation. Understanding the intricate web of factors influencing NEEVs is essential for policymakers^[1], industry experts, and researchers alike. This study delves into a comprehensive analysis of these factors, employing a structured approach that encompasses data-driven methods, correlation analysis, and regression modeling. By meticulously collecting data from various sources and selecting key indicators across policy, technology, economy, and market domains, we aim to unravel the intricate relationships driving NEEV development. Through rigorous correlation analysis, we seek to uncover the hidden dynamics between these indicators and their collective influence on NEEV sales, a crucial metric reflecting the industry's trajectory. Central to our methodology is the construction of a robust ridge regression model, tailored to handle the complexities and interdependencies within the dataset. This model not only enables us to quantify the impact of each factor but also provides insights into the relative importance of these factors in shaping NEEV sales trends^[2]. Through this multidimensional approach, we endeavor to contribute valuable insights that not only illuminate the current state of NEEVs in China but also inform strategic decision-making for sustainable transportation policies and industry practices^[3].

2. Analyze the main factors affecting new energy electric vehicles in China and their impacts.

Find out the factors of new energy electric vehicles in China, and their influence on the development. Firstly, we need to find the data, construct the indicators, analyze the correlation between the indicators, and then construct the ridge regression model to get the correlation coefficient of each indicator, and finally compare the absolute size of the normalized correlation coefficient to analyze the degree of influence of each indicator^[4].

2.1. Model building

2.1.1. Selection of indicators

The factors affecting China's new energy electric vehicles involve four areas: policy, technology, economy, and market, so we take the four areas as primary indicators, from which we select eleven factors as secondary indicators, as shown in Table 1 below.

Table 1. China's new energy electric vehicle influencing factor indicator system

Primary indicators	Secondary indicators	Unit	Description
Policy	Subsidy	ten thousand RMB/car	Subsidy amount per vehicle
	Policy quantity	/	The number of policies
Technology	Battery life	km	NEEC driving distance
	Energy Efficiency Ratio	%	Efficiency of NEEV divided by FV's
	Patent quantity	/	The number of Patents
Economy	Per capita GDP	RMB	GDP divided by population
	Gasoline price	RMB/t	one ton needed to fuel prices
Economy	Average price of NEEV	ten thousand RMB/car	The sale price of a NEEV
	Charging cost	RMB/kwh	One kilowatt-hour consumes RMB
Market	Industrial chain scale	/	Number of NEEV enterprises
	Charging pile quantity	/	The number of charging piles

The sales of new energy electric vehicles are used as the dependent variable to reflect the impact of the development of new energy electric vehicles, with higher sales volume indicating better development of new energy electric vehicles^[5-6].

2.1.2. Data collection

Part of 11 secondary indicators and sales data collected through domestic and foreign websites over the years are shown in Table 2 below.

Table 2. Influential Factors of New Energy Electric Vehicles in China

Month	Sales	Subsidy	Policy quantity	Battery life	Energy Efficiency Ratio	Patent quantity
2013-1	0.17	6.00	5.00	200.00	0.35	1234.00
2013-2	0.16	6.00	5.23	204.46	0.35	1232.46
2013-3	0.16	6.00	5.44	208.88	0.35	1238.69
2013-4	0.20	5.99	5.65	213.27	0.35	1252.29
2013-5	0.15	5.97	5.84	217.64	0.35	1272.85
2013-6	0.15	5.96	6.03	221.97	0.35	1299.98
				⋮		
2022-8	480.00	2.01	34.89	920.65	0.37	26859.35
2022-9	540.00	1.97	35.17	946.68	0.37	27370.32
2022-10	447.00	1.92	35.45	973.90	0.37	27881.54
2022-11	535.00	1.86	35.72	1002.33	0.37	28392.08
2022-12	713.00	1.80	36.00	1032.00	0.37	28901.00
2013-1	467.87	8830.00	35.00	1.20	230.00	0.50
2013-2	472.69	8830.00	34.87	1.20	233.65	0.58
2013-3	477.63	8520.00	34.78	1.20	236.40	0.66
2013-4	482.69	8125.00	34.72	1.20	238.34	0.73
2013-5	487.86	8220.00	34.70	1.20	239.54	0.80
2013-6	493.12	8125.00	34.69	1.20	240.10	0.86
				⋮		
2022-8	1107.25	9740.00	13.74	1.12	613.94	428.47
2022-9	1102.44	9725.00	13.55	1.12	617.75	452.37
2022-10	1093.22	9620.00	13.37	1.12	621.69	475.69
2022-11	1079.19	9775.00	13.18	1.12	625.77	498.28
2022-12	1060.00	9160.00	13.00	1.12	630.00	520.00

2.1.3. Data analysis

To test the correlation between the independent variables and the independent variables and the relationship between the independent variables and the dependent variable, we use Spearman correlation analysis^[7].

Spearman correlation analysis: The Spearman correlation coefficient is a nonparametric measure of ranked values. The degree of relationship between two variables is described using a monotonic function.

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \quad (1)$$

where ρ stands for the Spearman correlation coefficient and n represents the sample size, and n raw data are converted into grade data.

VIF: The VIF value represents the severity of multicollinearity and is used to test whether the model exhibits covariance or not^[8].

The heat map of the correlation coefficients is drawn through Python as shown in Figure 1 below.

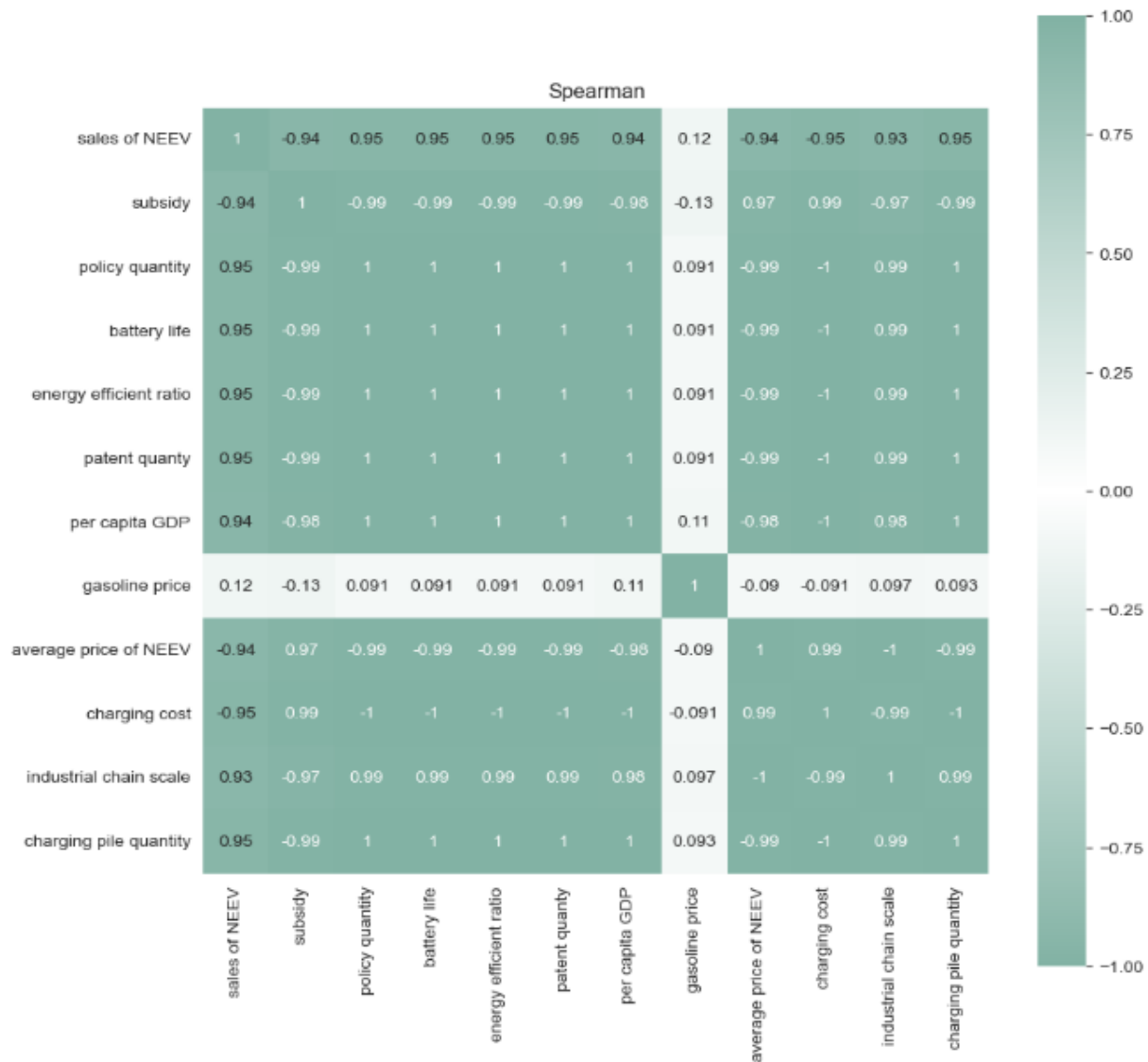


Figure 1. The heat map of the correlation coefficients

We find strong correlations between independent variables and independent variables, and between independent variables and dependent variables, and $VIF > 10$.

2.1.4. Construction of ridge regression model

Due to the high correlation between the independent variables and $VIF > 10$, we construct the ridge regression model.

Ridge Regression Model: Ridge regression is a biased estimation regression method dedicated to the analysis of covariate data, is a modified least squares estimation method, by giving up the unbiasedness of the least squares method, at the cost of losing some information and reducing the precision to obtain regression coefficients are more realistic and reliable regression method, the fit for pathological data is stronger than the least squares method. Below are the specific steps^[9-10].

Step 1: Input indicator data independent variable X (Subsidy, Policy quantity, Battery life, Energy Efficiency Ratio, Patent quantity, Per capita GDP, Gasoline price, Average price of NEEV, Charging cost, Industrial chain scale, Charging pile quantity) and dependent variable Y (sales) for normalization.

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (2)$$

Step 2: By means of the ridge trace plot (figure 2), the K value was determined, $K = 0.123$.

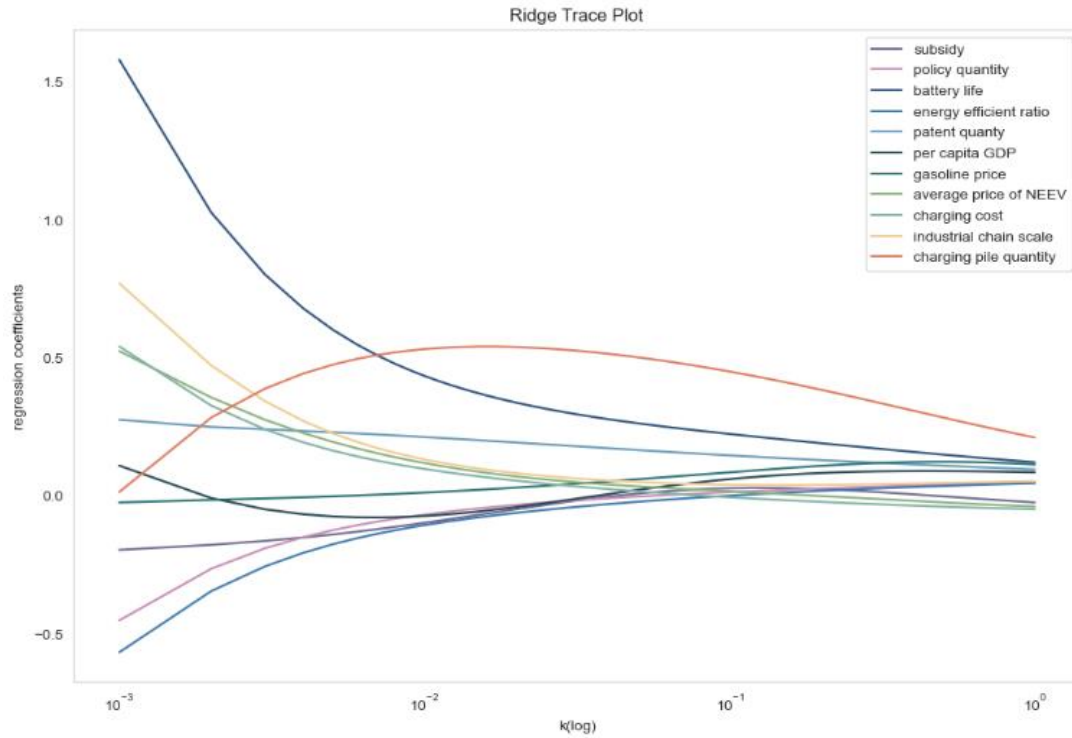


Figure 2. Ridge Trace Plot

Step 3: By analyzing the F value, whether the analysis model is meaningful ($P < 0.05$), if it is significant, it indicates that there is a regression relationship. Then the fitting of the model is analyzed by R^2 values. The significance of X is analyzed to explore the influence of X on Y.

Step 4: Combined with the value of the regression coefficient β_j , compare and analyze the degree of influence of X on Y. The final model equation is obtained.

$$y = \sum_{j=1}^p \beta_j x_j + \beta_0 \quad (3)$$

$$\hat{\beta}^{\text{bridge}} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^N \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_i \right)^2 + \lambda \sum_{j=1}^p \beta_j^2 \right\} \quad (4)$$

where β_j is Ridge regression formula coefficients.

2.2. Solution of the Model

2.2.1. Model result

Table 3. Results of ridge regression correlation coefficients

K=0.123	Non-standardized coefficient	Standardized coefficient
Constant	-61.491	-
Subsidy	12.71	0.03
Policy quantity	9.09	0.019
Battery life	134.801	0.215
Energy Efficiency Ratio	3.742	0.007
Patent quantity	71.584	0.142
Per capita GDP	37.986	0.07
Gasoline price	57.341	0.094
Average price of NEEV	5.472	0.013
Charging cost	-4.76	-0.01
Industrial chain scale	18.065	0.04
Charging pile quantity	275.367	0.433

We calculate the ridge regression correlation coefficient results as shown in Table 3 above.

Based on the F-test significance p-value of 0.000***, which presents significance at the level, the original hypothesis is rejected, indicating that there is a regression relationship between the independent variables and the dependent variable. At the same time, the model's goodness of fit R^2 is 0.888, the model performance is excellent. We thus derive the model regression equation:

Sales Volume (Thousand Units) = $-61.491 + 12.71 \times \text{Subsidy} + 9.09 \times \text{Policy quantity} + 134.801 \times \text{Battery life} + 3.742 \times \text{Energy Efficiency Ratio} + 71.584 \times \text{Patent quantity} + 37.986 \times \text{Per capita GDP} + 57.341 \times \text{Gasoline price} + 5.472 \times \text{Average price of NEEV} - 4.76 \times \text{Charging Cost} + 18.065 \times \text{Industrial chain scale} + 275.367 \times \text{Charging pile quantity}$.

The absolute value of the standardized correlation coefficient of the indicators can assess the degree of influence on the sales volume of new energy electric vehicles. The higher the standardized correlation coefficient, the higher the correlation with sales volume, i.e., the greater the influence on the development of new energy electric vehicles in China.

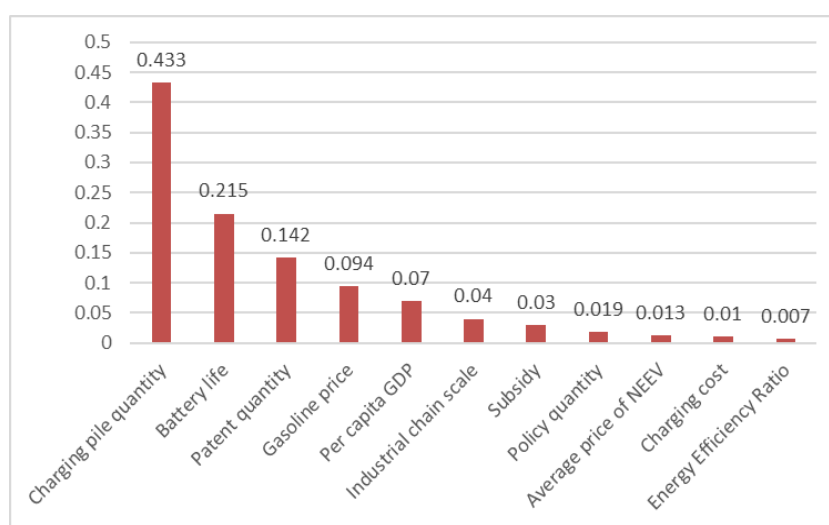


Figure 3. Standardized Correlation Coefficient Absolute Value Ranking Chart

The absolute values of the correlation coefficients of the indicators are ranked, as shown in Figure 3 above. (Charging pile quantity>Battery life>Patent quantity>Gasoline price>Per capita GDP>Industrial chain scale>Subsidy>Policy quantity>Average price of NEEV>Charging cost>Energy Efficiency Ratio). The results show that Charging pile quantity has the greatest impact on the development of China new energy electric vehicles. The energy efficiency ratio has the least impact.

2.2.2. Analysis of the results of model

We compare the regression equation of ridge regression to calculate the predicted sales with the actual statistical sales, plotted in Figure 4 below.

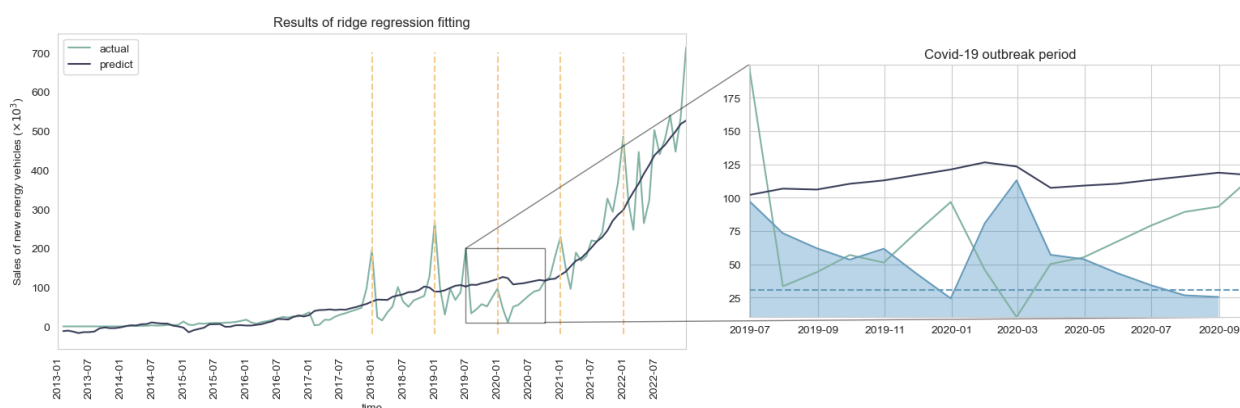


Figure 4. Results of ridge regression fitting

The overall effect is good, the model performs excellently, where the ridge regression does not make a good fit to the yellow dashed part and the epidemic period because of the end of the year paychecks and the epidemic shock.

3. Conclusions

Through a comprehensive analysis and modelling of the factors contributing to the development of new energy electric vehicles (NEEVs) in China, this study draws a number of important conclusions.

First, we find that policy, technology, economy and market are the key areas that influence the development of NEEVs. Among these areas, the number of charging piles has the most significant impact on NEEVs sales, while energy efficiency ratios have a relatively small impact. Second, by performing Spearman correlation analysis and VIF test on the data, we confirmed the correlation and multicollinearity problems among the indicators. This provided the basis for building a reliable regression model. Further, we constructed a model based on ridge regression, and determined the degree of influence of each factor on the sales of NEEVs by fitting the regression equation and analysing the absolute value of the standardised correlation coefficient. The results showed that the number of charging piles, battery life, and the number of patents had a greater influence on the sales of NEEVs. Finally, we compare the established model with the actual sales and demonstrate the effectiveness of the model fit through graphs. The results show that our model performs well in predicting the sales of NEEVs and has a high accuracy on the changing trends of the industry.

Taken together, this study provides an important reference and decision support for understanding and forecasting the NEEVs market in China. Our analytical methods and results can help policymakers, industry experts, and researchers better formulate sustainable development strategies and promote the healthy development of the NEEVs industry.

References

- [1] Wang Pin. Development status of new energy vehicles in China and countermeasure suggestions [J]. Automobile Practical Technology, 2024, 49 (08): 187-191.
- [2] Luo Jian,Zhao Erniu. Research on the Development of New Energy Vehicle Industry under the Background of Dual Carbon [J]. China Business Journal, 2024, (08): 128-131.
- [3] WU Weiguo. Development and application of intelligent manufacturing technology for new energy vehicles [J]. Heavy Duty Vehicle, 2024, (02): 43-44.
- [4] Qin Sbao. Development and planning of charging infrastructure for new energy vehicles [J]. Automotive Maintenance Technician, 2024, (08): 39-41.
- [5] Jingxin Li. Marketing Strategy of New Energy Vehicle Enterprises under the Concept of Sustainable Development--Taking BYD Auto as an Example [J]. Today's wealth, 2024, (12): 17-19.
- [6] Liu Tong. How to expand the advantage of new energy vehicle development? How to expand the development advantage of new energy vehicles?[J]. Automotive, 2024, (04): 42-46.
- [7] Zhu Lan,Gao Bai,Huang Mai. The "venture capital development" government and the rise of new energy vehicles in Hefei [J]. Culture, 2024, (02): 102-111.
- [8] ZHOU Shixing, WANG Jinwen, SHI Jingjing. Seeking technological breakthroughs to create an industrial highland[N]. Guangming Daily, 2024-04-09 (010).
- [9] Pu Zhenyu. Listed three months in Beijing is still no car Chery Fengyun build shop high requirements to slow development? [N]. Economic Observer, 2024-04-08 (022).
- [10] JIN Hui,JU Wenxia,HU Gou,et al. Changsha County automobile industry rides on new track[N]. Hunan Daily, 2024-04-05 (001).