The Relationship between New Energy Vehicle Price and Aftermarket-Parts Price Based on The GLM Regression Analysis

Shouchen Ding[1], Tianli Zu[2], Puchao Li[3], Bing Xue[4]

dingshouchen@catarc.ac.cn[1], zutianli@catarc.ac.cn[2], lipuchao@catarc.ac.cn[3], xuebing@catarc.ac.cn[4]

China Automotive Technology and Research Center Co.,ltd,Tianjin China

Abstract. Since the penetration rate of the new energy vehicles (NEVs) are gradually increasing in the vehicle market, the industry of NEVs have already entered an era of fast growth. This research aims to figure out the relationship between the price of NEVs and the price of aftermarket parts which is based on the CATARC self-owned data , that is collected from the real market of the China's NEV, via the methodology of the GLM Regression. The factors that are expected to affect vehicle prices includes the vehicle-brands, vehicle grade,vehicle body-structure, the differences of loading-part prices between separate categories, etc.

Keywords: New energy vehicles(NEVs); Price; Fittings; GLM

1. Introduction

There are many factors that is affecting the price of the vehicles in the new energy vehicles (NEV) market including the raw materials cost, the labor cost, the rental cost of the factories, the liabilities costs or the advertisement costs, etc. To be more specific, Tong Yang,Chen Xing, Xinyu Li(2021)[1] have pointed that according to the quantitative evaluation, more political implemented in China the lower the vehicle price will be reached and more willingness the customers would have to pay for the vehicle. Meanwhile, Guopan Zhang(2022)[2] has once mentioned that the quality and the goodwill are two of the most important factors that has the positive affect the on vehicle prices. However, there are fewer researches or researchs about the effect caused by the Aftermarket Parts-Pricing is scarce.Therefore this research is going to analyse the effect caused by the Aftermarket Parts-Pricing based on the CATARC self-owned data via GLM regression.

2. The Analysis for the current situation of NEV market in China

2.1. The positive policies drives the industry to the phase of rapid growth

The Ministry of Finance, the State Administration of Taxation, and the Ministry of Industry and Information Technology have issued the "Announcement on Extending and Optimizing the New Energy Vehicle Purchase Tax Reduction and Exemption Policy" on June 21. The purchase tax reduction and exemption policy for new energy vehicles purchased in China from

PMBDA 2023, December 15-17, Nanjing, People's Republic of China

DOI 10.4108/eai.15-12-2023.2345341

2024 to 2027 has been stipulated, and new energy vehicles have once again ushered in major benefits.

It is going to be a massive promotion for the Chinese new-energy vehicle brands in the global market[10] since the EU lawmakers voted to ban the sale of ICEVs by 2035. From Table 1 we can see that not only the Chinese NEV market but also the global NEV market will be on a rapidly growth track.

Table 1. The Prohibition Schedule for ICEV Brands

Category	Brand	Prohibition Year	Prohibition Schedule Announcement
Chinese Brand	Build Your Dream	2022	3rd Apr 2022
Chinese Brand	Hainan Mazda	2025	2025
Over Sea Brand	Bentley	2030	2030
Over Sea Brand	MINI	2030	2030
Joint Brand	Mercedes Benz	2030	Conditional prohibition by 2030
Joint Brand	Volvo	2030	2030

In the recent years, the market of China's NEV have been through a rapid development period[11]. The penetration rate for NEV in China market were increased by the guidance from the governmental support as well as the fiscal policies since 2010. The total consumption amount of the new energy passenger vehicle reached 2.94 million, 39.9%YOY and 28.3% as the penetration rate. The total amount of vehicle-on-road reached 14.68million, 46.6%YOY.

2.2. The current situation for the NEV aftermarket sector

There are so many changes in the aftermarkets service and maintenance process for NEVs, compared with ICEVs, due to technical differences. Therefore, the new changes about the common parts rang, the repair method and the pricing strategies shows up.

There are obvious differences between NEV and ICEV parts, which affect the vital process during the maintenance. Firstly, NEVs use battery packs and electric drive systems as power sources, while ICEVs use internal combustion engines, which leads to a big difference between the two in key components. Secondly, NEVs have fewer maintenance items and longer maintenance intervals with cheaper cost. For example, per 60,000 kilometers, the maintenance cost of the fuel version of the BYD Tang is 2.5 times higher then the pure electric version.Thirdly, due to the rapid development of the new energy vehicle market, the supply chain of core component, such as batteries, will be balanced at the oligopoly point on the supply side of related technologies and accessories,with a very high velocity, which could affects the price of new energy vehicle parts as well.

3. Methodology

3.1. Purpose and significance

The current researches in the vehicle pricing area are mostly focusing on the pricing strategies affected by manufacturing capitals rather then the aftermarket products[3]. Therefore, in this research, we are looking forward to study between the vehicle pricing and the aftermarket products pricing about their relationship and the factors that affects them.

3.2. Data Introduction

The Vehicle Brands Data: It Includes different vehicle brands that categorized based on the production nationalities such as Japanese, German or Chinese brands so that we can compare the pricing of different brands of different categories of vehicles and the impact on aftermarkets parts pricing; **The Vehicle Model and Configuration Data:** It includes more than 300 data fields such as complete vehicle body parameters, motor power, battery capacity etc., for different models. It could help to launch the characteristic analysis in different dimensions;**The Vehicle Price Data:** Vehicle price data is one of the vital data in this research and it is updated basing on the official guidance price in the CATARC "Vehicle-Parts" Data System, which could ensure the timing and efficiency; **The Vehicle Level Data:** It is specifically divided into 5 different ends, including lower-end, lower-to-mid-end, mid-end, mid-to-higher-end, and higher-end. In this way, it is possible to better analyze the impact of the pricing relationship between the whole vehicle and aftermarkets parts under different market positioning; **The Parts Price Data:** It includes data fields such as brand, parts encoding, parts prices, relying on the high-frequency iteration data maintained by China Automotive Technology and Research Center to ensure accurate parts price data and improve the reliability of model analysis results.**The Parts Adaptability Data:** The actual loading of accessories on the vehicle models such as the minimum dimension, covering the standard name of the accessory,etc. Accurate accessory fitting data is the core foundation to ensuring the effectiveness of the Parts Price Data; **The Parts Category Data:** It is the classification of parts basing on different attributes, competitiveness characteristics, frequency of use, and installation location. This research is focusing on the price of the power system parts, common maintenance parts, safety parts, appearance parts, interior parts, and configuration parts and their impact on the pricing of the entire vehicle;

3.3. Data process and model establishment

The total amount of the detailed distribution parts data for each vehicle are approximately 3,000 pieces on average. To avoid the possible problems during the modeling, such as over-long models, a unclear conclusions or difficult processing, the overall data was decided to be analyzed basing on the vehicle model dimension, and the price data of parts was grouped by the power system[4], the common-use maintenance parts, the safety parts, the exterior parts and the interior parts. Meanwhile, the vehicle model structure fields, the vehicle model level and the brand classifications are treated as virtual variables[5]. Furthermore, the arithmetic average price in each data group of different vehicle models were calculated as well[6]. Last but not least, to reduce the impact of extreme values on the regression results and follow the linear correlation & normal distribution assumptions, the logarithm was also launched for each group of parts price.

Table 2. Descriptive statistics of proposed variables

Variable name	Obs	Mean	Std. Dev	Min	Max
log_nsjpjj	422	7.970293	1.229167	4.19268	11.04628
log_dlxtpjj	422	10.39454	1.006145	6.945436	12.40726
log_wgjpjj	422	7.013579	0.9175657	3.89182	8.938427
log_cywbjpjj	422	5.123163	0.6028114	3.931826	6.461078
loc.pzjpjj	422	7.000355	1.475628	3.57655	9.576025
MPV	422	0.0165877	0.1278721	0	1
zxSVV	422	0.021327	0.1446435	0	1
zdxMPV	422	0.0308057	0.172996	0	1
zdxSUV	422	0.0663507	0.2491896	0	1
zdxc	422	0.021327	0.1446435	0	1
xxSUV	422	0.0876777	0.2831613	0	1
xxc	422	0.0521327	0.2225584	0	1
wxc	422	0.1800948	0.3847225	0	1
jcxSUV	422	0.1966825	0.397962	0	1
jcxc	422	0.2843602	0.4516446	0	1
zdd	422	0.6137441	0.4874683	0	1
zgd	422	0.1540284	0.3614045	0	1
rx	422	0.0236967	0.152283	0	1
mx	422	0.0592417	0.2363567	0	1
zz	422	0.6943128	0.4012447	0	1
hx	422	0.0023697	0.0486792	0	1

According to Table 2, there are 21 candidate independent variables have been chosen to join the initial model where the top 5 of them were parts group prices, 6th to 15th were vehicle structures,16th to 17th were vehicle levels, 18th to 21st were brand categories. The standard deviation is relatively acceptable and the distribution of the data is relatively concentrated.

3.4. Model design and independent variables selection

3.4.1. The GLM model

The Generalized Linear Model (GLM) assumes that the conditional probability distribution belongs to the family of exponential distribution[7].The exponential distribution satisfies the property of $E[y \mid x] = c'(\theta)$, $\mathrm{Var}(y \mid x) = c''(\theta)$, but the connection function of GLM g: $R \to R$ is used to covert the conditional expectation into a linear function of the independent variable. The Standard linear regression model is one of the GLM that assumes the conditional probability from dependent variable to independent variables are normally distributed. The connection function $g = id$ is an identity function which is $E[y \mid x] = \beta^T x$.

3.4.2. The model settings

This research is going to use the multiple linear regression model after the consideration of the relation between independent variables basing on factors that effect the vehicle price.The format of the model is assumed as following:

$$VP = Cons + \beta_1 Log\_dlxtpjj + \beta_2 Log\_cywbpjj + \beta_3 Log\_pzjpjj + \beta_4 \{Veh\_Type\} + residual \quad (1)$$

VP : vehicle price; **Cons**: constant; $\boldsymbol{\beta_i}$ **:** coefficients for independent variables

3.4.3. Independent variables

Table 3. Results of correlation test between parts group prices

	log_ns~j	log_dl~j	log_wg~j	log_cy~j	log_pz~j
log_nsjpjj	1.0000				
log_dlxtpjj	0.4376	1.0000			
log_wgjpjj	0.4351	0.3963	1.0000		
log_cywbjpjj	0.2924	0.2887	0.5171	1.0000	
log_pzjpjj	0.4231	0.295	0.4742	0.3994	1.0000

Before the establishment of the model, the correlation test between independent variables have been launched and the resulted with a strong correlation between part price groups(Table 3). Meanwhile the very first model we built, which involved every candidate independent variables, was ended with a very large p-value, after the F-test, and the collinearity was exists as well. Therefore to further optimize the model, we adopt forward stepwise regression, gradually introducing variables, observing changes in the goodness of fit of the model, retaining the newly introduced variables if R^2 improving, and excluding the variable when drop, until the optimal model is determined.

3.5. Establishment of the Model

Take the variable “Price of power system parts” as an example, once the log_ Dlxtpjj (average Price of power system parts Group) is introduced into the model with 0.05 as significant level, all price variables are significant, with a value of 0.8151. The fitting degree is improved by 0.0033, indicating a good fitting effect (Table 4). This variable is retained. When introducing log_ After nsjpjj (interior component group price), given a significance level of 0.05, log_ The nsjpjj is not significant, and there is no significant improvement in fit (Table 5. right), so the model excludes the price variable of interior component groups.

Table 4. The regression result with log_dlxtpjj

Source	ss	df	ms
Model	2.8799e+12	19	1.5157e+11
Residual	6.5337e+11	402	1.6253e+09
Total	8.5333e+12	421	8.3925e+09

AUTO_PRICE	Coef.	Std. Err.	t	P>|t|	[95%Conf.	Interval]
Log_pzjpjj	4899.886	1720.579	2.85	0.005	1517.43	8282.342
log_cywbjpjj	13931.98	5596.599	2.49	0.013	2929.724	24934.24

log_dlxtpjj	6262.429	2351.672	2.66	0.008	1639.317	10885.54
MPV	89148.65	22442.79	3.97	0.000	45028.76	133268.54
zxSVV	35478.05	17033.81	2.08	0.038	1991.586	68964.51
zdxMPV	170881.8	17339.72	9.85	0.000	136794	204969.7
zdxSUV	66522.12	16415.32	4.05	0.000	34251.53	98792.71
zdxc	98951.24	19667.85	5.03	0.000	60286.54	137615.9
xxSUV	-57233.29	15834.31	-3.61	0.000	-88361.69	-26104.9
xxc	-47840.52	16342.86	-2.93	0.004	-79968.66	-15712.38
wxc	-87709.01	14722.26	-5.96	0.000	-116651.2	-58766.77
jcxSUV	-2916.011	14140.56	-0.21	0.837	-30714.7	24882.68
jcxc	-42374.06	13629.88	-3.11	0.002	-69168.8	-15579.31
zdd	-11239.03	13490.48	-0.83	0.405	-37759.74	15281.68
zgd	61608.49	13288.31	4.64	0.000	35485.24	87731.74
rx	55317.51	14079.21	3.93	0.000	27639.44	82995.59
mx	13577.08	15894.03	0.85	0.393	-17669.72	44822.89
zz	-10134.38	11608.78	-0.87	0.383	-32955.88	12687.12
hx	39924.65	41068.67	0.97	0.332	-40811.53	120660.8
_Cons	43005.59	42094.42	1.02	0.308	-39747.1	125758.3

$\mathbf{R^2}$ = 0.8151; Adj R^2 = 0.8063

Table 5. The regression result with log_nsjpjj

Source	ss	df	ms
Model	2.812e+12	20	1.4406e+11
Residual	6.5204e+11	401	1.6260e+09
Total	3.5333e+12	421	8.3925e+09

AUTO_PRICE	Coef.	Std. Err.	t	P>|t|	[95%Conf.	Interval]
log_pzjpjj	4519.936	1771.367	2.55	0.011	1037.611	8002.261
log_cywbjpjj	13679.76	5604.775	2.44	0.015	2661.347	24698.17
log_dlxtpjj	5353.587	2557.34	2.09	0.037	326.1185	10381.06
log_nsjpjj	2154.509	2379.071	0.91	0.366	-2522.499	6831.518
MPV	90143.94	22474.7	4.01	0.000	45960.99	134326.9
zxSVV	36155.21	17054.02	2.12	0.035	2628.756	69681.66
zdxMPV	171524	17358.1	9.88	0.000	137399.8	205648.3
zdxSUV	66290.62	16420.98	4.04	0.000	34008.65	98572.59
zdxc	97931.17	19704.48	4.97	0.000	59194.19	136668.2
xxSUV	-54531.16	16116.47	-3.38	0.001	-86214.49	-22847.84
xxc	-45231.95	16598.36	-2.73	0.007	-77862.62	-12601.28
wxc	-86487.71	14787.18	-5.85	0.000	-115557.8	-57417.64

jcxSUV	-3058.773	14144.61	-0.22	0.829	-30865.62	24748.07
jcxc	-41951.39	13640.92	-3.08	0.002	-68768.03	-15134.75
zdd	-12857.76	13611.38	-0.94	0.345	-39616.34	13900.81
zgd	59775.93	13444.44	4.45	0.000	33345.54	86206.32
rx	54889.79	14090.28	3.90	0.000	27189.75	82589.84
mx	10759.64	16199.14	0.66	0.507	-21086.21	42605.5
zz	-10236.19	11611.92	-0.88	0.379	-33064.03	12591.66
hx	45128.17	41477.77	1.09	0.277	-36412.88	126669.2
_Cons	40045.04	42230.57	0.95	0.344	-42975.92	123066

$\mathbf{R^2}$ = 0.8155; Adj R^2 = 0.8063

There are 15 independent variables were chosen after the forward stepwise regression, where include different types of vehicles to see the effect , including the MPV, the medium-large MPV(zdxMPV), the compact SUV(jcxSUV), the small SUV(xxSUV), the medium SUV(zxSUV), the medium-large SUV(zdxSUV), the compact passenger car(jcxc), the small passenger car(xxc), the minicar(wxc),the medium-large car(zdxc). Then we finished the VIF test to make sure that the multicollinearity is not exist.

4. Model Results and Data Analysis

4.1. Model Results

Table 6 The result of multiple linear regression model

Source	ss	df	ms
Model	2.8761e+12	15	1.9174e+11
Residual	6.5715e+11	406	1.6186e+09
Total	3.5333e+12	421	8.3925e+09

AUTO_PRICE	Coef.	Std. Err.	t	P>|t|	[95%Conf.	Interval]
log_dlxtpjj	6071.923	2324.647	2.61	0.009	1502.074	10641.77
log_cywbjpjj	12219.64	4511.912	2.71	0.007	3350.01	21089.26
log_pzjpjj	5509.47	1627.441	3.39	0.001	2310.208	8708.732
MPV	90374.92	17306.39	5.22	0.000	56353.6	124396.2
zxSVV	37693.75	15641.88	2.41	0.016	6944.569	68442.93
zdxMPV	172574.4	12396.95	13.92	0.000	148204.2	196944.6
zdxSUV	68508.84	8864.422	7.73	0.000	51082.95	85934.73
zdxc	100374.8	15048.53	6.67	0.000	70792.03	129957.6
xxSUV	-56120.55	8437.812	-6.65	0.000	-72707.8	-39533.29

xxc	-45007.36	10136.27	-4.44	0.000	-64933.49	-25081.24
wxc	-88238.16	7221.396	-12.22	0.000	-102434.2	-74042.16
jcxc	-39215.56	5565.143	-7.05	0.000	-50155.66	-28275.47
zgd	73888.63	6860.831	10.77	0.000	60401.44	87375.82
rx	54710.65	13615.44	4.02	0.000	27945.09	81476.21
zz	-21753.15	5116.677	-4.25	0.000	-31811.64	-11694.66
_Cons	47778.58	32123.94	1.49	0.138	-15371.43	110928.6

$\mathbf{R^2}$ = 0.8140; Adj R^2 = 0.8071

After the further adjustments, the final multiple linear regression is resulted in the form below:

$$VP = \underset{(2.61)}{6071.923}\text{log\_dlxtpjj} + \underset{(2.71)}{12219.64}\text{log\_cywbjpjj} + \underset{(3.39)}{5509.47}\text{log\_pzjpjj} + \underset{(5.22)}{90374.92}MPV + \underset{(2.41)}{37693.75}zxSUV + \underset{(13.92)}{172574.4}zdxMPV + \underset{(7.73)}{68508.84}zdxSUV + \underset{(6.67)}{100374.8}zdxc - \underset{(-6.65)}{56120.55}xxSUV - \underset{(-4.44)}{45007.36}xxc - \underset{(-12.22)}{88238.16}wxc - \underset{(-7.05)}{39215.56}jcxc + \underset{(10.77)}{73888.63}zgd + \underset{(4.02)}{54710.65}rx - \underset{(-4.25)}{21753.15}zz + \underset{(1.49)}{47778.58} \quad (2)$$

n = 422 $R^2 = 0.8151$

According to the test result, the model is acceptable since the R^2 value is relatively high (81.4%) which mean it fits the relationship between the dependent and independent variables well. Meanwhile the model can also explain that the multiple linear relationship between dependent and independent variables is significant since the P-value of the F test is lower then 0.01. Furthermore, the t test also shows the positive result, except the constant, under 5% significant level for every independent variables (P values <0.05)(Table 6).

For the accuracy of GLM regression, it is necessary to strengthen the processing of multi-variable and multi-class classification problems by using the SVM. Therefore we have launched the SVC by using different kernels including liner, RBF(γ= 0.5 and γ= 50 respectively). Meanwhile we classified our data by the 4 different price ranges(P_i) which was divided into P_1<=100k CNY, 100k CNY<P_2<=300k CNY, 300k CNY<P_3<=500k CNY and 500k CNY<P_4 CNY. To be more specified, the model was launched by the one to one-set way and the final significance of SVC resulted with following :

1) Linear regression function SVM: 67.5%

2) RBF kernel function(γ = 0.5): 70.1%

3) RBF kernel function(γ = 50): 70.1%

According to the results we can see that the SVM model have showed a relatively significant results to support the point of view which made by the multiple linear regression model[9].

5. Discussion and Conclusion

5.1. The Relationship between aftermarket parts prices and vehicle prices

According to the model result we can see that the power system parts,common maintenance parts,configuration parts are significantly effecting the vehicle price in a positive way. To be more specific, every 1% price variation of the configuration parts will drive the vehicle price to increase about 55 CNY. This effect are also resulted in the common maintenance parts and the power system parts, where 1% increase will drive 122 CNY and 61CNY respectively.

5.2. The Relationship between the vehicle level and vehicle price

Ceteris paribus, the types of vehicles also shows a positive relation between the vehicle prices. The middle-large MPV(zdxMPV) have the highest price which is 260812 CNY higher then the minicar(wxc) on the average.

It is worth noting that, except the compact SUV(jcxSUV), the overall prices (including medium to large, medium, compact, small, and micro cars) are showing a decreasing trend.

The vehicle level also show a positive relation between the vehicle prices at the same time. The overall price of mid to high-end cars is approximately 73889 yuan higher than that of mid to low-end cars on average.

5.3. Other possible factors would effect the vehicle price

There are many other factors that haven't been involved in this model, for example the market situation, the position of the vehicle in the market, the political atmosphere,etc[8]. Those factors might be the key point to effect the vehicle price as well, but not been focused in this research. Therefore, they could be the deeper study direction in the future researches.

5.4. Conclusion

To sum up, this research is based on the self built data resource system of China National Automobile Corporation(CATARC), and uses multiple linear regression analysis to analyze the data of multiple vehicle models to deeply explore the influencing factors of the overall price of new energy vehicles. The empirical research results indicate that the terminal price of aftermarkets spare parts, vehicle type, and vehicle level all have a significant impact on the overall vehicle price, and show a positive correlation, which is consistent with the general view of the industry. In the subsequent research process, Zhongqi Data will provide more accurate pricing strategy support for the automotive industry by analyzing the correlation between market environment, vehicle positioning, and policy environment with the pricing of finished vehicles and aftermarkets products.

References

[1] Tong Yang, Chen Xing, Xinyu Li, Evaluation and analysis of new-energy vehicle industry policies in the context of technical innovation in China, Journal of Cleaner Production, Volume 281,2021,25126,ISSN 0959-6526, https://doi.org/10.1016/j.jclepro. 2020.125126(2021).

[2] Guopan Zhang, Chunling Liu, The Optimal Dynamic Analysis of New Energy Vehicle Price

Based on the Influence of Quality and Goodwill[J],Journal of Wuhan Textile University(2022)
[3] Yongchao Wang.The Analysis of the Current Situation for China's New Energy Vehicle development[J].New Energy Technology(2022)
[4] Xu.J,Chen.F, Ma.L,Tripartite Evolutionary Game Analysis of New Energy Vehicle Promotion Policy from the Perspective of Prospect Theory[J].Science and Technology Management Research,、(2022)
[5] Zhiqiang Luan, The Research on Automobile Pricing Strategy Based on Hedonic Price Model, China University of Mining and Technology, 2009
[6] Zhang.G,Wang.G,Song.X.Product pricing and operating decision of manufacturers under the sharing.Journal of Shandong University (Natural Science),2021
[7] Hong Gui,Multiple Linear Regression Model Analysis of Pricing in a Maturity Market, JOURNAL OF SHANGHAI UNIVERSITY (NATURAL SCIENCE), 2004
[8] The Research about the Influencing Factors of China's Vehicle Prices Based on the Hedonic Price Theory,Prices Monthly(2018)
[9] Ziying.Y, Manping.T.Welfare Analysis of Government Subsidy Programs for Fuel-Efficient Vehicles and New Energy Vehicles in China[J].Environmental and Resource Economics(2019)
[10] Qingyun.N, Lihui.Z, Zihao.T, Hubacek Klaus. Energy.Strategies for applying carbon trading to the new energy vehicle market in China: An improved evolutionary game analysis for the bus industry[J](2022)
[11] Panos M. Pardalos, Xiaoxi.Z, Minglun.R, Guangdong.W, Jun.P. Promoting new energy vehicles consumption: The effect of implementing carbon regulation on automobile industry in China[J].Computers & Industrial Engineering (2019)