International Journal of Innovation in Science and Mathematics
Volume 7, Issue 3, ISSN (Online): 2347–9051

Multiple Regression Models Based on Gray Correlation Analysis and Factor Analysis Predict the Ownership of Civil Automobiles

Liu Yingjie
Department of mathematics, yanbian university, yanji, China.
Corresponding author email id: 15011027040@163.com

Date of publication (dd/mm/yyyy): 18/06/2019

***Abstract* – With the development of China's economy, the ownership of civil automobiles is increasing. The research and analysis of the civil automobile ownership is helpful for us to make a judgment on the trend of China's macro economy and provide reference for the development of the automobile industry.**

In this paper, grey correlation analysis was used to select the impact indicators, and factor analysis was used to construct the economic growth factor (Fac1), energy and environment factor (Fac2) and household consumption factor (Fac3). Then, stepwise regression was used to obtain the prediction model. The final model contains 8 economic indicators in total, including two factors, energy and environment factor and resident consumption factor, which greatly improves the disadvantage that regression analysis in other literatures can only get 2-3 explanatory variables in the end, also can provide a good explanation for the change of civilian automobile ownership. Compared with the time series method and grey prediction method, the method in this paper also ensures high accuracy.

In this paper, a combined method, namely, selection of gray correlation analysis, factor analysis to construct impact factors, and regression analysis to obtain the final model, can not only predict China's civilian car ownership, but also explain the influencing factors.

***Keywords* – Grey Relational Analysis, Factor Analysis and Stepwise Regression.**

I. INTRODUCTION

With the rapid development of China's economy, the number of civilian cars is increasing. A series of promotion measures, such as bank mortgage loan to buy a car, also promoted the increase of civilian automobile ownership. The latest data from China's national bureau of statistics shows that the number of civilian cars in the country's big cities continues to rise sharply. The consumption of civil automobiles in China has developed rapidly in recent years, and its proportion in the overall consumption of automobiles is increasing year by year, which has an important impact on the automobile industry and its related industries.

Research on civil car ownership has a lot of, but most of the research [1] [2] is simply to civil car ownership univariate time series method is used to forecast, or neural network to predict, just the pursuit of the precision of prediction, although to do so are more accurate results, but can't reveal its internal factors. However, the regression analysis method can only select 2-3 indicators for analysis because most economic indicators have a strong correlation. In this paper, grey correlation analysis is used to select indicators, different impact factors are constructed, and then regression analysis is carried out to explain the comprehensiveness and ensure a certain prediction accuracy.

II. ECONOMIC INDICATOR SELECTION

By searching data and literature [3], this paper selects more than ten indicators such as gross national income and GDP. Then the grey relational degree is analyzed. Since most economic indicators are strongly correlated,

indexes with gray correlation degree greater than 0.8 are selected and other economic indicators are deleted. The details are as follows :

Table 1. Grey Relational Table of Selected Indexes

Economic indicators	Grey correlation
Gross national income	0.9610
GDP	0.9325
Added value of the secondary industry	0.9153
Total population	0.9002
Highway mileage	0.8831
Gasoline production	0.8750
Steel production	0.8628
GDP per capita	0.8530
Consumer level	0.8434
Total energy consumption	0.8361
Investment in environmental pollution control	0.8287
Added value of the tertiary industry	0.8254

III. Data Selection

Since the beginning of the 21st century, China's economy has entered the fast lane of development, with rapid growth of various economic indicators. Even though the country has made fine-tuning in recent years, the growth rate is far higher than that of the last century. Therefore, the data was selected as the data of the new century.

In addition, since the highway mileage was started in 2005, China's national bureau of statistics added the mileage of village roads into the highway mileage. In order to increase the reliability of the model, this paper selected relevant data from 2005 to 2018. The data from 2005 to 2016 were used for analysis and modeling, while the data from 2017 and 2018 were used for verification. (Data source: China National Statistical Year Book).

Table 2. Raw Data Table.

Year	Civilian Vehicle Ownership (y)	Gross National Income ($x1$)	GDP ($x2$)	Added Value of The Secondary Industry ($x3$)	Total Population ($x4$)	Highway Mileage ($x5$)
2005	3159.66	185998.9	187318.9	88084.4	130756	334.52
2006	3697.35	219028.5	219438.5	104361.8	131448	345.7
2007	4358.36	270704	270092.3	126633.6	132129	358.37
2008	5099.61	321229.5	319244.6	149956.6	132802	373.02
2009	6280.61	347934.9	348517.7	160171.7	133450	386.08
2010	7801.83	410354.1	412119.3	191629.8	134091	400.82
2011	9356.32	483392.8	487940.2	227038.8	134735	410.64

Year	Civilian Vehicle Ownership (y)	Gross National Income ($x1$)	GDP ($x2$)	Added Value of The Secondary Industry ($x3$)	Total Population ($x4$)	Highway Mileage ($x5$)
2012	10933.09	537329	538580	244643.3	135404	423.75
2013	12670.14	588141.2	592963.2	261956.1	136072	435.62
2014	14598.11	642097.6	641280.6	277571.8	136782	446.39
2015	16284.45	683390.5	685992.9	282040.3	137462	457.73
2016	18574.54	737074	740060.8	296547.7	138271	469.63

Table 2. Raw Data Sheet (Continued).

Year	Gasoline Production ($x6$)	Steel Production ($x7$)	GDP per Capita ($x8$)	Consumer Level ($x9$)	Total Energy Consumption ($x10$)	Investment in Environmental Pollution Control ($x11$)	Added Value of the Tertiary Industry ($x12$)
2005	5409.22	37771.14	14368	5771	261369	2388	425912.1
2006	5594.76	46893.36	16738	6416	286467	2566	383373.9
2007	5989.42	56560.87	20505	7572	311442	3387.3	346178
2008	6347.54	60460.29	24121	8707	320611	4937.03	308082.5
2009	7195.48	69405.4	26222	9514	336126	5258.39	277979.1
2010	7676.04	80276.58	30876	10919	360648	7612.19	244852.2
2011	8158.06	88619.57	36403	13134	387043	7114.03	216120
2012	8976.07	95577.83	40007	14699	402138	8253.46	182058.6
2013	9834.04	108200.5	43852	16190	416913	9037.2	154762.2
2014	11029.85	112513.1	47203	17778	425806	9575.5	136823.9
2015	12103.56	112349.6	50251	19397	429905	8806.3	115784.6
2016	12932.03	113460.7	53935	21228	436000	9219.8	91759.7

IV. Index was Selected for Factor Construction

Firstly, according to the similarity of these 12 indicators, this paper divides them into three aspects: $A1$ (economic growth), including $x1, x2, x3$ and $x12$; $A2$ (infrastructure, energy and environment), including $x5$, $x6$, $x7$, $x10$ and $x11$; $A3$ (resident purchasing power), including $x4, x8$ and $x9$.

Due to the strong correlation between these indicators, direct regression analysis will inevitably lead to multicollinearity problems, which will bring trouble for subsequent processing. If only one principal component can be extracted directly from principal component analysis, the explanation of residents' car ownership will be insufficient. Therefore, factor analysis was carried out for each aspect in this paper, and the main factors were extracted, respectively denoted as $Fac1$ of economic growth factor, $Fac2$ of energy and environment factor and $Fac3$ of household consumption factor. The relationship between each factor and its economic indicators is as follows:

$$Fac1 = 0.99818x1 + 0.99822x2 + 0.9975x3 - 0.99614x12$$

$$Fac2 = 0.99738x5 + 0.96571x6 + 0.99647x7 + 0.99407x10 + 0.97757x11$$

$$Fac3 = 0.99853x4 + 0.99919x8 + 0.99819x9$$

The explanatory degree of each factor to the economic indicator variables included is shown in the following table:

Table 3. Factor interpretation scale.

$Fac1$	$x1$	$x2$	$x3$	$x12$	
	0.9963	0.9964	0.9949	0.9922	
$Fac2$	$x5$	$x6$	$x7$	$x10$	$x11$
	0.9947	0.9325	0.9929	0.9881	0.9556
$Fac3$	$x4$	$x8$	$x9$		
	0.9970	0.9983	0.9963		

As can be seen from the above table, each factor extracted has the lowest degree of interpretation of the index variables contained, reaching more than 95%, indicating that the selected factors reflect the variable information well. The 12 selected economic indicators can be well illustrated using the three constructed factors.

V. The Factors of Economic Growth, Energy and Environment and Household Consumption were Analyzed

For the extracted three factors: economic factor, energy and environment factor and household consumption factor data, logarithmic transformation processing was carried out first, and then stepwise regression analysis [4] was carried out. The following logarithmic regression equation can be obtained:

$$y = -692224 - 20669 \ln Fac2 + 80238 Fac3$$

Then the significance test was carried out for the regression equation, $F = 903.8$, corresponding P value was 4.24e-11, and the adjusted determination coefficient was 0.9939, so it could be concluded that the model of logarithmic linear regression equation was an appropriate judgment. The corresponding regression coefficient test results are shown in the following table 4, According to the test results, the constructed factors $Fac2$ and $Fac3$ have passed the test of regression coefficient and are very significant. So it makes sense to choose these two factors.

Table 4. Regression coefficient test table.

variable	Estimate	Std. Error	T-statistic	P-value	significant
	-692224	27668	-25.019	1.25e-09	significant
$Fac2$	-20669	2943	-7.023	6.17e-05	significant
$Fac3$	80238	5306	15.122	1.05e-07	significant

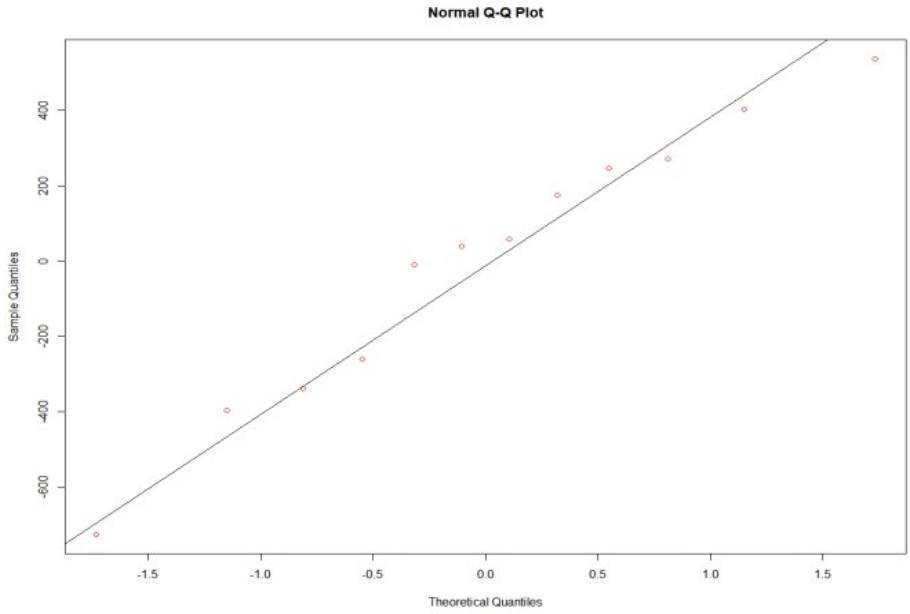

Fig. 1. Regression residual Q-Q diagram

As can be seen from the figure, the residual is approximately distributed around the straight line, indicating that the residual conforms to the assumption of normality. Therefore, it is reasonable to use this model to predict the ownership of civil automobiles.

VI. Comparative Analysis of Different Prediction Methods

According to the above model, the growth trend of the economic indicators included in the energy and environmental factors and household consumption factors that affect the ownership of civil automobiles was observed, and appropriate methods were selected for fitting, and the predicted values in the next five years were predicted.

Indicators x4, x5, x8 and x9 show a very stable linear growth trend, and the predicted value can be obtained by linear fitting with one variable. For x6, x7, x10 and x11, there is an increasing trend, but no obvious linear trend. Therefore, the method of time series is selected for fitting prediction.

The predicted values of various economic indicators obtained are calculated according to the above process, and finally the predicted values of civil automobile ownership are obtained. In order to compare the accuracy of prediction, this paper calculates the predicted values of ARIMA (0, 2, 0) model [5], quadratic moving average method and gray prediction method [6] [7] respectively for comparative analysis. The following table [8] [9] :

Table 5. Comparison Table of Prediction Results of Different Methods

year	real value	The regression model	ARIMA (0, 2, 0)	Grey prediction	Quadratic moving average method
2107	20906.67	20058.98	20062.08	21149	20299
2018	23122	23347.42	23154.72	23817	22206
error		4.05%	4.03%	1.16%	2.90%
		0.97%	0.14%	2.57%	3.96%

VII. CONCLUSION

From the above table, we can know that the multiple logarithmic regression model obtained through the use of gray correlation analysis to select indicators and factor analysis method to construct factors is no less accurate in the prediction than those methods that directly predict the civil automobile ownership [10]. Moreover, such a model can also explain the growth of civil automobiles through the selected indicators, so as to find the main reasons for the growth and provide a more accurate understanding of the future development of the industry.

From the positive and negative of the two constructed energy and environmental factors and residents' consumption factors, it can be seen that the increase of energy consumption and environmental governance will restrict residents' car purchase. Drivers of car purchases are population growth and rising per capita income. This is a good reference for the auto industry to make future development policies.

REFERENCES

[1] Ma jing. Econometric model analysis and Suggestions on urban automobile ownership in chongqing [J]. Modern commercial industry, 2018, v. 39(22):27-29.
[2] Zhu shenglei. Thoughts on the development of car ownership [J]. Shanghai automotive industry corporation, 2018, 340(12):28-30.
[3] Zhong Weizhou, Wang Jun, Factor Analysis of Factors Affecting China's Civil Automobile Ownership and its Policy Implications, Statistics and Information BBS, vol. 23, no. 11, 16-21, 2008.
[4] You Haolin, R language Prediction Practice, Beijing: China Industry and Information Technology Publishing Group, 2016, 358-367.
[5] Gao Tao, Xiao Nan, Chen Gang, R Language Practice, Beijing: People's Posts and Telecommunications Press, 2013, 180-197.
[6] Tang Lifang, Jia Dongqing, Meng Qingpeng, Grey Prediction GM11 Model Based on MATLAB, Journal of Cangzhou Teachers College, vol.24, no.2, 35-40, 2008.
[7] Huang Xianxian, Wang Fengxiao, Matlab Implementation of Multi-Variable Grey Prediction Model Algorithm, Journal of Sichuan University of Science and Technology, volume 21, issue 1, 45-47, 2008.
[8] Wang yunxin, wang yunxin, jiao jianling, et al. Short-term combined prediction of vehicle ownership in China [J]. Journal of hefei university of technology (social science edition), 2014, 28(5):41-48.
[9] Gu jibao, rong min, wu jian-lin. Combined prediction of vehicle ownership in China [J]. Journal of industrial technology and economics, 2010, 29(4).
[10] Yuan kun, lu jiang. Prediction of peak vehicle ownership in yunnan province [J]. Value engineering, 2017(11).

AUTHOR'S PROFILE

Liu Yingjie, is a student majoring in statistics, department of mathematics, school of science, yanbian university.