CORRELATION ANALYSIS OF INFLUENCING FACTORS ON TOTAL RETAIL SALES OF SOCIAL CONSUMER GOODS IN CHINA

Zhang Mi

Collage of Economics and Management,
Nanjing University of Aeronautics and Astronautics, China

ABSTRACT

The total retail sales of social consumer goods represent the total amount of consumer goods that each industry provides to residents and social organizations through commodity circulation channels, and it is an important part of China's national economy. The changes in the total retail sales of social consumer goods can reflect the changes in the domestic retail market and the economic prosperity. From the purpose of maintaining a stable and healthy development of the economy, our country's current focus is to expand domestic demand, especially consumer demand. Based on the economic data of the past ten years in our country, this paper chooses eight index variables that affect the total retail sales of social consumer goods, including retail price index of goods, household consumption level, per capita disposable income of urban residents, per capita net income of rural residents, year-end population, gross domestic product, the Engel coefficient for urban residents 'consumption, and Engel coefficient for rural residents' consumption. Then Pearson correlation analysis, partial correlation analysis, rank correlation analysis and grey correlation analysis were carried out. Comparing the results of these four analytical methods, the following conclusions are obtained: The total retail sales of consumer goods in China are mainly related to GDP, household consumption level, per capita disposable income of urban residents, and per capita net income of rural residents. However, the relationship between the total retail sales of social consumer goods and the retail price index, year-end population, the Engel coefficient of urban residents' consumption and the Engel coefficient of rural residents' consumption are relatively weak.

Keywords: Pearson correlation analysis, partial correlation analysis, rank correlation analysis, grey correlation analysis.

1. INTRODUCTION

Since the reform and opening up, China's economic development has made great achievements. The total economic output has achieved leaping development. Consumption, investment and net

exports of goods and services are also called the "three carriages" to boost economic growth, which has greatly promoted the economic growth. Investment and export play a major role in promoting economic development. However, under the current domestic economic situation, both exports and investment have cooled down. Expanding domestic demand has become the key to promoting economic growth, especially the expansion of consumption. The retail industry is a link between production and consumption. It not only promotes consumption but also directly affects the prosperity of the consumer market to a large extent.

As an important index to measure the consumption level of Chinese residents, scholars mainly focus on qualitative analysis of its influencing factors from a theoretical perspective, but quantitative studies on it are few. In recent years, scholars began to try to use a variety of different methods for empirical research. Liu Y N (2007) analyzed the data of China in recent 20 years by using the measurement software, and pointed out that the per capita net income of rural households is the main factor that affects people's consumption behavior. Farina (2010) used multivariate linear regression to test the factors influencing the total retail sales of social consumer goods. The cointegration test shows that there is a long-term stable equilibrium between the total retail sales of social consumer goods and its influencing factors. Peng B et al. (2012) made multiple linear regression analysis on the influencing factors of total retail sales of social consumer goods in Yangtze River Delta, and concluded that "increasing people's income substantially is the key to improving the overall consumption level of the whole society". Chen W H (2015) has carried out an empirical analysis of Guangdong on the basis of the multi linear regression model from the point of view of the region.

It has been proved that any model has some limitations, and the changing process of the trend of social consumption demand is influenced by a variety of factors. Therefore, this article takes the total retail sales of social consumer goods as the research object, and uses the four methods of gray correlation analysis, Pearson correlation analysis, partial correlation analysis and rank correlation analysis to explore the influencing factors of total retail sales of social consumer goods. By comparing these four methods, we can draw the main factors and the secondary factors that affect the total retail sales of social consumer goods in China.

2. CORRELATION ANALYSIS

2.1 Variable selection

Total retail sales of social consumer goods refer to the total amount of consumer goods supplied to residents and social groups by wholesale and retail trade, catering, manufacturing and other industries of various economic types. It reflects the speed of currency circulation, the people's life, the purchasing power of social consumer goods and so on. It is usually used to express the

total consumption level of a country. In real economic life, there are many factors affecting the total retail sales of social consumer goods. However, in terms of direct impact, there are mainly two levels (sales and purchase) and three subjects (residents, retail sectors and social groups). Therefore, based on the related theories of consumer economy, the explanatory variables are retail price index of goods (X_1), household consumption level(X_2), per capita disposable income of urban residents(X_3), per capita net income of rural residents(X_4), year-end population(X_5), gross domestic product(X_6), the Engel coefficient for urban residents' consumption(X_7), and Engel coefficient for rural residents' consumption(X_8). The explanatory variable is the total retail sales of social consumer goods(Y). This article uses the statistical data from 2006 to 2015, all from the data in the network statistics database. Due to the difference between the dependent variable and the independent variable's dimension and magnitude, we usually remove the unit limitation of data before the correlation analysis, that is data standardization. The standardized data is shown in Table 1.

Table 1: Data of each index after standardization

T	Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
2006	-1.199	0.1666	-1.214	-1.374	-1.2	-1.532	-1.292	1.472	1.742
2007	-1.056	-0.785	-1.06	-1.141	-1.048	-1.152	-1.116	0.331	1.15
2008	-0.881	-0.69	-0.868	-0.864	-0.854	-0.81	-0.897	-0.7	0.279
2009	-0.648	0.6426	-0.592	-0.506	-0.588	-0.474	-0.55	-0.13	0.314
2010	-0.305	1.6421	-0.286	-0.229	-0.343	-0.141	-0.213	1.701	0.523
2011	-0.017	-1.737	-0.053	0.1194	-0.061	0.1791	-0.025	0.103	-0.42
2012	0.3762	0.3094	0.2849	0.4249	0.3308	0.4959	0.399	-0.81	-0.38
2013	0.811	1.1662	0.844	0.7798	0.7853	0.8141	0.8999	-0.13	-0.63
2014	1.2373	-0.214	1.2644	1.2192	1.2603	1.1447	1.2244	-0.24	-1.01
2015	1.6816	-0.5	1.6799	1.5706	1.7178	1.4748	1.5698	-1.61	-1.57

2.2 Pearson correlation analysis

Pearson correlation analysis, also known as Pearson product difference correlation analysis, is the most common correlation coefficient calculation method. The Pearson correlation coefficient is a statistic for measuring the relationship (linear correlation) between the two variables X and

Y. It is often expressed as r or ρ. The absolute value of r indicates the degree of correlation between two variables. The larger the absolute value, the stronger the correlation. The general range of values is between [-1, +1]. The Pearson correlation coefficient for X and Y is calculated as follows:

$$r_{XY} = \frac{\sum\limits_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum\limits_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum\limits_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

Where X_i and Y_i are the sample observations of the variables X and Y. $\bar{X}$ and $\bar{Y}$ are the mean values of the samples of the variables X and Y.

Combined with the specific circumstances of the study, we selected 8 variables as the factors that affect the total retail sales of social consumer goods, and analyze the Pearson correlation with the total retail sales of social consumer goods respectively. The results are shown in Table 2.

Table 2: Pearson correlation test

	Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
Y	1	.034	.999***	.995***	.999***	.984***	.997***	-.594*	-.938***
	( )	(.926)	(.000)	(.000)	(.000)	(.000)	(.000)	(.070)	(.000)
X_1	.034	1	.048	.039	.030	.054	.083	.384	.126
	(.926)	( )	(.895)	(.915)	(.935)	(.883)	(.820)	(.273)	(.729)
X_2	.999***	.048	1	.994***	.999***	.983***	.997***	-.588*	-.938***
	(.000)	(.895)	( )	(.000)	(.000)	(.000)	(.000)	(.074)	(.000)
X_3	.995***	.039	.994***	1	.993***	.997***	.997***	-.597*	-.958***
	(.000)	(.915)	(.000)	( )	(.000)	(.000)	(.000)	(.068)	(.000)
X_4	.999***	.030	.999***	.993***	1	.981***	.996***	-.607*	-.940***
	(.000)	(.935)	(.000)	(.000)	( )	(.000)	(.000)	(.063)	(.000)
X_5	.984***	.054	.983***	.997***	.981***	1	.992***	-.598*	-.965***
	(.000)	(.883)	(.000)	(.000)	(.000)	( )	(.000)	(.068)	(.000)

X_6	.997*** (.000)	.083 (.820)	.997*** (.000)	.997*** (.000)	.996*** (.000)	.992*** (.000)	1 ( )	-.578* (.080)	-.941*** (.000)
X_7	-.594* (.070)	.384 (.273)	-.588* (.074)	-.597* (.068)	-.607* (.063)	-.598* (.068)	-.578* (.080)	1 ( )	.746** (.013)
X_8	-.938*** (.000)	.126 (.729)	-.938*** (.000)	-.958*** (.000)	-.940*** (.000)	-.965*** (.000)	-.941*** (.000)	.746** (.013)	1 ( )

Note: the value in brackets is the corresponding p value; ***, *** and * respectively represent 1%, 5% and 10% level.

When the significant level is 0.01, the factors that positively correlated with the total retail sales of social consumer goods are: household consumption level, per capita net income of rural residents, GDP, per capita disposable income of urban residents and year-end population, and their Pearson correlation coefficients are 0.999, 0.999, 0.997, 0.955 and 0.984, both of which are higher than 0.9. In addition, their P values are close to zero, showing a high correlation. The Engel coefficient of rural residents 'consumption and the total retail sales of social consumer goods have a significant negative correlation with the significance level of 0.01. The Pearson correlation coefficient is 0.938 and the P value is close to 0. The coefficients of commodity retail price index and the Engel coefficient for urban residents 'consumption are 0.034 and 0.594. The P values are 0.926 and 0.07, which are all greater than 0.05. It indicates that the correlation between these two factors and the total retail sales of social consumer goods is not significant.

2.3 Partial correlation analysis

Partial correlation refers to the relationship between two variables after eliminating the influence of other variables. Using the sample data to calculate the partial correlation coefficient reflects the degree of partial correlation between the two variables. After controlling the linear function of the variable X_3, the first-order partial correlation coefficient between X_1 and X_2 is defined as:

$$r_{12(3)} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}}$$

Therefore, under the control of other factors, we use SPSS software to make further partial correlation analysis on the total retail sales of social consumer goods and its influencing factors, which essentially reflect the importance of the various factors. The results are shown in Table 3.

Table 3: Partial correlation coefficient and sorting of each variable

Variable	Partial correlation coefficient	Score	Sort
X_1	-0.730	0.730	3
X_2	-0.821	0.821	1
X_3	-0.348	0.348	7
X_4	0.764	0.764	2
X_5	0.412	0.412	6
X_6	0.443	0.443	5
X_7	-0.015	0.015	8
X_8	0.594	0.594	4

Based on partial correlation analysis, it can be seen that the correlation between each factor and the total retail sales of social consumer goods are as follows: household consumption level, per capita net income of rural residents, retail price index of goods, Engel's coefficient of rural residents' consumption, GDP, year-end population, per capita disposable income of urban residents and Engel coefficient for urban residents' consumption. Among them, partial correlation coefficient and Pearson correlation coefficient of the four factors of retail price index of goods, household consumption level, per capita disposable income of urban residents and Engel coefficient for rural residents' consumption have a greater change. For example, the partial correlation coefficient between the per capita disposable income of urban residents and the total retail sales of social consumer goods is -0.348. However, according to the consumption function, the consumption should increase as the disposable income increases. Therefore, it is considered that the results of partial correlation analysis have a big deviation from the reality.

2.4 Spearman rank correlation analysis

Spearman rank correlation coefficient is a non-parametric measure of the correlation between two variables. It ranks the sample values of the two elements in order of the size of the data. Then, a correlation coefficient obtained by replacing the actual data with the order of each sample value. Without repeated data, if a variable is a strictly monotonic function of another variable, the Spearman rank correlation coefficient between the two is +1 or -1, which is called complete Spearman correlation. Generally speaking, the absolute value of Spearman rank correlation coefficient is closer to 1, the higher the degree of correlation. Conversely, the closer

the value is to 0, the lower the degree of correlation. The rank correlation coefficient is recorded as r_s, its calculation formula is as follows:

$$r_s = \frac{\sum d_i^2}{n(n^2-1)/3} = \frac{3\sum d_i^2}{n(n^2-1)}$$

Where d_i represents the difference between the variables X_i and Y_i, and n is the sample size.

The absolute value of rank correlation coefficient r_s is compared with the critical value w_p in the Spearman rank correlation coefficient statistical table. When the absolute value of r_s is greater than w_p, X and Y are considered to be related. When the absolute value of r_s is less than w_p, X and Y are considered as irrelevant. The Spearman rank correlation analysis was carried out by SPSS software, and the results were shown as shown in Table 4. (n=10, the w_p of the significant level 0.05 is 0.564, and the significant level 0.01 corresponding to the w_p is 0.746).

Table 4: Rank correlation coefficient and ranking of each variable

Variable	Rank correlation coefficient	Score	Sort
X_1	0.127	0.127	4
X_2	1.000[***]	1.000	1
X_3	1.000[***]	1.000	1
X_4	1.000[***]	1.000	1
X_5	1.000[***]	1.000	1
X_6	1.000[***]	1.000	1
X_7	-0.614[**]	0.614	3
X_8	-.939[***]	0.939	2

The results of rank correlation analysis show that the rank correlation coefficient of household consumption level, per capita disposable income of urban residents, per capita net income of rural residents, population at the end of the year, and gross domestic product are as high as 1 at a significant level of 0.01, which are greater than 0.746. These factors are completely related to the total retail sales of social consumer goods. In addition, the rank correlations between the Engel coefficient of consumption of urban residents and that of rural residents are -0.614 and -0.939. These two factors are also significantly related to the total retail sales of social consumer goods.

However, there is no obvious correlation between the retail price index of commodities and the total retail sales of social consumer goods.

2.5 Gray Correlation Analysis

Grey relational analysis belongs to a branch of grey system theory. It is a widely accepted method to conduct comprehensive evaluation of things and phenomena that are influenced by many factors. The basic idea of gray relational analysis is to measure the degree of closeness of each factor through the similarities or dissimilarities of the morphologies of the internal factors in the system. It is to judge the degree of correlation according to the similarity or difference degree of the geometric shape of the grey time series curve. The closer the gray time series curve is, the greater the correlation between the corresponding factors. Gray correlation analysis method is different from other related analysis, it makes up for the defects caused by the traditional system correlation analysis. Regardless of the quantity of the sample or the law of the sample, it is equally applicable. Moreover, it is not limited by the typical distribution and is convenient for calculation. There will not be a phenomenon that the quantitative results do not agree with the qualitative analysis. The grey correlation method is applied to analyze the selected indicators, and the total retail sales of social consumer goods as a reference sequence, and the other 8 indicators as a comparative sequence. The correlation coefficients of each comparison sequence and the reference sequence are calculated respectively, and the corresponding degree of association is obtained. The results of all correlation degrees are shown in Table 5.

Table 5: Grey relational degree and sorting of each variable

Variable	Grey relational degree	Sort
X_1	0.633	6
X_2	0.833	2
X_3	0.749	3
X_4	0.723	4
X_5	0.638	5
X_6	0.933	1
X_7	0.627	7
X_8	0.615	8

(a) The gray relational grade between gross domestic product and total retail sales of consumer goods reached as high as 0.933, indicating that GDP has a strong promotion effect on retail sales growth of social consumer goods. Since gross domestic product is the core indicator of China's new national economic accounting system, it reflects the economic strength and scale of a country, which is almost related to most variables in the economy. Therefore, the total retail sales of consumer goods are also inevitably affected by it, and there is a very high correlation between the two.

(b) The correlation between the total retail sales of consumer goods and the level of consumer consumption is up to 0.833, second only to GDP. This indicates that the consumption level of residents is closely related to the total consumption of consumer goods. The total amount of social consumer goods is divided into two categories according to the object of sale: the total amount of consumer goods and the total amount of consumer goods of social groups. The relationship between the level of household consumption and the total retail sales of consumer goods can be expressed by the level of household consumption and the total amount of consumer goods. At present, the living standard of living in our country is constantly improving. From this, the total amount of social retail goods may also be affected by the trend and will continue to rise.

(c) The per capita income of urban (rural) residents is also one of the important factors that affect the total retail sales of social consumer goods. The correlation between the per capita disposable income of urban households and the total retail sales of social consumer goods is 0.749. In addition, the gray correlation of per capita net income of rural residents is 0.723. There is a functional relationship between consumption and income, and consumption is directly restricted by the level of income. The rising income makes the purchasing power of residents increased, thus stimulating the willingness of consumers. However, under the environment of economic growth, the income growth of rural residents is much lower than that of urban residents. Therefore, the correlation degree between the per capita disposable income of urban residents and the total retail sales of social consumer goods is greater than that of the per capita net income of rural residents.

(d) The correlation between the year-end population and the total retail sales of social consumer goods is 0.638, and the correlation is not significant. At present, the number of rural population flowing into cities is increasing gradually. The number of newly added population is decreasing year by year, which makes the impact of changes in population structure more and more powerful. However, the impact of new population on the retail sales of social consumer goods will gradually weaken.

(e) The correlation between the total retail sales of consumer goods and the retail price index of commodity is only 0.633, and the degree of association is relatively small. Although the rise in

the retail price of commodities will cause an increase in consumption, which in turn will cause a change in the total amount of social consumer goods. But it will not be one of the important factors driving the growth of the retail sales of social consumer goods.

(f) The grey correlation coefficient between the Engel coefficient of urban (rural) residents and retail sales of social goods is 0.627 (0.615), and the correlation degree is the least. With the general improvement of people's living standard and gradual change of consumption concept, people's income is used not only for daily food consumption but also for other types of goods of different grades. These new consumption has gradually become an important part of the total retail sales of social consumer goods.

3. COMPARISON OF FOUR METHODS

The scores and sorting results of Pearson correlation analysis, partial correlation analysis, rank correlation analysis and grey correlation analysis are summarized, and the results are shown in Table 7. Pearson correlation analysis is one of the most common analysis methods. We can get some quantitative conclusions by the size of Pearson correlation coefficient. However, according to the analysis results of this paper, the scores of six explanatory variables are above 0.9 and the minimum difference between the variables is zero. The accuracy of the Pearson correlation index is limited by the requirement of each variable data conforming to the normal distribution, which makes the sorting result less ideal. Partial correlation analysis is to study the linear correlation between two variables in the case of controlling the other variables. According to the results of partial correlation analysis, it can be seen that the difference between the highest score and the lowest score is 0.806, and the minimum difference between variables is 0.031. The ranking result is relatively ideal. However, the result of partial correlation analysis is contrary to some economic theoretical knowledge, and we can't explain it reasonably. Rank correlation analysis does not require the original data to follow what kind of distribution, so it is more flexible and convenient to apply. However, according to the results of rank correlation analysis, the scores of residents' consumption level, per capita disposable income of urban residents, per capita net income of rural residents, population at the end of the year and GDP are all the same. It shows that these five factors have the same influence on the total retail sales of social consumer goods, and it can't distinguish which is the main factor and which is the secondary factor. The grey correlation analysis has no limitation on the sample size and the distribution of the sample, and the calculation is relatively small. According to the results of grey correlation analysis, the difference between the highest score and the lowest score is 0.318, and the minimum difference between variables is 0.005. The results are consistent with the qualitative analysis results, so the sorting results of grey correlation analysis are more reliable.

In summary, in accordance with the greater the difference between the variable score, the higher the degree of dissimilarity between variables, and the better the ranking criteria. Based on the qualitative analysis, the order of the four methods is as follows: gray relational analysis> partial correlation analysis> Pearson correlation analysis> rank correlation analysis.

Table 7: Scores and sort results of four correlation analysis methods

Variable	Grey relation		Pearson correlation		Partial correlation		Rank correlation	
	Score	Sort	Score	Sort	Score	Sort	Score	Sort
X_1	0.632	6	0.034	7	0.730	3	0.127	4
X_2	0.833	2	0.999	1	0.821	1	1.000	1
X_3	0.749	3	0.995	3	0.348	7	1.000	1
X_4	0.723	4	0.999	1	0.764	2	1.000	1
X_5	0.638	5	0.984	4	0.412	6	1.000	1
X_6	0.933	1	0.997	2	0.443	5	1.000	1
X_7	0.627	7	0.594	6	0.015	8	0.614	3
X_8	0.615	8	0.938	5	0.594	4	0.939	2

4. CONCLUSION

In this paper, grey correlation analysis, Pearson correlation analysis, partial correlation analysis and rank correlation analysis are used to analyze the correlation between 8 explanatory variables and total retail sales of social goods. These variables include retail price index of goods, consumption level of residents, per capita disposable income of urban residents, per capita net income of rural residents, year-end population, GDP, the Engel coefficient for urban residents 'consumption and Engel coefficient for rural residents' consumption. Through the first three analysis methods, the main factors that affect the total retail sales of social consumer goods are gross domestic product, consumption level, per capita net income of rural residents and consumption level of rural residents. However, rank correlation analysis indicates that the main factors affecting the total retail sales of social consumer goods are household consumption level, per capita disposable income of urban residents, per capita net income of rural residents, population at the end of the year, and GDP. After comparison, it is considered that the results obtained by grey correlation analysis are more reliable. According to the grey correlation

analysis, we get the correlation degree between the influencing factors and the total retail sales of social consumer goods in China, which can provide the theoretical basis for policy suggestions of expanding consumer demand.

REFERENCES

[1] Granger C.W.J..(1980)Forecasting in Business and Economics, London: Academic Press.

[2] Montgomery D C, Jenings C L, Kulahci M, et al. (2009)Introduction to time series analysis and forecasting: solution manual to accompany, Wiley-Interscience.

[3] Anonymous. (2010)Retail sales of consumer goods in China increased from US$719.6 bn in 2004 to US$1.5 trillion in 2008, M2 Presswire.

[4] Liu Y N. (2007)Analysis of the influential factors on the total retail sales of consumer goods, Journal of Times Trade,5 (84): 21-23.

[5] Falina. (2010)On the empirical analysis of China's total retail sales of social consumer goods continued to grow, Journal of Economic Problems, (5): 32-35.

[6] Peng B, Hu J, Deng J. (2012)Empirical analysis of the factors influencing the total retail sales of social consumer goods in the Yangtze River Delta, Journal of Knowledge Economy, (2): 93.

[7] Chen W H. (2015)Empirical analysis of the influencing factors on the total retail sales of social consumer goods in Guangdong , Journal of Market Economy and Prices, (11): 49-52.