Multiple Statistical Analysis of the Influencing Factors of the Retail Price Index

Xinying Wan[1]

[1]790847022@qq.com

[1]Management Science and Engineering, Sichuan University, Chengdu, Sichuan, China

Abstract—With the improvement of the income level of people, our country's economic level has continued to improve. However, due to the differences in geographical and historical conditions in various regions, there is a large gap in the level of economic development in various regions of our country, and there are also significant differences in the prices of different commodities. This article uses the multiple linear regression and Principal Component Analysis (PCA) by SAS to explore the impact of the Retail Price Index (RPI) and different factors, such as the Producer Price Index (PPI) of agricultural production materials and the PPI of industrial production, the Purchasing Price Index (PRI) of industrial producers, and the Price Index of Investment in Fixed Asset.

Keywords- multiple linear regression, Principal Component Analysis, Retail Price Index, SAS

1 INTRODUCTION

Since Economic Reform and open up, our country has transformed from a planned economy to a market economy with Chinese characteristics, and price controls have been gradually liberalized. However, after the liberalization of price controls, the retail prices of commodities have not reached a stable state, but have risen year by year. In recent years, residents' intuitive feeling is that prices have risen too fast, and residents' income cannot keep up with the rise of Costumer Price Index(CPI) and Retail Price Index(RPI), which has affected the quality of life and happiness of residents to a certain extent[1,2]. As the dominant economic intervention, the government's intervention methods have a huge impact on economic development. Its influence is not only reflected in direct intervention in market prices, but also in the formation of monopolies in some industries, and the market access system is not Looseness has led to insufficient market supply, and some commodities have formed monopoly prices. At the same time, my country's central government has different economic policies in different regions of our country, coupled with differences in geographical and historical conditions, leading to the level of economic development in various regions of our country. There is a big gap, and there are also significant differences in the prices of different commodities.

Stabilizing the price level is one of the ultimate goals of my country's macroeconomic regulation. How to set a reasonable retail price of goods to increase the turnover of stores has become a matter of concern to most retailers[3,4]. And there are many factors that affect the

ICIDC 2022, June 17-19, Qingdao, People's Republic of China

DOI 10.4108/eai.17-6-2022.2322819

retail prices of commodities, such as the prices of production materials, investment in fixed assets and so on[5]. The changes in the producer price index will cause changes in the consumer price index to a certain extent[6].

Changes in different price indices will be transmitted to the consumer price index to a certain extent, and ultimately affect the retail price index. This article aims to analyze the relationship between different parameters and Retail Prices Index through a multivariate statistical analysis method, and get the influence of different factors on the retail prices of commodities.

2 ANALYSIS PLAN

2.1Data Sources

The data comes from online data, that is, the changes in the Retail Price Index of a company A from 2006 to 2016 and other indexes (TABLE 1). The data fields include: the Retail Price Index (RPI) which reflects the trend and degree of changes in the retail prices of commodities in a certain period of time, the Producer Price Index (PPI) is an index that measures the trend and degree of changes in the ex-factory prices of industrial products. It is an important economic indicator that reflects the price changes in a certain period of production and is also an important basis for formulating relevant economic policies and national economic accounting. Since enterprises will eventually transfer their costs to consumers in the form of higher consumer prices, it is generally believed that changes in the Production Price Index are useful for predicting changes in the consumer price index. The Purchasing Price Index (PRI) reflects the trend in the price level that companies or investors pay when purchasing raw materials as investment and the Price Index of Investment in Fixed Asset is a relative number that reflects the trend and degree of changes in the price of fixed asset investment. It can truly reflect the price trend and range of changes of various commodities after removing the factors that affect fixed asset investment indicators

The specific data are as follows:

TABLE 1. RETIAL PRICE INDEX AND OTHER INDEXES OF COMPANY A FROM 2006 TO 2016

Years	Indexes of company A				
	RPI (y)	***PRI of agricultural production materials (x_1)***	***PRI of industrial production (x_2)***	***PPI of industrial producer (x_3)***	***Price Index of Investment in fixed asset (x_4)***
2006	3144.67	1595.48	4521.74	537.68	451.10
2007	3921.41	2032.00	5855.30	624.74	511.50
2008	4956.13	2216.15	7602.40	736.42	567.51
2009	5140.31	2240.61	12017.28	980.23	574.02
2010	6727.42	2482.89	13581.96	1158.66	639.42
2011	8691.90	2983.51	15124.09	1367.39	719.36
2012	9551.10	3297.21	18038.92	1537.07	791.43
2013	10447.52	3425.61	21049.15	1681.11	838.47
2014	10729.18	3594.17	23577.17	1787.12	1175.47
2015	11039.08	3745.32	25973.74	1871.55	1219.77
2016	11058.79	4005.42	19126.03	2138.45	1472.57

2.2Model assumptions and formulation

The retail price of commodities will be affected by many factors. This article focuses on the relationship between Retail Price Index y and the Producer Price Index of agricultural production material x_1, the Producer Price Index of industrial production price x_2, the Purchasing Price Index of industrial producer x_3 and the Price Index of Investment in fixed asset x_4.

Establish correlation images of RPI y with respect to the PPI of agricultural production materials x_1 and the PPI of industrial production x_2, the PRI of industrial producers x_3 and the Price Index of Investment in fixed asset x_4.

3 DATA ANALYSIS

This article uses SAS software to perform correlation analysis, regression analysis and principal component analysis on the selected data. The results are as follows.

3.1Regression analysis

3.1.1Analysis of Variance

TABLE 2. THE RESULT OF ANALYSIS OF VARIANCE

Source	Analysis of Variance				
	Degree of Freedom	***Sum of Square***	***Mean Square***	***Value of F***	$P_r > F$
Model	4	53.90006	13.47501	49.25	0.0001
Error	6	1.64176	0.27363		
Total Calibration	10	55.54182			

It can be seen that the value of F is 49.25 and the value of P_r is 0.0001, which is significant for the retail price index at the significance level of 0.05 (TABLE 2).

3.1.2R-square and Adjusted R-square

TABLE 3. THE RESULT OF R-SQUARE AND ADJUSTED R-SQUARE

Root Mean Square Error	0.5239	R-Square	0.9704
Mean of Dependent variable	101.92727	Adjusted R-Square	0.9507
Coefficient of Variation	0.51320		

The R-square and Adjusted R-square are 97.04% and 95.07%, respectively, indicating that the overall fit of the model is good and the correlation is high (TABLE 3).

3.1.3Parameter estimation

TABLE 4. PARAMETER ESTIMATION I

Parameter estimation				
Variable	***Label***	***Degree of freedom***	***Parameter estimation***	***Standard error***
Intercept	***Intercept***	***1***	***18.52608***	***12.40098***
X1	The PPI of agricultural production materials	1	0.13053	0.03808
X2	The PPI of industrial producion	1	0.61754	0.20194
X3	The PPI of industrial producer	1	-0.18445	0.10977
X4	The PPI of investment in fixed asset	1	0.26429	0.16571

TABLE 5. PARAMETER ESTIMATION II

Parameter estimation						
Variable	***Label***	***Value of t***	***$P_r > t$***	***Tolerance***	***Confidence interval***	
Intercept	***Intercept***	***1.49***	***0.1858***	***.***	***-11.8180***	***48.8702***
X1	The PPI of agricultural production materials	3.43	0.0140	0.28031	0.03735	0.22371
X2	The PPI of industrial producion	3.06	0.0223	0.09851	0.12342	1.11167
X3	The PPI of industrial producer	-1.68	0.1439	0.07220	0.07220	0.08414
X4	The PPI of investment in fixed asset	1.59	0.1619	0.10799	0.10799	0.66977

As can be seen from the above table (TABLE 4 and TABLE 5), the regression equation is

$$y = 18.52608+0.13053x_1+0.61754x_2-0.18445x_3+0.2649x_4 \quad (1)$$

In equation (1), the P_r value of the corresponding variable agricultural production material price index x_1 is 0.0140, which is significant to the retail price index y at the significance level of 0.05, and its 95% confidence interval is (0.03735, 0.22371). The P_r value of the industrial production price index x_2 is 0.0223. It is significant to the retail price index y at the significance level of 0.05, and its 95% confidence interval is (0.12342, 1.11167). And the P_r value of the industrial producer purchase price index x_3 is 0.1439 at the significance level of 0.05 to the retail price index y is not significant, and its 95% confidence interval is (-0.45304, 0.08414) and the P_r value of the total fixed asset investment price index x_4 is 0.1619. At the significance level of

0.05, it is not significant to the retail price index *y*, and its 95% confidence interval is (-0.14120, 0.66977).

Therefore, x_3 and x_4 can be removed and analyzed again in the subsequent analysis.Principal component analysis.

3.1.4Corllinearity diagonosis

TABLE 6. CORLLINEARITY DIAGONOSIS

Eigenvalues	Condition Index	intercept	x_1	x_2	x_3	x_4
4.99599	1.00000	0.00000647	0.00006281	0.00000245	0.00000801	0.00000349
0.00322	39.37677	0.01583	0.27909	0.00106	0.00218	0.0005744
0.000686	85.97752	0.04678	0.50909	0.00169	0.16771	0.00005362
0.0000678	271.07364	0.25117	0.18735	0.05179	0.15195	0.99733
0.0000433	339.70087	0.68621	0.02441	0.94545	0.67815	0.00204

The above is the diagnosis result of multicollinearity (TABLE 6), the maximum condition index is 339.70087, and the feature value of multiple dimensions is about 0, indicating that there is serious collinearity between the independent variables.

3.1.5Simple statictical analysis

TABLE 7. SIMPLE STATISTICAL

Simple statistical				
	x_1	x_2	x_3	x_4
Mean	104.20000	99.73636	101.43636	101.85455
StD	8.20427	2.63677	5.609826	3.03755

TABLE 8. CORRELATION MATRIX

		X_1	X_2	X_3	X_4
X_1	The PPI of agricultural production materials	1	0.13053	0.03808	3.43
X_2	The PPI of industrial producion	1	0.61754	0.20194	3.06
X_3	The PPI of industrial producer	1	-0.18445	0.10977	-1.68
X_4	The PPI of investment in fixed asset	1	0.26429	0.16571	1.59

The above table reflects the correlation coefficient matrix between all the original indicator variables (TABLE 7 and TABLE 8).

TABLE 9. EIGENVALUE OF CORRELATION MATRIX

	Eigenvalue	**Difference**	**Proportion**	***Cumulation***
X_1	The PPI of agricultural production materials	3.58976	3.29028	0.8952
X_2	The PPI of industrial producion	0.29048	0.20938	0.9678
X_3	The PPI of industrial producer	0.08111	0.03346	0.9881
X_4	The PPI of investment in fixed asset	0.04765		

The above table shows all the eigenvalues calculated from the correlation coefficient matrix(TABLE 9), the difference between two adjacent eigenvalues, the contribution rate of each principal component and the cumulative contribution rate.

The contribution rate of the first principal component to the variance was 89.52%, the contribution rate of the second principal component to the variance was 7.26%, and the contribution rate of the third principal component to the variance was 2.03%. The subsequent principal component contribution rate was 1.19%. The cumulative contribution rate of the first principal component is 89.52%. Therefore, this article takes the first component as the main analysis object.

TABLE 10. EIGENVECTOR

		Comp1	**Comp2**	***Comp3***	***Comp4***
X_1	The PPI of agricultural production materials	0.47015	0.83409	0.27116	0.09862
X_2	The PPI of industrial producion	0.50214	-.48172	0.39238	0.60154
X_3	The PPI of industrial producer	0.51398	-.26762	0.21769	-.78537
X_4	The PPI of investment in fixed asset	0.51249	-.02479	-.85154	0.10779

The output of the above table(TABLE 10) is the eigenvectors corresponding to all eigenvalues, which are linearly independent unit variables. The first column represents the score coefficient of the first principal component *comp1*, the second column represents the score coefficient of the second principal component *comp2*, and so on. Based on this, the relationship between the principal components expressed by the standardized variables can be written, and Comp1 is the main component here(formula (2)).

$$Comp1 = 0.470146\, x_1 + 0.502138\, x_2 + 0.513983\, x_3 + 0.512485\, x_4 \quad (2)$$

3.1.6Principal component regression

3.1.6.1Variance analysis

TABLE 11. ANALYSIS OF VARIANCE

Source	Degree of Freedom	Sum of Square	Mean Square	Value of F	$P_r > F$
Model	4	53.90006	13.47501	49.25	0.0001
Error	6	1.64176	0.27363		
Total Calibration	10	55.54182			

We can see that the value of F is 59.15 and P_r is less than 0.0001(TABLE 11), indicating that the principal component *comp1* has a significant effect on the dependent variable *y*.

3.1.6.2R-square and Adjusted R-square

TABLE 12. THE RESULT OF R-SQUARE AND ADJUSTED R-SQUARE

Root Mean Square Error	0.66312	R-Square	0.9367
Mean of Dependent variable	101.92727	Adjusted R-Square	0.9208
Coefficient of Variation	0.65058		

The above table(TABLE 12) shows all the eigenvalues calculated from the correlation coefficient matrix, the difference between two adjacent eigenvalues, the contribution rate of each principal component and the cumulative contribution rate.

The contribution rate of the first principal component to the variance was 89.52%, the contribution rate of the second principal component to the variance was 7.26%, and the contribution rate of the third principal component to the variance was 2.03%. The subsequent principal component contribution rate was 1.19%. The cumulative contribution rate of the first principal component is 89.52%. Therefore, this article takes the first component as the main analysis object. The R-square and adjusted R-square are 93.67% and 92.08%, respectively, indicating that the correlation between the principal component and the dependent variable is relatively high, and the dependent variable can be well explained by the principal component.

3.1.6.3Parameter estimation

TABLE 13. PARAMETER ESTIMATION III

Variable	***Label***	***Degree of freedom***	***Parameter estimation***	***Standard error***
Intercept	***Intercept***	***1***	***101.92727***	***0.19994***
Comp1		1	1.20084	0.11082

TABLE 14. PARAMETER ESTIMATION IV

Variable	*Label*	*Value of t*	$P_r > t$	*Standardized estimate*
Intercept	***Intercept***	***509.79***	***< .00001***	***0***
Comp1		10.84	< .0001	0.96419

The above table shows that the P value is less than 0.0001(TABLE 13 and TABLE 14), which is significant for the dependent variable at the significance level of 0.05. Therefore, the regression equation (3) is:

$$\hat{Y} = 101.92727 + 1.20084comp1 \tag{3}$$

Combination formula (1), marked with X_i is

$$\hat{Y} = 0.56457X_1 + 0.60299X_2 + 0.61721X_3 + 0.61541X_4 + 101.92727 \tag{4}$$

And:

$$X_1 = \frac{x_1 - 104.20000}{8.20427}, X_2 = \frac{x_2 - 99.73636}{2.63677}, \tag{5}$$

$$X_3 = \frac{x_3 - 101.43636}{5.60826}, X_4 = \frac{x_4 - 101.85455}{3.03755} \tag{6}$$

The equation (4) is expressed by x_i:

$$\hat{Y} = 0.06881x_1 + 0.22869x_2 + 0.11005x_3 + 0.20260x_4 + 40.14935 \tag{7}$$

4 DATA ANALYSIS

Based on basic statistical analysis, regression analysis and principal component analysis, we found that most of the factors affecting the retail price index are interrelated, and different price index factors will have a certain impact on the retail price index. Through the analysis of the main component, the first main component comp1 in the model can explain the original 4 arguments information 89.52%. Retail Price Index *y*, the Producer Price Index of agricultural production material x_1, the Producer Price Index of industrial production price x_2, the Purchasing Price Index of industrial producer x_3 and the Price Index of Investment in fixed asset x_4 are all positively correlated. The collinearity diagnosis result performed in the regression analysis part shows that there is serious collinearity among the independent variables;

And the multiple regression model of the PPI of agricultural production materials x_1 and the PPI of industrial production x_2, the PRI of industrial producers x_3, the Price Index of Investment in Fixed Asset x_4 and dependent variable commodity Retail Price Index y is

$$\hat{Y} = 0.06881x_1 + 0.22869x_2 + 0.11005x_3 + 0.20260x_4 + 40.14935 \quad (7)$$

Based on the above conclusions, we found that most of the factors affecting the retail price index are interrelated, and different price index factors will have a certain impact on the retail price index. Only by controlling each price index can we better realize the growth of retail sales and have a more positive and far-reaching impact on the development of my country's market economy.

REFERENCES

[1] Muhammad Shahbaz, Najeeb M Nasir. Producer & Consumer Prices Nexus: ARDL Bounds Testing Approach. International Journal of Marketing Studies. 2009, 1(2).

[2] SarantisTsiaplias. The CPI and Other Measures of Australian Inflation[J]. Australian Economic Review. 2008 (1)

[3] Bank W. Producer price index manual: theory and practice. International Monetary Fund, 2004.

[4] Espejo M R. Consumer Price Index Manual: Theory and Practice[J]. Journal of the Royal Statistical Society: Series A (Statistics in Society), 2005.

[5] Dorestani, A., & Arjomand, L. H. An empirical study of the relationship between consumer and producer price index: a unit root test and test of cointegration. Coastal Business Journal.

[6] Kyrtsou C, Labys W C. Evidence for Chaotic Dependence between US Inflation and Commodity Prices[J]. Journal of Macroeconomics, 2006, 28(1):256-266.