

Analysis And Prediction of Influencing Factors of Total Retail Sales of Consumer Goods in China——Based on Multiple Linear Regression Models and ARIMA Models

Yan Zhang^{1, a}

¹School of MinZu University of China

^azy08285043@163.com

Abstract: This paper takes the annual data of total retail sales of consumer goods in China from 2001 to 2020 as the research object. Using methods such as principal component estimation, ridge regression and Lasso regression, a multivariate linear regression model is established, and the regression results of the three models are compared and analyzed, and the significant factors affecting the total retail sales of consumer goods in China are the income level of residents, the level of consumption of residents, and the development of the tertiary industry, and the less significant factors are demographic factors. In addition, this paper takes the annual data of china's total retail sales of social consumer goods from 1973 to 2020 as the research object, and uses the ALAMA (0,1,1) model to predict and analyze the total retail sales of social consumer goods in China in the next ten years, and concludes that the total retail sales of social consumer goods in China will gradually rise from 2021 to 2030, which is consistent with the actual situation. Thus, it puts forward rationalization suggestions for increasing the total retail sales of consumer goods in China and promoting economic development.

Keywords: Total Retail Sales of Social Consumer Goods, Multiple Linear Regression Model, Influencing Factors, ARIMA Model, Forecast.

1. Introduction

Since the outbreak of the new crown pneumonia epidemic in 2019, due to the impact of the global epidemic, the world economy has been in a serious recession, the circulation of the industrial chain supply chain has been blocked, international trade and investment have shrunk, and the commodity market has been turbulent. Domestic consumption, investment, and exports have declined, employment pressure has increased significantly, enterprises, especially private enterprises, small and medium-sized and micro enterprises, have become more difficult, risks in the financial and other fields have accumulated, and the contradiction between grass-roots fiscal revenue and expenditure has intensified. In the face of changes in the domestic and international economic situation, the party and the state pointed out in the 2020 government work report that China's domestic demand potential is large, and it is necessary to deepen the structural reform of the supply side, highlight the orientation of people's livelihood, and effectively combine and promote each other by boosting consumption and expanding investment. It is necessary to implement the strategy of expanding domestic demand and accelerate the transformation of the mode of economic development. Therefore, it has become very important to expand China's domestic demand and stimulate domestic consumption.

The total retail sales of social consumer goods is one of the most important indicators to observe the level of domestic consumption, it is from the sales transaction end of the enterprise, through the sum of the sales of goods to individuals, social groups and the provision of catering services to obtain the amount of income obtained; it reflects the total amount of consumer goods supplied by various industries to residents and social groups through a variety of commodity circulation channels, is an important indicator to study the changes in the domestic retail market and reflect the degree of economic prosperity. Research and analysis of the significant factors affecting the total retail sales of social consumer goods, through the adjustment of these factors, can stimulate consumption to a certain extent, and at the same time is of great significance to the design of economic and industrial development strategies, optimize the allocation of social resources, strengthen effective supply, and better meet the needs of the people for material life.

2. Literature Review

A great deal of research has been conducted by domestic scholars on the issue of retail sales of social consumer goods. Feng Haiqi, Zhao Yumei [1] et al. used the Lasso regression model to improve the influence of multiple collinearity, and analyzed the influencing factors of the total retail sales of social consumer goods in Shandong Province; Comfortable Hua [2] used Ridge regression to overcome the multiple collinearity between the data to analyze the influencing factors and the degree of influence of the total retail sales of social consumer goods in China; Xue Song, Zhou Jing [3] et al. used quantitative analysis, using unit root test and other econometric methods combined with the actual situation and model results. Yang Yang and Yang Guiyuan [4] used the monthly data of total retail sales of social consumption in China to establish a multivariate linear regression model and a seasonal fluctuation prediction model to predict the data of January and February 2021, and compared with the actual values to find that the model fits well. Wu Long [5] using the principal component regression model. The factors affecting the growth of total retail sales of consumer goods in China are discussed. Although the above studies have achieved certain results, due to the small number of indicators, they cannot fully reflect the influencing factors of the retail sales of social consumer goods, so the research in this area still needs to be explored from the analysis methods and the breadth of influencing factors. Wang Zhijian et al. [6] used the annual data of total retail sales of consumer goods for ARMA modeling, using the model to predict future data, and proved that the model fit well by comparing it with the actual values. Chen Yunjie et al. [7] Using the knowledge of time series, the Holt two-parameter exponential smoothing prediction method is carried out on the total retail sales of consumer goods in China, and then the ARIMA model is established to make short-term predictions for future series values. Fan Liang [8] By establishing an ARIMA model that can effectively fit its law of change, the development trend of total retail sales of social consumer goods in various counties and districts of Longnan City is given, and the theoretical basis is provided for relevant departments to make correct decisions.

3. Data description and pre-processing

3.1 Data sources

The National Bureau of Statistics will publish the total retail sales of consumer goods in the current month, the cumulative value of the current year and their year-on-year changes, and list the specific conditions of each sub-item, according to which the changes in the total consumption volume and growth rate of each month can be observed. It is the most frequently released data reflecting consumption; it covers a comprehensive system and category, subdivided into catering services and commodity sales, in which commodity sales also list the retail sales of units above designated size and their sub-items. This paper uses annual data of the total retail sales of social consumer goods in China from 2001 to 2020 released by the National Bureau of Statistics and the annual data of related influencing factors.

3.2 Data standardization

Before data analysis, we usually need to normalize the data and use the standardized data for data analysis. In this paper, the Z-score standardization method is selected to use SPSS to standardize the data.

Z-score normalization normalizes data based on the mean and standard deviation of the original data. Normalize the original value of an x to z using z-score. New data = (original data - mean) / standard deviation, i.e. $X = (x_{ij})^T$ the original data, x_i the mean of the i th independent variable, s_i the standard deviation of the i th argument, the new data, then

$$z_{ij} = \frac{x_{ij} - x_i}{s_i}$$

$$i=1, \dots, p, j=1, \dots, n$$

4. The Selection of Indicators

Through a review of existing literature and economic theories, factors influencing total retail sales of consumer goods in society may include the size and structure of the population, household income, consumption level and prices. Therefore, this paper preliminarily selects the number of people (x_1), urbanization rate (x_2), per capita disposable income of urban residents (x_3), per capita disposable income of rural residents (x_4), retail price index of goods (x_5), consumer price index (x_6), value added of the tertiary industry (x_7), national fiscal expenditure (x_8) eight regression independent variables, that is, explanatory variables, the remaining influencing factors are classified as random perturbations. And the total retail sales of consumer goods in the country (y) are determined as the dependent variable, that is, the explanatory variable.

The standardized data are first tested for The Person correlation, so that the explanatory variables with significant correlation with the dependent variable are screened out and included in the subsequent regression model, in order to ensure that the selected independent variable can explain the dependent variable to a certain extent. As shown in Table 1, it can be seen that there is a strong correlation between the selected eight independent variables and the dependent variables, indicating that the selected independent variables are highly linearly related to y , and it is reasonable to use y and independent variables for multiple linear regression. Therefore, eight indicators were finalized.

Table 1 Correlation tests

	y	x ₁	x ₂	x ₃	x ₄	x ₅	x ₆	x ₇	x ₈
y	1								
x ₁	0.982**	1							
x ₂	0.983**	0.999**	1						
x ₃	0.997**	0.988**	0.991**	1					
x ₄	0.995**	0.970**	0.974**	0.995**	1				
x ₅	0.965**	0.982**	0.985**	0.974**	0.950**	1			
x ₆	0.982**	0.993**	0.995**	0.990**	0.973**	0.995**	1		
x ₇	0.996**	0.973**	0.976**	0.996**	0.999**	0.953**	0.975**	1	
x ₈	0.999**	0.985**	0.987**	0.999**	0.995**	0.972**	0.987**	0.996**	1

**At level 0.01 (two-tailed), the correlation is significant.

5. Based on The Analysis of Influencing Factors of Multiple Linear Regression Model

5.1 Principal component regression model

5.1.1 Determination of principal components

Using R to analyze the principal components of the selected independent variables, Table 2 is the extracted principal components, Figure 1 is the lithotripsy diagram, because the curve changes gently after the principal component 2, so it is reasonable for us to select the first two principal components, that is, the first principal component and the second principal component, the contribution rate of these two principal components has been as high as 99.708%, almost retaining all the information of the original indicator, from which the linear expression of these two principal components has been calculated:

$$z_1 = 0.354x_1 + 0.354x_2 + 0.356x_3 + 0.352x_4 + 0.350x_5 + 0.355x_6 + 0.353x_7 + 0.355x_8$$

$$z_2 = 0.245x_1 + 0.230x_2 - 0.163x_3 - 0.485x_4 + 0.543x_5 + 0.292x_6 - 0.450x_7 - 0.210x_8$$

The above equation results in a linear expression of two uncorrelated principal component z_1, z_2 , thus eliminating the problem of multicollinearity between variables. Principal component z_1 generally covers all the information of the original variables and can be considered a comprehensive indicator. Principal component z_2 mainly contains information about the original variables of x_4, x_5, x_7 , mainly explains the price factors, and can be considered as a comprehensive price indicator.

Table 2 Principal component analysis results

Principal Components	Characteristic Value	Contribution Rate (%)	Cumulative Contribution Rate (%)
1	2.808	98.590	98.590
2	0.300	1.118	99.708
3	0.139	0.242	99.950
4	0.049	0.030	99.980
5	0.027	0.009	99.989
6	0.023	0.007	99.996
7	0.015	0.003	99.999
8	0.010	0.001	100

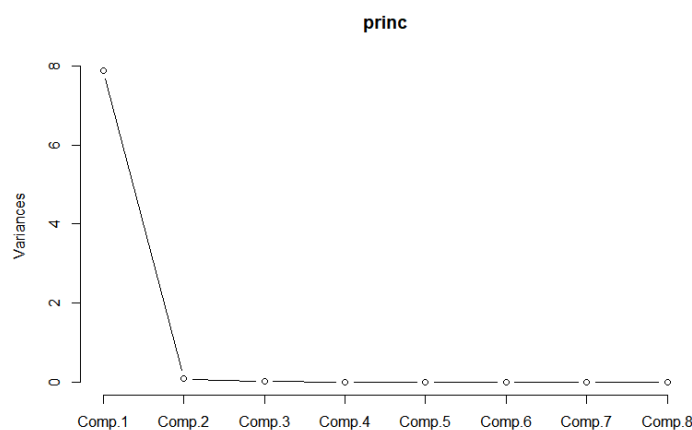


Fig .1 Gravel diagram

5.1.2 Estimation of regression parameters

For the normalized dependent variable, take z_1, z_2 as the independent variable and use the OLS regression method to perform multiple linear regression analysis to obtain the PCR expression of the dependent variable:

$$y = 0.345z_1 - 0.270z_2$$

Table 3 Principal component regression coefficients

Model	Not standardized Coefficients		Normalized coefficients	t	Significance	Collinear statistics	
	B	standard error	Beta			Tolerance	VIF
constant	-1.000E-6	0.015		0.000	1.000		
comp1	0.345	0.005	0.995	64.857	0.000	1.000	1.000
comp2	-0.270	0.050	-0.083	-5.397	0.000	1.000	1.000

a. Dependent variable: y

The fitting effect of the model is very good, the adjusted coefficient of determination $R^2 = 0.996$, and the model passed the $F(2117.771)$ test and the $t(t_1 = 65.587, t_2 = -5.397)$ test, the coefficients of the model as a whole and the variables are statistically significant, and the model setting is reasonable. The model is then converted to a regression equation represented by standardized independent variables, which is calculated from the matrix of principal constituent coefficient vectors and the estimation of the principal component regression coefficients, from which a new regression equation is obtained:

$$y = 0.056x_1 + 0.06x_2 + 0.167x_3 + 0.252x_4 - 0.026x_5 + 0.044x_6 + 0.243x_7 + 0.179x_8$$

From the standardized regression equation of principal component conversion, it can be seen that it is the retail price index of goods that have a negative effect on the total retail sales of social consumer goods, and the rest of the variables has a positive impact. Among them, the per capita disposable income of rural residents has the greatest impact on the total retail sales of social consumer goods; every 1% increase in the per capita disposable income of rural residents will increase by an average of 0.252%. Secondly, the added value of the tertiary industry, the national fiscal expenditure, and the per capita disposable income of urban residents have a greater impact, which will cause an average increase of 0.243%, 0.179%, and 0.167% for each 1% growth of the dependent variables, respectively; the population, urbanization rate, and consumer price index will have a smaller impact on the total retail sales of social consumer goods, and each 1% increase will cause the dependent variables y to increase by an average of 0.056%, 0.06%, and 0.044%. Rural and urban residents' income growth will promote consumption, people's income can be divided into consumption and savings from the point of view of use, income growth will make consumption increase, but because the marginal consumption tendency of rural residents is greater than that of urban residents, in the case of the same amount of income increase, rural residents have greater influence than urban residents, because the same income rural residents will spend more income on consumption, which means that improving the income level, especially for rural residents, is an important way to increase the total retail sales of social consumer goods. Fiscal expenditure in the use of government purchases and transfer payments, government purchases will form a direct demand for social retail consumption, the government's transfer payments will be converted into residents' disposable income and thus have an impact; the more added value of the tertiary industry indicates that its activity is higher, and the increase in the total retail sales of social consumer goods has a certain pulling effect on it; people will have subsistence consumption in order to survive. The growth of the population base is also a pull on the total retail sales of social consumer goods, and other phenomena can illustrate the positive impact of these variables. Finally, the retail price index has a negative impact on total retail sales of consumer goods in society, but it is small, and every 1% increase will reduce the dependent variable y by 0.026%. Judging from the number of goods that residents can buy, if the retail price index of goods is too high, the number of purchases will decrease, and the total retail sales of social consumer goods will decline. But this does not correspond to the scatter plot matrix of the actual data, so the principal component regression model is not perfect.

5.2 Ridge regression model

Estimate the regression coefficient according to the ridge regression method, use R software tools to achieve, and set the iterative step size to take 0.01 to determine the best ridge regression parameter α , run the program to obtain the ridge rail map, as shown in Figure 2, where the ordinate is the regression coefficient and the abscissa is the optimal ridge regression parameter.

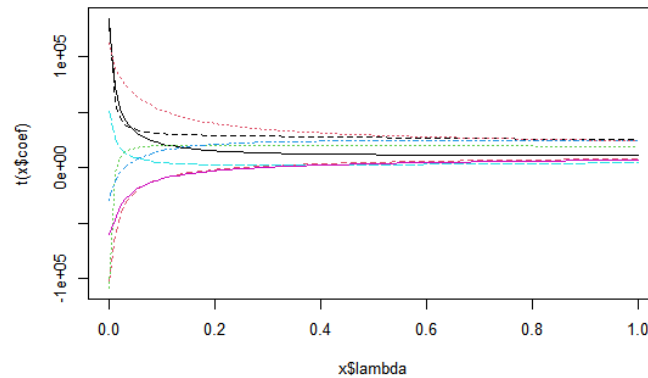


Fig. 2 Ridge regression ridge rail map

As can be seen from Figure 2, as α gradually increases, the coefficients of the respective variables also gradually stabilize, and when $\alpha = 0.8$, the coefficients of the respective variables are basically unchanged, so the optimal ridge regression parameter is $\alpha = 0.8$. Bring the ridge regression parameter $\alpha = 0.8$ into the program to obtain the regression parameter, and the result is shown in Table 4.

Table 4 Ridge regression results

Model	Unnormalized coefficients	Standardized coefficients
constant	-437364.6	-0.352
x_1	2.652	0.092
x_2	881.430	0.007
x_3	1.634	0.157
x_4	5.118	0.195
x_5	84.515	0.030
x_6	73.101	0.048
x_7	0.156	0.209
x_8	0.337	0.207

Table 5 Model goodness of fit

Sample size	R^2	Adjusted R^2	RMSE
20	0.982	0.969	16628.501

Based on the non-normalized systems and constants of regression, we can get the final ridge regression equation, i.e.:

$$y = -437364.6 + 2.652x_1 + 881.430x_2 + 1.634x_3 + 5.118x_4 + 84.515x_5 + 73.101x_6 + 0.156x_7 + 0.337x_8$$

From Table 4, it can be seen that the standardized regression coefficients of the eight independent variables are consistent in order of magnitude and are all positive, which can objectively reflect the influence of the eight explanatory variables on the dependent variable. From Table 5, it can be seen that the adjusted complex correlation coefficient $R^2 = 0.969$, compared with the least squares regression, R^2 has decreased, mainly to solve the collinearity, ridge regression discards some information, resulting in a decrease in fitting accuracy, but the decline is not large, still relatively high, or acceptable. The normalized coefficient values of the ridge regression estimate can objectively reflect the degree of influence of the independent variable on the dependent variable. Among the 8 standard ridge regression regression coefficients, due to $\beta_7 > \beta_8 > \beta_4 > \beta_3 > \beta_1 > \beta_6 > \beta_5 > \beta_2$, it is possible to judge the degree of influence of the 8 independent variables on the dependent variables, that is, the factors affecting the retail sales of consumer goods in China, according to the degree of influence, from large to small: the added value of the tertiary industry, the national fiscal expenditure, the per capita disposable income of rural residents, the per capita disposable income of urban residents, the number of population, the consumer price index, the retail price index of commodities, and the urbanization rate, which is basically consistent with the actual situation.

The realization of the total retail sales of social consumer goods should be achieved through the wholesale and retail trade, accommodation and catering industry in the tertiary industry. It reflects the degree of realization of the purchasing power of social goods to a certain extent can reflect the operation of the tertiary industry, in 2020, due to the impact of the epidemic, there is no doubt that the economic development of the accommodation and catering industry has had an impact, so in 2020, the total retail sales of social consumer goods in China have declined, so the added value of the tertiary industry has the most significant impact on the total retail sales of social consumer goods, ranking first. In 2019, national resident national income and consumption expenditure will increase steadily, and the per capita disposable income exceeded 30,000 yuan. Rural incomes continue to grow faster than urban dwellers. In recent years, the per capita disposable income of rural residents in China has always been faster than that of urban residents, so it is reasonable that the impact on the total retail sales of social consumer goods is greater than the per capita disposable income of urban residents. It is worth noting that what we call the impact on the retail sales of social consumer goods refers to the degree to which the change of the independent variable affects the change of the dependent variable, rather than the absolute contribution of the independent variable to the dependent variable. If the absolute contribution of urban residents' consumption expenditure and rural residents' consumption expenditure to the retail sales of social consumer goods is calculated, it should still be a large proportion of urban residents.

5.3 Lasso regression model

This section uses the Lasso regression algorithm to estimate the regression coefficients. The Lasso regression path diagram is obtained by R software operations, as shown in Figure 3. Where the bottom horizontal axis represents the model coefficient ratio, and the left vertical axis represents the coefficient; The dotted line represents the variable and the vertical line represents the penalty value. The variables screened in Figure 3 are the penalty values at the minimum of the cross-validation indicator (CV), and the variables that intersect the penalty values are the variables that the model eventually incorporates, as can be seen from the figure, and the filters are 8 variables.

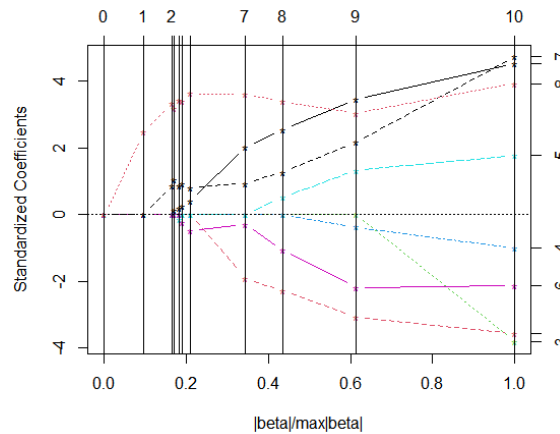


Fig. 3 Lasso regression path

The optimal subset is selected by the C_p (Mallow's C_p) statistic of the model, and the smaller the C_p -value, the optimal subset. Table 6 is about the change in the C_p -value of the number of variables, and it can be seen that when the number of variables is 8, the C_p -value is the smallest. Therefore, 8 variables are selected to enter the model.

Table 6 C_p -values for the number of different variables

The number of variables	0	1	2	3	4	5	6	7	8
C_p -value	41435.858	7838.896	132.569	82.759	79.277	77.525	31.818	18.378	9.000

Since the coefficients of multiple independent variables in the R regression result model are negative, which is inconsistent with the actual situation, we use SPSSAU for Lasso regression below. $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$ is used as the independent variable, and y is used as the dependent variable for lasso regression (Lasso regression) analysis to obtain a trajectory plot, when the K value is 0.96, the normalized regression coefficient of the independent variable tends to stabilize at this time, so SPSSAU recommends setting the optimal K value to take 0.96. That is, we take $\lambda = 0.96$ for regression, and the regression coefficient is found as shown in Table 7.

Table 7 Lasso regression analysis results

Model	Not standardized Coefficients		Normalized coefficients	t	p	R^2	Adjusted R^2	F	p
	B	standard error	Beta						
constant	-2059219.018	8817.89	-	-233.527	0.000**	0.999	0.999	1660.549	0.000
x1	19.683	0.04	0.106	492.638	0.000**				
x2	-9747.598	19.559	0.105	-498.38	0.000**				
x3	4.019	0.011	0.116	362.362	0.000**				
x4	0.322	0.041	0.122	7.86	0.000**				
x5	862.904	5.045	0.098	171.031	0.000**				
x7	0.125	0.001	0.122	112.124	0.000**				
x6	-925.437	1.9	0.104	-487.075	0.000**				
x8	1.121	0.002	0.119	588.757	0.000**				

Dependent variable: y, * $p < 0.05$ ** $p < 0.01$

For the normalized dependent variables, from Table 7 we can obtain the Lasso regression model as follows:

$$y = 0.106x_1 + 0.105x_2 + 0.116x_3 + 0.122x_4 + 0.098x_5 + 0.122x_6 + 0.104x_7 + 0.119x_8.$$

At the same time, $R^2 = 0.999$, test statistic $F=1660.549$, and $p=0.000$ verify the validity of the Lasso regression model. From the regression model, we can see that the 8 independent variables all have a positive effect on the dependent variable y , and the regression coefficients are of the same order of magnitude, so the degree of influence is almost the same. This is similar to the results of ridge regression. However, in the 8 standard Lasso regression regression coefficients, due to $\beta_6 = \beta_4 > \beta_8 > \beta_3 > \beta_1 > \beta_2 > \beta_7 > \beta_5$, it is possible to judge the degree of influence of the 8 independent variables on the dependent variables, that is, the factors affecting the retail sales of consumer goods in China, according to the degree of their influence, from large to small: consumer price index, per capita disposable income of rural residents, national fiscal expenditure, per capita disposable income of urban residents, population number, urbanization rate, added value of the tertiary industry, and commodity retail price index, which is basically consistent with the actual situation.

6. Based on the ARIMA model, the total retail sales of consumer goods in China are predicted

6.1 Data sources

This paper selects the annual data of total retail sales of social consumer goods in China from 1973 to 2020 as the research object (see Table 8), and the data comes from the database website of the National Bureau of Statistics. A total of 48 sample data were used to build an ARIMA model using R software to predict the total retail sales of consumer goods in China in the next ten years.

Table 8 Total retail sales of consumer goods in China from 1974 to 2021

Annual	Total Retail Sales/100 Million	Annual	Total Retail Sales/100 Million	Annual	Total Retail Sales/100 Million
1973	1106.7	1989	8101.4	2005	66491.7
1974	1163.6	1990	8300.1	2006	76827.2
1975	1271.1	1991	9415.6	2007	90638.4
1976	1339.4	1992	10993.7	2008	110994.6
1977	1432.8	1993	14270.4	2009	128331.3
1978	1558.6	1994	18622.9	2010	152083.1
1979	1800	1995	23613.8	2011	179803.8
1980	2140	1996	28360.2	2012	205517.3
1981	2350	1997	31252.9	2013	232252.6
1982	2570	1998	33378.1	2014	259487.3
1983	2849.4	1999	35647.9	2015	286587.8
1984	3376.4	2000	39105.7	2016	315806.2
1985	4305	2001	42240.4	2017	347326.7
1986	4950	2002	47124.6	2018	377783.1
1987	5820	2003	51303.9	2019	408017.2
1988	7440	2004	58004.1	2020	391980.6

6.2 Model analysis

6.2.1 Stationarity test of the model

According to the sample data, the time series curve of the total retail sales of social consumer goods in China is plotted, and it is found that the total retail sales of social consumer goods in China have a clear growth trend, and the sequence is judged to be a non-stable time series. From Figure 4, we see that the p-value is greater than the significance level of 0.05, and the null hypothesis cannot

be rejected, then the sequence is non-stationary. Therefore, the time series needs to be smoothed before the model can be established.

```

Augmented Dickey-Fuller Test

data: y
Dickey-Fuller = -3.4354, Lag order = 3, p-value = 0.06179
alternative hypothesis: stationary
    
```

Fig. 4 ADF test results

6.2.2 Identification of the model

Use R to do Log smoothing first, then do a difference, the time series is shown in Figure 5. From Figure 5, we see that the sequence becomes a stationary series after the 1st order difference, and the autocorrelation plot shows that the autocorrelation coefficient is rapidly reduced to 0 after lagging 1st order. Further indicate that the sequence is stable. Preliminary determination of $d=1$, below we use `auto.arima` to find the best model, from Figure 6 we can see that the best model is ARIMA(0,1,1).

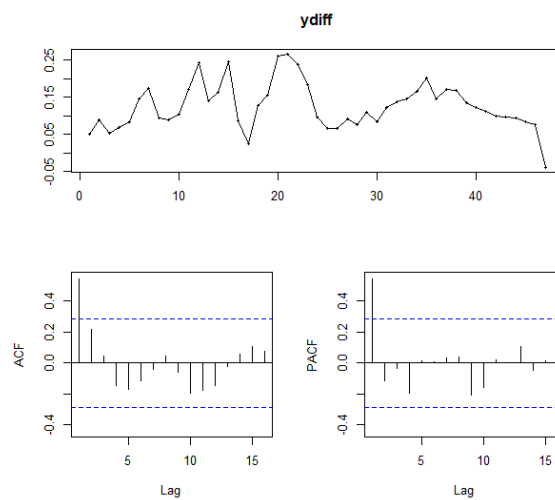


Fig. 5 Results of one smoothing and one differential

```

ARIMA(2,1,2) with drift          : Inf
ARIMA(0,1,0) with drift         : -123.6655
ARIMA(1,1,0) with drift         : -142.2324
ARIMA(0,1,1) with drift         : -143.1308
ARIMA(0,1,0)                    : -49.71524
ARIMA(1,1,1) with drift         : -142.4621
ARIMA(0,1,2) with drift         : -141.7587
ARIMA(1,1,2) with drift         : -140.4707
ARIMA(0,1,1)                    : -96.17479

Best model: ARIMA(0,1,1) with drift

Series: ylog
ARIMA(0,1,1) with drift

Coefficients:
      ma1      drift
    0.7438  0.1215
s.e.  0.1063  0.0123

sigma^2 = 0.002489:  log likelihood = 74.84
AIC=-143.69  AICC=-143.13  BIC=-138.14
    
```

Fig. 6 Optimal model selection results

6.2.3 Estimation and diagnosis of model parameters

R is used to estimate the model parameters, and the estimation results are shown in Figure 7. The AIC value is the criterion for judging whether the model fits well or badly, the smaller the better, and the AIC value of ARIMA (0,1,1) is -81.91.

```

Call:
arima(x = ylog, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 1), period = 12),
      method = "ML")

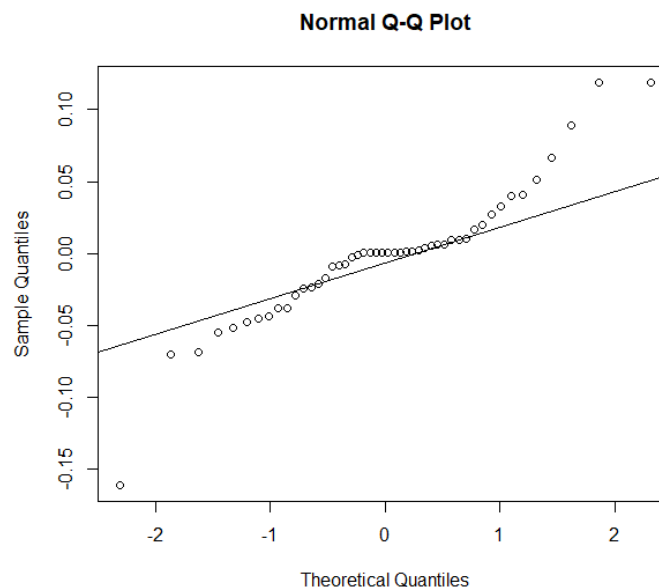
Coefficients:
      ma1      sma1
    0.7192 -0.9993
s.e.  0.1394  0.5314

sigma^2 estimated as 0.002916: log likelihood = 43.95, aic = -81.91

```

Fig. 7 ARIMA(0,1,1) model parameter estimation results

The model parameter test consists of two tests: a significance test for parameters and a normality and independence test for residuals. Significance test of the parameter: the quotient obtained by dividing the estimated coefficient by its standard deviation (se) is compared with the critical value of 5% of the T statistic (1.96), and the absolute value of the quotient is greater than 1.96, the null hypothesis is rejected, and the coefficient is considered to be significant if it is not 0, otherwise the coefficient is not significant. Here $0.7192/0.1394 \approx 5.16 > 1.96$ rejects the null hypothesis, and the coefficient is significantly not 0. Normality test of residuals: Draw a QQ plot of residuals to judge, qq plots in which the residuals basically fall completely at the 45° line, that is, in line with the normality hypothesis. Otherwise, the model may have errors. Residual independence test: Also known as the residual white noise test, from the previous definition of white noise, the residual (= estimated value - true value) should be an uncorrelated sequence. LB statistics are commonly used to test for residuals. Follows a chi-square distribution with K-p-q degrees of freedom, where n is the sample size.



Box-Ljung test

```

data: arimal$residuals
x-squared = 0.11442, df = 1, p-value = 0.7352

```

Fig. 8 Residual QQ plot and LB test results

From the results of the QQ plot and LB test of Figure 8, $p > 0.05$, accept the null hypothesis, the residuals conform to the normality hypothesis and are not correlated, then the model fitting data is considered to be sufficient and can be used for the next prediction.

6.2.4 Future value prediction

Through the fitted ARIMA model (0, 1, 1), the total retail sales of social consumption in the next decade are forecasted, and the forecast results are shown in Table 9.

Table 9 Model prediction results

Year	Actual forecast value	Lo 99.5	Hi 99.5
2021	397086.7303	334248.4976	471738.4596
2022	441639.8456	314849.4919	619495.0564
2023	498390.906	318905.4682	778901.5992
2024	568183.6654	333397.2497	968302.6768
2025	620456.0183	337975.9439	1139044.146
2026	673612.2988	343416.7489	1321291.202
2027	754673.1345	362242.8057	1572252.716
2028	865471.7388	392923.4154	1906329.074
2029	1002101.647	431900.2769	2325114.865
2030	1159744.175	475950.9111	2825935.449

According to Table 9, it can be seen that in 2020, due to the epidemic, although the total retail sales of consumer goods in China have declined, the total retail sales of consumer goods in China will gradually increase between 2021 and 2030.

7. Conclusions and recommendations

The per capita disposable income of rural households is constantly increasing. The increase in the per capita disposable income of rural household residents can increase the total social consumption, thereby expanding domestic demand and stimulating consumption, while helping to reduce the current urban and rural income gap and achieve social equity.

Promote the steady growth of per capita disposable income of urban household residents. The disposable income of urban household residents occupies the main position, and by raising income and enhancing the willingness to purchase power, thereby expanding consumption, promoting the gradual increase in the total retail sales of social consumer goods, and thus maintaining the vitality of the domestic consumer market.

To a certain extent, the intensity of national fiscal expenditure has been increased. The results of the analysis of the four models show that the impact of national fiscal expenditure on the total retail sales of consumer goods in China is still very large. Increasing national fiscal expenditure can not only meet more consumption demand, but also increase the disposable income of residents, so a reasonable increase in national fiscal expenditure will not only help to promote the gradual increase of the total retail sales of social consumer goods, but also play an important role in improving people's livelihood.

Formulate relevant policies and vigorously promote the development of the tertiary industry. Since the more active the tertiary industry, the higher the total retail sales of social consumer goods, and the retail sales of social consumer goods can reflect the development level of some industries in the third night in a region, so while promoting the development of the tertiary industry, it can also increase the total retail sales of social consumer goods in China to a certain extent.

References

- [1] Feng Haiqi,Zhao Yumei,Wang Linlin. Analysis of influencing factors of retail sales of social consumer goods in Shandong Province——Based on Lasso regression model[J].Journal of Ludong University (Natural Science Edition),2021,37(04):310-314.
- [2] Comfortable Hua. Analysis of influencing factors of retail sales of social consumer goods in China——Based on Ridge Regression[J].Journal of Taiyuan Normal University (Social Science Edition), 2019,18(06):44-50.
- [3] Xue Song,Zhou Jing,Zhang Qingkai,Li Xu,Meng Yuwei. Quantitative analysis of the influencing factors of total retail sales of consumer goods in China[J].Times Finance, 2020(13):28-29.

- [4] Yang Yang, Yang Guiyuan. Evaluation of the impact of the new crown pneumonia epidemic on the total retail sales of consumer goods in China[J]. Value Engineering, 2020, 39(18): 31-33. DOI: 10.14018/j.cnki.cn13-1085/n.2020.18.015.
- [5] WU Long. Principal Component Regression Analysis of Influencing Factors of Total Retail Sales of Social Consumer Goods[J]. Development reform theory and practice, 2017(10): 49-51. DOI: 10.13814/j.cnki.scjyjg.2017.10.010.
- [6] Wang Zhijian, Wang Binhui. Forecast of Total Retail Sales of Social Consumer Goods Based on ARMA Model[J]. Statistics and Decision, 2014(11): 77-79.
- [7] Chen Yunjie, Xu Ping, Bai Tongyuan. Analysis and Prediction of Total Retail Sales of Social Consumer Goods Based on ARIMA Model[J]. Modern Commerce and Trade Industry, 2021, 42(34): 26-27. DOI: 10.19311/j.cnki.1672-3198.2021.34.011.
- [8] Fan Liang. Prediction analysis of total retail sales of social consumer goods based on ARIMA model: A case study of counties in Longnan City, Gansu Province[J]. Journal of Tonghua Normal University, 2020, 41(08): 29-34. DOI: 10.13877/j.cnki.cn22-1284.2020.08.006.
- [9] Lin Tao. The short-term impact of the epidemic on the retail market is large, and the long-term trend of consumption has not changed. http://www.stats.gov.cn/tjsj/sjjd/202003/t20200316_1732420.html.
- [10] Wang Chunjie, Wen Linan, Ma Yuanjia. Empirical analysis of rebar futures prices based on Ridge Regression and Lasso Regression[J]. Journal of Jilin Normal University (Natural Science Edition), 2020, 41(01): 36-41. DOI: 10.16862/j.cnki.issn1674-3873.2020.01.007.
- [11] Hu Yuwen. Analysis of influencing factors of R&D investment intensity of universities in Jiangxi Province based on LASSO regression[J]. Science and Technology and Industry, 2020, 20(05): 84-88.