The Predicting Model of Urban Residents' Pension Income and Expenditure based on Multiple Linear Regression

Liu Yuanjing [a] , Shi Xiaotao [b], Gang Jiatai [c*]

Dalian University, Dalian, Liaoning province, China

[a]liuyuanjing_vince@163.com, [b]sxt0915@163.com, [c]gjt1960@126.com

* Corresponding author: Gang Jiatai.

Keywords: multiple linear regression; pension gap ;gross error analysis; gray correlation degree

Abstract: In order to keep the social security system have sustainable development and solve the problem of population aging, this paper adopts the method of grey correlation degree analysis among the influencing factors of pension income and expenditure, picking out the best index to reflect the income and expenditure of urban residents' pension, using multiple linear regression method, establishing the predicting model of income and expenditure of urban residents' pension, so as to test the validity and rationality of the model. The results showed that this model had high fitting degree with small error, and the results were reasonable and reliable. It can be knew from the prediction results of the model that: by 2020, it will exist pension gap, the pension gap in 2025 will exceed 500 billion. Thus, it can provide the theoretical basis for the relevant departments to develop the relevant policies.

Introduction

In recent years, with the aging population of Chinese era, the problem of social pension insurance fund gap becomes a hot issue. China has entered the era of an aging population in the 90 years of the last century, the speed of population aging is increased quickly, therefore, it must carry on the reform corresponding to the basic old-age insurance system, so as to ensure the basic protection of the elderly after retirement. The study found out that the basic pension fund faced a huge revenue and expenditure pressure [1]. How to design a basic old-age insurance system that can bring benefit to the country's future economic development has become an important issue in front of us [2]. As for the prediction of the gap of China's pension problem, document [3] adopted the support vector machine method to establish prediction model, document [4] used grey GM (1,1) model to predict, ignoring the multiple correlation of index data of pension, which can affect the estimation of the parameters, therefore, it can make the model error increased, the predicted results can be inaccurate and affect the overall judgment of pension situation of urban residents' income and expenditure in china. In this paper, it adopts the method of grey correlation degree analysis among the influencing factors of pension income and expenditure, picking out the best index to reflect the income and expenditure of urban residents' pension, using multiple linear regression method to establish the predicting model for urban residents of pension payments, so as to provide a theoretical basis for the relevant departments to formulate corresponding policies.

Problem Analysis

Factors that can affect the income and expenditure of urban residents' pension are [5]: rate of substitution, rate of payment, the number of the insured staff in service, the number of the insured retirees and other 19 factors, which can be shown in Table 1. Among them, rate of substitution refers to the basic old-age insurance pension per capita accounted for the average wage rate of the

on-the-job workers in the unit of the town,while the rate of payment refers to the basic pension per capita payment accounted for the average wage rate of on-the-job workers in the unit of the town.

According to the released data of National Bureau of Statistics from 2007 to 2015, first of all, it can have gross error analysis on the data [6] and make standardization data processing; secondly, it can adopt the grey correlation degree analysis method [7], so as to determine the degree of correlation between the indexes and then select the indexes that can best reflect the income and expenditure of urban residents' pension; finally, using the selected indexes, adopting multiple linear regression method, establishing the predicting income and expenditure model of urban residents' pension, so as to make analysis on the validity and accuracy of the model.

Data preprocessing

Original data extraction. The index data from 2007 to 2015 is much more completed [8,9], so it can adopt the data to be used to study the prediction of the income and expenditure of the pension. After sorting out the primary index of the data, it can be shown in Table 1.

Table 1 Summary of annual data of primary index variables

Year		2007	2008	2009	2010	2011	2012	2013	2014	2015
Urban pension income	(Billion yuan)	3680	4258	5093	6301	7834	9740	11941	13420	16895
Urban pension expenditure	(Billion yuan)	3122	3502	4040	4897	5965	7390	8894	10555	12765
Rate of substitution	(%)	0.58	0.53	0.5	0.5	0.48	0.48	0.47	0.45	0.44
Rate of payment	(%)	0.19	0.18	0.18	0.18	0.17	0.17	0.16	0.15	0.15
The number of the insured staff in service	(Ten thousand people)	11646	12250	13120	14130	15183	16587.5	17743	19402.3	21565
Number of the insured retirees	(Ten thousand people)	38602	4103	4367.5	4635	4953.7	5303.6	5806.9	6305	6826
The death rate of people from 0-20 years old	（ppm）	1124.8	1204.5	1232.47	1148.1	1188.4	1054.6	1035.71	1293.6	983.7
The death rate of people from 21-60 years old	(‰)	21.34	20.9	20.73	16.1	15.71	14.26	15.04	16.1	15.29
The death rate of people over 61 years old	(‰)	411.86	402.65	398.12	353.49	333.94	310.26	325.9	394.49	319.01
The growth rate of GDP	(%)	10	10.1	11.3	12.7	14.2	9.6	9.2	10.4	9.3
The growth rate of GDP per captia	(%)	9.3	9.4	10.7	12	13.6	9.1	8.7	9.9	8.8
Financial subsidy	(Billion yuan)	530	614	651	971	1157	1437	1646	1954	2272
Wage level	(Yuan)	14040	16024	18364	21001	24932	29229	32736	37147	42452
CPI(Taking the above year as the base period)	-	101.2	103.9	101.8	101.5	104.8	105.9	99.1	103.3	105.4
Inflation rate	(%)	1.2	3.9	1.8	1.5	4.8	5.9	-0.9	3.3	5.4
CPI(Taking 2007 as the base period)	-	101.09	105.04	106.93	108.53	113.74	120.45	119.61	123.55	130.23
Commodity retailing price index	(%)	99.9	102.8	100.8	101	103.8	105.9	98.9	103.1	104.9
Rate of return on investment	(%)	3.56	2.61	4.16	29.01	43.19	-6.79	16.12	4.23	0.84
Investment income	**(Billion yuan)**	**44.71**	**36.72**	**71.22**	**619.79**	**1453.5**	**-393.72**	**850.43**	**321.22**	**73.37**

Gross Error Analysis. Due to the impact of human factors on the process of having statistical data, abnormal data may be occurred, if it is deleted, the accuracy of the measurement results will be low. Therefore, this paper adopts the method of the residual of student-oriented [10] to have gross error analysis on the original data, excluding the impact of abnormal data on the measurement results. "Student-oriented" is mainly aimed at the common residual that does not have the variance homogeneity [11], but it can carry on one kind of standardization processing, causing it to be expected to be 0, with 1 as the variance, which is is easy to make comparison.

According to the definition of the student- oriented residual error, the annual data for each index can be calculated, Fig. 1 is the income and expenditure of urban residents' pension,as well as the index data of the student- oriented residual in 2007-2015. When the absolute value of the student-oriented residual is greater than 3, the data can be considered as the variation data, namely, the gross error data, which can be eliminated. But it can be seen from Fig. 1, the absolute value of the majority of indexes of the student-oriented residual is less than 3, so that no gross errors are existed in the above index data, it can directly carry out the subsequent analysis as well as processing.
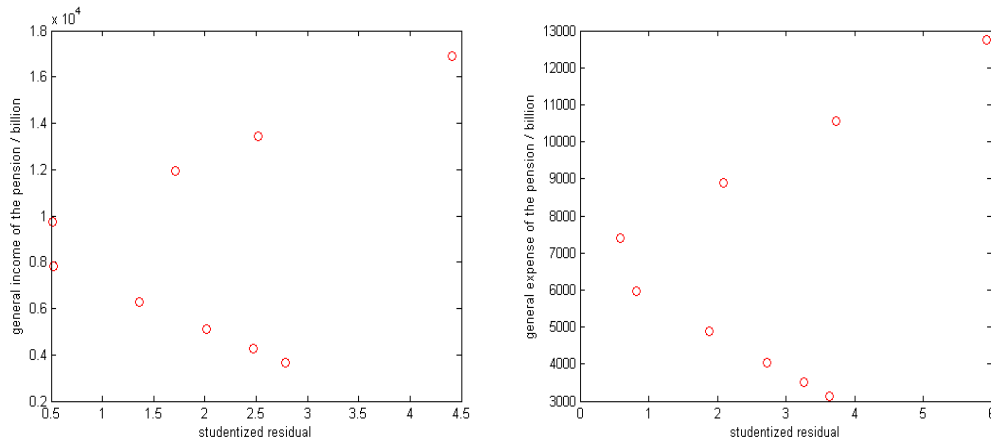
(a) The income of the student-oriented residual (b)The expenditure of the student-oriented residual

Fig.1 Annual income and expenditure of urban residents' pension of the student- oriented residual in 2007-2015

Analysis on Index Variables Based on Grey Relational Degree. This paper adopts gray correlation analysis method, calculating out the close degree of the relationship between the income and expenditure of the urban residents' pension with the various indexes, then it can delete the index that the relationship is not close, so as to achieve the purpose of reducing the dimension of the model.

In the process of system development, if the trend of the two factors change is consistent, that is, the degree of synchronization change is higher, then the relationship between the two is higher; otherwise, it will be lower. Therefore, the grey correlation analysis method, which is based on the similarity or dissimilarity degree of the development trend of the factors, namely, "gray correlation degree", can be regarded as a method to measure the degree of correlation between factors.

Using SPSS to calculate the grey relational degree. In Table 1, the grey correlation degree of the 14 indexes is less than 0.8, so we can know that they are not closely related to the income and expenditure of the urban residents' pension [12]. Therefore, the first to do is to eliminate these indexes, using the remaining indexes to set up modeling. After selecting the indexes before the remaining, 5 indexes are left, which can be shown in Table 2.

Table 2 Part of the indexes are related to the grey correlation degree of income and expenditure of urban residents' pension

	Rate of substitution	Rate of payment	Rate of substitution	The number of insured staff in service	The number of insured retirees	Financial subsidy
Income of pension	0.932	0.889		0.915	0.864	0.875
Expenditure of pension	0.937		0.852	0.921	0.847	0.877

Based on the multiple linear regression model of income and expenditure of urban residents' pension

Multiple linear regression model refers to a linear regression model with multiple explanatory variables, which can be used to represent the linear relationship between the explanatory variables and the other variables. Its regression equation is as follows:

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p \qquad (1)$$

Multiple linear regression used SPSS software to have implementation, the specific process is as follows:

The remaining 5 indexes after the previous screening are as follows: rare of substitution, rate of payment, the number of insured staff in service, the financial subsidies as well as the price index of consuming. Extracting the original data of these 5 indexes correspondingly in Table 1, adopting the method of stepwise regression [13], so as to establish the multiple linear regression model, thus the results can be shown in Table 3:

Table 3 The Parameters of Model and Testing Parameters

Model		Coefficient	Adjusting RParty	$Sig.$
Income of pension	(constant)	-42909		0.005
	The number of the insured staff in service	1.823	0.998	0.000
	Rate of payment	135299.066		0.023
Expenditure of pension	(constant)	-17257.155		0.000
	The number of the insured staff in service	1.14	0.998	0.000
	Rate of substitution	12412.33		0.009

The result of Table 3 showed us that the model of pension income took the number of the insured staff in service and the rate of contribution as the independent variables, while the expenditure model took the number of the insured staff in service and the rate of substitution as the independent variables. After the adjustment,both of their R party was 0.998, which told us that the fitting degree of these two models are very high; each coefficient of the variables can pass the test of significance ($p < 0.05$), the level of significance of the model is much higher.

According to the above results of the regression model, it can visually see from the figure, and the model of pension income as well as the model of urban residents' expenditure model can be shown as follows:

The income of urban residents' pension=-42909.367+1.823×The number of the insured staff in service+135299.066×Rate of payment
The expenditure of urban residents' pension=-17257.155+1.140×The number of the insured staff in service+12412.33×Rate of substitution
The fitting results can be shown in Fig. 2, it can visually see from the figure that the fitting degree of these two models is very high.It means that the accuracy of this model is very high, and the error is within the reasonable range.

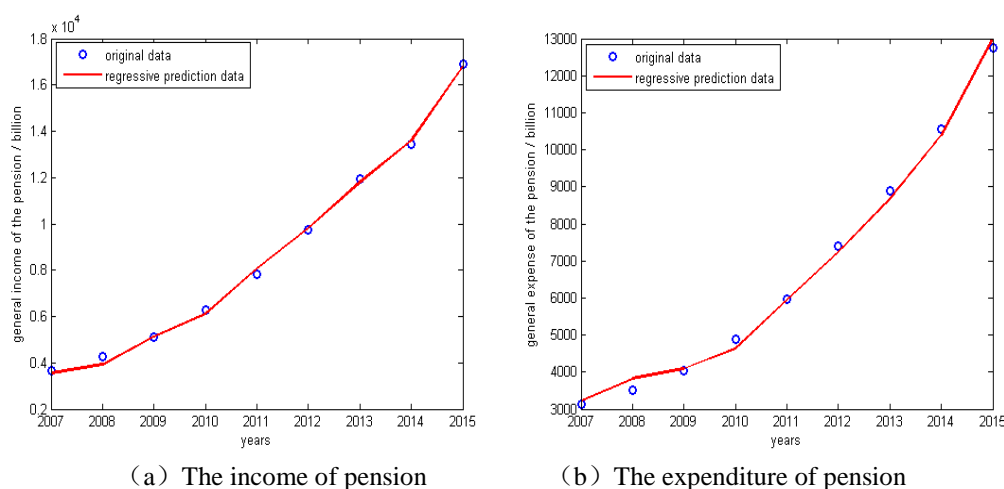（a）The income of pension　　　　　（b）The expenditure of pension

Fig.2 Multiple linear regression model of urban residents' income and expenditure of pension

Using the trained model to forecast the income and expenditure of the pension, which can be shown in Table 4.

Table 4 The prediction of income and expenditure of pension based on multiple linear regression model

particular year	Revenue (100 million)	Expenditure (100 million)	Revenue - Expenditure (100 million)
2016	16919	13665	3254
2017	18544	15733	2811
2018	20169	18480	1689
2019	21793	21247	546
2020	23418	23743	-325
2021	25043	26275	-1232
2022	26667	28841	-2174
2023	28292	31408	-3117
2024	29917	33975	-4059
2025	31541	36542	-5001

From Table 4,it can be seen that: by 2020, the expenditure will exceed the income, there will be a gap existing in pension, after that, the gap will be bigger and bigger, by 2025, it will exceed 500 billion, therefore, the social pooling account of pension will bring a huge burden to finance.

Conclusions and Suggestions

Through calculation, $R^2=0.998$ in the income and expenditure model, the coefficients of each independent variables have passed the significance test ($p<0.05$). The results can show us that the fitting degree of model is high, with small error, and the results of the prediction are reasonable and reliable. From the prediction results of the model, it can show that: from 2020,the pension requirements of the financial subsidies will continue to rise, which will occur the pension gap, thus, by 2025, it will exceed 500 billion, with is estimated by the World Bank [14], if you do not change the current system and model, by 2075, the gap of China's pension payment will reach 9150 billion.

Based on the above, this paper put forward some suggestions, which can provide some theoretical basis for the relevant departments to develop appropriate policies: (1) to postpone the age of taking the pension , which can slow down the pension gap; (2) to extend the endowment insurance lowest age so as to expend the fixed number of year; (3) to increase the individual payment, personal accounts can be available ; (4) to increase the strategic reserves.

Reference

[1] Ai Hui, Zhang Yang, Yang Changyu, Wu Yandong. Study on the financial sustainability of China's endowment insurance whole account: Based on the prediction and calculation of the open system [J]. *Financial Research*, 2012,02:91-101.

[2] Liu Xueliang. The income and expenditure gap of China's pension insurance and sustanabile research [J]. China Industrial Economy, 2014,09:25-37.

[3] Wang Chunlan, Ye Shangbin. Modeling and forecasting of the pension gap of urban residents in China [J]. Statistics and Decision Making, 2015,08:160-162.

[4] Ma Yuan. Research on the the small and medium-sized enterprise financing efficiency in Xi'an city by taking gray correlation analysis [J]. Accounting Monthly Periodicals, 2010,36:66-68.

[5] Wang Yuxi, Wang Hong, Xiao Jun. Prediction on Shanghai urban pension insurance population distribution by means of gray correlation analysis based on the grey model GM (1,1)[J]. System Engineering Theory and Practice, 2010,12:2244-2253.

[6] Fu Xinping, Zou min, Zhou Chunhua, Gao Zhuqiao. Analysis on influencing factors ofpension insurance fund balance under the new policy [J]. Journal of Wuhan University of Technology (SOCIAL SCIENCE EDITION), 2007,02:150-154.

[7] Liu Zonbao, Gao Shiqiao, Du Jingqing. Algorithm of measurement data with eliminating gross error and smooth processing [A]. Technical Committee on Control Theory,Chinese Association of Automation、Systems Engineering Society of China.The paper set of the thirty first Committee on Control Theory,China D volume[C].Technical Committee on Control Theory,Chinese Association of Automation、Systems Engineering Society of China,2012:6.

[8]National Bureau of Statistics. Yearbook of China Statistics[Z].http://www.stats.gov.cn/tjsj/ndsj/.

[9] National Bureau of Statistics. Yearbook of China Labor Statistics [Z]. CNKI; statistics database of Chinese economic and social development

[10] Wang Hong, Tong Hengqing, Wei Ping. Stepwise regression modeling of trade volume forecast of import and export [J]. Statistics and Decision, 2006,02:80-82.

[11] Wang Zhongyu, Zhang Haibin, Liu Zhimin. A new method to eliminate outliers of the student-oriented of the residual[J].Journal Instrument, 2006,06:624-628+637.

[12] Liu Yingan, Wei Bocheng. The non-linear regression model with bilinear BL (1,1,1,1) error correlation and homogeneity of variance test [J].System Science and Mathematics,2005,03:366-377.

[13] Tian min, Liu Sifeng, Bu Zhikun. Review of the research on algorithm model of grey relational degree [J]. Satistics and Decision Making, 2008,01:24-27.

[14] Zhang Xiantong, Hao Yong, Wang Shanshan. Research and study on the pension fund raising and counting methods [J]. Labor Security World, 2009,10:69-73.