Analysis of Factors Influencing the Demand for Agricultural Insurance Based on Grey Correlation and Principal Component Analysis - An Example from Liaoning Province

Xiaoting Dong[1,a], Hui Che[1,b], Tingrui Liu[1,c]

596008988@qq.com[a], 1289266932@qq.com[b], 1544250485@qq.com[c]

School of Safety Engineering, Shenyang Aerospace University, Shenyang, Liaoning, China 110136[1]

Abstract: Tn this paper, grey correlation analysis was conducted on 17 characteristic indicators from various perspectives, such as socio-economic and agricultural production, and 12 indicators with higher correlation were selected to provide a basis for demand forecasting. Secondly, principal component analysis was used to reduce the dimensionality of the 12 indicators, and finally the expressions of the five principal components and the comprehensive scores were obtained. Subsequently, principal component regression analysis was used to empirically analyse the panel data of 14 cities in Liaoning Province from 2004 to 2020, and a multiple linear regression model was derived and the regression equation was obtained. The regression equation is derived from the regression model.

Keywords: Agricultural insurance demand, Grey correlation, Principal component analysis, Panel data.

1 INTRODUCTION

Covering an area of 148,000 square kilometres, Liaoning Province is one of the second batch of 10 pilot provinces in China for agricultural insurance [1]. Liaoning is also a large agricultural province, with an arable land area of 4,217.1 kilo hectares and the second highest grain production in China with 50.77 billion kg in 2021. In November 2021, the "Implementation Plan for Accelerating the High-Quality Development of Agricultural Insurance in Liaoning Province" was promulgated, signifying a qualitative leap in support for agricultural insurance in Liaoning Province [2]. Measured in terms of insurance density and depth, it is also evident that agricultural insurance in Liaoning Province has experienced leaps and bounds in recent years, reflecting the fact that the insurance awareness of agricultural practitioners has been significantly raised in recent years through the active efforts of all parties [3]. Therefore, accurate forecasts of insurance influencing factors are very important in analysing future trends in the agricultural insurance industry [4].

The subject of insurance influencing factors has been paid attention to by a wide range of scholars in the study of agricultural insurance development. Xie Qian et al. conducted a meta-analysis of the empirical findings of the existing domestic agricultural insurance literature and found that the influencing factors were easily influenced by the "number of independent variables", "study year", "whether government subsidies were considered" and "sample size"[5].

ICBBEM 2023, May 19-21, Hangzhou, People's Republic of China

DOI 10.4108/eai.19-5-2023.2334429

Han Ke et al. established a logistic regression model to analyse the factors influencing farmers' willingness to demand agricultural machinery insurance, and found that farmers' household income, farmers' education level, government premium subsidies and external working area were important factors influencing farmers' willingness to demand [6]. Feng Hao et al. used a logit model to construct a meta-regression analysis of the significance of mediating variables on the factors influencing agricultural insurance demand [7]. Cai Guiquan et al. proposed a multicore support vector machine based agricultural insurance demand forecasting method to explore a new model for agricultural insurance demand assessment in China [8]. Ezdini Sihem used Logit model to analyse the economic and social factors influencing agricultural insurance demand in various countries [9]. Lv Caifei selected the panel data of 31 provinces and cities in China from 2008 to 2018. A random effects model was used to study the direction and intensity of the impact of agricultural insurance and its compensation on agricultural production [10]. Therefore, this paper uses grey correlation method and principal component analysis to predict the impact factors of agricultural insurance in order to enrich the current theoretical research on agricultural insurance demand, and also to provide reference and reference for the research on agricultural insurance in other provinces.

2 GREY CORRELATION METHOD AND SELECTION OF INDICATORS

2.1 Analysis based on grey correlation forecasting method

Firstly, the data are unrigidly quantified. In order to study the correlation size of the data, the degree of correlation between the data is measured by the correlation degree, which provides a theoretical basis for the analysis of the results. In this paper, the selected indicators are normalised and the very small, intermediate and interval indicators are uniformly transformed into very large indicators, and the final matrix of the normalisation process is obtained as X,the expression of matrix is as formula (1),

$$\left(X_1^{'}, X_2^{'} \ldots, X_n^{'}\right) = \begin{pmatrix} x_1^{'}(1) & x_2^{'}(1) & \ldots & x_n^{'}(1) \\ x_1^{'}(2) & x_2^{'}(2) & \ldots & x_n^{'}(2) \\ \ldots & \ldots & \ldots & \ldots \\ x_1^{'}(m) & x_2^{'}(m) & \ldots & x_n^{'}(m) \end{pmatrix} \tag{1}$$

where m is the number of indicators, and its expression is as formula (2)

$$X_i^{'} = \left(x_i^{'}(1), x_i^{'}(2), \ldots, x_i^{'}(m)\right)^T \quad i = 1,2,\ldots,n \tag{2}$$

In this paper, 17 indicators such as agricultural insurance expenditure, gross product per capita and household Engel coefficient are selected as characteristic indicators, and the data are summarised in Table 1. The household Engel coefficient is a very small indicator, and the remaining 16 indicators, such as rural insurance payout expenditure and gross product, are very large indicators.

We set the parent series to agricultural insurance premium income and the characteristic series to the 17 indicators mentioned above to obtain grey correlation coefficient values. The correlation coefficients of the corresponding elements of each comparison series and the reference series are calculated separately. The calculation formula is shown in equation (3).

$$z_i(k)=\frac{\min_i \min_k |x_0(k)-x_i(k)|+r\times\max_i \max_k |x_0(k)-x_i(k)|}{|x_0(k)-x_0(k)|+r\times\max_i \max_k |x_0(k)-x_i(k)|} \quad (3)$$

In equation (3) is the ρdiscrimination factor,$0 < \rho < 1$,k=1...m, If ρ is smaller, the greater the difference between the correlation coefficients, the greater the differentiation ability. After solving for the value of the grey linkage between the parent series and the feature series. At this point, the resolution coefficient is 0.5, and the value of the correlation coefficient can be calculated by combining with the formula for calculating the correlation coefficient, and then calculating the correlation degree value for evaluation. The next step is to carry out grey correlation analysis for the above 17 evaluation items and related data of Liaoning Province from 2008 to 2020, and use the agricultural insurance premium income of Liaoning Province as the "reference value" (parent series) to study the correlation degree of the 17 evaluation items.

2.2 Analysis of relevance results

The mean values of the correlation coefficients between each evaluation object and the corresponding elements of the reference series were calculated separately for each evaluation object to reflect the correlation relationship between each evaluation object and the reference series, and were called correlation order. Based on the correlation order of each observation object, the results of the above 17 observation objects were firstly weighted to obtain the correlation degree value, and the 17 observation objects were ranked. Table 1 shows that, for the 17 items evaluated, the total expenditure per capita of farmers' households/yuan was the highest rated (correlation degree: 0.806), which was close to 1, implying that it was the highest rated. This is followed by total income per farmer household/yuan (correlation: 0.791).

Table 1 Results of the correlation analysis

Relevance results		
Evaluation items	Relevance	Ranking
Total per capita expenditure of farmers' households/$	0.806	1
Farmers' gross household income per capita/$	0.791	2
Agricultural insurance payout ratio/%	0.786	3
Disposable income per farmer household/$	0.785	4
GDP per capita/billion	0.726	5
Total output value of agriculture, forestry, animal husbandry and fishery/billion yuan	0.696	6
Age structure/year	0.671	7
Total crop area sown/thousand hectares	0.657	8
Years of schooling per farmer/year	0.656	9

Relevance results		
Evaluation items	Relevance	Ranking
Area of crops affected/thousands of hectares	0.639	10
Total power of agricultural machinery / million kW	0.631	11
Engel's coefficient for farmers' households/%	0.6	12
Area of crops affected/thousands of hectares	0.583	13
Area of crop damage/thousand hectares	0.532	14
Payout ratio/%	0.501	15
Number of rural population / 10,000 people	0.483	16
Growth rate/%	0.469	17

2.3 Selection of indicators

According to the grey correlation analysis, the correlation degree of each influencing factor was obtained as shown in the table, and ranked, and the 12 objects with correlation degree of 0.5 or above were selected as the research indicators of this paper. That is, agricultural insurance payout expenditure, per capita gross product, Engel coefficient of farmers' households, per capita gross income of farmers' households, per capita disposable income of farmers' households, per capita gross expenditure of farmers' households, total sown area of crops, area affected by crops, total output value of agriculture, forestry, animal husbandry and fishery, total power of agricultural machinery, per capita years of education of farmers, and age structure.

3 ANALYSIS OF FACTORS INFLUENCING DEMAND FOR AGRICULTURAL INSURANCE

3.1 Principal component analysis

The value of KMO was tested to be 0.544. Bartlett's spherical test of significance p-value was 0.000***, showing significance at the level. It was concluded that both df and p-values satisfy the conditions for which principal component analysis can be performed, the variables are correlated and can be used for principal component analysis, and the principal component analysis is valid. From the gravel plot analysis, it was found that the slope tends to level off at the point where the principal component is 5, thus judging that the number of extracted principal components should be 5.

Table 2 shows the component matrix, which can illustrate the factor score coefficients, or principal component loadings, contained in each of the five principal components. Using them to calculate the component scores leads to the factor formula, which is calculated as: linear combination coefficient * (variance explained÷cumulative variance explained). Finally, it is normalised to the factor weight score, and the linear combination factor formula is: factor loading coefficient÷ corresponding to the characteristic roots (coefficients of the component matrix).

Table 2 Table of component matrices

Component Matrix					
Name	Ingredients				
	Component 1	Component 2	Component 3	Component 4	Component 5
Agricultural insurance payout expenditure/$m	0.1	0.13	-0.254	0.066	1.183
GDP per capita/billion	0.112	-0.081	0.061	-0.144	0.314
Engel's coefficient for farmers' households/%	-0.105	0.018	0.18	0.209	1.4
Farmers' gross household income per capita/$	0.114	-0.1	0.014	-0.025	0.068
Disposable income per farmer household/$	0.113	-0.113	0.026	-0.033	0.421
Total per capita expenditure of farmers' households/$	0.112	-0.12	-0.03	-0.022	-0.502
Total crop area sown/thousand hectares	0.072	0.348	0.262	1.275	-0.259
Area of crops affected/thousands of hectares	0.013	0.359	-0.752	0.065	-0.068
Total output value of agriculture, forestry, animal husbandry and fishery/billion yuan	0.109	0.09	0.194	-0.292	0.379
Total power of agricultural machinery / million kW	0.019	0.638	0.305	-0.744	-0.145
Years of schooling per farmer/year	0.113	-0.09	-0.075	0.053	-0.05
Age structure/year	0.114	-0.051	0.079	-0.108	-0.173

For ease of observation and calculation, the 12 factors are denoted by letters: agricultural insurance payout expenditure (X_1), gross product per capita (X_2), Engel's coefficient of farmers' households (X_3), total income per capita of farmers' households (X_4), disposable income per capita of farmers' households (X_5), total expenditure per farmer household (X_6), total crop area sown/1000 ha (X_7), area affected by crop damage (X_8), total output value of agriculture, forestry, animal husbandry and fishery ((X_9), total agricultural machinery power (X), years of education per farmer (X)$_{10}$ X_{11}), and age structure (X_{12}). The five principal components were used as F_1 , F_2 , F_3 , F_4 and F_5 , we can obtain formula (4)

$$F = (0.725/0.989)\ F_1+(0.113/0.989)\ F_2+(\ 0.095/0.989)F_3 +(0.035/0.989)F_4 +(0.02/0.989\ F_5 \quad (4)$$

The weights for the principal component analysis were calculated as shown in Table 3, with a weight of 73.373% for principal component 1, 11.436% for principal component 2, 9.585% for principal component 3, 3.58% for principal component 4 and 2.026% for principal component 5. The maximum value of the indicator weights is principal component 1 (73.373%) and the minimum value is principal component 5 (2.026%).

Table 3 Factor weighting analysis

Name	Explanation of variance	Cumulative variance explained	Weighting
Main component 1	0.725	0.725	73.373%
Principal Component 2	0.113	0.838	11.436%
Principal Component 3	0.095	0.933	9.585%
Principal Component 4	0.035	0.969	3.58%
Main component 5	0.02	0.989	2.026%

3.2 Principal component regression analysis

FUsing SPSS 26.0 to bring the standardised raw big data into the five principal component expressions, five principal components can be calculated values of F_1 , F_2 , F_3 , F_4 and F_5 as the explanatory variables. The multiple linear regression analysis was carried out with the premium income of Liaoning Province from 2008 to 2020 as the explanatory variables. The adjusted R^2 value of the regression model is 0.955, which is close to 1, and the DW values are 2.113, 2.113≈ 2, thus indicating that the residuals are independent and the overall goodness of fit of the model is high. The table of variance results for the model shows an observed value of 47.970 for the F-statistic and a probability p-value of 0.000, suggesting that the five principal components are considered to be linearly correlated with premium income at a significance level of 0.05.

Table 4 Table of regression coefficients

Coefficient[a]							
	Unstandardised factor		Standardisation factor	t	Significance	Covariance statistics	
	Biased regression coefficient B	Standard errors	Beta			tolerance	VIF
(Constant)	-166.156	208.72		-0.796	0.456		
Main component 1	0.016	0.036	0.563	0.449	<0.01	0.003	35.871
Principal Component 2	-0.037	0.029	-1.069	-1.284	<0.01	0.006	17.599
Principal Component 3	-0.023	0.020	-0.207	-1.105	<0.01	0.116	8.651
Principal Component 4	0.056	0.033	1.177	1.701	0.004	0.009	11.619
Main component 5	0.027	0.006	0.734	4.861	0.003	0.178	5.604

According to the table 4, the coefficients of principal component 1, principal component 2 and principal component 3 are significant <0.01, while the coefficients of principal component 4 and principal component 5 are all <0.05, with a 95% probability of rejecting the original hypothesis, indicating that the regression coefficients of all five principal components are significant. The constant coefficient of 0.456 > 0.05 indicates that its value does not have a significant degree of influence on the regression equation. Therefore, from the coefficient analysis table, it can be seen that the multiple linear regression model ai formula (5):

$$y=0.016 F_1 -0.037 F_2 -0.023 F_3 +0.056 F_4+0.027 F_5 \quad (5)$$

This in turn gives the total regression equation as formula (6)

$$Y=0.038062X_1 +0.001905 X_2-0.163418 X_3+0.005638 X_4+0.01491 X_5 -0.007864X_6 +0.045325X_7 -0.763271X_8 -0.012167X -_9 0.075896 X_{10}+0.024006 X_{11}-0.008825 X_{12} \quad (6)$$

4 EMPIRICAL ANALYSIS

4.1 Model selection

n terms of the selection of variables, according to the regression equation obtained from the principal component regression analysis, the gross social agricultural product (set as P_1), the average annual disposable income per rural resident (set as P_2), the sown area of agricultural crops (set as P_3), the gross output value of agriculture, forestry, animal husbandry and fishery (set as P_4), the compensation expenditure of agricultural insurance (set as P_5) and the years of education per farmer (set as P_6) were selected as the source variables, and the model obtained was used to fit the actual situation in Liaoning Province. The model was fitted to the actual situation in Liaoning Province.

For the choice of FE or RE model, this paper uses the Hausman test, and the value of the panel data statistic obtained by the test is 10.915, with a significance p-value equal to 0.09103930256396667 greater than 0.05, and the list of results is shown in Table 5. The analysis shows that the results do not present significance and the original hypothesis cannot be rejected, i.e. the RE model should be selected.

Table 5 Hausman model test results

Type of test	Statistical quantities	p-value	Conclusion
Hausman test	10.915	0.091	RE Model

Note: ***, ** and * represent 1%, 5% and 10% significance levels respectively

4.2 Regression analysis

The F-test results of the RE model showed a significance p-value of 0.000***, which showed significance at the level of 0.000**, rejecting the original hypothesis, therefore the model is valid. The regression analysis using the RE (random effects) model on the panel data panel data of 14 cities in Liaoning Province from 2008-2020, the regression results are shown in Table 6 below:As can be seen from Table 6, the model fit is good with an excellent fit of approximately 0.806.

Table 6 RE model regression results

Variables	Coefficient	Standard error	t-statistic	p-value	R^2	F
Const	-137.455	19.824	-6.934	0.000***	within=0.806 between=0.907 overall = 0.839	F=133.416 p=0.000***
P_1	1.162	0.396	2.934	0.004***		
P_2	0.013	0.001	10.392	0.000***		
P_3	0.191	0.046	4.153	0.000***		
P_4	0.527	0.215	2.456	0.015**		
P_5	0.528	0.077	6.818	0.000***		
P_6	0.089	0.096	0.93	0.034**		
Dependent variable: Agricultural insurance premium income ($ million)						
Note: ***, **, * represent 1%, 5%, 10% level of significance respectively						

5 CONCLUSIONS

(1) The model shows a positive correlation between Liaoning's gross domestic product (GDP) and demand for agricultural insurance: for every 1 unit increase in agricultural GDP, premium income will increase by 1.162 percentage points. The overall level of economic development in Liaoning Province can be seen to be positive over a long period of time. Along with economic growth, economic relationships drive plus national policies fueling a gradual increase in insurance depth, and the rate of agricultural insurance development will steadily advance.

(2) The model shows that the average annual disposable income per rural resident is positively correlated with the demand for agricultural insurance: for every 1 unit increase in the average annual disposable income per urban resident, premium income will increase by 0.013 percentage points. Recognition of agricultural insurance in Liaoning Province is currently not as high as in economically developed regions. Its demand will rise along with the increase in spending power.

(3) The model shows a positive correlation between crop sown area and insurance demand: for every 1 unit increase in crop sown area in each region, premium income increases by 0.191 percentage points.

(4) The model shows that there is a positive relationship between the total output value of agriculture, forestry, animal husbandry and fishery in Liaoning Province and the demand for insurance: for every 1 unit increase in the total output value of agriculture, forestry, animal husbandry and fishery, the premium income will increase by 0.527 percentage points. The increase in the total output value of agriculture, forestry, animal husbandry and fishery

represents an increase in the scale of agriculture in Liaoning Province, and the higher the output value means the greater the risk of loss in the event of an accident, the greater the acceptance of agricultural insurance by farmers, and especially the greater the demand for agricultural insurance with higher coverage, which will significantly lead to an increase in the demand for agricultural insurance.

(5) The model shows a positive correlation between agricultural insurance payout and insurance demand in Liaoning Province: for every 1 unit increase in agricultural insurance payout, premium income will increase by 0.528 percentage points. The annual payout of agricultural insurance represents, in a sense, the level of insurance protection and is an important reference for the trust of the majority of policyholders. The increase in payout will result in a more effective transfer of risk in the minds of policyholders and a more "secure" interest for themselves, which will lead to a gradual increase in their recognition of and demand for agricultural insurance.

(6) For the overall social context of the model, the average number of years of education per rural resident in Liaoning Province in the model is 10.34 years (from the 7th Census in 2020). As rural residents become more educated, they will become more aware of the economic significance of agricultural insurance recognition, and under the same conditions will be more inclined to purchase agricultural insurance - incurring a small cost to make their economic interests more secure. The demand for agricultural insurance in general and accelerate the promotion of an overall prosperous agricultural insurance market.

Acknowledgements: The authors appreciate the financial support from the Provincial Undergraduate Innovation and Entrepreneurship Training Program Support Project (Grant No. 202210143045)

REFERENCES

[1]Han Xudong, Liu Shuang, Wang Ruonan, Zheng Fengtian. The impact of agricultural insurance on household operating income: an empirical analysis based on a national survey of three types of farmers [J]. Agricultural Modernization Research, 2020,41 (06): 946-956. DOI: 10.13872/j.1000-0275.2020.0075

[2]Jia Baohong, Wang Xiaorong, Xin Liyuan, Wang Lijuan, Liu Gengchun, Huo Wenjuan. Research on the influencing factors of environmental evaluation for agricultural independent innovation: A case study of Tianjin City [J]. China Agricultural Science Bulletin, 2010,26 (23): 433-436

[3]Lin Yizheng. Empirical Study on the Development Level and Influencing Factors of China's Agricultural Economy [J]. Science and Technology and Industry, 2015,15 (02): 57-61+143

[4]Liu Jinxia Research on Theoretical Methods and Applications of Agricultural Risk Management [D]. Tianjin University, 2004

[5]Xie Qian, Luo Jian. A meta-regression analysis of factors influencing demand for agricultural insurance[J].Economic Commentary,2019(02):113-124.DOI:10.19361/j.er.2019.02.09.

[6]Han Ke. A study on China's agricultural machinery insurance model based on the willingness of supply and demand[D].China Agricultural University,2014.

[7]Feng Hao, Zhu Meirong, Pan Jie, He Sixian. A study on the factors influencing the demand for agricultural insurance based on meta-regression analysis[J]. Journal of Economic Research,2022(04):39-42.

[8]Cai Guiquan, Tao Jianping. Agricultural insurance demand forecasting based on principal component analysis and multicore support vector machine[J]. Journal of Jinan University (Natural Science Edition),2021,35(02):138-143.DOI:10.13349/j.cnki.jdxbn.20201010.005.
[9]Ezdini Sihem. Economic and socio-cultural determinants of agricultural insurance demand across countries[J]. Journal of the Saudi Society of Agricultural Sciences,2019,18(2).
[10]Lv Caifei. Empirical Analysis on the Effect of Agricultural Insurance on Production——Based on panel data of 31 provinces and cities in China from 2008 to 2018[J]. E3S Web of Conferences,2020,214.