An Empirical Study on the Satisfaction with Big Data Based Rural Pension Insurance

Li Ma1, 2

1. Economics and Management School of Wuhan University, Wuhan 430072, China
2. Guang Dong University of Finance and Economics, Guangzhou 510320, China

Abstract

With the constant advancement of urban and rural integration and the continuous expansion of urban scale, the area of arable land in China is witnessing sharp decline, a situation extremely unfavorable to rural residents that rely on land resources for a living. To guarantee the quality of their life, and to solve the series of social problems thus follow, in 2009, China officially adopted a new type of pension insurance for rural residents. Expanding in coverage and constantly lifting its benefits, this new pension insurance is gradually replacing the traditional family pension mode. Pension insurance is one of the five major insurance types of social insurance, a basic security system that offers a stable and reliable source of life to the elderly that have lost their labor capacity. Based on this background, field survey and questionnaires are combined in the preparation of this paper to collect substantial information on people’s satisfaction with pension insurance in rural areas. Then, relevant influencing factors and survey data are studied with binary logistic regression analysis and SPSS22.0. Last but not least, recommendations are given on the empirical analysis, which is expected to serve as a reference in improving the new rural social pension insurance.

Keywords: Big Data, Rural Pension Insurance, Satisfaction, Empirical Research.

1. RESEARCH BACKGROUND

1.1 Literature review

Pension insurance, or basic social pension insurance, is a major concern of national audit for social security. It can guarantee the basic living of the elderly, and thus a long-term stability of the society (Qiu et al., 2014). Generally, the satisfaction of urban and rural residents with the basic pension insurance system is applied to measure how effect this system has been implemented. Indicators such as personal profile, pension situation, policy satisfaction, education, political outlook, etc. are probed into to see their impact on people’s satisfaction with the basic pension insurance (Hu and Peng, 2017) After revising the ACSI model, the research model of satisfaction is constructed based on the reality of the new rural pension insurance system. The empirical analysis shows that there is big room for the new rural pension insurance to improve, and both the government support and the institutional service ability have positive influence on satisfaction (Hu et al., 2014). Meanwhile, to watch on the pension situation of the disadvantaged groups, it is necessary to summarize problems that demand prompt solution and put forward recommendations for further improvement of the new rural pension insurance system. For example, improve the pension benefits, set up a dynamic pension insurance adjustment mechanism, improve the government’s service capacity in this area, introduce incentive policies and publicity system, and further enhance China’s new rural pension insurance system (Zhou et al., 2013). The research reveals that many scholars have studied the satisfaction with the new rural pension insurance system, which proves great reference in the preparation of this paper. However, the existing research results are mainly from some pilot areas or regions, and therefore certain deviation exists. Moreover, as the system penetrates deeper in provinces and cities, more problems and countermeasures thus arising require more in-depth and comprehensive studies.

1.2 Purpose of research

After nearly 3-year practice and implementation, China’s new rural social pension insurance system has realized full coverage, and basically formed its institutional framework and operational mechanism. The achievement is
remarkable. As of the beginning of 2013, the new pension insurance had covered 460 million people and widely recognized by rural residents. Nevertheless, it has to be further improved and developed. To this end, problems encountered in the promotion of the new rural pension insurance system is probed into with results and data from field survey, and it is of great significance to enhancing the satisfaction with rural pension insurance (Zhang and Zhang, 2013). Research on satisfaction is generally done by constructing multiple linear regression mode, then influencing factors are analyzed, and countermeasure and recommendations proposed accordingly (Liu and Cao, 2010). We can also classify the influencing factors to the characteristics, and analyze the satisfaction and its influencing factors according to Likert Scale (Liao and Tao, 2014). As the most important initiative for people’s livelihood, the new rural insurance system has a direct impact on the living standards of rural residents. Therefore, it is of practical significance to understand and clarify people’s satisfaction with the new rural medical insurance system, and it is also essential to the long-term and orderly operation of the new rural cooperative medical system (Jiang and Ma, 2011). Given on this, the paper probes into the satisfaction with the rural pension insurance system for its further improvement.

2. DATA AND SELECTION OF VARIABLES

2.1 Data sources

The data of this paper are mainly from rural household survey questionnaire, and villages and households remain unchanged once selected to relevant codes and sort. The questionnaire involves the following aspects: (1) personal average annual income; (2) source of money to pay pension insurance; (3) reasons for participating in pension insurance; (4) main sources of living for the elderly at home. In this paper, the samples (1,333 farmers, but the data of some samples need improving) are selected from 113 administrative villages of 5 counties in Liaoning Province, as is shown in Table 1.

<table>
<thead>
<tr>
<th>County name</th>
<th>Number of administrative villages</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fafu County</td>
<td>26</td>
<td>215</td>
</tr>
<tr>
<td>Lvtu County</td>
<td>26</td>
<td>226</td>
</tr>
<tr>
<td>Tangping County</td>
<td>22</td>
<td>224</td>
</tr>
<tr>
<td>Xifeng County</td>
<td>20</td>
<td>235</td>
</tr>
<tr>
<td>Qingyuan County</td>
<td>19</td>
<td>216</td>
</tr>
</tbody>
</table>

2.2 Data processing principles

2.2.1 Missing value

There are missing values in the original data, and the possible reasons are as follows: One, questions are missed, for example, the respondents do not answer the question, so there is no related data; two, the data value is zero and the reason for that is raw data is not recorded in a standardized manner. The paper has handled missing data in the original data as appropriate. If other data from the survey can provide supplementary support, and the variable can be treated as a continuous variable, calculation can be done or the mean of the variable used as a substituted (Liu et al., 2012). For instance, some farmers do not provide their agricultural income in 2016, then it is replaced by the difference of the total income in 2016 minus non-agricultural income. Missing values, without any corresponding data to provide the support, will be taken as zero. If no data is received from the farmer, then there is no sample observation value from him/her. The main purpose of this data processing is to ensure sufficient sample data, though there may be errors that are hard to measure.

2.2.2 Outlier

Before data analysis, the original data is investigated according to the principle of experience. For example, a farmer from Anmin Village, Anmin Town, Xifeng County gets about RMB 550 million if he sells his houses, livestock, land, agricultural production materials and all other property. It is not possible in reality and the reason for such mistake is that when filling in the questionnaire, the respondent fails to correctly understand the relevant values. To ensure adequate survey sample data, it is replaced by RMB 56,000. Relevant data is also adjusted, as appropriate, when found unreasonable in the analysis.
2.3 Selection of variables

2.3.1 Explained variable

Take farmers’ satisfaction with rural social pension insurance as the explained variables. Satisfaction refers to how the farmers like the social pension insurance, including satisfaction and dissatisfaction. The explained variable is replaced by the D5 variable in the questionnaire. This paper mainly uses binary logistic regression analysis, where satisfaction is set to be 1, and dissatisfaction, 0.

2.3.2 Explanatory variable

This paper chooses five explanatory variables and mainly studies the satisfaction of rural residents with rural pension insurance. The five variables are age (B2_1), household income, population aged above 60 (B2_3), household income in 2016 (B8_2), household expenditure on daily consumption (B9_2_1), satisfaction with rural social cooperative medical insurance (C5).

3. HYPOTHESIS

3.1 Age (B2_1)

If there is a positive correlation between the age of the selected variable and the satisfaction of the rural pension insurance, then people’s satisfaction with rural pension insurance increase with their age. Compared with the young, the elderly concern themselves more with rural pension insurance, which is quite consistent with the reality.

3.2 Rural household agricultural income in 2016 (B8_2)

Assume that there is a positive relationship between agricultural income and satisfaction. As the new rural insurance attaches great importance to insurance scope and coverage, their insurance benefits are low. Higher household income indicates that its main source of income is agricultural product. Families with relatively difficult economic conditions are still quite receptive to existing rural pension insurance and show higher satisfaction (Zhang, 2012), while families with better economic conditions are less satisfied with the existing low level insurance and therefore choose commercial insurance or other to improve they are content with.

3.3 Population above 60 (B2_3)

Assume there is a positive correlation between the number of elderly population and the satisfaction. Rural social pension insurance provides that people above 60 can receive pension funds. Moreover, the hypothesis variables of the population and the national policy conditions are closely connected and will directly affect rural residents’ willingness to buy insurance and satisfaction with it.

3.4 Satisfaction with rural social cooperative medical insurance (C5)

Assume that farmers are very satisfied with the new rural cooperative medical insurance and the new rural insurance. As the pilot work of Satisfaction with Rural Social Cooperative Medical Insurance started earlier than the introduction of the new rural insurance policy, the former will boost farmers’ willingness to buy the latter.

3.5 Household consumption in 2016 (B9_2_1)

Assume that there is a positive relationship between household daily consumption expenditure and satisfaction, that is, the higher farmers’ consumption expenditure is, the higher their satisfaction with rural pension insurance will be. Of the average consumption of n rural households, food makes up the largest part. Therefore, the higher the Engel coefficient and the more the daily consumption of the farmer is, the less the household savings will be. Such households are relatively impoverished and show higher satisfaction with rural pension insurance (See Table 2).
Table 2 Explanatory Variables

<table>
<thead>
<tr>
<th>Variable type</th>
<th>Variable type</th>
<th>Calculation formula</th>
<th>Expected symbols</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explained variable</td>
<td>Satisfaction</td>
<td>Dissatisfied is equal to 0, satisfied is equal to 1</td>
<td>/</td>
</tr>
<tr>
<td>Age (B2_1)</td>
<td></td>
<td>/</td>
<td>+</td>
</tr>
<tr>
<td>Rural agricultural income (B8_2)</td>
<td></td>
<td>/</td>
<td>+</td>
</tr>
<tr>
<td>Population of the elderly (B2-3)</td>
<td></td>
<td>/</td>
<td>+</td>
</tr>
<tr>
<td>Household expenditure (B-9-2-1)</td>
<td></td>
<td>/</td>
<td>+</td>
</tr>
<tr>
<td>Debt or not (B18-1)</td>
<td>No debt is equal to 0, if there is debt, 1</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Satisfaction with medical insurance (C5)</td>
<td></td>
<td>Dissatisfied is equal to 0, satisfied is equal to 1</td>
<td>+</td>
</tr>
</tbody>
</table>

4. MODEL BUILDING

SPSS22.0 method is used to analyze the data and the satisfaction with pension insurance. Moreover, maximum likelihood and estimating methods are required to solve parameters of the logistic regression equation. General multiple regression model is then applied to build probability model where the explained variable satisfaction value is selected to be 1, and the explained variable, between 0 and 1. That is \( F_y = \beta_0 + \beta_1 X_1 \). Since the selection P value is difficult to meet the requirement of \((-\infty, +\infty)\), it needs a 2-step conversion.

First, turn P into \( \Omega \), the relation of P and \( \Omega \) is \( \Omega = p_1 - p \).

Then, turn \( \Omega \) into \( \ln \Omega = \ln \frac{p}{1-p} \), where \( \ln \Omega \) is called LogitP. There is a consistency between LogitP and \( \Omega \), and the range of values is \((-\infty, +\infty)\), as is the case with that of dependent variables in general regression models. Then, use the dependent model between explained variable and explanatory variable, \( \logit P = \beta_0 + \beta_1 X_1 \), and further deduce:

\[
P = \frac{1}{1 + \exp[-\beta_0 + \beta_1 X_1]}
\]

The above formula is a Logistic function that effectively reflects the non-linear relationship between P and explanatory variables. According to regression model coefficient, when the explanatory variables remain unchanged, the increase in explanatory variable \( X_i \) will lead to increase (or reduce) of \( \logit P \) by \( \beta_i \). Therefore, when \( X_i \) increases, P also increases (or decreases), but the increase or decrease is non-linear, depending on the interaction between the selected value range of the explanatory variable and the explanatory variable.

5. EMPIRICAL RESULTS ANALYSIS

In the regression equation, collinearity between the explanatory variables indicates that there is information overlap between the explanatory variables, which may seriously affect the regression structure. At the same time, Tolerance and the variance expansion factor VIF are usually adopted to examine collinearity in explanatory variables. It is represented as \( TOL = 1 - R_i^2 \), where \( R_i^2 \) is discriminant coefficient, or the so-called goodness of fit, got in the regression of any variable of \( X_i \) (i=1, ...,13) as the explanatory variable and the other five variables. The value of tolerance is between 0 and 1, and the larger the value is, the weaker the collinearity between the variables will be. When the TOL is lower than 0.2, there are multiple collinearity problems between the explanatory variables. The variance expansion factor is the reciprocal of tolerance, and the range of the values is \( 1^{-\infty} \), and the smaller the value is, the smaller the possibility of collinearity between the explanatory variables will be. As is shown in the following Table 3, the minimum value of tolerance is 0.764 and above 0.2, the maximum value of variance expansion factor is 1.309, indicating weak collinearity between explanatory variables, and it’s reasonable for explanatory variables to enter an equation.
Table 3 Collinearity

<table>
<thead>
<tr>
<th>Model</th>
<th>Collinearity Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tolerance</td>
</tr>
<tr>
<td>Age</td>
<td>.899</td>
</tr>
<tr>
<td>Population of the elderly</td>
<td>.963</td>
</tr>
<tr>
<td>Household expenditure</td>
<td>.870</td>
</tr>
<tr>
<td>Household agriculture income</td>
<td>.896</td>
</tr>
<tr>
<td>Whether included as impoverished by the government</td>
<td>.763</td>
</tr>
<tr>
<td>Whether enjoy the government’s relief subsidies</td>
<td>.899</td>
</tr>
</tbody>
</table>

6. CONCLUSION AND RECOMMENDATION

Based on big data analysis technology, China’s rural pension insurance should give full consideration to the interests of rural residents to improve their satisfaction with it. All the pension insurance related work in China should be done from the perspective of rural residents. Use big data technology to ensure that the rural social pension insurance system is well enacted, and constantly meet the needs of rural residents. Using logistic empirical model, the study finds that there is high satisfaction with China’s rural pension insurance system. The elderly population has a positive effect on the rural pension insurance satisfaction, so do family agricultural income and family expenditure. The impact of other variables is less evident. To improve the satisfaction with rural pension insurance, recommendation are given as follows: First, against the backdrop of big data technology, constantly improve the rural family pension insurance mechanism (with traditional family pension mode as an important basis) and continuously improve the quality of life of rural residents, so that the elderly will be taken good care of; second, continue to narrow the gap between urban and rural areas, improve people’s skills so that they can better adapt to the work and life in cities, and the integration of pension insurance; third, change the concept with pension insurance. Only by constantly improve the new rural pension insurance system can we substantially improve rural residents’ satisfaction with it and in turn contribute to China’s endeavor to become a moderately prosperous society.

REFERENCES


