

Analysis of the Impact of Different Family Structures on the Number of Commercial Insurance Purchases

Lingcheng Kong*

College of Finance and Statistics, Hunan University, Changsha, China

**Corresponding author: Lingcheng Kong.*

Abstract

In recent years, the global insurance industry has developed rapidly, and families, as an important part of insurance purchasing, play a crucial role in its development. Studying the impact of family structure on the number of commercial insurance policies purchased allows the insurance industry to make reasonable predictions about the insurance configuration of target families. Currently, scholars have conducted extensive research in this field, but there are still gaps in the research. Some studies have selected small samples or fewer dimensions. This paper aims to study the demographic structure of large-sample families to fill these gaps. Based on the framework of regression analysis, this study conducts Heckman regression analysis on family samples from a database to obtain the parameter relationships between different independent and dependent variables. These relationships indicate that the probability of families purchasing commercial insurance significantly decreases with an increase in the proportion of minors and the elderly population, and increases with an increase in family size and male ratio, providing a reference for the insurance industry in designing and pricing family insurance

Keywords

family structure, family insurance, commercial insurance

1. Introduction

Currently, China's economic and demographic structures are undergoing accelerated transformation. The aging family population structure, coupled with the continuous expansion of economic risks, has put pressure on social security pension funds. For example, against the backdrop of China's "three-child policy" and "active aging policy", the demand for caregiving services for infants and disabled elderly has surged, elevating family risks to societal risks [1]. Commercial insurance plays a crucial role in different families. However, there are significant differences in the understanding and demand for commercial insurance among families with different economic situations and structures in China and even globally. Studying the commercial insurance configurations of different families is beneficial for assessing the overall resilience and ability to withstand risks of families, and facilitates the implementation of different policies by the government. For families with poor risk resistance, inclusive finance can be implemented to improve their ability to withstand risks. Additionally, understanding the commercial insurance configurations of different families enables insurance

companies to more accurately develop and market insurance products, facilitate actuarial analysis of insurance products, and implement precise pricing. Therefore, studying the impact of family structure on family life insurance is crucial. This paper primarily explores whether and how family size, proportion of elderly population, proportion of minor population, and proportion of males affect a family's insurance configuration, whether positively or negatively. Based on existing literature and practical analysis, it can be hypothesized that an increase in the number of elderly and youth, as well as the number of families with commercial insurance such as commercial health insurance, will occur [2]. A smaller family size can increase demand for commercial insurance [3]. As the backbone of Chinese families, males tend to enhance family risk management awareness and increase demand for commercial insurance.

To identify the magnitude of the impact of different factors in family structure, this paper will construct a multiple regression model and utilize Stata software for empirical analysis. Finally, based on the analysis results, suggestions for optimizing family insurance will be proposed.

Currently, more scholars are focusing on the research of household insurance allocation, mainly believing that different family structures have different impacts on household insurance allocation, but the scope and content of the research as well as the selected samples vary. Scholars Li, Tan, and He empirically explored the impact of changes in rural household population structure on insurance allocation through two dimensions: population age structure and population gender structure, using Ologit and Tobit models based on field survey data from 794 rural households in Henan, Hubei, and Hunan provinces in China. They concluded that an increase in the proportion of the elderly population has a significant negative impact on the basic insurance needs and investment of households [4]. Yin, Yan, and Zhang focused on the impact of families with children on their household insurance allocation due to child-rearing responsibilities. They believed that under the background of family planning, families would be more active in utilizing external resources from the commercial insurance market to better cope with child safety risks and provide better development opportunities for their children. The power of the family ultimately manifests as changes in household commercial insurance demand [5]. Zhang analyzed a large number of factors that affect household insurance demand, including not only family structure but also conducted detailed empirical analysis on economic factors using Eviwes [6]. However, the current research conducted by various scholars has its limitations. Some are limited to rural family structure, ignoring the vast number of urban families; some mainly focus on families with children, neglecting other families with richer structures; and some provide a relatively cursory analysis of the impact of family structure on household insurance allocation.

Currently, scholars generally agree that families that are married, entrepreneurial, have high incomes, and are covered by social insurance may have higher demand for commercial insurance. Conversely, families with more elderly members, risk-averse attitudes, and located in rural and remote areas may have lower demand for commercial insurance [7]. Financial literacy and knowledge significantly promote the probability and degree of participation in household commercial insurance, protection-type insurance, and investment-type insurance [8]. Furthermore, with the development of digital finance, the participation rate and premium expenditure level of household insurance have also increased [9]. The marginal contribution of this paper lies in specifically examining the impact of family structure- whether in rural or urban areas- on household insurance allocation. It enriches the content related to household commercial insurance and family structure.

2. Empirical Analysis

2.1 Data Sources

This paper utilizes data from the China Household Finance Survey (CHFS) for the year 2021[10], which was established by the China Household Finance Survey and Research Center at Southwestern University of Finance and Economics. The database primarily includes information on personal basic details, employment, financial assets, non-financial assets (agricultural, industrial and commercial, housing, land, household durable goods, etc.), household liabilities, household income and expenditure, social security, and insurance. This paper utilizes a sample of 22,027 households observed in 2021

2.2 Variable Definition

The table 1 shows the explanatory variables, the dependent variables, and the control variables used in this paper.

Table 1: Variable Definition

Type of Variable	Symbol	Construction
Explained variable	total_premium	f1008a (Annual payment amount for social pension insurance), f1029(The monthly payment amount for occupational annuity or supplementary pension insurance last year), f2004 (Last year's medical insurance payment for family members), f6110a (Last year's commercial life insurance premium payment), f6203_1 (Last year's commercial life insurance premiums), f6502 (Last year's other commercial insurance premiums), f6601 (Last year's annuity insurance contribution amount). The logarithm of the annual commercial insurance premium expenditure for the variable family each year. The calculation method is as follows: $\ln Y = \ln(f1008a + f1029 * 12 + f2004 + f6110a + f6203_1 + f6502 + f6601)$
Explanatory Variable	under18 ratio male ratio household_size male ratio	the proportion of family members under 18 years old to the total number of family member the proportion of family members over 60 years old to the total number of family members the number of family members the proportion of males in the family
Control Variable	Area	The economic levels and the development of the insurance industry vary significantly across different regions. Therefore, it is necessary to include the region as a control variable in the analysis

2.3 Model Setting

The model is set as follows:

$$\ln Y = \ln \beta_0 + \ln \beta_1 X_1 + \ln \beta_2 X_2 + \ln \beta_3 X_3 + \ln \beta_4 X_4 + \varepsilon \quad (1)$$

Where X_1 represents the proportion of family members under 18 years old to the total number of family members, X_2 represents the proportion of family members over 60 years old to the total number of family members, X_3 represents the proportion of males in the family, X_4 represents the number of family members, and Y represents the annual insurance premium expenditure of the family.

2.4 Descriptive statistical analysis

Table 2: Descriptive statistics

variable name	average	median	standard deviation	range
Y	788.802	0	6938.31	[0,480000]
X1	0.113221	0	0.17216	[0,1]
X2	0.381286	0.25	0.41669	[0,1]
X3	0.500324	0.5	0.206674	[0,1]
X4	3.10151	3	1.55943	[1,16]

As shown in Table 2, among the 22,027 sampled household samples, there are a large number of 0 values for Y , which means that most households had no insurance premiums in the past year. This can lead to selection bias. Households that purchase insurance may differ from those who do not in terms of risk preference, income, and health attitudes. These differences affect both the decision to purchase and the amount/premium of insurance. However, ordinary regression only considers those who have purchased, omitting the role of unobservable factors. Therefore, simple multiple linear regression analysis is no longer applicable, and the Heckman model should be chosen.

2.5 Selection Equation (Full Sample)

Among the total of 22,027 valid household samples, 16,158 households had no insurance premium expenditure in the past year, while 5,869 households had insurance premium expenditure in the past year. Calculations show that the proportion of households without purchasing commercial insurance is approximately 73.39%, while the proportion of households that have purchased commercial insurance is approximately 26.61%. This indicates that during the sample period, most households did not purchase commercial insurance products. When estimating the Probit model using Stata, maximum likelihood estimation (MLE) iteration was employed. After 4 iterations, the change in the log-likelihood value was extremely small (-11289.835→-11289.835, actual change < 0.001), and the iteration was stopped. The entire convergence process was completed with only 4 iterations, indicating that the model setting is reasonable and the numerical calculation is stable. The final log-pseudo-likelihood value is -11289.835, which will be used for subsequent model goodness-of-fit comparisons and the calculation of the inverse Mills ratio.

Table 3: Probit regression

Number of obs	22,027
Wald chi2(4)	2689.87
Prob > chi2	0.0000
Log pseudolikelihood	-11289.835
Pseudo R2	0.1158

The overall fit statistics of the Probit regression model show that the model is estimated based on 22,027 household samples, with a log-pseudo-likelihood value of -11289.835. The Wald joint significance test has a chi-square statistic of 2689.87 (degrees of freedom 4), with a corresponding p-value less than 0.0001, indicating that all independent variables (household size, proportion of minors, proportion of elderly, proportion of males) in the model jointly have a highly significant predictive ability for whether households purchase commercial insurance. The pseudo-R² is 0.1158, indicating that the model can explain about 11.58% of the variation and has a certain explanatory power.

2.6 Regression Results

Table 4: Regression results

total_premium	Coefficient	std. err	z	P> z	[95% conf. interval]
household_size	0.0257503	0.0075082	3.43	0.001	0.0110345 0.0404662
under18_ratio	-0.9006432	0.0670633	-13.43	0.000	-1.032085 -0.7692015
over60_ratio	-1.423969	0.0283646	-50.20	0.000	-1.479563 -1.368376
male_ratio	0.2463198	0.0480156	5.13	0.000	0.152211 0.3404287
_cons	-0.289365	0.036189	-8.00	0.000	-0.3602941 -0.2184358

According to the regression results, the coefficient of household size (household_size) is positive and significant (0.0258, $p < 0.01$), indicating that the higher the household size, the higher the probability of purchasing commercial insurance. This may be due to the higher risk exposure associated with a larger population, leading to a corresponding increase in insurance demand. The coefficient of the proportion of population under 18 years old (under18_ratio) is negative and significant (-0.9006), suggesting that a higher proportion of minors significantly decreases the probability of households purchasing insurance. This may be because minors have relatively lower health risks and are less likely to purchase general medical insurance. Although minors are more likely to purchase accident insurance, such insurance often relies more on basic coverage such as social security or school health insurance, with lower premiums. Additionally, due to the limited economic conditions of minors, insurance agents are less likely to promote high-priced products to such households. The coefficient of the proportion of population over 60 years old (over60_ratio) is also negative, and its absolute value is greater than that of under18_ratio, indicating that the elderly population has a stronger negative impact on the purchasing probability. The main reason is that elderly premiums are expensive, and most elderly individuals are unemployed or rely on pensions, so they usually do not spend a large amount on insurance and rely more on retirement medical insurance (employee/resident medical insurance). Moreover, elderly individuals may already have illnesses or be too old to purchase insurance. In addition, the stronger negative impact of over60_ratio compared to under18_ratio also suggests that decision makers (the younger generation) may prioritize purchasing insurance for themselves or their children. The coefficient of the proportion of males (male_ratio) is positive and significant, indicating that households with

a higher proportion of males are more likely to purchase commercial insurance. This is usually because males are often the main source of income for the household and have a greater need for risk management, requiring life insurance or accident insurance to cover income risks.

Continue the analysis after adding the control variable (region):

Table 4: Adding the control variable (region)

variable name	Coef.	(Std.Err.	z	P> z	Lower limit of 95% confidence interval	Upper limit of 95% confidence interval
household size	-687.695	174.8749	-3.93	0.000	-1030.528	-344.8618
under18_ratio	5617.361	7824.618	0.72	0.473	-9722.397	20957.12
over60_ratio	3746.89	12168.66	0.31	0.758	-20109.14	27602.92
male_ratio	1627.633	2788.674	0.58	0.559	-3839.417	7094.683

The results indicate that the dummy variable $P>|t|$ for the vast majority of provinces falls between 0.05 ~and 0.2, with only a few (such as Hainan and Guangxi) approaching significance. A positive coefficient indicates that the premium level of the province is higher than that of the reference province; a negative coefficient indicates that it is lower than the reference province (the reference province is Beijing). There are certain differences in premiums among regions, but most of these differences are not statistically significant (due to limited sample size or large standard errors). Adding regions increases R^2 to 0.04, indicating that regional differences can explain approximately 3.3 percentage points of variation.

3. Conclusion

The probability of households purchasing commercial insurance significantly decreases as the proportion of minors and elderly increases, and increases with the increase in household size and male ratio. The model is overall significant but has limited explanatory power (pseudo $R^2 \approx 0.12$).

The family structure of different households has an impact on their purchase of commercial insurance. However, due to the small Pseudo R^2 , it indicates insufficient explanatory power. The impact of different family structures on the purchase of commercial insurance is relatively small. Among the four explanatory variables, the impact on the purchase volume of commercial insurance is ranked from highest to lowest as follows: the proportion of family members aged 60 and above in the total number of family members, the proportion of family members under 18 years old in the total number of family members, the proportion of males in the family members, and the number of family members (family size).

This research model possesses the following advantages: Firstly, the p-values for each variable in the model are greater than the absolute value of z, indicating that all variables are significant and the coefficients of the variables are reasonable and in line with expectations. Secondly, the model has a large sample size, stronger statistical power, and covers various types of households, providing broad coverage. However, this research model also has the following shortcomings: The primary drawback is limited explanatory power, with Pseudo $R^2 = 0.1158$ in a large sample, and approximately 0.04 after adding control variables. Limited explanatory power suggests that many important variables have not been included (such as income, education, type of medical insurance, region, etc.), or indicates that the impact of household structure on the purchase volume of commercial insurance is weaker, not stronger than the impact of other variables on the purchase volume of commercial insurance. Furthermore, there is multicollinearity among the explanatory variables: both under18_ratio and over60_ratio are variables related to age, exhibiting approximate collinearity. Future research could incorporate more control variables (such as income, education, type of medical insurance, etc.) and add squared terms for under18_ratio and over60_ratio or interaction terms between elderly×children to reduce collinearity, while also examining whether there is a threshold effect (possibly with negative effects intensifying when the proportion of elderly exceeds 0.5).

References

- [1] Wu, W. F., & Wang, R. H. (2023). The construction of family as a principal part in social insurance law: Family care in an aging society challenges and responses. *Social Security Review*, 7(5), 36–53.
- [2] Lu, Y. J., & Zhang, W. H. (2020). Research on the impact of family structure on family participation in the insurance market. *Modern Economic Exploration*, (5), 25–35.
- [3] Zhai, G. Y., & Xue, Y. K. (2026). The downsizing of family size and commercial insurance demand: Impact effect, transmission mechanism, and evolution trends in change. *Financial Research*, (3), 114–132.
- [4] Li, T., Tan, S. F., & He, J. (2025). The impact of rural population structure on insurance allocation and policy recommendations. *Chinese Agricultural Resources and Regional Planning*, 46(7), 243–253.
- [5] Yin, Z. C., Zhang, Z. X., & Yan, Y. (2022). The impact of child-raising on household insurance demand. *Financial Forum*, 27(4), 60–69.
- [6] Zhang, C. (2020). Analysis of the influencing factors of household insurance demand. *China Industry & Economy*, (14), 39–40.
- [7] Sun, W. J., & Gao, Y. (2018). Financial knowledge, liquidity constraints, and household commercial insurance demand. *Journal of Finance*, 12(2), 53–75.
- [8] Wang, Y. K., & Liu, L. J. (2025). Research on the impact of financial literacy on household insurance decision-making. *Journal of Central China Normal University (Humanities and Social Sciences Edition)*, 64(2), 70–82.
- [9] Zhang, X., Li, Y., & Jing, H. (2026). Impact of digital finance on household insurance demand. *Finance Research Letters*, 104, 110165.
- [10] Gan, L., Yin, Z. C., Jia, N., Xu, S., & Ma, S. (2013). The status of household assets and housing demand in China: Analysis. *Journal of Financial Research*, (4), 1–14.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Open Access

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

