Analysis of household demographic structure on risky financial asset allocation based on probit and tobit models

Lingling Zeng[a], Xiaomin Wang[*], Ruili Zhang[b]

[a]2955926860@qq.com, [*]xiaominwang1022@163.com, [b]2325792165@qq.com

School of Economics, Wuhan University of Technology, Wuhan, China

Abstract: This study employs Probit and Tobit models to investigate the impact of household demographic structure on the allocation of risky financial assets. Given the evolving Chinese societal landscape, demographic shifts hold profound implications for financial asset allocation, particularly in the context of gradually relaxed new reproductive policies. Utilizing data from the "China Financial Survey" in 2019, this paper conducts empirical research from two perspectives: the propensity of households to engage in risky financial assets and the proportion of investment in such assets. The Probit model uncovers relationships between factors such as family size, elderly dependency ratio, child dependency ratio, and gender of family children with the participation rate in risk assets. On the other hand, the Tobit model reveals connections between household demographic structure and the proportion of investment in risky financial assets. The findings contribute to offering more precise and effective recommendations for household asset allocation, introducing novel perspectives and methods for family financial planning and risk management.

Keywords: Probit model, Tobit model, household demographic structure, risky financial asset allocation

1 Introduction

As China's society evolves, a range of issues has surfaced, including the "low fertility rate trap," exacerbated aging challenges, and imbalanced demographic structures. Against this backdrop, understanding the influence of household demographic structure on the allocation of risky financial assets becomes crucial. This is especially significant given the progressively relaxed new reproductive policies, as changes in household demographic structure will have far-reaching impacts on financial asset allocation. The implementation of the third-child policy has triggered shifts in social structure, the interplay between low fertility rates and aging, and changes in population structure, all of which could alter household consumption patterns and savings behavior, thereby affecting decisions regarding the allocation of risky financial assets.[1]

In this context, the present study utilizes data from the "China Financial Survey" in 2019 as its foundation, employing Probit and Tobit models to empirically explore the impact of household demographic structure on the allocation of risky financial assets. The study aims to delve into the mechanisms by which household demographic structure influences participation rates and holding proportions of risky financial assets, offering more specific and effective

ECIT 2024, March 15-17, Changsha, People's Republic of China

DOI 10.4108/eai.15-3-2024.2346185

recommendations for household asset allocation. This paper will elaborate on several aspects: introducing the fundamental principles of the Probit and Tobit models, detailing the design and analytical process of the empirical study, and finally summarizing research outcomes and presenting conclusions. Through this research, we anticipate providing fresh perspectives and methodologies for family financial planning and risk management.[2][3]

2 Model fundamentals elaboration

2.1 Probit model

The Probit model is a binary regression model, implying that its dependent variable can only take two possible values, such as 0 and 1. For instance, if we intend to examine whether an individual participates in a certain activity, we can use 0 to denote non-participation and 1 to denote participation. The underlying premise of the Probit model assumes the existence of a latent continuous variable y*, which is a linear function of the independent variable X added to a random error term u, as depicted in Equation (1).

$$y^{*} = X\beta + u \tag{1}$$

Here, y* represents an individual's latent propensity or willingness to engage in the activity, albeit unobservable. Instead, we observe a binary variable y, determined based on the sign of y *: if $y^*>0$, then $y=1$; if $y^*\leq 0$, then $y=0$. Thus, the challenge of a regression problem involving a continuous variable is transformed into a binary variable regression problem. The crux of the Probit model lies in the assumption that the error term u follows a standard normal distribution, i.e., $u \sim N(0,1)$. Consequently, we exploit the properties of the normal distribution to derive the conditional probability distribution of the dependent variable y, as illustrated in Equation (2).

$$\underline{P}(y = 1 \mid X) = \Phi(X\beta) \tag{2}$$

Where Φ represents the cumulative distribution function of the normal distribution. Subsequently, we resort to maximum likelihood estimation (MLE) to estimate the parameter β.

The computational and operational process of the Probit model broadly follows these steps:

Firstly, we gather or acquire a dataset comprising binary dependent variables and independent variables, such as whether an individual participates in an activity, purchases a product, or contracts a certain ailment.

Secondly, we engage in preprocessing and descriptive analysis of the data, including checking for completeness, handling missing values, outliers, correlations, alongside visualization and summarization.

Next, we construct the Probit model and employ the MLE method to estimate the parameter β. This step can be implemented using statistical software or programming languages such as Stata, R, Python, etc.

Subsequently, we subject the model to diagnostics and tests, encompassing fit assessment, significance, residual analysis, and inference on parameter β.

Finally, we evaluate and apply the model, encompassing comparisons of model performance, predictions for new data, analysis of influencing factors and policy effects, among other aspects.

The specific model computation process is illustrated in Figure 1.

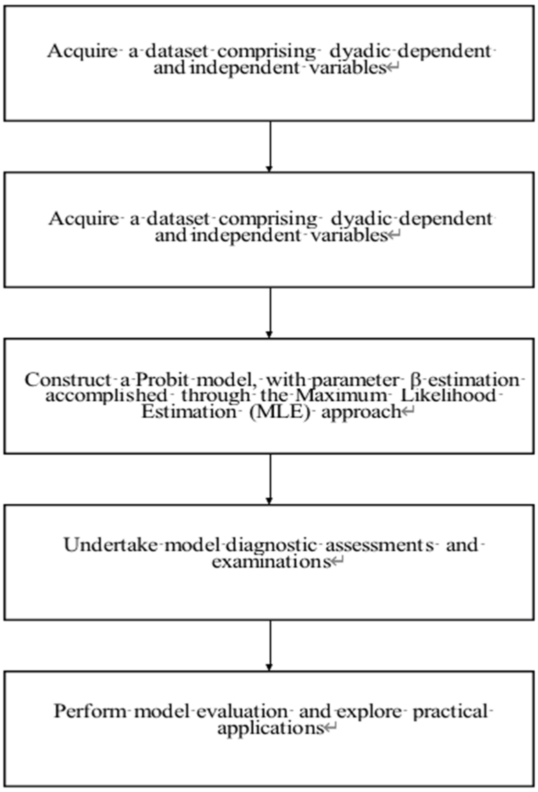

Fig. 1. Probit Model Computation Process

2.2 Tobit model

The Tobit model is a regression model employed to analyze truncated or censored dependent variables, implying that the dependent variable has either a lower or upper limit, beyond which values are truncated or become missing. The fundamental idea of the Tobit model is that a latent continuous variable y* exists, which is a linear function of the independent variable X combined with a random error term u that follows a normal distribution, as illustrated in Equation (3).

$$y^{*} = X\beta + u \tag{3}$$

Here, y* signifies the genuine level or quantity of an individual or event's outcome, yet remains unobservable. We observe a truncated or censored variable y, which is determined based on y* and a threshold c: if y*>c, then y=c or is missing; if y*≤c, then y=y*. This transformation thus converts a regression problem with continuous variables into one involving truncated or censored variables.

The crux of the Tobit model hinges on employing the properties of the normal distribution to derive the conditional probability distribution and conditional expected value of the dependent variable y, while accounting for selection bias introduced by truncation or censoring. Subsequently, we employ maximum likelihood estimation to estimate the parameters β and σ^2.

The computational and operational process of the Tobit model is broadly as follows:

Initially, we need to gather or acquire a dataset containing truncated or censored dependent variables and independent variables, such as an individual's monthly expenditure on a certain item, an enterprise's annual investment in a project, a country's annual research and development expenditure in a particular field, etc.

Subsequently, we conduct preprocessing and descriptive analysis of the data, including checks for completeness, missing values, outliers, correlations, and visualizing and summarizing the data.

Next, we construct the Tobit model and employ the MLE method to estimate the parameters β and σ^2. This step can be carried out using statistical software or programming languages such as Stata, R, Python, etc.

We then subject the model to diagnostics and tests, including assessing fit, significance, residual analysis, and interpretation and inference of parameters β and σ^2.

Finally, we evaluate and apply the model, comparing model superiority, predicting outcomes for new data, analyzing influencing factors and policy effects, among other tasks.

The specific model computation process is illustrated in Figure 2.

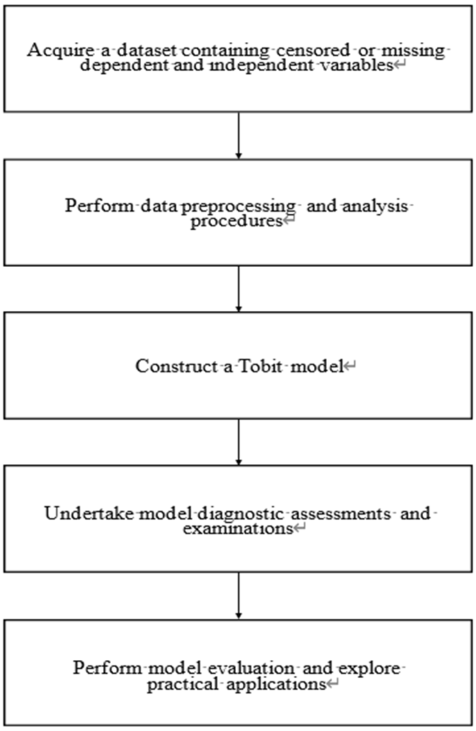

Fig. 2. Boxplots of different indicators

Probit model errors follow a standard normal distribution, while Tobit model errors follow a general normal distribution. Probit model observations are determined by the sign of latent variables, while Tobit model observations are determined by the magnitude of latent variables and a threshold. Probit models solely require coefficient parameter estimation, while Tobit models additionally require variance parameter estimation. Probit models are immune to selection bias, whereas Tobit models require correction for selection bias.[4]

Simultaneously, Probit and Tobit models are both latent variable regression models, assuming the existence of an unobservable continuous variable y∗, which is a linear function of the independent variable X combined with a random error term u. Both models utilize properties of the normal distribution to derive the conditional probability distribution of the dependent variable y and employ the maximum likelihood estimation method for parameter estimation. Ultimately, both models can be applied to the regression problems of finite-value dependent variables, such as participation behavior, expenditure amounts, etc.

3 Empirical study design and analysis

3.1 Empirical study design and analysis

Building upon existing literature, this study examines the impact of household demographic structure on the allocation of risky financial assets from two dimensions: first, considering the propensity of households to engage in risky financial assets, and second, contemplating the proportion of investment in such assets. Initially, regarding the inclination of households to participate in risky financial assets, where the explanatory variable can only take on "yes" or "no" values, conforming to a typical binary choice model, the Probit model is employed. Subsequently, for the investment proportion in risky financial assets, a substantial portion of the samples are compressed at a value of "0". To address this data censoring issue, the Tobit model is utilized for exploration. The fundamental models set by this study are depicted in Equations (4) and (5).

$$RA_i^* = \alpha + \beta \cdot p\text{-structure} + \gamma \cdot \text{control} + \varepsilon \tag{4}$$

$$RA_i = \begin{cases} 1, \text{ if } RA_i^* < 0 \\ 0, \text{ if } RA_i^* \geq 0 \end{cases} \tag{5}$$

Where RA^* signifies the participation rate of households in the risky financial market, assigned 1 if they hold investments in risky financial assets and 0 otherwise. $p - structure$ represents the core explanatory variable, denoting household demographic structure, evaluated through indicators such as family size, elderly dependency ratio, child dependency ratio, and gender distribution of household children. $control$ encompasses a series of control variables affecting household participation in risky financial assets.

Subsequently, the Tobit model is applied to quantify the influence of household demographic structure on the proportion of investment in risky financial assets, as detailed in Equations (6) and (7).

$$\text{rate}_i^* = \alpha + \beta . p - \text{structure}\gamma\text{.control}\varepsilon \tag{6}$$

$$rate_i = \begin{cases} 1,\text{if} rate_i^* < 0 \\ 0,\text{if} rate_i^* \geq 0 \end{cases} \tag{7}$$

Here, $rate_i^*$ represents the latent variable, while $rate_i$ signifies the actual proportion of household investment in risky financial assets. When $rate^*$ assumes a negative value, $rate^*$ is set to 0. The $p - structure$ and $control$ variables in Equation (7) correspond to the definitions provided above.

3.2 Data source, explanation, and analysis

The data for this study is derived from the "China Household Finance Survey and Research Center," specifically the 2019 dataset. The survey encompassed 29 provinces, autonomous regions, and municipalities directly under the central government, spanning 343 districts and counties, along with 1360 village committees or neighborhood committees. The data is representative of the entire country, as well as provincially and sub-provincially significant cities. After excluding samples with missing key variables and consolidating data, a final dataset of 27,052 households was obtained.[5][6]

Aligned with the research objectives, the central explanatory variables in this study include household demographic structure, household age structure, and household children's gender structure. Household demographic structure primarily refers to family size, indicating the total number of family members cohabiting. Household age structure entails indicators such as the elderly dependency ratio (proportion of individuals aged 65 and above to the total labor force population) and the child dependency ratio (proportion of children aged 14 and below who are financially dependent to the total labor force population). As for household children's gender structure, this variable is treated as a dummy variable, with households having boys assigned a value of "1" and others assigned "0."

Numerous control variables are addressed in this study, predominantly categorized into three groups: head of household characteristics, background risks, and regional disparities.[7] The head of household characteristics encompass elements such as age, gender, years of education, marital status, and risk attitude of the head of the household. Background risks encompass the overall income level of the household, property ownership, health status, and social security measures. Regional disparities involve the distinction between rural and urban areas, geographic regions ("central, eastern, and western"), as well as GDP and housing prices for research purposes.

Variable definitions and descriptive statistical results are outlined in Table 1, illustrating the comprehensive scope of variables considered in the study.

Table. 1. Variable specification and descriptive statistical analysis.

Variable Names	Variable Descriptions	Observations	Mean Values	Standard Deviations
RA	Participation Rate in Risky Financial Assets	27,052	0.422	0.494
rate	Proportion of Investment in Risky Financial Assets	27,052	0.00842	0.0354
size	Household Size	27,052	3.367	1.481
PE	Elderly Dependency Ratio	27,052	0.266	0.369
PC	Child Dependency Ratio	27,052	0.193	0.194
GC	Children's Gender	27,052	0.424	0.494
age	Head of Household Age	27,052	32.92	14.45

GH	Head of Household Gender	27,052	0.792	0.406
edu	Head of Household Education Years	27,052	9.208	3.933
marriage	Head of Household Marital Status	27,052	0.925	0.263
attitude	Head of Household Risk Attitude	27,052	0.414	0.589
lnincome	Total Household Income Level	27,052	10.71	1.403
NE	Number of Real Estate Properties	27,052	1.232	0.529
health	Health Condition	27,052	0.401	0.490
ss	Social Security Coverage	27,052	1.739	0.872
AT	Rural vs Urban Residence	27,052	0.377	0.485
region	Geographic Region	27,052	1.869	0.831
GDP	Per Capita GDP	27,052	12.35	5.551
HP	Housing Prices	27,052	14,098	7,496

From Table 1, it becomes evident that in terms of the propensity to participate in the risky financial market, the 2019 sample mean is merely 0.42. Among households engaged in the allocation of risky financial assets, the proportion of investment in such assets relative to the total household assets is a mere 0.84%, indicating a state of "limited participation" and risk aversion in the asset allocation process among Chinese households.

Regarding demographic quantity structure, two-person households constitute the smallest group, while three-person households are the norm, indicating a leaning towards smaller family units. In terms of age structure, the mean elderly dependency ratio is 0.266, and the mean child dependency ratio is 0.193, offering preliminary insights into the challenges faced by the majority of middle-aged individuals in the current society. In terms of gender structure, the preference for male offspring in Chinese households is gradually diminishing, with the 2019 mean reaching 42.4%.

Analyzing head of household characteristics, the average age is 56 years, with a predominance of males, individuals with at least a junior high school education, married individuals, and risk-averse attitudes. Regarding background risks within households, the logarithmically transformed total household income is not subjected to mean analysis, but it has a standard deviation of 1.403, further emphasizing the uneven distribution of wealth among Chinese households. Within the statistical data, household property ownership, social security, and per capita GDP mean values are nearly identical. On average, each household owns one property; however, there is a considerable disparity between the minimum and maximum values. Furthermore, the general health status of Chinese households is relatively moderate, social security measures are relatively balanced, and the average per capita GDP is 12.35 with a standard deviation of 5.51, further underscoring the uneven regional economic development in China and the presence of spatial heterogeneity.

3.3 Empirical results analysis

The impact of household demographic structure on participation rates in the risk market was modeled using the Probit regression, while investment depth was modeled using the Tobit regression, as presented in Table 2.

Table. 2. Baseline regression findings

Variables	Probit		Tobi	
	(1) Regression Coefficients	(2) Marginal Effects	(3) Regression Coefficients	(4) Marginal Effects
size	-0.0491***	-0.0190***	-0.000786***	-0.0003321***
	(0.00716)	(0.00276)	(0.000185)	0.0000782
PE	-0.323***	-0.124***	-0.00237***	-0.0010025***
	(0.0346)	(0.0133)	(0.000837)	（0.0003536）
PC	0.657***	0.253***	0.00394**	0.0016659***
	(0.0637)	(0.0246)	(0.00164)	（0.0006933 ）
GC	-0.188***	-0.0723***	-0.00213***	-0.0008986***
	(0.0247)	(0.00943)	(0.000638)	（0.0002696）
age	-0.0197***	-0.00762***	0.000533	2.25e-06
	(0.000891)	(0.000344)	(0.00222)	（9.36e-06）
GH	-0.0483**	-0.0187**	-0.00222***	-0.000938***
	(0.0217)	(0.00844)	(0.000545)	（0.0002302）
edu	0.0539***	0.0208***	0.000811***	0.0003424***
	(0.00268)	(0.00103)	(0.000652)	（0.0000276）
marriage	-0.0115	-0.00445	0.000597	0.000252
	(0.0334)	(0.0129)	(0.000842)	（0.0003557 ）
attitude	0.0881***	0.0340***	0.00298***	0.0012589***
	(0.0145)	(0.00559)	(0.000361)	（0.0001526）
lnincome	0.144***	0.0554***	0.00282***	0.0011929***
	(0.00743)	(0.00286)	(0.000179)	（0.0000759）
NE	0.133***	0.0513***	0.000769*	0.0003246*
	(0.0164)	(0.00632)	(0.000413)	（0.0001744）
health	0.130***	0.0504***	0.000220	0.0000929
	(0.0176)	(0.00684)	(0.000447)	（0.0001889）
ss	0.0506***	0.0195***	0.000564**	0.0002383**
	(0.0103)	(0.00398)	(0.000257)	（0.0001087）
AT	-0.366***	-0.139***	-0.00207***	-.0008763***
	(0.0194)	(0.00718)	(0.000485)	（0.0002048）
region	-0.00815	-0.00314	-0.000104	-0.0000438
	(0.0125)	(0.00481)	(0.000312)	（0.0001316）
GDP	0.00492**	0.00190**	0.000141**	0.0000594**
	(0.00244)	(0.000942)	(0.000613)	（0.0000259）
HP	0.000383**	0.000148**	6.71e-08	2.83e-08
	(0.000193)	(0.0000746)	(4.86e-08)	（2.05e-08）
Constant	-2.022***		-0.0311***	
	(0.0952)		(0.00233)	
Observations	27,052	27,052	27,052	27,052

In Table 2, ***, **, and * respectively indicate significance at the 1%, 5%, and 10% levels, with standard errors shown in parentheses. From Table 2, it is evident that household size, elderly dependency ratio, and households with male children exhibit negative correlations with participation rates in risky financial assets. For each additional unit, the participation rate in risky assets decreases by 1.9%, 36%, and 10.1% respectively. Conversely, the child dependency ratio exhibits the opposite trend, with a one-unit increase leading to a 41.7% rise in the likelihood of participation in risky financial assets.

Regarding control variables, most of them significantly influence the allocation of household financial risk assets. Within head of household characteristics, older male heads of households tend to have a stronger risk-averse attitude.[8] Marital status has minimal impact on asset allocation. Higher levels of education, risk attitude, and family assets positively influence participation and the extent of participation in risk assets. In terms of background risks, improvements in family health status, total income, and the number of owned properties enhance the household's risk resistance capacity, thereby intensifying their inclination towards investing in risky financial assets and increasing participation rates.[9][10]

Concerning regional disparities, residing in rural areas displays a negative correlation in both regression coefficients and marginal effects, indicating that rural households exhibit a lower willingness to participate in risky financial assets compared to urban households. This aligns with traditional patterns of asset allocation, as rural households are more inclined to store assets in banks for security purposes. The testing outcomes of other variables are generally in line with expectations.

4 Conclusion

Through the utilization of the Probit and Tobit models, this study delved into the influence of household demographic structure on the allocation of risky financial assets. Through empirical analysis, we have discerned that household size, elderly dependency ratio, and child dependency ratio, among other demographic factors, play vital roles in influencing both the participation rate and investment proportion of households in risky financial assets. The Probit and Tobit models have demonstrated remarkable efficacy and explanatory power in unveiling these mechanisms of influence.

By delving deeply into the relationship between household demographic structure and financial asset allocation, we have provided more specific recommendations for household asset allocation and introduced novel perspectives and methodologies for risk management and financial planning. These research findings hold crucial practical significance for addressing demographic structural changes, optimizing financial asset allocation, and enhancing household financial stability.

References

[1]. Calvet, L. E., & Sodini, P. (2014). Twin Picks: Disentangling the Determinants of Risk-Taking in Household Portfolios. The Journal of Finance, 69(2), 226-248.

[2]. Coile, C., & Milligan, K. (2009). How Household Portfolios Evolve After Retirement: The Effect of Aging and Health Shocks. Review of Income and Wealth, 55(2), 226-248.

[3]. Fan, E., & Zhao, R. (2008). Health Status and Portfolio Choice: Causality or Heterogeneity? Journal of Banking and Finance, 33(6), 1079-1088.

[4]. Robinson, C. D., & Schumacker, R. (2019). Interaction Effects: Centering, Variance Inflation Factor, and Interpretation Issues. Multiple Linear Regression Viewpoints, 35(1), 6-11.

[5]. Cao, Q., Niu, X., & Li, J. (2021). Are Women Better Stock Traders? An Experimental Study. Southern Economic Journal, 87(1), 28-44.

[6]. Jia, X., Wang, A., & Hu, H. (2019). Financial Education Investment and Household Investment Behavior: An Empirical Analysis Based on China Urban Resident Family Financial Consumption Survey Data. Financial Forum, (12), 27-37.
[7]. Shi, G., & Shen, T. (2021). Population Structure, Household Asset Allocation, and Spatial Heterogeneity: An Analysis Based on China Family Finance Survey Data. Social Science Front, (07), 313(07).
[8]. Shu, J., Wu, Y., & Tang, W. (2021). Per Capita Income and Household Financial Asset Allocation: Influence Effects and Heterogeneity. Western Forum, 31(03), 57-71.
[9]. Wu, W., Shen, T., & Jiang, T. (2014). Does Real Estate Crowd Out Risky Financial Asset Allocation? An Empirical Analysis Based on Micro Survey Data. Scientific Decision Making, (11), 52-69.
[10]. Wu, W., Yi, J., & Zheng, J. (2010). Household Investment Structure in China: An Empirical Analysis Based on Lifecycle, Wealth, and Housing. Economic Research Journal, 45(S1), 72-82.