Savings, Asset Holding and Debt:

New Evidence from Chinese Household Data

Jialong Li

A thesis submitted in partial fulfilment of

the requirements for the degree of

Doctor of Philosophy

October 2021

Department of Economics

University of Sheffield

Abstract

This thesis consists of three empirical studies that investigate contemporary topics related to household finance. Specifically, this thesis aims to contribute to the literature relating to household savings, household risky assets and household debt by examining three distinct but related topics in the context of China.

The first empirical study (Chapter 2) examines the relationship between planning for overseas education and household saving behaviour in China by using a household-level dataset, the China Household Finance Survey (CHFS) covering 2011, 2013, 2015 and 2017. This chapter also examines the role of planning for overseas education on wider measures of household assets such as household financial assets and household net wealth. The results indicate that households where parents plan to send their children to study abroad hold more household savings, household financial assets and household net wealth than those who do not plan to do so for their children. Furthermore, such a positive effect of planning for overseas education on household savings is revealed after dealing with potential endogeneity issues. In addition, different effects of planning to send children to study abroad on household savings are found across the whole savings distribution.

The second empirical study (Chapter 3) examines the relationship between financial literacy and risky asset holding in China using a panel dataset from the CHFS covering 2013, 2015 and 2017 in order to control for unobserved heterogeneity. Risky asset holding is captured in three ways: the probability of holding risky assets; the log level of risky assets; and the share of risky assets in total household financial assets. Then, household risky assets are split into high risk assets and low risk assets. In addition, this chapter explores the relationship between financial illiteracy and household risky asset holding. The findings indicate that financial literacy is positively associated with risky asset holding. The importance of the role of financial literacy on household risky asset holding remains once time invariant effects have been accounted for. Furthermore, financial literacy has been found to be positively associated with high risk asset holding and low risk asset holding but the size of the positive effect of financial literacy differs across high risk assets and low risk assets. Finally, these findings have been found to

be robust after dealing with the potential endogeneity of financial literacy and the results have also revealed a negative relationship between financial illiteracy and household risky asset holding.

The third empirical study (Chapter 4) examines the association between risk attitudes and household debt using a household-level dataset from the CHFS (2011, 2013, 2015 and 2017). Household debt is captured by the probability of holding household debt and the amount of total household debt held. Then, household debt is split into housing debt and non-housing debt to explore how risk attitudes affect the two types of debt. In addition, households are split into urban and rural households in order to explore whether the effect of risk attitudes on household debt differs across urban and rural households. Finally, this chapter investigates the two-part process related to holding total household debt: (1) the decision to hold debt; and (2) the decision over the amount of debt held. The results indicate that risk tolerance is positively associated with household debt. The findings also indicate a positive relationship between risk tolerance and non-housing debt. In addition, we have found differences in the effect of risk tolerance across total household debt, housing debt and non-housing debt by rural and urban households. For example, the magnitude of the marginal effect of risk attitudes on the probability of holding total household debt is larger for rural households than for their urban counterparts. Finally, the findings are robust to using the double hurdle approach thereby providing further evidence that the risk tolerance of the head of household is positively associated with household debt.

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the University of Sheffield is solely my own work, other than where I have clearly indicated otherwise (in which case the extent of any work produced jointly by myself and any other person is clearly identified in it).

I warrant that this authorisation does not, to the best of my knowledge, infringe the rights of any third party.

I certify that this thesis consists of approximately 91,000 words.

Acknowledgement

I would like to express my greatest sincere gratitude to my supervisors, Karl Taylor and Sarah Brown for their time, guidance and unwavering support at every stage of my PhD. Special thanks to Daniel Gray and Alberto Montagnoli for their helpful comments and suggestions during my confirmation review.

I would also like to express my thanks to everyone at the Department of Economics at the University of Sheffield for their friendly support and encouragement. Thanks to my former colleagues Raslan Alzuabi and Celia Wallace who were very helpful in the early stages of my research.

I am also grateful to my friends and peers who have made my PhD experience extremely enjoyable. Thanks goes to Jiao Li, Dongzhe Zhang, Zizhou Luo, Shanshan You, Bingxue Wang, Meng Xu, Ta An Thu, Gomez Cardena Myriam Alejandra and many others.

Lastly, but most importantly, I wish to express my very profound gratitude to my family. I wish to thank my father, Yuebing Li, my gratitude to him is beyond words. My sincerest appreciation goes to my mother, Jing Huang.

Contents

Abstract i

Declaration iii

Acknowledgement iv

1 Introduction 1

1.1 Motivation and Aims 2

1.2 Structure and Content of the thesis 6

1.2.1 Chapter 2 7

1.2.2 Chapter 3 8

1.2.3 Chapter 4 9

2 Household Saving in China: Parental Investment in their Children's Human Capital 11

2.1 Introduction 12

2.2 Literature Review 16

2.2.1 Keynesian theory 16

2.2.2 The life-cycle model 18

2.2.3 The theory of precautionary saving 20

2.2.4 Economic and pension reform 23

2.2.5 Population structure and the one-child policy 24

2.2.6 Housing 27

2.2.7 Overseas education 28

2.3 Data and Methodology 31

2.3.1 Data 31

2.3.2 The Tobit model 33

2.3.3 The Lewbel IV approach 39

2.3.4 Censored and uncensored quantile analysis 40
2.4 Results 41
2.4.1 The Tobit Model 41
2.4.2 The OLS Model: Household net wealth 48
2.4.3 The Lewbel IV Approach 49
2.4.4 The Uncensored and Censored Quantile Regression Analysis 51
2.5 Conclusion 53
2.6 Figures 56
2.7 Tables 57
Appendix to Chapter 2 67
3 Financial Literacy and Risky Asset Holding: Evidence from Chinese Households 74
3.1 Introduction 75
3.2 Literature Review 79
3.2.1 The U.S. and other Developed Countries 79
3.2.2 China 85
3.3 Data and Methodology 97
3.3.1 Data 97
3.3.2 Cross-sectional Analysis 99
3.3.2.1 The Logit Model 99
3.3.2.2 The Tobit model 100
3.3.3 Panel Analysis 102
3.3.3.1 The Fixed Effects Logit Model 102
3.3.3.2 The Random Effects Tobit Model 103
3.3.4 IV Probit Analysis 104
3.3.3.1 Cross-sectional IV Probit Model 104
3.3.3.2 Pooled IV Probit Model 105
3.3.5 The Financial Literacy and Financial Illiteracy Measures 105

3.3.6 Other Explanatory Variables 108

3.4 Results 111

3.4.1 Cross-sectional Analysis 111

3.4.1.1 The Logit Analysis 111

3.4.1.2 The Tobit Analysis 115

3.4.2 Panel Analysis 117

3.4.2.1 The Fixed Effects Logit Analysis 117

3.4.2.2 The Random Effects Tobit Analysis 117

3.4.3 IV Probit Analysis 119

3.4.3.1 The Cross-sectional IV Probit Analysis 119

3.4.3.2 The Pooled IV Probit Analysis 120

3.4.4 Cross-sectional Logit Analysis for Financial illiteracy 121

3.5 Conclusion 123

3.6 Figures 126

3.7 Tables 129

4 Household Debt and Attitudes towards Risk: Evidence from China 143

4.1 Introduction 144

4.2 Literature Review 147

4.2.1 The U.S. and other Developed Countries 147

4.2.2 China 159

4.3 Data and Methodology 169

4.3.1 Data 169

4.3.2 The Random Effects Logit Model 171

4.3.3 The Random Effects Tobit Model 172

4.3.4 Additional Robustness Checks 174

4.3.4.1 The Fixed Effects Logit Model 174

4.3.4.2 The Double Hurdle Model 174

4.3.5 The Risk Attitudes Measure 176

4.3.6 Other Explanatory Variables 177

4.4 Results 180

4.4.1 Random Effects Logit Analysis 180

4.4.2 Random Effects Tobit Analysis 186

4.4.3 Additional Robustness Checks 188

4.4.3.1 Fixed Effects Logit Analysis 188

4.4.3.2 Double Hurdle Analysis 189

4.5. Conclusion 192

4.6. Figures 195

4.7. Tables 197

Appendix to Chapter 4 204

5 Conclusion **210**

5.1 Conclusion 211

5.1.1 Thesis Summary 211

5.1.1.1 Summary of Chapter 2 211

5.1.1.2 Summary of Chapter 3 212

5.1.1.3 Summary of Chapter 4 213

5.1.2 Policy implications 214

5.1.3 Shortcomings of the research and areas for future research 216

References **218**

List of Figures

2.1 Distribution of Ln(Savings) 56

2.2 Distribution of Ln(Financial Assets) 56

2.3 Distribution of Ln(NS-Financial Assets) 56

2.4 Distribution of Ln(Net Wealth) 56

3.1 Distributions of the log level of risky assets in 2013, 2015 and 2017, i.e. Ln(Riksy Assets)>0 126

3.2 Distributions of the log level of high risk assets in 2013, 2015 and 2017, i.e. Ln(High Risk Assets)>0 126

3.3 Distributions of the log level of low risk assets in 2013, 2015 and 2017, i.e. Ln(Low Risk Assets)>0 127

3.4 Distributions of the share of risky assets in 2013, 2015 and 2017, i.e. Share of Risky Assets>0 127

3.5 Distributions of the share of high risk assets in 2013, 2015 and 2017, i.e. Share of High Assets>0 128

3.6 Distributions of the share of low risk assets in 2013, 2015 and 2017, i.e. Share of Low Assets>0 128

4.1 Distribution of the log level of total household debt in 2011, 2013, 2015 and 2017 (panel), i.e. Ln(Total Debt)>0 195

4.2 Distribution of the log level of housing debt in 2011, 2013, 2015 and 2017 (panel), i.e. Ln(Housing Debt)>0 195

4.3 Distribution of the log level of non-housing debt in 2011, 2013, 2015 and 2017 (panel), i.e. Ln(Non-housing Debt)>0 196

List of Tables

2.1 Definition of Variables 57

2.1 Definition of Variables(continued) 58

2.2 Summary Statistics - All Variables; Cross-section (t=2011, 2013, 2015, 2017) 59

2.3 Tobit analysis: Household Savings, Financial Assets, NS-Financial Assets and Plans to send children to study abroad 60

2.4 Tobit analysis - Marginal effects at the extensive margin (M.E.): Household Savings, Financial Assets, NS-Financial Assets and Plans to send children to study abroad 61

2.5 Tobit analysis - Marginal effects at the intensive margin (M.E.): Household Savings, Financial Assets, NS-Financial Assets and Plans to send children to study abroad 62

2.6 OLS Regression Analysis: Net Wealth and plans to send children to study abroad 63

2.7 Lewbel IV Approach: Household Savings, Financial Assets, NS-Financial Assets, Net Wealth (Study Abroad is the endogenous variable) 64

2.8 Lewbel IV Approach: Household Savings, Financial Assets, NS-Financial Assets, Net Wealth (Ln(Income) is the endogenous variable) 65

2.9 Censored Regression: Household Savings, Financial Assets, NS-Financial Assets and Uncensored Quantile Regression: Net Wealth, and Plans to send children to study abroad 66

2.10 Random effects Tobit analysis: Household Savings, Financial Assets, NS-Financial Assets and Plans to send children to study abroad 67

2.11 Random effects Tobit analysis - Marginal Effects at the extensive margin (M.E.): Household Savings, Financial Assets, NS-Financial Assets and Plans to send children to study abroad 68

2.12 Random effects Tobit analysis - Marginal Effects at the intensive margin (M.E.): Household Savings, Financial Assets, NS-Financial Assets and Plans to send children to study abroad 69

2.13 Random effects Regression Analysis: Net Wealth and plans to send children to study abroad 70
2.14 Fixed-Effects Lewbel IV Approach Model: Household Savings, Financial Assets, NS-Financial Assets, Net Wealth (Study Abroad is the endogenous variable) 71
2.15 Fixed-Effects Lewbel IV Approach Model: Household Savings, Financial Assets, NS-Financial Assets, Net Wealth (Ln(Income) is the endogenous variable) 72
2.16 Censored Regression: Household Savings, Financial Assets, NS-Financial Assets and Uncensored Quantile Regression: Net Wealth, and Plans to send children to study abroad 73
3.1 Definition of Variables 129
3.1 Definition of Variables (Continued) 130
3.2.A Summary Statistics - All Variables; Cross-section (t = 2013 and 2015) 131
3.2.B Summary Statistics - All Variables; Cross-section & Unbalanced (t = 2017, 2013 & 2015& 2017) 132
3.3 Summary Statistics - Financial Literacy; Cross-section (t = 2013, 2015 and 2017) 133
3.4 The determinants of the probability of Risky Asset, High Risk Asset and Low Risk Asset Holding - Cross-sectional Logit analysis 134
3.5 The determinants of the log level of Risky Assets, High Risk Assets and Low Risk Assets - Cross-sectional Tobit analysis 135
3.6 The determinants of the share of Risky Assets, High Risk Assets and Low Risk Assets - Cross-sectional Tobit analysis 136
3.7 The determinants of the probability of Risky Asset, High Risk Asset and Low Risk Asset Holding in 2013, 2015 and 2017 - Fixed-effects Logit analysis 137
3.8 The determinants of the log level of Risky Assets, High Risk Assets and Low Risk Assets in 2013, 2015 and 2017 - Random effects Tobit analysis 138
3.9 The determinants of the share of Risky Assets, High Risk Assets and Low Risk Assets in 2013, 2015 and 2017 - Random effects Tobit analysis 138
3.10 The determinants of the probability of Risky Asset, High Risk Asset and Low Risk Asset Holding - Cross-sectional IV Probit analysis 139

3.11 The determinants of the probability of Risky Asset, High Risk Asset and Low Risk Asset Holding in 2013, 2015 and 2017 - Pooled IV Probit analysis 139

3.12.A The determinants of the probability of Risky Asset Holding - Cross-sectional Logit analysis - financial illiteracy 140

3.12.B The determinants of the probability of High Risk Asset Holding - Cross-sectional Logit analysis - financial illiteracy 141

3.12.C The determinants of the probability of Low Risk Asset Holding - Cross-sectional Logit analysis - financial illiteracy 142

4.1 Definition of Variables 197

4.1 Definition of Variables (Continued) 198

4.2 Summary Statistics - All Variables; Panel (t = 2011, 2013, 2015 and 2017) 199

4.3 The determinants of the probability of Total Debt, Housing Debt and Non-housing Debt Holding - Random effects Logit analysis 200

4.4 The determinants of the log level of Total Debt, Housing Debt and Non-housing Debt - Random effects Tobit analysis 201

4.5 The determinants of the probability of Total Debt, Housing Debt and Non-housing Debt Holding - Fixed effects Logit analysis 202

4.6 The determinants of the log level of Total Debt, Housing Debt and Non-housing Debt - double hurdle analysis (pooled) 203

A4.3 The determinants of the probability of Total Debt, Housing Debt and Non-housing Debt Holding - Random effects Logit analysis (Risk Attitudes Dummy variables).................. 205

A4.4 The determinants of the log level of Total Debt, Housing Debt and Non-housing Debt - Random effects Tobit analysis (Risk Attitudes Dummy Variables)............................... 206

A4.4 The determinants of the log level of Total Debt, Housing Debt and Non-housing Debt - Random effects Tobit analysis (Risk Attitudes Dummy Variables) (Continued) 207

A4.5 The determinants of the probability of Total Debt, Housing Debt and Non-housing Debt Holding - Fixed effects Logit analysis (Risk Attitudes Dummy variables) 208

A4.6 The determinants of the log level of Total Debt, Housing Debt and Non-housing Debt - Pooled double hurdle analysis (Risk Attitudes Dummy variables)................................ 209

Chapter 1

Introduction

1.1 Motivation and Aims

The term “Household finance” was used by Campbell (2006) to describe the area of financial economics that explores how households use financial instruments to attain their objectives, which has been attracting substantial academic attention over the last two decades. As stated by Guiso and Sodini (2013), household finance is now a thriving, vibrant and self-standing field of research and household finance is distinct from the traditional fields of asset pricing and corporate finance. Households have to make many financial decisions, which are irrelevant to asset pricing and corporate finance but instead are focused on household finances and financial wellbeing. For example, households have to manage means of payment by cash or credit cards, forms of debt (informal or formal loans), insurance contracts (such as accident, property and health insurance), savings behaviour (such as bank deposits) and investment decisions (such as deciding between stocks or mutual funds).

Additionally, households have many distinct features that give the field its specific character. For example, human capital, the main determinant of lifetime income for most households, is typically non-traded, accumulates very slowly and is hard to predict. In addition, households generally hold illiquid assets, notably housing, which is probably the most important tangible asset for most households. Finally, households often face constraints in borrowing, which impair consumption smoothing over time (Campbell, 2006; Guiso and Sodini, 2013). In short, household finance takes into account the differences in household characteristics and the changing financial environments in which households operate. It takes into account investment decisions but, unlike asset pricing, it focuses on the median, instead of the marginal, household. It explores the financing of household consumption and investment but, unlike corporate finance, it does not deal with issues such as the separation between ownership and control. A growing number of studies related to household finance have been conducted over the past two decades, focusing on topics such as household saving behaviour, risky asset holding and debt. Many studies have focused on the U.S. and the U.K., with a smaller number of studies exploring household finance in China.

Regarding household saving behaviour, China has the highest household savings rate in the world. For example, data from the OECD (2015) shows that from 2000 to 2012, household savings as a proportion of net household income in China was the highest compared to all OECD countries and, furthermore, in 2010 this rate reached 42.1%. In contrast, households in the U.S. saved only 5.8% of household net disposable income in 2010 and, similarly, the household saving rate in the U.K. was 6.1% in 2010 (OECD, 2015). Such a phenomenon in China has attracted some attention among academics. For example, Modigliani and Cao (2004) find evidence in favour of the life-cycle hypothesis, according to which households attempt to smooth their expected consumption over their life cycle. They analyse aggregate (national level) data from the China Statistics Year Book for the period 1953 to 2000 and find that income growth has a positive and significant impact on household savings. However, Kraay (2000) finds different results using a panel of Chinese province-level saving data between 1978 and 1995. Specifically, Kraay (2000) finds that the savings rate of rural households in China falls with expectations of income growth. Meng (2003) finds that urban households in China are likely to smooth their total consumption and increase their savings when temporary income shocks occur.

However, the existing literature on household saving in China has ignored an important educational factor, which is related to parental investment in their children's human capital. This is the focus of the first empirical study (Chapter 2). China has experienced rapid economic development and Chinese households have started to invest more in the human capital of young people and, in particular, in overseas education. For example, data from the Ministry of Education of the People's Republic of China (2018) shows that the number of Chinese students who chose to study abroad grew by 8.83% between 2016 and 2017. Furthermore, with respect to finance, in 2018, 90% of Chinese students studying overseas were financed by their parents, with the rest being educated overseas through government support (the Ministry of Education of the People's Republic of China, 2019). Thus, the ability of parents to finance such investments in human capital relies on accumulating savings. Thus, the first empirical study presented in this thesis contributes to the existing literature on household saving in China by analysing a motive for household saving behaviour that has been neglected in the existing literature. Specifically, Chapter 2 investigates the relationship between whether parents in China plan to have their children educated overseas and their saving behaviour from an empirical perspective.

Turning to risky asset holding, there has been considerable interest in the role of financial literacy in household risky asset holding in a number of countries, which is examined in the second empirical study (Chapter 3). For example, Van Rooij et al. (2011) find that financial literacy is positively associated with stock market participation for the Dutch population in 2005. Arrondel et al. (2015) obtain similar findings for France, specifically that basic financial literacy is positively associated with stock market participation, but not related to the share of stocks held in financial assets. Similarly, Thomas and Spataro (2018) provide micro-level evidence supporting a positive association between financial literacy and stock market participation using the individual-level dataset from waves 3 and 4 of the Survey of Health, Aging and Retirement in Europe (SHARE) for nine countries (Austria, Belgium, Denmark, Germany, Italy, France, Switzerland, Sweden, and the Netherlands).

A small yet growing literature has focused on the implications of financial literacy for household asset holding in China. For example, Liao et al. (2017) use cross-sectional household-level data for China from the 2014 Survey Consumer Finances (SCF) and find that low financial literacy is a widespread phenomenon in China and that, the higher the level of financial literacy, the higher is the probability of holding risky financial assets for urban households in China. Similarly, Zou and Deng (2019) present relatively recent empirical evidence suggesting that financial literacy is positively associated with the probability of household financial market participation (stock market participation, fund market participation and bond market participation) in China using a household-level dataset from the 2012 SCF.

Although these recent studies present some interesting results relating to China, it should be noted that they focus on cross-sectional data, which cannot be used to control for unobserved heterogeneity. That is, there may be unobserved effects that could affect both financial literacy and risky asset holding. Thus, Chapter 3 contributes to further investigating the relationship between financial literacy and participation in household financial markets in China using panel data thereby controlling for unobserved heterogeneity across households. In addition, this chapter investigates the relationship between financial illiteracy and risky asset holding, which has been largely ignored in the existing literature in China. Finally, this chapter is the first study for China to split risky assets into high-risk assets and low-risk assets to explore how financial literacy is associated with the likelihood of holding high-risk and low-risk assets in China.

The focus of the final empirical study (Chapter 4) is on debt and risk attitudes in China. Regarding household debt, once again, a number of studies exist, which have explored debt holding amongst U.S. households. For example, Brown et al. (2013) investigate the correlation between household debt and attitudes towards risk based on a household-level unbalanced panel dataset from the U.S. Panel Study of Income Dynamics (PSID). They find that households are more likely to hold debt if the household head is more tolerant towards risk. Wildauer (2016) explores household debt in the U.S. based on a household-level pooled dataset from the Survey of Consumer Finances (SCF) between 1995 and 2007 and finds evidence of a negative relationship between household income and household debt. More recently, Coibion et al. (2020) find that debt accumulation in the U.S. is relatively higher for high-income households in high-inequality regions than in low-inequality regions.

It is perhaps not surprising that there is a relatively small literature on household debt in China as the level of household debt in China is relatively low as compared with developed countries. For example, the level of total household debt in China was around $517.7 billion in 2007 and increased to $9,683.5 billion in 2020.[1] To place this into context, the gross domestic product (GDP) in China was $3,550 billion in 2007 and increased to $14,723 billion in 2020.[2] This means that the proportion of household debt to GDP in China was 14.6% in 2007 and has increased to 65.8% in 2020. In contrast, in the U.S., the level of total household debt was around $12,000 billion in 2007 and has increased to $14,559 billion in 2020, and the GDP in the U.S. has increased from $14,452 billion in 2007 to $20,937 billion in 2020, which indicates that the proportion of household debt to GDP was 83.0% in 2007 and has actually decreased by 13.5% in 2020.[3,4] Although such a large increase in the proportion of household debt to GDP in China has led to policy-makers being concerned about financial vulnerability and risk at the household level, there remains a shortage of academic research into the determinants of debt at the household level in China. In other words, there is a relatively small literature on household debt in China. For example, Fan et al. (2017) find that the household's social network is positively associated with the amount of informal borrowing for house-purchase based on a household-level dataset from the second wave (2013) of the China Household Finance Survey (CHFS). In a similar vein,

---

[1] Data source: https://www.ceicdata.com/en/indicator/china/household-debt.
[2] Data source: https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?locations=CN.
[3] Data source: https://www.ceicdata.com/en/indicator/united-states/household-debt.
[4] Data source: https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?locations=US.

Cull et al. (2019) find that households are more likely to take out a bank loan if anyone in the household is a Communist Party member and that the number of siblings is positively associated with the probability of having loans, bank loans, and non-bank loans.

Given the limited number of studies on household debt in China, Chapter 4 contributes to the existing literature by exploring the relationship between risk attitudes and household debt in China. Risk attitudes have been identified as an important determinant of household debt in the U.S. but the role of risk attitudes has not been explored in the context of household debt in China. It is interesting to explore this relationship in China because, as Brown et al. (2013) argue, debt repayments are usually financed from household income. Hence, it is apparent that, if there exists uncertainty in household income, then attitudes towards risk will potentially influence household debt holding, given the distribution of future income and interest rates. Moreover, there exists uncertainty in Chinese household income. For example, Yu and Zhu (2013) find evidence supporting the existence of uncertainty in Chinese household income and, furthermore, they find that, compared with U.S. households, Chinese household income is characterised by much more uncertainty (i.e. volatility). Similarly, Chamon et al. (2013) find that income uncertainty substantially has increased from the 1990s to the 2000s in China, which further supports the existence of household income uncertainty in China.

1.2 Structure and Content of the Thesis

Chapters 2, 3 and 4 present the original empirical analysis of this thesis and each chapter is a standalone self-contained study, although they are linked by the focus on household finance as well as the dataset analysed, the China Household Finance Survey. Chapter 2 examines the relationship between whether parents plan to have their children educated overseas and household saving behaviour in China. Chapter 3 provides further empirical evidence of the relationship between financial literacy and household risky asset holding in China. Chapter 4 presents an empirical analysis of the determinants of household debt in China, focusing on risk attitudes. Finally, Chapter 5 provides a conclusion to the thesis and discusses key policy implications and areas for future research. A brief overview of each of the empirical chapters is given below.

1.2.1 Chapter 2

Chapter 2 aims to examine the relationship between planning for overseas education and household saving behaviour in China. This chapter contributes to existing knowledge by examining a motive for household saving behaviour that has been neglected in the existing literature on China. Specifically, the chapter investigates the relationship between whether parents in China plan to have their children educated overseas and their saving behaviour by using a household-level dataset from China Household Finance Survey (CHFS) covering 2011, 2013, 2015 and 2017. We mainly focus on pooled cross-sectional models because there is only a small panel element to the data analysed in this chapter given the focus on households with children. Household saving behaviour is captured by the amount of household savings. The chapter also examines the role of planning for overseas education on wider measures of household assets such as household financial assets and household net wealth.

From a methodological point of view, a number of econometric models are employed in this chapter in order to explore the robustness of the findings. Following the existing literature, Tobit models are used for the censored outcomes, namely household savings and household financial assets. An Ordinary Least Squares (OLS) model is used to estimate the role of planning for overseas education on household net wealth. In addition, the Lewbel approach is employed to explore the robustness of the findings to dealing with the potential endogeneity of our key explanatory variable of interest, i.e. planning for overseas education. Finally, a censored quantile regression approach is used to explore the effect of parents planning to send their children to study abroad across the entire distribution of household savings and household financial assets, while the uncensored quantile regression approach is used for household net wealth.

The results of the Tobit and OLS analysis suggest that households where parents plan to send their children to study abroad hold more household savings, household financial assets and household net wealth than those who do not plan to do so for their children, after controlling for a wide set of covariates including household income. The results from using the Lewbel approach provide further evidence of a positive relationship between household savings and parents planning to have their children educated overseas. Finally, the results from the censored quantile regression analysis shed light on the effects of planning overseas education across the whole savings distribution. For example, planning to send children to study abroad has a statistically insignificant effect on household savings below the 60th percentile but in contrast a statistically positive effect at and above the 60th percentile.

1.2.2 Chapter 3

The aim of Chapter 3 is to provide further insights into the relationship between financial literacy and risky asset holding in China. Chapter 3 sheds further light on the association between financial literacy and household risky asset holding using household-level data from the CHFS covering waves 2013, 2015 and 2017.[5] We first analyse waves 2013, 2015 and 2017 as three separate cross-sections for comparison with the existing literature, then employ panel models. Financial literacy is based on three financial knowledge questions about interest rates, inflation, and risk diversification, which were first devised by Lusardi and Mitchell (2008). Thus, financial literacy is defined as an index ranging from 0 to 3, which is increasing in the level of financial literacy, where 0 denotes the lowest level of financial literacy and 3 denotes the highest level of financial literacy. Risky asset holding is captured in three ways: the probability of holding risky assets; the log level of risky assets; and the share of risky assets in total household financial assets. Thus, the Logit model is used for modelling the probability of holding risky assets, which is a binary outcome and the Tobit model is employed for modelling the log level of risky assets, and the share of risky assets in total household financial assets, which are censored outcomes. We also split risky assets into high risk assets and low risk assets based on the different risk-levels of financial products in order to investigate how financial literacy influences the holding of assets characterized by different levels of risk. To explore the robustness of the findings, the instrumental variable (IV) Probit model is used to deal with the potential endogeneity related to our key explanatory variable of interest, i.e. financial literacy. Finally, this chapter explores the relationship between financial illiteracy and risky asset holding because it has been argued that financial illiteracy is distinct from choosing incorrect answers to financial literacy questions (Yoong, 2011). For example, if an individual is ambiguity-averse, he/she prefers known risks over unknown risks. In this case, an individual without any financial knowledge, i.e. he/she is financially illiterate, is less likely to hold any stocks than those with some financial knowledge, i.e. those who choose incorrect answers to financial literacy questions

[5] There is no information on financial literacy provided in wave 2011. Financial assets in Chapter 3 relate to risky assets as opposed to safe assets such as savings and cash.

(Gollier, 2011). In other words, those who choose incorrect answers to financial literacy questions and believe that stocks are not risky and have high returns, will be more likely to participate in the stock market than those who have no financial knowledge (Yoong, 2011).

The contribution of this chapter to the existing literature is threefold. The chapter investigates the relationship between financial illiteracy and risky asset holding, which has been largely ignored in the existing literature in China. This chapter is the first study for China to split risky assets into high-risk assets and low-risk assets to explore whether financial literacy is associated with the likelihood of holding high-risk and low-risk assets in China. Finally, in contrast to existing studies on China, the use of panel data enables investigation of the association between financial literacy and household risky asset holding whilst controlling for unobserved heterogeneity.

The results of the Logit and Tobit models verify the findings obtained in the existing literature on China as well as for other countries in that financial literacy is found to be positively associated with the probability of holding risky assets and the amount of risky assets held. The results are robust to using the IV Probit approach and suggest that the level of financial literacy is positively associated with the probability of holding risky assets, high risk assets and low risk assets and that the instrumental variable, specifically whether the household head received any economics or finance education during school, has a positive impact on financial literacy. Finally, this chapter also finds a negative relationship between financial illiteracy and household risky asset holding.

1.2.3 Chapter 4

Chapter 4 aims to examine the relationship between the head of household's attitudes towards risk and household debt in China using panel data. We use the CHFS, as used in the previous chapters, focusing on waves 2011, 2013, 2015 and 2017, where Logit and Tobit models are used for the panel analysis while the double hurdle model is based on pooled cross-sectional analysis. Risk attitudes are measured based on the question *'in which project below would you want to invest most if you have adequate money?'* The answers include: (1) a project with high risk and high return; (2) a project with slightly high risk and slightly high return; (3) a project with average risk and average return; (4) a project with slight risk and return; and (5) unwilling to carry any risk. Following Hu et al. (2015), we assign a value of 0 to 4 to each of the above five options and the index is increasing in risk-tolerance. To the best of our knowledge, this chapter is the first attempt to explore the association between attitudes towards risk and household debt in China, which may provide further understanding of which factors affect household debt behaviour in China.

A range of econometric methods are used in this chapter, depending on the nature of the dependent variables analysed. The random effects Logit and fixed effects Logit model are used for binary outcomes, i.e. total debt holding, housing debt holding and non-housing debt holding, whereas, following the existing literature, random effects Tobit models are employed for the analysis of the role of risk attitudes on the amount of total household debt, housing debt and non-housing debt, which are all censored outcomes. Finally, we use the double hurdle approach for an additional robustness check to allow for the two-part process related to holding total household debt: (1) the decision to hold debt; and (2) the decision over the amount of debt held.

The findings from the random effects Logit and fixed effects Logit models indicate a positive relationship between the head of household's risk tolerance and household debt in China. The results of the random effects Tobit model suggest that risk tolerance is positively associated with the amount of household debt and non-housing debt held by the household. The findings are robust to using the double hurdle approach and suggest that the risk tolerance of the head of household is positively associated with the amount of household debt and that financial illiteracy has a positive impact on the probability of holding total household debt, housing debt and non-housing debt. Finally, some interesting differences are found across the three types of debt as well by rural and urban residence. For example, attitudes towards risk play a more important role in determining the probability of holding total household debt for rural households than their urban counterparts.

Chapter 2

Household Saving in China: Parental Investment in their Children’s Human Capital

2.1 Introduction

Chinese household saving behaviour has attracted much attention among academics and policy-makers, which is not surprising as China has the highest household savings rate in the world. For example, data from the OECD (2015) shows that from 2000 to 2012, the household net savings rate in China was the highest compared to all OECD countries and, furthermore, in 2010 this rate reached 42.1%. In contrast, households in the U.S. saved only 5.8% of household net disposable income in 2010 and, similarly, the household savings rate in the U.K. was 6.1% in 2010 (OECD, 2015). In addition, China's gross domestic savings (as a % of GDP) in 2017 reached an extraordinary 46.36%, while the world gross domestic savings (as a % of GDP) was 25.23% (World Bank, 2017).[6]

A number of theories have been put forward to explain the extraordinary saving phenomenon in China including Keynesian theory, Life-cycle models and models of precautionary savings. For example, Modigliani and Cao (2004) focus on the Life-cycle model and use estimates of national savings based on aggregate (national level) data from the China Statistics Year Book published by the National Bureau of Statistics for the period 1953 to 2000 to investigate household saving behaviour in China. They find that income and the population structure are key factors in explaining the high savings rate in China. Chamon and Prasad (2010) find that the saving rate is especially high for younger and older households using household-level panel data from the annual Urban Household Surveys conducted by China's National Bureau of Statistics over the period 1990 to 2005. There is also substantial evidence suggesting that the savings rate in China differs between rural and urban households (see Kraay, 2000; Cristadoro and Marconi, 2012) and that, in general, urban households save more (see Lugauer et al., 2019). Du and Wei (2013) provide both macro-level evidence from different countries and micro-level evidence from China on the association between the gender-ratio (male-to-female) and the savings rate using cross-country data from the United Nations' Population Division over the period 1990-2010 and cross-sectional household-level data from a survey of rural households (the Chinese Household Income Project in 2002). They find that the gender-ratio has a significant impact on savings in China.

[6] Gross domestic savings (as a % of GDP) are defined as GDP minus final consumption expenditure, expressed as a percentage of GDP. Gross domestic savings (as a % of GDP) consist of savings of the household sector, the private corporate sector and the public sector.

At the microeconomic level, the education level, age and the gender of the household head are important explanatory variables in recent research. Specifically, both the age and education level of the household head are positively related to household savings while the gender of the household head has no significant impact on household savings (Kong and Dickinson, 2016; Lin and Lai, 2018; Gruber, 2018). Moreover, national policies in China such as the pension reform, the state-owned enterprise reform and the one-child policy have had significant impacts on household savings (Feng et al., 2011; He et al., 2018; Lugauer et al., 2019). All these issues are discussed in detail in Section 2.2 below.

Focusing on household structure, the existing literature has found that the number of dependent children in a family has played an important role in household saving behaviour in China. Specifically, Chinese families with fewer dependent children have a significantly higher savings rate (Lugauer et al., 2019). However, existing studies have ignored the importance of expenditure on education for children and, in particular, expenditure on overseas education for children in influencing household saving behaviour in China. Household expenditure on children's education is an important component of investment in human capital, which is generally defined as the skills and abilities embodied in individuals. Such investments can help individuals find a 'decent' job with higher income, as well as, help individuals more generally in terms of their decisions, attitudes and behaviours related to numerous aspects of life, such as a healthy lifestyle. In addition, there may be more investment in children's human capital in China as compared to in the U.S. or U.K. since less attention is paid to out-of-school expenditure in the U.S. or U.K. while the phenomenon of developing children's extra-curricular skills or improving their academic performance through private tutoring and after-school classes is more commonly observed in Chinese households (Chi and Qian, 2016). Moreover, parents planning to send children to study abroad seems to be a non-negligible issue in China. For example, data from the Ministry of Education of the People's Republic of China (2018) shows that the number of Chinese students who chose to study abroad was 179,800 (0.0135% of the total Chinese population) in 2008 and after only one decade, this number had reached 662,100 (0.0474% of the total Chinese population), with a growth rate of 8.83% over the previous year in 2017. Between 1978 and 2018, the cumulative number of Chinese overseas students studying abroad had reached 5,857,100, i.e. 0.4197% of the total Chinese population (the Ministry of Education of the People's Republic of China, 2019). The phenomenon of such an increase in the number of students studying abroad indicates that sending children to study abroad is getting more and more important for Chinese parents. Most Chinese parents appear to send their children to study abroad for postgraduate education due to the fierce competition in postgraduate entrance exams in China and, to some extent, a master's degree or above is a necessary condition for a student to find a good job with a relatively high salary in China. This is especially the case in first-tier cities, where lots of 'decent' jobs are only open to postgraduate students. This is reflected in the increase in the number of students enrolled for masters degrees by over tenfold from 145,433 in 1995 to 1,793,953 in 2013.[7] In addition, investment in children's education can be regarded as insurance for support in later life, especially in developing countries where the capital market is underdeveloped and important social welfare institutions are lacking (Nugent, 1985). Such an old-age security motive for investment in children's education is more credible when the number of children is limited in a household by family planning policies, such as the one-child policy in China. In this case, parents may enhance human capital when they have fewer children by allocating more resources to each child (Becker and Lewis, 1973). So, sending children to study abroad may be a good way to invest in children's human capital for Chinese parents.

With respect to finance, in 2018, the total number of students studying overseas was 662,100 and among them, there were 596,300 Chinese students studying overseas financed by their parents, accounting for 90.06%, while the rest of the Chinese students were educated overseas through government support (the Ministry of Education of the People's Republic of China, 2019). The report in the Value of Education series, *Higher and Higher*, conducted by HSBC in 2017 (HSBC, 2017) shows that the country with the highest proportion of parents currently paying for private tuition (or who have done so in the past) and financially preparing for their children's education through general savings, investments or insurance is China, at 93% and 55%, respectively. In contrast, these proportions for the U.K. are 23% and 5%, respectively. Additionally, over two-fifths (41%) of parents from the 15 countries or territories plan to send their children to study abroad and parents are more likely to consider postgraduate education (36%) than undergraduate education abroad (34%).[8] Parents in the Middle East and Asia are the most likely to have their children educated overseas, with over half of parents in the United

---

[7] Data source: http://www.stats.gov.cn/tjsj/ndsj/2014/.
[8] The 15 countries/territories are China, Indonesia, Egypt, Hong Kong (China), India, Singapore, Malaysia, Taiwan (China), the United Arab Emirates (UAE), Mexico, the U.S., France, Canada, Australia and the U.K..

Arab Emirates (65%), Indonesia (60%), India (55%), and China (54%) considering this for their children. However, the proportions of parents who consider university education abroad for their children in developed countries such as the U.S., France, the U.K. and Australia are relatively low, with the proportions being 36%, 22%, 22% and 17%, respectively (HSBC, 2017).

The reason why sending children to study abroad seems to be attractive to Chinese parents includes three main aspects. First, the quality of overseas higher education, especially in the U.S. and the U.K., is regarded as better than that in China. For example, there are 32 and 18 world top 100 universities from the U.S. and the U.K. respectively, compared to only 6 universities from mainland China (excluding Hong Kong, Taiwan and the Macau areas) in the QS World University Rankings, 2019.[9] Second, parents from the 15 countries or territories including China listed above consider foreign language skills, gaining international work experience and exposure to new experiences, ideas and cultures to be the main benefits that cannot be easily gained domestically (HSBC, 2017). Finally, 91% of Chinese parents believe that postgraduate study may help their children find a 'decent' job. To summarise, overseas postgraduate education is regarded as a good choice for a variety of reasons with more than half of Chinese parents having considered overseas education for their children (HSBC, 2017).

This chapter analyses a motive for household saving behaviour that has been neglected in the existing literature on China. Specifically, the chapter investigates the relationship between whether parents in China plan to have their children educated overseas and their saving behaviour from an empirical perspective. Intuitively, one might predict that the households where parents plan to send their children to study abroad have more household savings than those in which parents do not plan to do so, given their additional motive for saving. Data from a recent household-level survey, the China Household Finance Survey, conducted in 2011, 2013, 2015 and 2017, is analysed. The number of households surveyed in each of the years is as follows: 8,438, 28,141, 37,289 and 40,011. To explore the robustness of the findings, a range of econometric techniques are applied in this chapter including the Tobit model, the OLS model, the Lewbel IV model, and uncensored and censored quantile models.

[9] https://www.topuniversities.com/university-rankings/world-university-rankings/2019.

We find that households where parents are planning to send their children to study abroad have higher household savings than those who do not plan to do so.

The rest of the chapter is organized as follows. Section 2.2 reviews the relevant literature. Section 2.3 discusses the data and methodology. The empirical results are discussed in Section 2.4 and Section 2.5 concludes.

2.2 Literature Review

This section reviews the relevant literature on household saving behaviour in China. As stated in Section 2.1, the phenomenon of the high savings rate in China has been explored from the perspective of approaches such as Keynesian theory (Modigliani and Cao, 2004; Liu and Hu, 2013), Life-cycle theory (Modigliani and Brumberg, 1954; Kraay, 2000; Horioka and Wan, 2007; Chao et al., 2011; Kong and Dickinson, 2016; Pan, 2016) and precautionary savings theory (Meng, 2003; Liu and Hu, 2013; Chamon et al., 2013; He et al., 2018). The existing literature has explored different factors that may affect Chinese household saving behaviour such as economic and pension reform (Kraay, 2000; Feng et al., 2011), population structural factors (Banerjee et al., 2010; Du and Wei, 2013; Wei and Zhang, 2011; Chamon and Prasad, 2010; Ge et al., 2018; Lugauer et al., 2019), and housing (Wang and Wen, 2012; Chen et al., 2013). In contrast, only a limited number of papers have discussed the role of overseas education for a household in China (Davey, 2005; Zweig et al, 2004; Mazzarol and Soutar, 2002; Yang, 2007; Qian and Smyth, 2011). The rest of this section discusses each of the above approaches in turn.

2.2.1 Keynesian Theory

From a standard Keynesian theory perspective, Chinese household saving behaviour is entirely determined by current household income (Modigliani and Cao, 2004; Liu and Hu, 2013). Modigliani and Cao (2004) propose that people living in a poor country with low income may not accumulate enough wealth whilst working in order to support consumption after their retirement in the future and saving behaviour in relatively poor countries can be explained by the Keynesian savings model. Although China has experienced rapid economic development, it is still a developing country. In order to verify standard Keynesian theory, Liu and Hu (2013) construct a province-level panel dataset from 1990 to

2009 in China based on data available in the China Statistical Year Books to investigate the relationship between the level of income and the household saving rate with a fixed-effects model for estimation. Standard Keynesian theory considers income to be the only determinant of the household saving rate. Therefore, they use per capita income (unit: ¥100/person which approximates £10/person) as the only explanatory variable in estimations for urban households and rural households and find that income, indeed, has a positive impact on the household saving rate in urban households but the effect is not statistically significant in the rural household cohorts.[10,11] Specifically, they find that each ¥100/£10 increase in per capita income would increase the urban household saving rate by 12%.

Keynesian theory emphasizes the importance of income in household saving behaviour and several studies focus on the influence of income on the savings rate (e.g. Chu and Wen, 2017; Gruber, 2018). For instance, the level of household income and its corresponding coefficient of variation of disposable income (income inequality) are found to be positively related to the aggregate savings rate by Chu and Wen (2017).[12] They first analyse the data from the Flow of Funds Table of China (1992-2012) and the China Year Book of Household Survey (2013) and find that China has experienced a remarkable increase in the aggregate savings rate at the beginning of 2000s and that the household sector is the main contributor to aggregate savings since the 2000s. Those households with the highest income are found to save the most. Then, they use the community-level panel data from the China Family Panel Studies (CFPS) covering 2010 and 2012 to explore how income inequality affects the aggregate savings rate of China with quantile regression at the median rather than OLS regression due to the fact that the dependent variable, the aggregate savings rate, has a large number of extreme values. They find that the higher the income inequality, the greater is the marginal effect of income inequality on the

---

[10] Liu and Hu (2013) use real disposable income, with the benchmark year 2000=100. Rural per capita income is measured by the rural per capita real net income and the per capita income of all households uses the weighted average mean value of urban per capita real disposable income and the rural per capita real net income with the ratio of their respective populations.

[11] ¥ represents the Chinese currency symbol, with unit Yuan/RMB. We convert the Chinese currency into sterling throughout the thesis based on the average exchange rate in 2013: 9.6182 CNY/GBP from https://www.exchangerates.org.uk.

[12] Chu and Wen (2017) define the household-specific consumption function as $c_i = b_i * y_i = (\alpha + \beta y_i) * y_i$, where c_i is the consumption of household $i \in N$, b_i is the corresponding propensity to consume out of income, $\beta < 0$ captures the feature that the propensity to consume is a decreasing function of disposable income, and $\alpha > \beta y_i$ ensures that the propensity to consume is strictly positive. Total consumption of this economy is defined as $C = \sum_{i \in N} c_i$, total income of this economy is defined as $Y = \sum_{i \in N} y_i$, the aggregate saving rate is given by $s = 1 - \frac{C}{Y} = 1 - \sum_{i \in N} c_i / \sum_{i \in N} y_i = \tilde{\alpha} - \beta(CV^2 * \mu + \mu)$, where $\tilde{\alpha} = 1 - \alpha$ is a constant term and $CV = \sqrt{Var(y)}/E(y)$ is the coefficient of variation of income, which captures income inequality.

aggregate savings rate, which implies that income inequality is an important determinant of China's rapidly rising savings rate.

More recently, Gruber (2018) uses the household-level cross-sectional dataset from the Universities Service Centre for China Studies in the Chinese University of Hong Kong from 1993 to 1997. The saving rate is modelled using OLS and the key independent variable is relative income (defined as the ratio of household income to the average income in its location).[13] However, the error term is likely to be related to both saving rate and income.[14] So, the education dummy variables are employed as instruments in IV regression analysis to solve the problem. He finds that the Chinese household savings rate is determined by a relative, rather than absolute, income level and that Chinese households with higher than the average income in their locality save a larger fraction of their income.

2.2.2 The Life-cycle Model

The life-cycle model, which was first developed by Modigliani and Brumberg (1954), is based on the hypothesis that households will attempt to smooth their expected consumption over their life cycle. This theory has been used in a number of studies to investigate Chinese household saving behaviour, and, interestingly, the studies have come to different conclusions. For example, Modigliani and Cao (2004) find evidence in favour of the life-cycle hypothesis using aggregate (national level) data for the period 1953 to 2000 employing OLS. However, Kraay (2000) finds different results using a panel of province-level saving data between 1978 and 1995 and applying a two-stage least squares estimator because there exists measurement error in the proxy for the explanatory variable capturing expected future income growth. It is also possible that this measurement error is correlated with the dependent variable (i.e. savings). Moreover, Modigliani and Cao (2004) show that income growth has a positive and significant impact on household savings. In contrast, Kraay (2000) finds that the savings rate of rural households falls with expectations of income growth. Modigliani and Cao (2004) find that inflation has a significant and positive effect on household savings. Horioka and Wan (2007) use province-level

[13] I only control for the level of household income rather than permanent and transitory income since the focus is household savings, i.e. a stock rather than flow variable.
[14] Household income contains both permanent and transitory elements and, theoretically, permanent income has no effect on the saving rate but the transitory part is likely to be related to the saving rate. Yearly earnings in logs, y_{it}, are specified as a function of permanent income, age and a deviation term,$y_{it} = y_i^p + h(age_{it} - age_0) + u_{it}$, where y_i^p is permanent income, $h(age_{it} - age_0)$ represents the age-earnings profile, and u_{it} is the deviation term which is likely to be correlated with the saving rate due to transitory income and can be decomposed into two parts: transitory income, y_{it}^{tr}, and measurement error, me_{it}

panel data for the 1995-2004 period from the China Statistics Year Book, the China Population Statistics Year Book and the International Monetary Fund's International Financial Statistics to investigate the determinants of China's high savings rate. They employ generalised-method-of-moments (GMM) dynamic models for three reasons: (1) inertia is likely to be present in annual data, and a dynamic specification seems appropriate; (2) it seems desirable to control for the endogeneity of explanatory variables; (3) there exists the possibility of unobserved province-specific effects correlated with the regressors and it is desirable to control for these effects. The GMM model results show that inflation is negatively related to rural household savings. The studies described above find different results relating to the determinants of Chinese savings rates. The differences in the results may be due to the different types of data such as province-level data, aggregate-level data, and household-level data, as well as due to different methods such as OLS, Tobit, IV and two-stage least squares (2SLS) models.

In order to calibrate the life-cycle model, some scholars have considered additional factors based on the life-cycle model. For example, Chao et al. (2011) build a structural model of household saving behaviour based on the life-cycle hypothesis for the Chinese economy and find that only 35% of the growth in Chinese household savings can be explained by the life-cycle hypothesis. However, after allowing for the motivation of young adults to buy a house and the remittances received from their parents, the model can explain the high and increasing level of savings since the mid-nineties.[15]

Chinese households can be divided into two types of residents, rural households and urban households because of the "hukou" system. The rural group differs from the urban group in terms of saving motives and the amount of savings (see Pan, 2016). Pan (2016) uses the household-level panel dataset from the Chinese Household Income Project (CHIP) in 1995 and 2002 to investigate household saving behaviour for both rural and urban households and employs the DiNardo-Fortin-Lemieux decomposition analysis (DFL) proposed by DiNardo et al. (1996) to decompose the total saving rate into two parts for rural and urban households separately. The first part represents the changes in household characteristics (the "endowment effect") while the second one is attributed to changes in 'returns' to household

[15] Kong and Dickinson (2016) extend the life-cycle model and use the first wave of the household-level panel dataset from the China Household Finance Survey in 2011, separating household income into permanent and transitory income. They also find that factors such as household wealth (including housing), self-employment, age and urban/rural dwelling as well as the region of residence influence the household saving rate.

characteristics (the "return effect").[16] He finds different shifts for an increase in the saving rate for rural and urban households through analysing the "endowment effect" and the "return effect" for rural and urban households, respectively. In detail, the rural saving rate increases the most at lower percentiles while the urban saving rate experienced a large shift at higher percentiles. Then, unconditional quantile regression analysis, as proposed by Firpo et al. (2009), is employed to explore each specific household characteristic's contribution to the total saving rate increase. This approach enables the evaluation of the marginal effect of a change of each observation on the unconditional quantiles. The results of the unconditional quantile regression analysis reveal that a large part of the rural household saving rate increase can be explained by changes in household characteristics such as household income, the number of children studying at school and there is evidence that households at higher income quantiles have strong incentives to save for education as a result of increased tuition fees for higher education. Specifically, the education variable measured by the amount of tuition is statistically significant and positively associated with the saving rate at the 90th income percentile.[17] However, only a small portion of the urban household saving rate increase from 1995 to 2002 can be explained by changes in household characteristics.

2.2.3 The Theory of Precautionary Saving

The precautionary saving model is an extension of the basic intertemporal optimization model that extends the life-cycle framework by introducing uncertainty about future resources (Yilmazer and Scharff, 2014). The theory of precautionary saving plays an important and different role in different countries. Taking the U.S. for example, Starr-McCluer (1996) uses the cross-sectional household-level dataset from the 1989 Survey of Consumer Finances (SCF) and applies an OLS model to investigate the effect of uncertainty in health costs on U.S. household savings.[18] They find that it is difficult to establish that households facing greater risks in health save more as a result. Similarly, Yilmazer and

---

[16] DFL analysis calculates the distribution of a variable of interest for two groups (A and B), then returns the relevant percentiles for the factual distributions of groups A and B, as well as the counterfactual distribution of group B based on the assumption that group B has the same characteristics as group A. In Pan (2016), group A includes the households surveyed in 2002 and group B contains the households surveyed in 1995, which means that the counterfactual simulates the saving rate in 1995 assuming households have identical characteristics as in 2002. The gap between the counterfactual and the true saving rate in 1995 measures the proportion of the savings increase that can be explained by changes in household characteristics (the endowment effect) and the gap between the counterfactual and the true saving rate in 2002 represents the 'returns' part that cannot be explained by household attributes (i.e. the return effect).

[17] According to the Chinese Household Income Survey (CHIP) 2002 survey, 42.9% of rural households mainly saved for education.

[18] We focus on the U.S. because U.S. households seem to have different saving behaviour to Chinese households. The existing literature on precautionary saving theory has been discussed extensively in the U.S., and hence provides an interesting basis for comparison.

Scharff (2014) use the household-level panel dataset from the U.S. Health and Retirement Study, which has been conducted in 1992, 1994, 1996, 1998, 2000 and 2002, to estimate an OLS model to explore the relationship between health risks and U.S. household savings as well as median quantile regressions. This approach reduces the sensitivity of the results to outliers and avoids having to throw out outlier observations on the basis of subjective judgement about the validity of the data. They find no evidence that savings increase with health risks. Although uncertainty in health has been found to have little effect on the U.S. household saving motive, uncertainty in income has been found to affect the U.S. household precautionary saving motive. For instance, Kazarosian (1997) initially uses an unbalanced individual-level panel dataset covering 1965, 1966, 1968, 1970, 1972, 1974, 1975 and 1980 from the National Longitudinal Survey to create measures of permanent income and income uncertainty for each individual.[19] He then regresses the cross-sectional wealth to permanent income ratio (i.e. precautionary savings) on these measures and a vector of personal characteristics to investigate whether precautionary savings due to uncertainty in income exist for U.S. households. The estimation results show that precautionary motives are important for U.S. household savings.

Focusing on the literature on China, the effects of income risk and employment uncertainty on saving behaviour are taken to be evidence of household precautionary saving. For example, Meng (2003) uses cross-sectional household-level data from the 1999 Urban Household Income, Expenditure and Unemployment Survey (UHIEE) with OLS for the estimation to investigate the effect of income uncertainty, measured by the variance of past income, and the probability of household workers being unemployed in 1999, on household total consumption, which is a continuous variable.[20] He finds that urban households are likely to smooth their total consumption and increase their savings when temporary income shocks occur. He also finds evidence of the precautionary saving motive among Chinese urban households, which is consistent with consumption theory in that those households with precautionary motives will consume less (more) when future income uncertainty is higher (lower).

[19] Permanent income can be estimated by the equation $Y_i^P = Z_i\beta + \delta_i$ using panel data, where Y_i^P is annual income with no transitory component, evaluated at the same age for everyone, Z_i is a vector of observable characteristics, with the parameter β. δ_i is the time invariant individual-specific error. Income uncertainty is defined as the standard deviation of the residual of each individual's estimated age versus log-income profile.

[20] The 1999 Urban Household Income, Expenditure and Unemployment Survey (UHIEE) asked households to report their last five years' income. There is no information on the households' past income in the CHFS, and the first wave of the CHFS is used. Thus, there are no previous waves to construct a measure of income uncertainty.

Similarly, Liu and Hu (2013) construct a province-level panel dataset from 1990 to 2009 for China based on data available in the China Statistical Year Books to investigate whether Chinese households have a precautionary saving motive. They use an autoregressive model to calculate the determined income trend and use the difference between real income and the determined income trend to measure income uncertainty. They then estimate the precautionary saving model using 2SLS because future income uncertainty and future income growth are endogenous independent variables. The 2SLS results show that the impact of income uncertainty on household saving is positive and statistically significant, which supports the existence of precautionary saving motives among Chinese households.

Chamon et al. (2013) evaluate the effects of macroeconomic transformation on income uncertainty at the household level in China. They use a panel of urban Chinese households containing 7 waves, which are 1989, 1991, 1993, 1997, 2000, 2004 and 2006, with a sample of about 4400 households from the Urban Household Survey (UHS). They estimate a Mincer earnings function and use the residuals as the measurement of the permanent and transitory components of income and use the variances of permanent and transitory shocks to capture the income risk.[21,22] Then, they calibrate a precautionary savings model, which provides a quantitative measure of how an increase in the variance of transitory shocks to household income can translate into a rise in savings.

In terms of employment uncertainty, He et al. (2018) present relatively recent empirical evidence for precautionary savings among Chinese households between 1995 and 2002. Prior to the state-owned enterprises (SOEs) reform, the SOEs workers were treated similarly to government employees who never lose their jobs. Post-reform, many loss-making SOEs were closed or privatised, which led to large scale unemployment of SOEs workers starting in 1997. However, this reform had no impact on government employees. He et al. (2018) use cross-sectional household-level data from the Chinese Household Income Survey (CHIP) in 1995 and 2002, which allows them to identify the groups prior to

[21] Chamon et al. (2013) use the residuals to estimate the permanent and transitory components of income: $y_{iat} = u_{iat} + v_{iat}$, $u_{iat} = u_{i,a-1t} + \omega_{iat}$, where y_{iat} denotes the log earnings residuals for the head of household i aged a in year t from a Mincer earnings regression, u_{iat} is the permanent component and v_{iat} is the transitory component. They assume $Var(\omega_{iat}) = \text{б}_{\omega t}^2$ and $Var(v_{iat}) = \text{б}_{\gamma t}^2$ as the variances of permanent and transitory shocks.

[22] A Mincer earnings regression was developed by Mincer (1958), which is a single-equation model that explains earnings as a function of schooling and experience, $\ln(w) = f(s,x) = \ln w_0 + \rho s + \beta_1 x + \beta_2 x^2$, where w is earnings (the intercept w_0 is the earnings of someone with no education and no experience); s denotes years of schooling; x denotes years of potential labour market experience. The parameters ρ, and β_1, β_2 can be interpreted as the returns to schooling and experience.

the state-owned enterprises (SOEs) reform and post-reform, respectively, and to create a dummy variable to capture the unemployment risks for SOE workers. This variable takes a value of one, if the household head works for a SOE and zero if the household head works for a government or public institution. They use a Difference in Difference (DiD) approach to estimate a model of precautionary savings and apply an IV-Tobit model, where the instruments include education dummy variables and interactions of education with age and age-squared to estimate the effect of unemployment risks as an endogenous variable on household precautionary savings.[23] In general, the DiD model is used to obtain causal estimates of a policy change that affects different subgroups at different points in time. They find significant evidence of precautionary savings caused by an increase in unemployment risk for SOE workers relative to that for government employees.

2.2.4 Economic and Pension Reform

In China, some policy reforms have had a considerable impact on household saving behaviour such as the reform of the economy from planned to market as well as the pension reform. Under the planned economy, economic activity, resource allocation, and even the number of products is governed by the government. Post-reform, markets and enterprises dominate most of these economic activities. Kraay (2000) considers the transition period from a planned to a market economy and shows that a standard life-cycle hypothesis model is unable to explain Chinese household saving behaviour during the transition period from a planned to a market economy. He suggests that disequilibrium factors, especially shortages and rationing in goods and credit markets, can explain China's saving experience. The pension reform started in the mid-1990s and prior to the reform, the arrangements were the same for public sector employees and those in private enterprises, which are based on the pay-as-you-go (PAYG) system. This PAYG system covered about 75-90% of a worker's wage and provided housing, medical care and social security to their workers. But post-reform, a new three-pillar pension system and a new framework were established. The new three-pillar pension system includes a pooling account

[23] He et al. (2018) use the empirical specification $\frac{W_i}{P_i} = \beta_0 + \beta_1 SOE_i + \beta_2 RISK_i + \beta_3 \log(P_i) + \beta_4 Z_i + v_i$, where the dependent variable is the ratio of financial wealth W_i to permanent income P_i for household i, SOE_i captures unemployment risks, which is a dummy variable that takes a value of one if the household head works for a SOE and zero if the household head works for a government or public institution, $RISK_i$ measures income risks conditional on being employed, vector Z_i includes a number of demographic control variables. They use $\Delta\beta_1 = \beta_1^{2002} - \beta_1^{1995}$ to capture the magnitude of the precautionary savings of the SOE workers caused by increases in their unemployment risk, where β_1^{1995} is the coefficient for the SOE dummy in 1995 and β_1^{2002} is the coefficient for the SOE dummy in 2002,

to redistribute to all beneficiaries; compulsory individual accounts; and voluntary supplementary pensions provided via commercial insurance. The first pillar imposes a payroll tax of 17% (paid by employers) to ensure that employees who have worked for more than 15 years have a replacement ratio of 20%. The second pillar (paid jointly by employers and employees) establishes an individual account for each employee. The contribution rate for this is 11% of an individual's wage, of which the employer contributes 3%. After retirement, the employee receives a monthly benefit from this account amounting to the accumulated value divided by 120. Under the new framework, those who had retired before 1997 (old workers) remained in the original PAYG system, those who entered the labour market in or after 1997 (new workers) came under the new three-pillar pension system, and those who started work before 1997 and are retired or will retire after 1997 (middle workers) were covered by a transitional plan.

Feng et al. (2011) use the household-level cross-sectional dataset from the Chinese Household Income Project covering 1995 and 1999 to estimate the relationship between public pension wealth, computed at the individual level, and the household savings rate.[24] They treat the pension reform of 1995-1997 as the source of exogenous variation in pension wealth and use an OLS estimator. They find a statistically significant negative relationship between pension reform, as measured by the ratio of adjusted pension wealth to household income, and household savings. Then they use the IV approach because the calculation of pension wealth is based on projected (rather than actual) future earnings, which are clearly unobserved. Hence, pension wealth is inevitably measured with error and this may bias the OLS estimation. The interactions of the time controls and the age dummies are treated as instruments and the IV results are consistent with those of OLS.

2.2.5 Population Structure and the One-child Policy

Turning to household structure, the family planning policies started in 1972 and the one-child policy announced in 1979, which was designed to restrict the rapid growth of the Chinese population and lower fertility, plays an important role in the change of the demographic structure, especially in terms of the number of children in the household. Moreover, the one-child policy affected the saving behaviour

[24] Pension wealth is defined as the present value of expected pension income from retirement age to the expected age at death. There is no information on pension wealth in the CHFS.

of Chinese households by reducing household consumption needs (and so the income share of consumption) for families with children and removing what was in the Chinese tradition a substitute for savings for retirement because children are obliged to take care of the elderly, not only by social norms but also by law in China (Cristadoro and Marconi, 2012).

Banerjee et al. (2010) adopt a life-cycle model and focus on the specific observation that Chinese households rely on their children to support their parents as they grow older. They consider the endogeneity between fertility and savings decisions and address it by exploiting the variation in fertility caused by a shift in family planning policies in 1972. The household-level cross-sectional dataset from the Urban Households Survey (UHS) portion of the larger 2008 Rural-Urban Migration in China and Indonesia (RUMiCI) for China survey in 2008 is used and OLS and 2SLS estimation techniques are applied to explore the correlation between household size and savings. There may be reverse causality between savings and fertility, so 2SLS estimation is applied to address such issues. There are two instruments included in the 2SLS estimation, which are a dummy variable indicating whether the first child was born during 1972 or after 1972 and an interaction term between the first instrument and a dummy variable indicating if the first child is a son. They find that savings rise with a fall in the number of children in a household and that this effect is lower when the single child (i.e. the one child in the household) is a boy.

Chamon and Prasad (2010) analyse a household-level panel dataset over the period 1990 to 2005 from the annual Urban Household Survey (UHS) conducted by the National Bureau of Statistics using quantile regression at the median with the household savings rate as the dependent variable to explain why Chinese households are deferring consumption regardless of rapid income growth. This is the first detailed investigation of Chinese household saving behaviour using micro-data over a long time span. Their findings show that there is a U-shaped pattern of savings over the life-cycle for Chinese households, which is the opposite of the traditional "hump-shaped" profile of savings over the life-cycle, where young workers save very little, which is a common phenomenon for U.S. households, see e.g. Browning and Lusardi (1996). The U-shaped age-profile shows that the households headed by young persons and old persons have the highest savings rates. Du and Wei (2013) consider gender differences and incorporate the gender ratio (male-to-female ratio) in the pre-marital cohort. They develop a theoretical model (an overlapping generations model) to analyse whether and how a rise in the gender ratio in China may lead to an increase in the male savings rate. Specifically, men increase their savings rate to improve their relative status in the marriage market, while women may reduce their savings rate because of the increased savings from their future husbands if intra-household bargaining is ignored. If intra-household bargaining is incorporated, then the woman's response becomes ambiguous because of their incentive to increase their savings rate to protect their bargaining power within a family.[25]

Wei and Zhang (2011) suggest that people save to improve their relative standing in the marriage market. They use one wave of the household-level panel dataset from the Chinese Household Income Project (CHIP) of 2002 and apply OLS and quantile regression techniques to explore the relationship between the gender ratio and household savings for two groups, where one is the households with a son and the other group comprises those households with a daughter. The quantile regression analysis is applied as a comparison and robustness check with the OLS model and, according to both estimation results, they find that savings appear to be related to whether the child is a boy or not. Specifically, they find that households save more when the (only) child is a boy. The rising proportion of boys born, as a result of the one-child policy, contributes to high savings since parents wish to increase their children's chances of securing a wife.

Lugauer et al. (2019) relate the number of dependent children to a life-cycle model. They analyse a household-level unbalanced panel of data from the China Household Finance Survey (CHFS) covering two waves (2011 and 2013) using a two-stage least squares Tobit regression approach. The one-child policy in China allows them to deal with the possible endogeneity between household savings and fertility decisions because the policy has reduced family size for the whole country. They apply 2SLS Tobit regression techniques to deal with the endogeneity issue of the main explanatory variable, i.e. the total number of dependent children in the household, and use the county birth rate as the instrument because the enforcement of the population control policy differs from place to place. For example, some

[25] Wei and Zhang (2011) suggest that people save to improve their relative standing in the marriage market. They use one wave of the household-level panel dataset from the Chinese Household Income Project (CHIP) of 2002 and apply OLS and quantile regression techniques to explore the relationship between the gender ratio and household savings for two groups, where one is the households with a son and the other group comprises those households with a daughter. The quantile regression analysis is applied as a comparison and robustness check with the OLS model and, according to both estimation results, they find that savings appear to be related to whether the child is a boy or not. Specifically, they find that households save more when the (only) child is a boy. The rising proportion of boys born, as a result of the one-child policy, contributes to high savings since parents wish to increase their son's chances of securing a wife.

areas have allowed additional children if both parents work in high-risk occupations, or are minorities, or if both parents are single children themselves, while other areas have not.[26] Their control variables include the number of elderly people, the number of workers, age, years of education and the health status of the household head. Their results show that Chinese families with fewer dependent children have a significantly higher savings rate. Moreover, the savings rate is found to be higher for households with more workers, higher education, better health, and more assets.

Similarly, Ge et al. (2018) develop an overlapping generations (OLG) model to illustrate the effect of population control policies (reflected by birth rates) and demographic structural changes such as the number of adult children in the household, the number of siblings, educational attainment and the age of the individual, the gender-ratio (the male-to-female ratio) and birth rates in the previous year on household savings decisions between 1990 and 2005. They match the 1989-1991 and 2004-2006 household-level cross-sectional data from the Urban Household Survey (UHS) with the 1990 and 2005 population censuses for each single year-of-birth cohort in each province. They find that those older households with fewer adult children saved more, middle-aged households with fewer dependent children experienced an increase in savings, and younger households with fewer siblings also saved more.

2.2.6 Housing

Expenditure on housing usually accounts for a large fraction of household spending and, in general, households must save more and for a longer period of time to buy a house when house prices go up. However, Wang and Wen (2012) find a weak relationship between aggregate saving rates and house prices in China. They argue that the view that the high aggregate household saving rate in China resulted from the rapidly rising house prices is a popular misconception because this view ignores the saving-expenditure-cancellation effect across cohorts and the offsetting population effects from rising income and house prices. The saving-expenditure-cancellation effect means that the housing expenditure of current homebuyers in the aggregate saving ratio always cancels out the income saved by potential future homebuyers for future house purchases. This means that if the population and house

[26] Single child means only one child in the household.

prices are constant, the average saving rate across all cohorts at any point in time is zero and independent of house prices. The offsetting population effect is defined as follows. If house prices increase over time and grow faster than income, then the population share of the potential future homebuyers is greater than that of the current homebuyers. In this case, the expenditure of the current homebuyers cannot cancel out the savings of potential future homebuyers. Chen et al. (2013) construct a Life-cycle model incorporating housing demand and incomplete markets to investigate the relationship between housing demand and the household saving rate in China. They find that the change in the demand for housing does not affect the aggregate household savings rate since the higher down payment ratio leads to substitution between housing and non-housing assets.

2.2.7 Overseas Education

Although the papers reviewed in this section all present some interesting evidence on household saving in China, they ignore an important educational factor, which is related to overseas education. China has experienced rapid economic development and Chinese households have started to invest more in children's human capital. Since the reform and opening in 1978, the Chinese government has promulgated a series of policies, which aim to allocate part of government funding to a small number of 'elites' to study abroad and also to encourage self-funded overseas study.[27] A good education can guarantee a better future for children, which is regarded as common-sense for Chinese parents (Jiang and Ashley, 2013). Specifically, a student who has successfully completed higher education can find a 'decent' job and, furthermore, overseas higher education can lead to a better quality of education, especially that provided by the world top 100 universities. In addition, a graduate with a foreign degree is regarded as having better skills and being more employable in the market place of industry (Davey, 2005). Moreover, overseas study is a formal requirement for people who seek employment in a prestigious university in China. Many local governments have enacted a series of returnee talent plans to

[27] Elites are defined as those researchers who can push the development of their research fields. These researchers and students can enjoy the benefits of these policies if they succeed in finishing their overseas study and choose to work in China. The policies include living costs and house-purchasing subsidies. In addition, a small number of overseas students from China become 'elites' who can gain funding from the government. Only approximately 10% of total overseas students in 2018 were government-funded, which means that funding is allocated on a highly competitive basis (see the Ministry of Education of the People's Republic of China, 2019).

attract those who have been educated overseas, such as students with foreign PhDs because foreign PhDs are regarded as being worth more than domestic PhDs in China (see Zweig et al. 2004).[28]

Only a limited number of papers have analysed the issue of overseas higher education for Chinese households. For example, Mazzarol and Soutar (2002) analyse a questionnaire collected from four countries or territories (Taiwan, China, India and Indonesia) and argue that the motives of the Chinese students studying abroad are influenced by "push and pull" factors. The "push" factors relate to the economic, social and political forces within the source country. Specifically, the "push" factor is fourfold. The first one is the perception that overseas education is "better" than home, i.e. Chinese, education. The second one relates to a student's ability to gain entry into local programs because if it is difficult for them to gain entry into particular study programs within their own country or the programs they wish to enter are unavailable, then they would choose to study abroad because it is easier to get onto courses overseas as long as they have the solid foundation of knowledge for the course and are able to pay the overseas education fee. The third and final factors relate to a desire to gain a better understanding of the "west" and an intention to migrate after graduation, respectively. The "pull" factors include the characteristics of the host country such as the reputation for the quality of education of the destination country and that the qualifications from the destination country are recognised in China. Other important factors include the influence of relatives, parents and friends, the cost of tuition fees and living expenses, the climate, lifestyle, crime rate, safety and racial discrimination, the importance of geographic proximity to a destination country and whether family or friends have studied there. Similarly, Yang (2007) proposes four factors motivating Chinese students to study abroad. First, fast economic growth in China has caused household income growth enabling families to be able to afford to send their children to study abroad. Second, going abroad to study has become a trend in Chinese society. A third factor is that recently Chinese government policy has changed to a more positive attitude towards supporting international education in terms of providing more funding. However, such funding is still somewhat limited. The fourth factor relates to an inadequate supply of university places in China's higher education system.

[28] For example, the local governments in Beijing, Shanghai, Guangzhou, Shenzhen and other prosperous cities provide housing and venture subsidies to overseas educated students and even allow them to apply for the local registered permanent residence. The local registered permanent residence application is under the severe restrictions in rich cities, especially in Beijing and Shanghai.

Turning to empirical analysis, Qian and Smyth (2011) investigate the determinants of household educational expenditure on educating children abroad and highlight that in recent years there has been a significant increase in the number of Chinese studying abroad, and that, although significant progress has been made in higher education in China, resources are still inadequate with the result being that not all demand can be met domestically. They use a cross-sectional household-level dataset collected by the China Mainland Marketing Research Company (CMMRC) in 2002 with a sample of 6,383 observations, and use a Tobit estimator since households with no domestic education expenditure are censored at zero. They also examine the factors determining the probability that households have incurred expenditure on overseas education. However, there are only a small number of respondents who reported having incurred expenditure on overseas education. Thus, the sample used to analyse expenditure on overseas education is skewed with a very high proportion of 'zeros'. They find that families with children under college age (less than 18 years old) are less likely to incur expenditure on overseas education for their children, whereas parents are more likely to incur such expenditure if their children are older than 19. In other words, unsurprisingly, the demand is greater for overseas education at the ages associated with undergraduate and postgraduate education. The household income level has also been found to have significant effects on the magnitude of the overseas educational expenditure. Specifically, household income is divided into 20 categories ranging from the lowest income group to the highest income group. The findings indicate that the households in the fourth income quintile spent 130% more on overseas education than the households in the bottom two income quintiles, while those at the top income level spent 200% more on overseas education than the households in the bottom two income quintiles. Finally, households with a college-educated father, a mother who is a middle professional and live in a coastal area are significantly more likely to send their children to study abroad.

In addition, Li and Feng (2018) investigate how the experience of overseas education changed students' evaluation of China's development and prospects based on an analysis of individual-level panel data from the Beijing College Students Panel Survey (BCSPS) from 2009 to 2013. Students' attitudes toward China are measured by two indicators where the first one is the satisfaction with China's overall current situation using the following question: "In general, how satisfied are you with China's overall current situation?". The second one is confidence in China's future prospects using the following question: “In general, how optimistic are you about China’s future prospects?” As both questions were asked in each wave of the BCSPS, these two indicators are time-variant and as the responses for each question are given on a scale from 0 to 100, these two indicators are treated as continuous outcomes. Regarding overseas education, the key explanatory variable, this is defined as a dummy variable, which equals 1 if the student i at time t is studying abroad otherwise studying in China. They employed a fixed effects estimator to investigate how overseas education affects students’ attitudes toward China and find that, compared with those with no experience of overseas study, students who pursue an overseas postgraduate education tend to be less satisfied with China’s current situation and less confident in China’s future prospects.

From the studies reviewed in this section, it seems appropriate to conclude that overseas education is an important issue for Chinese households and an issue which may affect the financial situation of Chinese households. Hence, this chapter aims to explore how parental investment in their children’s human capital, as indicated by whether parents are planning to send their children to study abroad, affects the saving behaviour of Chinese households.

2.3 Data and Methodology

2.3.1 Data

The dataset analysed in this chapter is from the China Household Finance Survey (CHFS) conducted by the Southwestern University of Finance and Economics in China, which conducts a national survey every two years, starting in 2011.[29] The CHFS collects particularly detailed information about households' demographic characteristics, financial literacy, assets and debt, insurance and social welfare, income and consumption and has a relatively low non-response rate (for example, 10.9% in 2013). The CHFS employed a stratified three-stage probability proportion to size (PPS) random sample design. Taking the first wave as an example, the first stage selected 82 counties (including county-level cities and districts) from 2,585 counties (primary sampling units, or PSUs) from 25 provinces and municipalities in Mainland China. The second stage selected 3 to 4 neighbourhood committees/villages from each of the selected PSUs at the first stage. The third stage selected 20 to 50 households (depending

[29] Gan et al. (2014) use the CHFS dataset to report on Chinese household finance development including household demographics, work characteristics, non-financial assets, financial assets, household debts, insurance, social welfare, expenditure, income and wealth.

on the level of urbanization and economic development) from each of the neighbourhood committees/villages chosen at the previous stage. Every stage of sampling was carried out using the PPS method and weighted by population size.

Although there is a small panel element to the CHFS dataset when focusing on households with at least one child, we treat it as a pooled cross-sectional dataset across all waves completed in 2011, 2013, 2015 and 2017.[30] Specifically, the number of households increases over these years to 8,438, 28,141, 37,289 and 40,011, respectively. As the total sample size is clearly increasing across each wave, we have a very limited number of households (about 600), which responded in all four waves and 4,804 households responded in 2 waves (24.04% of the sample) and 2,288 households responded in 3 waves (14.13% of the sample) after allowing for missing responses.[31] The reason for such a large increase in sample size between waves is that the sampling frame changed over time in order to ensure the national representativeness of the survey. Specifically, the first wave of the survey in 2011 was distributed in 25 provinces (autonomous regions and municipalities directly under the central government), 82 counties (districts and county-level cities), and 320 village committees and residential committees, with a sample size of 8,438 households. The second wave of the survey in 2013 covered 29 provinces, 267 counties, and 1,048 village committees and residential committees, with a sample size of 28,141 households. The second wave of the survey extended the province coverage on the premise of ensuring the national representativeness of the data. In 2015, the third wave covered 29 provinces, 351 counties, and 1,396 village committees and residential committees, with a sample size of 37,289 households. Finally, in 2017, the fourth wave covered 29 provinces, 355 counties, and 1,428 village committees and residential committees, with a sample size of 40,011 households. In sum, the number of sampled areas has increased over time to enhance the national representativeness of the survey. Thus, we treat the CHFS dataset as a pooled cross-sectional one in our econometric analysis, given the small panel element of the data, with standard errors clustered at the household level. It is important to note that our econometric findings are generally unchanged if we use econometric approaches, which

[30] The version of the 2017 CHFS used in this chapter includes information on parents' planning to send their children to study abroad. This is no longer available in the latest public version of the 2017 CHFS, which has removed such information.
[31] The official Household Financial Survey Data Description in 2011 and 2013 states that after the data is collected, the centre appends the tracked households and newly added households, which means that the dataset is essentially an unbalanced panel, but, as indicated above, many households are observed for only a limited number of times over the panel.

allow for the panel component. For completeness, Tables 2.10 to 2.16 in the Appendix present these results for the purposes of comparison.

We focus on the households with at least one child but no more than three children aged under 25 because our study focuses on parent's investment in their dependent children's human capital (66.58% of observations are omitted due to this restriction). In detail, 66.50% of observations have no children and only 0.08% of observations have more than 3 children. In addition, we focus on households with a head of working age, i.e. aged between 20 and 60 (5.55% of observations are omitted due to this restriction). After allowing for missing responses, we have 19,977 households (N) and 32,228 observations (NT) in our pooled dataset. All monetary variables in the 2013, 2015 and 2017 waves are deflated using China's yearly CPI, with the benchmark year 2011=100.

There are two main reasons why this dataset has been chosen. The main one is that other household-level surveys do not include information about whether parents plan to send their children to study abroad, which is the focus of this chapter. For example, the Chinese Household Income Project Survey (CHIP) mainly focuses on household income, consumption, employment and production over four years (1988, 1995, 2002 and 2006) and 10 provinces. This dataset is not suitable for the aim of this chapter because it does not contain information about overseas education. Similarly, the China Health and Nutrition Survey (CHNS) contains information mainly on health and it includes 10 provinces with the study conducted in 1989, 1991, 1993, 1997, 2000, 2004 and 2006. Finally, the Rural Household Survey (RHS) contains information from rural households only and the Urban Household Survey (UHS) focuses on information from urban households only. In contrast, the CHFS allows us to consider both types of household. Finally, the CHFS includes almost all provinces of China, and so is representative of the Chinese population, while the CHIP and CHNS only cover 10 provinces, and the UHS and RHS only include 6 provinces.[32]

2.3.2 The Tobit Model

To estimate the household savings model, we use a Tobit model, where households with no savings are censored at zero. We use the logarithmic transformation of household savings as the dependent

[32] China has 34 provinces in total.

variable in the model. Taking the logarithm of household savings causes a problem because some households have no savings. Therefore, a value of 1 is added to the household savings variable when taking the logarithm. Thus, our dependent variable, i.e. household savings, is defined as the natural logarithm of the total amount of money the household saves in the bank plus 1, which is a stock variable and is also called bank deposits and includes both current and savings accounts, following Lin and Lai (2018).[33] The Tobit estimator was first proposed by Tobin (1958) and it can be defined as follows:

$$y_{it}{}^{*} = X_{it}\beta + \varepsilon_{it}, \quad i = 1, 2, \dots, N \ \ and \ t = 1, 2, 3, \dots, T \tag{2.1}$$

$$y_{it} = y_{it}{}^{*} \ \ if \ \ y_{it}{}^{*} > 0 \tag{2.2}$$

$$y_{it} = 0 \quad if \ \ y_{it}{}^{*} \ \leq 0 \tag{2.3}$$

where NT is the number of observations, X_{it} is a vector of independent variables, β is a vector of unknown coefficients, and ε_{it} is a normally and independently distributed error term with zero mean and constant variance, σ^2. It is assumed that there is an implicit, stochastic index (latent variable) equal to $y_{it}{}^{*}$ which is observed only when positive.

In our pooled cross-sectional specification, we model household savings as follows:

$$Ln(Savings_{it}) = \alpha + \beta_1 Study\ Abroad_{it} + \beta_2 X_{it} + \varepsilon_{it} \tag{2.4}$$

where

$$Ln(Savings_{it}) = Ln(Savings_{it}^{*}) \quad if \ \ Ln(Savings_{it}^{*}) > 0 \tag{2.5}$$

$$Ln(Savings_{it}) = 0 \qquad\qquad\qquad otherwise \tag{2.6}$$

and the stock of savings of household i at time t is given by $Ln(Savings_{it})$ such that $i = 1, 2, \dots, N$ and $t = 2011, 2013, 2015, 2017$. The key explanatory variable $Study\ Abroad_{it}$ is a dummy variable, which equals one if the parents plan to send their children to study abroad in household i at time t based on the survey question *'Do you plan to send your child (children) to study abroad?'* The parameter α is a constant and the coefficient β_1 captures the relationship between the dependent variable,

[33] We focus on household savings rather than the household saving rate because there are some issues regarding the measurement of consumption in the CHFS dataset which means that it is not possible to accurately define the household saving rate.

$Ln(Savings)$, and our key independent variable, $Study\ Abroad$. ε_{it} is a normally and independently distributed error term with zero mean and constant variance, σ^2.

The vector X_{it} contains all control variables generally used in the existing literature on household savings, focusing on the literature on China. Specifically, $No.Pre\ School\ Kids$ denotes the number of dependent children aged below 6 (preschool age) in the household. $No.School\ Age\ Kids$ denotes the number of children aged between 6 and 16 (the compulsory school age) in the household. $No.Continued\ Study$ denotes the number of dependent children aged between 16 and 25, and still studying at school (after 16 years old, in China the child can choose to work or continue studying at school). We assume that parents will support their children at each stage of study including high school, undergraduate and postgraduate education, which is typical in China (see, for example, Lugauer et al., 2019). Age is the age of the household head. The $Education$ variable is classified into six categories: $No\ Schooling$ (the omitted category) is a dummy variable, which equals 1 if the household head never attended school; $Primary\ School$ is a dummy variable, which equals 1 if the highest educational attainment of the household head is primary school; $Junior\ High$ is a dummy variable, which equals 1 if the highest educational attainment of the household head is junior high school; $Senior\ High$ is a dummy variable, which equals 1 if the highest educational attainment of the household head is senior high school or technical school; $College/Bachelor$ is a dummy variable, which equals 1 if the highest educational attainment of the household head is vocational college or a bachelor degree; and $Master/PhD$ is a dummy variable indicating if the highest educational attainment of the household head is a master's degree or PhD.[34,35] $Ln(Income)$ is the natural logarithm of the total amount of disposable annual income of the household.[36] $Rural$ is a dummy variable, which equals 1 if the household resides in a rural area or equals 0 if the household lives in an urban area. As discussed in the previous section, there is a large disparity between rural and urban households in terms of health care, unemployment insurance and pensions (Pan, 2016). $No.Workers$ indicates the number of workers in the household excluding

[34] According to the Chinese Education System, primary school includes 6 years, junior high school includes 3 years, senior high school / specialized secondary school includes 3 years, higher vocational school includes 3 years, Bachelor education includes 4 years, Masters education includes 3 years, PhD education includes 4 years. After junior high school, students can choose a specialized secondary school or senior high school and after graduation, they can choose a higher vocational school or regular university e.g. Bachelor education.

[35] There is no information on the head of household's overseas education in the CHFS.

[36] The CHFS defines household disposable income as: Salary net income after tax; net income from agricultural products after-tax; net income from business after-tax; net income from investment after-tax (rent, stock markets; interest from bank deposits, etc); and net transfer income after-tax (social security, social insurance, annuity, etc.).

the household head. Following Lugauer et al. (2019), who use the following question to capture the risk attitudes of the head of the household, we include a dummy variable to measure the risk attitudes of the household head according to the survey question: *'in which project below do you want to invest most if you have adequate money?'* The answers include: (1) a project with high-risk and high-return (5.19%); (2) a project with slightly high-risk and slightly high-return (5.65%); (3) a project with average risk and return (22.03%); (4) a project with slight risk and return (15.77%); and (5) unwilling to carry any risk (51.36%). We classify the answers (1) and (2) as being the most risk-tolerant households. Hence, the dummy variable $Risk\ Attitudes$ equals 1 if (1) or (2) is chosen, and otherwise equals 0. The dummy variable $Male$ equals 1 if the household head is male. $Self\ Assessed\ Health$ is a 5 point index from 0 to 4, where the index equals 0 if the self-assessed health status of the household head is very poor; it equals 1 if the self-assessed health status of the household head is poor; it equals 2 if the self-assessed health status of the household head is normal; it equals 3 if the self-assessed health status of the household head is good and it equals 4 if the self-assessed health status of the household head is very good.[37] Labour market status is captured by the following four categories: $Self\ Employed$ is a dummy variable, which equals 1 if the household head is self-employed; $Employed$ is a dummy variable, which equals 1 if the household head is an employee; $Farming$ is a dummy variable, which equals 1 if the household head is a farmer and $Not\ Woking$ (the omitted category) is a dummy variable, which equals 1 if the household head is retired, unemployed, volunteering, incapacitated or unwilling to work. In addition, we control for housing tenure: $House\ No\ Mortgage$ is a dummy variable, which equals 1 if the household has a house without a mortgage; $House\ Mortgage$ is a dummy variable, which equals 1 if the household has a house with a mortgage and $House\ Renters$ (the omitted category) is a dummy variable, which equals 1 if the household rents a house. Finally, we also control for regions and years represented by $Region$ and $Year$, respectively. Specifically, China has been separated into seven regions in this chapter: $North\ East$, $North$, $East$, $Central$, $South$, $South\ West$, $North\ West$ (the omitted category) and the data covers four years: 2011, 2013, 2015, 2017 (the omitted category).

[37] The results are unchanged if we replace the index variable $Self\ Assessed\ Health$ with a set of dummy variables.

Table 2.1 provides full definitions and Table 2.2 presents the summary statistics for all variables in our estimation sample. The minimum value of household savings, $Ln(Savings)$, is zero which indicates that some households have no savings (approximately 38%). The number of children studying at school measured by $No.School\ Age\ Kids$ and $No.Continued\ Study$ is greater than the number of preschool children ($No.Pre\ School\ Kids$). Turning to the educational attainment of the household head, 52.14% of household heads have at least junior high school education, which indicates that the nine-year compulsory education system has been successful in China.[38] A large household income disparity exists between households comparing the minimum value and the maximum value, and the standard deviation of income. Approximately 10% of households indicate that they would invest in a project with high-risk and high-return, and a project with slight high-risk and slight high-return, which suggests that most Chinese households are risk-averse, which is in line with findings for U.S. households (see, for example, Avery and Kennickell, 1991). 79.05% of household heads are male. Finally, the proportion of parents planning to send their children to study abroad is 10.96% across all waves, reaching 13.75% in 2013 but then decreasing to 8.19% in 2017. This may be due to the changes in sampling across each wave. We also provide summary statistics for all variables split by $Study\ Abroad$ and find that the households where parents plan to send their children to study abroad have, on average, more household savings and household income than those in which parents do not plan to do so.

Although the main focus of this chapter relates to the determinants of household savings, for robustness, we also estimate the relationship between parents planning to send their children to study abroad and total household financial assets, $Ln(Financial\ Assets)$, which is defined as the natural logarithm of the total amount of household financial assets plus 1, including household savings, stocks, bonds, funds, financial derivatives, banking wealth management products, non-banking wealth management products, foreign currency assets, gold, cash and lending (see Table 2.1). We follow the approach taken by Yin (2011), who uses household financial assets as an alternative dependent variable for a robustness check to explore Chinese household saving behaviour.[39] We also explore the natural

[38] The Law on Nine-Year Compulsory Education, which took effect on July 1st, 1986, established requirements and deadlines for attaining universal education tailored to local conditions and guaranteed school-age children the right to receive at least nine years of education (six years of primary education and three years of secondary education).

[39] Our focus in this chapter lies in analysing the relationship between household assets and parent's plans to send their children to study abroad. Hence, we do not focus on modelling the liabilities side of the household balance sheet separately. Our focus on savings and assets reflects the high savings rate typically observed in China.

logarithm of another dependent variable called household non-savings financial assets, which is defined as household financial assets (excluding household savings) plus 1, represented by $Ln(NS\ Financial\ Assets)$. We explore the determinants of $Ln(NS\ Financial\ Assets)$ in order to shed light on the relationship between parents planning to send children to study abroad and the holding of risky financial assets and high-liquidity assets (i.e. cash and lending). We also use a Tobit estimator to model these alternative dependent variables, i.e. household financial assets and household non-savings financial assets, because they are censored with a lower bound of zero (see Table 2.2 and Figures 2.2 and 2.3).

In addition, household net wealth represented by $Ln(Net\ Wealth)$ is an alternative dependent variable, which is defined as the natural logarithm of total assets minus total debt plus 1 if household total assets are equal to and greater than household total debt, otherwise it is equal to minus one times the natural logarithm of the modulus of total assets minus total debt (see Table 2.1). We explore household net wealth as this variable captures the household's saving behaviour in an alternative way and controls for liabilities, which has attracted attention in the existing literature (see, for example, Yin, 2011). We use an ordinary least squares estimator to model household net wealth because the dependent variable $Ln(Net\ Wealth)$ is continuous and contains both negative and non-negative values (see Table 2.2 and Figure 2.4).

The distributions of the four dependent variables, household savings, household financial assets, household non-savings financial assets and household net wealth, are shown in Figures 2.1, 2.2, 2.3 and 2.4 for our pooled dataset (2011, 2013, 2015, and 2017). It is apparent from Figures 2.1, 2.2 and 2.3 that approximately 38% of households report zero household savings, approximately 3% of households report zero household financial assets and approximately 4% of households report zero non-savings financial assets, respectively. Such a large disparity in the proportion of reported zero household savings and the proportion of reported zero household financial assets is likely to be due to the fact that a large proportion of households in China typically hold cash in their house, rather than in a formal institution like a bank. From Figure 2.4, we can see that household net wealth is a continuous variable. Hence, in the analysis which follows, savings, total financial assets and non-savings financial assets are treated as censored dependent variables, whilst net wealth is modelled as a continuous outcome.[40]

2.3.3 The Lewbel IV Approach

In the Tobit analysis described above, we treat parents planning to send their children to study abroad, $Study\ Abroad$, as an exogenous variable. It may be the case that omitted variables related to $Study\ Abroad$ are correlated with the dependent variable and this endogeneity issue will cause the estimator to be biased. However, it is difficult to find an appropriate instrumental variable for parents planning to send their children to study abroad, i.e. a variable which is related to $Study\ Abroad$ but not related to $Ln(Savings)$, $Ln(Financial\ Assets)$, $Ln(NS\ Financial\ Assets)$ and $Ln(Net\ Wealth)$. Thus, we explore the robustness of our findings using an alternative approach, specifically, the Lewbel IV model proposed by Lewbel (2012). This is used to address the potential endogeneity issue of the main explanatory variable, i.e. $Study\ Abroad$. However, there is a precondition of applying this method, which is the existence of heteroskedasticity in the data. Thus, the Breusch-Pagan test is used, as suggested by Lewbel (2012), to test for heteroskedasticity. The estimation approach taken in this subsection can be summarized as follows (omitting individual i and time subscripts t for brevity):

$$Y_1 = \beta_1 X' + \gamma Y_2 + \varepsilon_1 \quad \varepsilon_1 = \alpha_1 U + V_1 \tag{2.7}$$

$$Y_2 = \beta_2 X' + \varepsilon_2 \qquad \varepsilon_2 = \alpha_2 U + V_2 \tag{2.8}$$

Y_1 is the dependent variable and Y_2 is the independent variable in the above set of equations. U, V_1 and V_2 are unobserved variables that are uncorrelated with X and are conditionally uncorrelated with each other. Here, V_1 and V_2 are idiosyncratic errors in the equations for Y_1 and Y_2, respectively, while U denotes the unobserved factors, which affect both Y_1 and Y_2. In our case, Y_1 is the dependent variable, household savings, Y_2 is the main independent variable, i.e. parents planning to send children to study abroad and the vector X includes all other covariates also used in the Tobit and OLS analysis, as detailed above.

[40] The double hurdle model is an alternative modelling approach, which separates the decision to save and the amount saved. However, due to the difficulties related to finding a suitable variable to identify the decision to save, we focus on the Tobit approach but we do present marginal effects at the intensive and extensive margins as detailed below.

The Lewbel (2012) estimator is calculated by two steps. The first one is to estimate $\hat{\beta}_2$ by an OLS regression of Y_2 on X, and obtain the estimated residuals $\widehat{\varepsilon_2} = Y_2 - \hat{\beta}_2 X'$. Then, the next step is to let Z be defined as some or all elements of X, estimate $\hat{\beta}_1$ and γ by an ordinary linear two stage least squares regression of Y_1 on X, using X and $(Z - \bar{Z})\widehat{\varepsilon_2}$ as instruments, where $\bar{Z}$ is the sample mean of Z. In addition to the standard exogenous X assumptions, the Lewbel (2012) estimator also requires the assumptions that $E(X\varepsilon_1) = 0$, $E(X\varepsilon_2) = 0$, $E(XX')$ is nonsingular, $Cov(Z, \varepsilon_1, \varepsilon_2) = 0$ and $Cov(Z, {\varepsilon_2}^2) \neq 0$, where either $Z = X$ or Z is a subset of the elements of X. In our case, we use all elements of vector X as the instruments in the analysis. However, there exists a potential issue when using this approach in our case. As mentioned above, our main dependent variables, household savings, household financial assets, household non-savings financial assets are censored and this approach is suggested for continuous dependent variables or where the regressor is binary or both the outcome and regressor are binary (Lewbel, 2018). Therefore, the results of this approach may be biased in our case. Hence, this analysis is included for robustness purposes only in order to explore the robustness of the sign of the estimated coefficient of our main explanatory variable.

2.3.4 Censored and Uncensored Quantile Analysis

To further explore the robustness of the findings, we use the censored quantile regression approach to explore the effect of the key explanatory variable, parents planning to send their children to study abroad, at different points of the distribution of the dependent variables (household savings, household financial assets and household non-savings financial assets). In contrast, the Tobit approach only reveals the effect of the study abroad variable at the mean of the respective dependent variable. Specifically, we set deciles at the 10^{th}, 20^{th}, 30^{th}, 40^{th} 50^{th}, 60^{th}, 70^{th}, 80^{th}, and 90^{th} levels. We employ uncensored quantile regression in the case of household net wealth because household net wealth is a continuous variable rather than a censored one.

We briefly introduce the censored quantile and uncensored regression approaches as follows. The censored quantile regression model was introduced by Powell (1984, 1986). The latent variable can be defined as:

$$y_{it}^* = x_{it}'\beta_\theta + u_{\theta_{it}} \text{ with } Quant_\theta(u_{\theta it}|x_{it}) = 0 \quad (2.9)$$

Let $y_{it} = y_{it}^{*}$ if $y_{it}^{*} \leq y_{it}^{0}$ and $y_{it} = y_{it}^{0}$ if $y_{it}^{*} > y_{it}^{0}$, where y_{it}^{0} is the censoring value. Then the conditional quantile of y_{it}, given x_{it}, is given by $Quant_{\theta}(y_{it}|x_{it}) = \min(y_{it}^{0},\ x'_{it}\beta_{\theta})$. y_{it}^{*} is the conditional quantile of household savings and x_{it} is a vector, which contains our key variable of interest, whether parents are planning to send their children to study abroad, and the other covariates as defined above. Similarly, the censored quantile regression approach is also employed in the case of household financial assets and household non-savings financial assets.

The quantile regression model (uncensored) proposed by Koenker and Bassett (1978) can be written as

$$y_{it} = \beta_{\theta}x'_{it} + u_{\theta_{it}} \text{ with } Quant_{\theta}(y_{it}|x_{it}) = \beta_{\theta}x'_{it} \qquad i(= 1, \dots, n), t(= 1, \dots, T) \tag{2.10}$$

Where β_{θ} and x_{it} are $K * 1$ vectors, and $x_{it1} \equiv 1$. $Quant_{\theta}(y_{it}|x_{it})$ denotes the θth conditional quantile of y_{it} given x_{it}. Also, let $f_{u_{\theta_{it}}}(\cdot\,|x_{it})$ denote the density of $u_{\theta_{it}}$ given x_{it}.[41] This approach is used to model net wealth.

2.4 Results

2.4.1 The Tobit Model

The results from estimating the Tobit model are shown in Table 2.3, Table 2.4 and Table 2.5, where the estimated coefficients, and marginal effects at the intensive and extensive margins are presented, respectively, for household savings, household financial assets and household non-savings financial assets. We find a large disparity between the effects at the extensive margin and at the intensive margin. This is not surprising as a marginal effect at the intensive margin relates to the portion of the variation of the explanatory variable that is correlated with the variation of the expected value of the dependent variable conditional on being non-zero. In contrast, the marginal effect at the extensive margin relates to the change in the probability that the dependent variable is greater than zero. We initially focus the discussion on the effects of our independent variable of interest, namely whether parents plan to send their children to study abroad, on household savings. It is apparent from Table 2.3 that the

[41] It follows that $Quant_{\theta}(u_{\theta_{it}}|x_{it}) = 0$. It is assumed that the distribution function of $u_{\theta_{it}}$ given x_{it}, $F_{u_{\theta_{it}}}(\cdot\,|x_{it})$, is continuously differentiable with density function $f_{u_{\theta_{it}}}(\cdot\,|x_{it}) > 0$.

estimated coefficient of the explanatory variable, $Study\ Abroad$, is positive and statistically significant in the case of household savings, which is in accordance with our expectations in that it captures an additional motive for saving. In addition, in the case of total household financial assets and household non-savings financial assets, the estimated coefficient on $Study\ Abroad$ is positive and attains statistical significance at the 1% level for both outcomes.[42]

We now analyse the effects of $Study\ Abroad$ on household savings in more detail by exploring the marginal effects at the extensive margin. It can be seen from Table 2.4 that the effect of whether parents plan to send their children to study abroad at the extensive margin is 0.0149 in the case of household savings. This means that the households where parents plan to send their children to study abroad have a 1.49% higher probability of having non-zero household savings than those in which parents do not plan to do so. Similarly, those households where parents have such plans have a higher probability of holding non-zero total household financial assets and household non-savings financial assets. Specifically, the effects of parents planning to send their children to study abroad on total household financial assets and household non-saving financial assets at the extensive margin are 0.04% and 0.15%, respectively, which are far smaller in magnitude than that on household savings. This indicates that parents planning to have their children educated overseas has a relatively small influence on the probability that households hold non-zero total household financial assets and household non-saving financial assets, which is probably because only a small number of households have zero total household financial assets and household non-saving financial assets (3% and 4%, respectively).

For brevity, we do not discuss all covariates here but instead focus on the marginal effects of total household income at the extensive margin, for purposes of comparison as income has attracted considerable attention in the existing literature on household saving. From Table 2.4, it is apparent that an increase of one percent in household income is associated with a 2.55% increase in the probability that the household holds non-zero household savings. Furthermore, the size of the marginal effect of household income at the extensive margin is slightly greater than that of parents planning to send children to

[42] We discuss the results relating to the other covariates at the end of this subsection.

study abroad on household savings, which indicates that $Study\ Abroad$ plays an important role in determining households saving behaviour as household income has been found to be an important determinant of household saving behaviour (Modigliani and Cao, 2004; Liu and Hu, 2013; Gruber, 2018). In the case of total household financial assets, the marginal effect of household income at the extensive margin is 0.0003, which means that an increase of one percent in household income is associated with a 0.03% increase in the probability that the household holds non-zero total household financial assets. Similarly, we can see that an increase of one percent in household income is associated with a 0.07% increase in the probability that the household holds non-zero household non-savings financial assets, which means that household income has a small influence on the probability that households hold non-zero total financial assets and non-saving financial assets.

Additionally, we shed further light on the determinants of household saving by analysing the marginal effects at the intensive margin, as presented in Table 2.5. The estimated marginal effect of $Study\ Abroad$ in the case of household savings is 0.2165, which indicates that among the households with non-zero household savings, the households in which parents plan to send their children to study abroad have 21.65% more household savings than those where parents do not plan to do so. This positive effect of $Study\ Abroad$ on household savings indicates that planned parental investment in children's human capital, especially in children's education, plays an important role in household saving behaviour. For example, a high proportion of Chinese parents (93%) financially prepare for their children's education through savings and Chinese parents pay more attention to overseas education for their children than some other countries (see, HSBC, 2017). Thus, parents' plans to send their children to study abroad may lead them to accumulate household savings. However, it is important to acknowledge here that the findings in this chapter indicate a positive correlation between $Study\ Abroad$ and $Ln(Savings)$ rather than a causal effect.

As stated above, household income is also an important determinant of household savings and it is interesting to compare the effects of parents planning to send their children to study abroad with that of income at the intensive margin. Specifically, the findings in Table 2.5 suggest that an increase of one percent in household income is associated with a 0.3703% increase in household savings, which is in accordance with our expectations, the higher the household income, the greater are household savings.

Such a finding is not surprising: households with high household income are able to accumulate more household savings than those with low household income, which is consistent with the findings of Chu and Wen (2017), who argue that China's rising aggregate saving rate is mainly due to the household sector and they also find that rich households save at a higher rate.

Turning to the household financial assets regression in Table 2.5, among the households with non-zero household financial assets, the estimate for the marginal effect at the intensive margin on $Study\ Abroad$ is 0.4282, which means that the households where parents plan to have their children educated overseas have, on average, 42.82% more financial assets than those where parents do not plan to do so. The magnitude of the marginal effect at the intensive margin stemming from $Study\ Abroad$ is noticeably higher for household financial assets compared to household savings. This reflects the fact that the stock of household savings is part of household financial assets. We also compare the magnitude of the marginal effect of parents planning to send children to study abroad on household financial assets at the intensive margin with that of household income and we can see from Table 2.5 that the size of the positive effect of $Study\ Abroad$ is higher than that of $Ln(Income)$ on household financial assets. This demonstrates the importance of including the $Study\ Abroad$ variable in the household assets equation.

As indicated by the results in Table 2.5, we can see that the marginal effects of parents planning to send children to study abroad and household income at the intensive margin are characterized by relatively large changes from column 1 to column 2. Specifically, the marginal effect of $Study\ Abroad$ at the intensive margin increases from 21.65% in the case of household savings to 42.82% in the case of total financial assets, while the marginal effect of $Ln(Income)$ at the intensive margin decreases from 0.3703% to 0.3071%. Such an increase in the effect of parents planning to send children to study abroad implies that the positive effect of $Study\ Abroad$ increases when including risky financial assets such as stocks, bonds, funds, financial derivatives, banking wealth management products, non-banking wealth management products, foreign currency assets, gold as well as the high-liquidity financial assets represented by cash and lending. In order to explore this further, we turn to the dependent variable referred to as non-savings household financial assets, which is household financial assets net of household savings. Hence, this dependent variable captures the risky and the high-liquidity part of household financial assets.

Turning to the determinants of household non-savings financial assets, it can be seen from column 3 in Table 2.5 that the marginal effect relating to $Study\ Abroad$ at the intensive margin is still positive and statistically significant and that the size of this positive effect increases to 51.75% in column 3 as compared to 21.65% in column 1. Such a finding suggests that the effect of parental investment in their children's human capital, i.e. parents planning to send their children to study abroad, has a greater effect on household non-savings financial assets than on household savings, which may be due to the fact that household savings are less risky than household non-saving financial assets since household savings refer to bank deposits, which are regarded as risk-free assets in China, which may lead to different motivations behind household savings and household non-savings financial assets. Taking the effect of household income as a comparison, the estimated marginal effect of household income at the intensive margin is 0.2649%, see column 3 in Table 2.5, which is smaller than that in the case of household savings. Such a result suggests that the positive effect of household income is stronger for household savings than for household non-savings financial assets. This finding suggests that when an increase in household income occurs, households are more likely to save rather than invest in risky financial assets.

Turning to the other covariates, for brevity, we only discuss the marginal effects of education at the extensive margin in the case of household savings because it has a relatively larger effect than the other control variables in terms of magnitude (see Table 2.4). We find that the higher is the educational attainment of the household head, the higher is the probability of households having non-zero household savings. Specifically, the households, where the highest educational attainment of the household head is primary school, have a 11.11% higher probability of having non-zero household savings than those households in which the household head never attended school. The households, where the highest educational attainment of the household head is junior school, have a 17.40% higher probability of having non-zero household savings than those households in which the household head never attended school.[43] In the cases of household financial assets and household non-savings financial assets, the marginal effects of the covariates at the extensive margin are generally small or statistically insignificant (e.g. the number of workers and gender have coefficients equal to zero). Thus, we focus on the marginal effects of the control variables at the intensive margin shown in Table 2.5.

In Table 2.5, we firstly discuss the marginal effects at the intensive margin for the control variables conditional on those households with non-zero household savings. Our findings accord with Lugauer (2019) in that those households with more children save less. The control for whether the household resides in a rural area has a large negative effect on household savings, which is in line with Qian (1988), who provides evidence of a substantial difference in saving behaviour between urban and rural sectors. With respect to the education variables, we can see that the educational attainment of the household head is positively related to household savings. In detail, conditional on having non-zero household savings, the households, where the highest education level of the household head is primary school, have 161.18% higher household savings than those households in which household head never attended school, i.e. double the amount of savings. In those households, where the highest education level of the household head is junior school, savings are 252.50% higher than those households in which the household head never attended school. Chamon and Prasad (2010) find a U-shaped age-savings profile for China, where households headed by a young person and those headed by an old person have the highest saving rates. In contrast, we find an inverted U-shaped age-savings profile, where having a young or old household head is negatively related to household savings. This may be because the focus here is on the stock of household savings whereas the aforementioned study looks at the household savings rate (i.e. a flow). We also find that households have 17.08% more household savings if the household is relatively risk tolerant and both the number of workers and the self-assessed health status of the household head are positively associated with household savings. A self-employed head of household has 65.72% more household savings compared to those where the household head is not working, while the household has a lower level of household savings if the household head is a

[43] The household head having never attended school is the omitted group.

farmer relative to households where the head of household is not working.[44] Interestingly, we find that the households living in the $South$ and $East$ regions of China have more household savings than those living in the $North\,West$ region, while residing in the $North\,East$ region is inversely associated with household savings. This accords with our expectations given that the provinces in those four regions are, on average, more prosperous than those belonging to the $North\,West$ region. Finally, we also find that, compared with households responding in 2017, those heads of household responding in 2011, 2013 and 2015 have a statistically significantly lower level of household savings.

We now turn to the marginal effects at the intensive margin for the controls in the case of household financial assets conditional on non-zero household financial assets (see Table 2.5). The number of preschool-aged (below age 6) children has an impact on household financial assets, and for those children aged over 6, each additional school-aged child decreases the total amount of household financial assets by 11.92%. This negative influence also exists for those households with children aged between 16 and 25 who continue studying at school. We also find an inverted U-shaped relationship between the age of the household head and household financial assets and that household financial assets are positively associated with the educational attainment, the self-assessed health status of the household head but are inversely associated with living in a rural area. A self-employed household head is positively associated with total financial assets, whilst conversely employees and farmers, as household heads, are inversely associated with financial assets, compared to those not working. Living in the $East$, $North$, $Central$ and $South$ regions is positively associated with total household financial assets relative to living in the $North\,West$ region, while residing in the $North\,East$ and $South\,West$ regions is inversely associated with total households financial assets. Responding to the survey in 2011 and 2013 is negatively associated with total household financial assets relative to responding in 2017.

Finally, we briefly discuss the marginal effects at the intensive margin for the other covariates in the case of household non-saving financial assets shown in Table 2.5. The higher the educational attainment of the household head, the higher are household non-savings financial assets. Households have

[44] It is important to note that the category of $Not\,Working$ heads of household includes a range of labour market states, including the retired, homemakers and the incapacitated, and only a small proportion of this category being unemployed. Given the sample size, in this chapter, we do not disaggregate this group further.

75.56% more household non-savings financial assets if the head of household is relatively risk tolerant, which accords with intuition. Living in the $East$, $North$, $Central$, and $South$ regions is positively associated with household non-savings financial assets relative to living in the $North\ West$, while residing in the $South\ West$ region is inversely associated with household non-savings financial assets. Responding to the survey in 2011 and 2013 is inversely associated with household non-savings financial assets relative to responding in 2017.

2.4.2 The OLS Model: Household Net Wealth

For robustness, we use the OLS regression approach to estimate the effects of whether parents plan to send their children to study abroad on household net wealth. The Tobit model is not employed in this case because household net wealth is a continuous dependent variable. It is apparent from Table 2.6 that the coefficient of $Study\ Abroad$ is positive and statistically significant and the estimated effect of $Study\ Abroad$ is 0.2254, which means that the households where parents plan to send their children to study abroad have, on average, 22.54% more household net wealth than those in which parents do not plan to do so. Such a finding is in accordance with our expectations that those households planning overseas education for their children have higher net wealth and ties in with the findings from the Tobit analysis. There is only a small disparity between the size of the effects of $Study\ Abroad$ and $Ln(Income)$ on household net wealth, which provides further evidence that parents planning to send children to study abroad is an important determinant of household net wealth.

The findings related to the other explanatory variables are generally in line with those found for the other three dependent variables. For example, the age-net wealth profile is characterised by an inverted U-shape, which means that the youngest and the oldest households have the lowest household net wealth. This may reflect the fact that young households often need time to accumulate wealth and may hold debt at early states of the life cycle and that the older ones have transferred part of wealth to the next generation. The educational attainment of the household head has a large positive influence on household net wealth except in the case where the household head has only attended primary school. It is also apparent that household income, the self-assessed health status of the household head and the number of workers in the household are positively related to household net wealth. In addition, we find that households have more household net wealth if the head of household is relatively risk tolerant or self-employed. Whether the household resides in a rural area has a large negative effect on household net wealth, which reflects the large economic disparity between rural and urban areas, as discussed above. Households living in the $East$, $North$ and $South$ regions have, on average, more household net wealth than those living in the $North\ West$, while the households living in the $North\ East$ have less. Finally, households responding to the survey in 2013 and 2015 have, on average, more household net wealth than those responding in 2017, which may be due to total household debt increasing from 2013 to 2017.

2.4.3 The Lewbel IV Approach

Turning to the estimates from the Lewbel (2012) IV approach used to deal with the potential endogeneity related to our main explanatory variable, i.e. parents planning to send their children to study abroad, the Breusch-Pagan tests (see Table 2.7) for heteroskedasticity for household savings, household financial assets, household non-savings financial assets and household net wealth rejected the null hypothesis of constant variance for the dependent variables and other covariates. This indicates that the precondition of using the Lewbel IV approach is met because this method requires the existence of heteroskedasticity in the data.[45]

We focus the discussion on the effects of the $Study\ Abroad$ variable. The effect on household savings is shown in column 1 of Table 2.7, where the coefficient of $Study\ Abroad$ is statistically significant and positive in the case of household savings, which provides further evidence that households, where parents plan to send their children to study abroad, have more household savings than those who do not plan to do so. In general, the IV estimates for the other variables are consistent with the findings from the Tobit models and OLS regressions discussed in the previous two subsections. Hence, for brevity, we do not discuss them in detail here.

The result of the under-identification test shows that the model is not under-identified and from the weak identification test, we can see that the instruments are strong. The p-value of the over-identification test is 0.4129 (>0.1), so we cannot reject the null hypothesis that the instruments are valid and the model is correctly specified. Therefore, according to the three identification tests above, i.e. the under-

[45] Baum and Schaffer (2020) provide the Lewbel IV method command in Stata, specifically ivreg2h.

identification test, the weak identification test and the over-identification test, we can conclude that the Lewbel IV model can generate appropriate instruments to deal with the endogeneity issue of our key explanatory variable, i.e., $Study\ Abroad$ in the case of the household savings model.

Additionally, the estimated coefficients of $Study\ Abroad$ on household financial assets, household non-savings financial assets and household net wealth are all positive and statistically significant, which is consistent with the results from the Tobit and the OLS analysis. However, the instruments in the household financial assets equation, the household non-savings financial assets equation and the household net wealth equation do not pass the over-identification test, which means that the instruments are invalid. One possibility is that, as discussed above, the Lewbel IV approach is proposed for continuous dependent variables or where the regressor is binary or both the outcome and regressor are binary.

As a basis for comparison, it could be argued that household income is also an endogenous variable in our model. Thus, we repeat the analysis and treat household income as an endogenous variable using the Lewbel approach, with $Study\ Abroad$ assumed exogenous. The results are shown in Table 2.8, where the coefficient of $Ln(Income)$ is statistically significant and positive in all cases except in the case of household non-savings financial assets, which suggests that an increase in household income will not drive households to invest in risky financial assets. The coefficient of $Study\ Abroad$ is still positive and statistically significant in all cases, which in accordance with the previous findings endorses the argument that $Study\ Abroad$ is an important regressor in household savings models. Compared with the effect of household income treated as an exogenous covariate as in Table 2.7, the effect of $Ln(Income)$ in terms of size becomes smaller in the case of household savings, household financial assets and household net wealth when household income is the endogenous independent variable. Finally, it is apparent that all models reject the hypothesis that the instruments are valid through the over-identification test, which means that the Lewbel approach is not correctly specified in all cases when treating household income as endogenous. Hence, the findings in Table 2.8 should be treated with caution and it should be noted that they are presented here for comparison purposes only.

From the results of the Lewbel IV approach, we can conclude that our main independent variable, whether parents plan to send their children to study abroad, is an important determinant of household savings in China, a finding which is robust to treating this variable as exogenous and endogenous. Hence, we go on to further explore the effects of whether parents plan to send their children to study abroad at different parts of the household saving distribution in the next section.

2.4.4 The Uncensored and Censored Quantile Regression Analysis

In order to explore the effect of our main independent variable ($Study\ Abroad$) across the entire distributions of the four dependent variables, we conduct uncensored quantile regression analysis for household net wealth and censored quantile regression analysis for household savings, household financial assets and household non-savings financial assets. For brevity, Table 2.9 only presents the results for $Study\ Abroad$ as well as for $Ln(Income)$ in order to provide a basis for comparison. Household income is selected as a basis for comparison as the existing literature has indicated that it is an important determinant of household savings (see, for example, Kong and Dickinson, 2016; Gruber, 2018). We firstly discuss the censored quantile regression results for household savings, household financial assets and household non-savings financial assets and then turn to the uncensored quantile regression results for household net wealth.

The estimated effects of $Study\ Abroad$ and $Ln(Income)$ are shown in Panel A and Panel B in Table 2.9, respectively. It can be seen from Panel A that the effect of parents planning to send their children to study abroad on household savings is not statistically significant at the 1% level until the 60th percentile (at the 20th percentile it only attains statistical significance at the 10% level), which implies that parents planning to send their children to study abroad has no impact on household savings at the bottom of the savings distribution. In addition, the magnitude of the positive effect of $Study\ Abroad$ is increasing monotonically above the 40th percentile. Specifically, the effect of $Study\ Abroad$ is zero on household savings at the 10th percentile and this effect increases to 31.32% at the highest decile (90th).

Turning to household financial assets and household non-savings financial assets, the results are somewhat different. In both cases, the estimated coefficients of $Study\ Abroad$ are all statistically significant and positive across all deciles. In addition, in the case of household financial assets, the positive effect of $Study\ Abroad$ increases from 34.77% at the 10th percentile to 43.62% at the 20th percentile but then decreases to 40.70% at the 70th percentile and increases to 52.42% at the 90th percentile. In the case of household non-savings financial assets, the positive effect of $Study\ Abroad$ increases from

19.13% at the 10th percentile to 59.22% at the 50th percentile and from the 60th percentile, the effect shows a monotonic increase to 63.55% at the 90th percentile. In the case of household net wealth, the effect of $Study\ Abroad$ is statistically insignificant at the 10th percentile and then shows a monotonic increase in terms of magnitude across deciles.

It is interesting to compare the quantile regression results with those in Table 2.5 and Table 2.6, where the Tobit model and OLS model are used, respectively. To recall, in the previous empirical analysis, the marginal effect at the intensive margin and the coefficient of $Study\ Abroad$ are positive and statistically significant at the 1% significance level for all four dependent variables at the mean (household savings, household financial assets, household non-savings financial assets and household net wealth). We find consistent evidence from the quantile analysis for parents planning to send their children to study abroad for all dependent variables at the higher deciles (specifically above the 60th percentile), while no strong evidence is found in the case of household savings at the lower deciles. These findings indicate the additional information revealed by the quantile analysis. Both the Tobit model and the OLS model reveal the effects of the $Study\ Abroad$ variable at the mean of the dependent variable, whereas the quantile regression allows us to compute the effects across the entire distribution of the dependent variables. Thus, the quantile regression approach provides a comprehensive understanding of the effect of whether parents plan to send their children to study abroad over the entire distributions of household savings, household financial assets, household non-savings financial assets and household net wealth. In summary, the results of the quantile analysis reveal that the effect of $Study\ Abroad$ is not just confined to the mean of the alternative dependent variables.

In addition, it is apparent from Panel B in Table 2.9 that there exists a positive relationship between household income and household savings above the 20th percentile, and that, in contrast to the effect of $Study\ Abroad$, the effects of household income in terms of magnitude diminish monotonically after the 40th percentile. This indicates that household income is less important in determining the amount of household savings at higher deciles. Similarly, this declining, albeit positive effect, of household income also occurs in the case of household financial assets and household net wealth from the 10th percentile.

Although the effect of household income varies across the different deciles for all four dependent variables, the positive and statistically significant effect of household income across all distributions is consistent with the results from the Tobit model and OLS model.

Thus, we can conclude from the uncensored and censored quantile analysis that parents planning to send children to study abroad indeed has different impacts across the distribution of household savings. Specifically, $Study\ Abroad$ is found not to be statistically significant at the 1% level until the 60th percentile of our various dependent variables.

2.5 Conclusion

Using household-level pooled cross-sectional dataset from the CHFS (2011, 2013, 2015 and 2017), this chapter has examined the effect of parents planning to send their children to study abroad on household saving behaviour in China. We have focused on household savings, which is a stock variable measuring the accumulation of savings for a household. For robustness, we have also investigated how parents planning to send their children to study abroad affects household financial assets, household non-savings financial assets and household net wealth.

We have first used the Tobit model in the case of household savings, household financial assets and household non-savings financial assets, and the OLS model in the case of household net wealth. Our findings suggest that households where parents plan to send their children to study abroad hold more household savings, household financial assets, household non-savings financial assets and household net wealth than those who do not plan to do so for their children. In addition, we find that the effect stemming from parents planning to send children to study abroad is greater in magnitude than that of household income in the amount of household savings held, which suggests that parents planning to send their children to study abroad does influence household savings even when household income is controlled for.

The Lewbel approach was then employed to explore the robustness of the findings to dealing with the potential endogeneity problem of our key explanatory variable of interest, i.e. $Study\ Abroad.$ Our results in the case of household savings pass the under-identification test, weak identification test and over-identification test, which endorses the validity of the instruments. This approach provides further evidence supporting a positive relationship between household saving and parents planning to have their children educated overseas. Finally, we have explored the effect of parents planning to send their children to study abroad across the entire distribution of household savings, using the censored quantile regression approach. Our findings from the censored quantile regression analysis suggest that parents planning to send children to study abroad is statistically insignificant at the 1% level for household savings below the 60^{th} percentile and that there is a statistically positive relationship with household savings at and above the 60^{th} percentile.

To summarise, the results reveal a positive relationship between parents planning to send their children to study abroad and household savings, household financial assets, household non-savings financial assets and household net wealth. This relationship is found to be robust to a range of econometric approaches. Hence, this chapter has explored the role of overseas education in household saving behaviour in China and has identified an influence on household saving behaviour, which has been ignored in the existing literature.

The reason why studying abroad is an important area to study for China is that Chinese parents invest significant amounts of money in their children's human capital accumulation, especially in education including overseas education, which can be particularly expensive. With respect to policy implications, our findings suggest that the Chinese government should help to relieve the saving pressure for Chinese households. For example, this could be achieved by providing more funding or subsidies to improve the quality and reputation of domestic higher education. This could disincentivise households from sending children abroad (if domestic tuition fees are less expensive than international fees), and thus leave households with more money for other areas such as housing, raising young children, and insurance against retirement. In detail, house prices in China have been increasing over time, especially in the large cities such as Beijing, Shanghai and Guangzhou. For example, the average property prices have tripled within the last 10 years from 2010 to 2019.[46] In addition, household savings may be used to raise young children in terms of living costs, domestic education and housing or to finance retirement. Moreover, if studying overseas is regarded as beneficial in terms of developing the skills and knowledge

[46] The data source of house prices is: https://www.anjuke.com.

of individuals, providing Government financial support may ultimately lead to a more productive and skilled workforce in the future thereby enhancing economic growth. Alternatively, if one assumes that saving is foregone consumption, reducing the attractiveness of studying abroad may free up more income for consumption, which would increase household welfare.

In terms of shortcomings, this study focuses on how parents planning to send children to study abroad affects household saving behaviour but we do not know whether parents actually send their children to study abroad. Thus, it would be interesting to investigate household saving behaviour among those parents who have sent their children to study abroad and those who have not, i.e. employing panel data to compare parents' plans and their actual behaviour. Unfortunately, as yet, there is no longitudinal data available to explore this. As there is only a small panel element to the data analysed in this chapter, we have to rely on pooled cross-sectional analysis, which means that we cannot control for unobserved heterogeneity. Such issues remain interesting areas for future research subject to data availability.

2.6 Figures

Figure 2.1

Distribution of $Ln(Savings)$

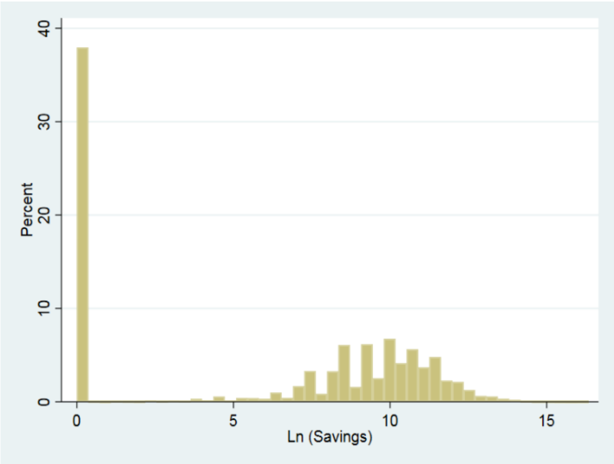


Figure 2.2

Distribution of $Ln(Financial\ Assets)$

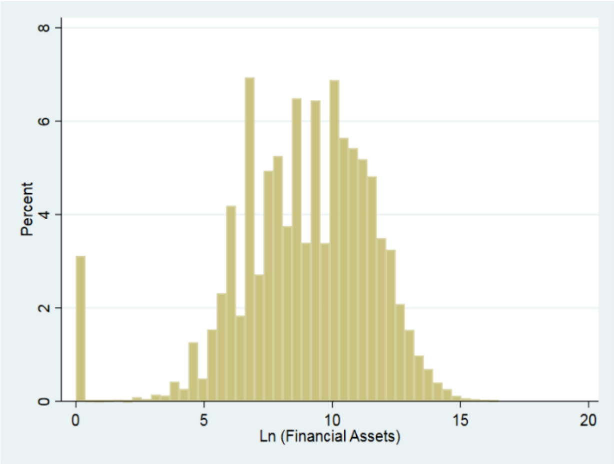


Figure 2.3

Distribution of $Ln(NS\ Financial\ Assets)$

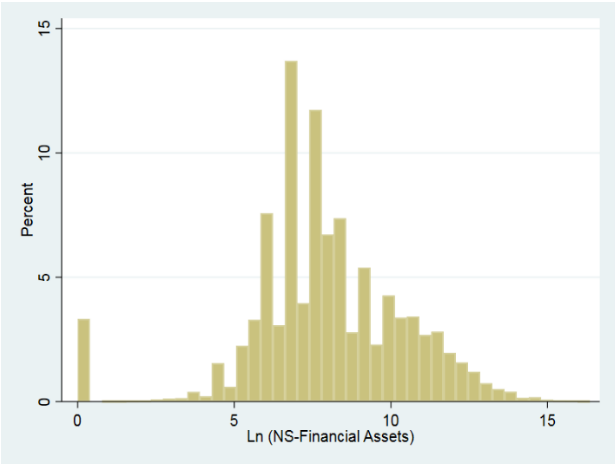
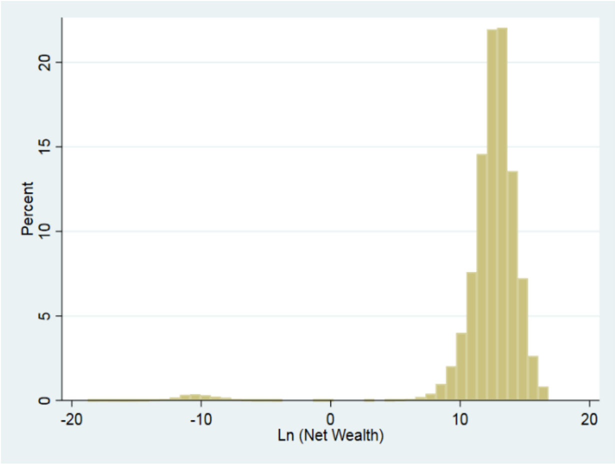


Figure 2.4

Distribution of $Ln(Net\ Wealth)$



2.7 Tables

Table 2.1 Definition of Variables

Variable	Description
Ln(Savings)	Household savings, defined as the natural logarithm of the total amount of money the household saves in the bank plus one, which is a stock variable and is also called bank deposits and includes both current and savings accounts.
Ln(Financial Assets)	Household financial assets, defined as the natural logarithm of the total amount of household financial assets plus one including household savings, stocks, bonds, funds, financial derivatives, banking wealth management products, non-banking wealth management products, foreign currency assets, gold, cash and lending.
Ln(NS Financial Assets)	Household non-savings financial assets, defined as the natural logarithm of the total amount of the household financial assets excluding household savings (bank deposits) plus one, which is called non-savings financial assets.
Ln(Net Wealth)	Household net wealth, defined as the natural logarithm of the total amount of the household net wealth plus one, equals total household assets minus total household debt if assets are equal to and greater than debt, otherwise equals minus one times the natural logarithm of the modulus of total assets minus total debt. Total household assets include agricultural and business assets, land and real estate, vehicles, stocks, financial derivatives, non-RMB assets, gold, funds, bonds and financial wealth-management products, savings and cash etc. Total household debt includes agricultural/business debt, vehicle-purchasing debt, house-purchasing debt, education debt, credit debt and other debt.
Ln(Income)	Household income, defined as the natural logarithm of the total amount of the household disposable annual income.
No. Pre School Kids	Number of Children aged below 6 (preschool age) in the household.
No. School Age Kids	Number of Children aged between 6 and 16 (compulsory school entrance age) in the household.
No. Continued Study	Number of Children aged between 16 and 25, and are still a student in the household (after 16 years old, children can choose to work or continue studying in the school).
No. Workers	Number of workers in the household excluding the household head.
Age	Age of the household head.
Age^2	Age squared of the household head.
Self Assessed Health	A 5-point index for the head of household ranging from 0 to 4 where 0 denotes very poor; 1 denotes poor; 2 denotes normal; 3 denotes good and 4 denotes very good.
No Schooling (Omitted)	Dummy variable (0/1) equals 1 if the household head never attended school.
Primary School	Dummy variable (0/1) equals 1 if the highest education level the household head is primary school.
Junior High	Dummy variable (0/1) equals 1 if the highest education level the household head is junior high school.
Senior High	Dummy variable (0/1) equals 1 if the highest education level the household head is senior high school or technical school.
College/Bachelor	Dummy variable (0/1) equals 1 if the highest education level the household head is vocational college or bachelor.
Master/PhD	Dummy variable (0/1) equals 1 if the highest education level the household head is above bachelor (master or PhD).
Rural	Dummy variable (0/1) equals 1 if the household resides in a rural area, equals 0 if the household resides in an urban area.

[a] All monetary variables in the 2015 and 2017 waves are deflated using China's yearly CPI, with the benchmark year 2011 = 100, year 2013 = 102*102, year 2015 = 102*102*101.4*102 and year 2017 = 102*102*101.4*102*102*101.6.

Table 2.1 Definition of Variables(Continued)

Variable	Description
Risk Attitudes	Dummy variable, which equals 1 if the household is risk-tolerant. Based on the question: 'in which project below do you want to invest most if you have adequate money?' The answers to question include (1) project with high-risk and high-return, (2) project with slightly high-risk and slightly high-return, (3) project with average risk and return, (4) project with slight risk and return, (5) unwilling to carry any risk. The household is risk-tolerant if the household head's answer is (1) or (2), otherwise equals 0.
Male	Dummy variable (0/1) equals 1 if the household head is male.
House No Mortgage	Dummy variable (0/1) equals 1 if the household owns a house without a mortgage
House Mortgage	Dummy variable (0/1) equals 1 if the household owns a house with a mortgage
House Renters (Omitted)	Dummy variable (0/1) equals 1 if the household rents a house.
Self Employed	Dummy variable (0/1) equals 1 if the household head is self-employed.
Employed	Dummy variable (0/1) equals 1 if the household head is an employee, i.e. employed by someone else.
Farming	Dummy variable (0/1) equals 1 if the household head is a farmer.
Not Working (Omitted)	Dummy variable (0/1) equals 1 if the household head is not working including retired, un-employed, volunteer, incapacitated, homemaker, unwilling to work.
Study Abroad	Dummy variable (0/1) equals 1 if parents plan to send their children to study abroad.
North East	Dummy Variable (0/1) equals 1 if the household lives in the northeastern region of China including 3 provinces: Heilongjiang, Jilin, Liaoning.
East	Dummy Variable (0/1) equals 1 if the household lives in the eastern region of China includ-ing 7 provinces: Shanghai, Jiangsu, Zhejiang, Anhui, Fujian, Jiangxi, Shandong.
North	Dummy Variable (0/1) equals 1 if the household lives in the northern region of China includ-ing 5 provinces: Beijing, Tianjin, Shanxi, Hebei, Neimenggu.
Central	Dummy Variable (0/1) equals 1 if the household lives in the central region of China including 3 provinces: Henan, Hubei, Hunan.
South	Dummy Variable (0/1) equals 1 if the household lives in the southern region of China in-cluding 3 provinces: Guangdong, Guangxi, Hainan.
South West	Dummy Variable (0/1) equals 1 if the household lives in the southwestern region of China including 4 provinces: Chongqing, Sichuan, Guizhou, Yunnan.
North West (Omitted)	Dummy Variable (0/1) equals 1 if the household lives in the northwestern region of China including 4 provinces: Shaanxi, Gansu, Qinghai, Ningxia.
2011 Year	Dummy Variable (0/1) equals 1 if the household responded in 2011.
2013 Year	Dummy Variable (0/1) equals 1 if the household responded in 2013.
2015 Year	Dummy Variable (0/1) equals 1 if the household responded in 2015.
2017 Year (Omitted)	Dummy Variable (0/1) equals 1 if the household responded in 2017.
NT	Total number of observations
N	Total number of households
CN	Censored number of observations

[a] All monetary variables in the 2013, 2015 and 2017 waves are deflated using China's yearly CPI, with the benchmark year 2011 = 100, year 2013 = 102*102, year 2015 = 102*102*101.4*102 and year 2017 = 102*102*101.4*102*102*101.6.

Table 2.2 Summary Statistics - All Variables; Cross-section (t=2011, 2013, 2015, 2017)

	Mean	Std. Dev	Min	Max	Study Abroad = 0 Mean	Study Abroad = 1 Mean
Ln(Savings)	6.6045	4.7310	0	16.2562	6.4931	7.5080
Ln(Financial Assets)	9.2345	2.7401	0	16.8493	9.1243	10.1285
Ln(NS Financial Assets)	8.0595	2.7455	0	16.8156	7.9395	9.0332
Ln(Net Wealth)	12.0738	4.1654	- 18.7065	17.0960	11.9750	12.8752
Ln(Income)	10.3519	2.1078	0	15.3399	10.2865	10.8823
Age	41.8136	7.2838	20	60	41.9313	40.8589
Self Assessed Health	2.4177	1.0291	0	4	2.4101	2.4795
No. Pre School Kids	0.2468	0.4347	0	2	0.5417	0.2881
No. School Age Kids	0.6551	0.6844	0	3	0.6580	0.6313
No. Continued Study	0.4288	0.5478	0	2	0.4348	0.3800
No. Workers	0.9805	0.8197	0	5	0.9829	0.9610
No. Schooling	0.0245	0.1577	0	1	0.0261	0.0110
Primary School	0.1633	0.3722	0	1	0.1689	0.1176
Junior High	0.3583	0.4791	0	1	0.3715	0.2508
Senior High	0.2099	0.4076	0	1	0.2117	0.1951
College/Bachelor	0.2276	0.4176	0	1	0.2090	0.3780
Master/PhD	0.0166	0.1250	0	1	0.0128	0.0475
Rural	0.2774	0.4474	0	1	0.2851	0.2146
Risk Attitudes	0.1078	0.3110	0	1	0.0987	0.1815
Male	0.7905	0.4084	0	1	0.7951	0.7526
House No Mortgage	0.7187		0	1	0.7231	0.6828
House Mortgage	0.1051		0	1	0.0988	0.1563
House Renters	0.1762		0	1	0.1781	0.1609
Self Employed	0.1849		0	1	0.1804	0.2217
Employed	0.5450		0	1	0.5434	0.5581
Farming	0.1611		0	1	0.1659	0.1216
Not Working	0.1090		0	1	0.1103	0.0987
Study Abroad	0.1097		0	1		
Observations		32,228			28,691	3,537

[a] The mean of *Study Abroad* is 0.1020, 0.1375, 0.1189 and 0.0819 in 2011, 2013, 2015 and 2017, respectively.

Table 2.3 Tobit Analysis: Household Savings, Financial Assets, NS-Financial Assets and Plans to Send Children to Study Abroad

	Ln(Savings)		Ln(Financial Assets)		Ln(NS-Financial Assets)	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Study Abroad	0.3549***	3.21	0.4305***	9.39	0.5265***	10.52
Age	0.2288***	4.62	0.1458***	7.52	0.1740***	8.78
Age^2	- 0.0029***	- 4.90	- 0.0018***	- 7.52	- 0.0020***	- 8.60
Male	- 0.1163	- 1.23	- 0.0335	- 0.88	- 0.0310	- 0.78
Risk Attitudes	0.2799**	2.54	0.4982***	11.66	0.7688***	16.67
Self Assessed Health	0.4330***	10.84	0.2561***	15.93	0.2170***	13.32
Primary School	2.6416***	7.95	0.7745***	6.08	0.5959***	4.78
Junior High	4.1382***	12.74	1.4621***	11.78	1.1386***	9.33
Senior High	4.9505***	14.95	1.8695***	14.74	1.5569***	12.44
College/Bachelor	5.7341***	17.13	2.5014***	19.41	2.2759***	17.77
Master/PhD	6.6827***	16.79	3.2747***	20.26	3.0536***	16.86
Self Employed	1.0770***	7.10	0.7750***	12.86	0.7785***	12.80
Employed	0.2122	1.57	- 0.0933*	- 1.69	- 0.2297***	- 4.18
Farming	- 0.8092***	- 4.57	- 0.2878***	- 4.15	- 0.2061***	- 3.04
Rural	- 0.9818***	- 8.16	- 0.5140***	- 11.40	- 0.4280***	- 9.82
Ln(Income)	0.6068***	24.63	0.3088***	32.44	0.2695***	28.83
House No Mortgage	0.3229***	3.16	0.2552***	6.48	0.2239***	5.52
House Mortgage	- 0.5233***	- 3.78	- 0.0656	- 1.18	0.1211**	2.04
No. Workers	0.1138**	2.12	0.0184	0.84	0.0022	0.10
No. Pre School Kids	- 0.4503***	- 4.43	- 0.1198***	- 3.10	- 0.0269	- 0.67
No. School Age Kids	- 0.6395***	- 7.72	- 0.2224***	- 6.82	- 0.1249***	- 3.84
No. Continued study	- 0.4419***	- 4.43	- 0.1636***	- 4.12	- 0.1091***	- 2.78
North East	- 0.9786***	- 5.58	- 0.2802***	- 4.01	- 0.0382	- 0.54
East	0.3157**	2.18	0.4501***	7.63	0.5893***	9.58
North	0.0685	0.42	0.1986***	3.05	0.2697***	4.02
Central	0.1805	1.08	0.2122***	3.19	0.2769***	4.03
South	0.6229***	3.99	0.2483***	3.86	0.4026***	5.98
South West	- 0.0846	- 0.50	- 0.2205***	- 3.17	- 0.1544**	- 2.16
2011 Year	- 1.5536***	- 10.69	- 0.4012***	- 7.63	- 0.2598***	- 4.90
2013 Year	- 1.2601***	- 13.45	- 0.2654***	- 7.37	- 0.2663***	- 7.05
2015 Year	- 0.3629***	- 4.50	0.0390	1.20	0.0540	1.55
Constant	- 9.5562***	- 8.88	0.8030*	1.94	- 0.4960	- 1.16
$F\ (31, 32197)$	162.81		282.57		221.58	
Pseudo R^2	0.0364		0.0648		0.0527	
Uncensored obs	22,182		31,245		30,852	
Left censored obs	10,046		983		1,376	
Observations (NT)			32,228			

[a] *, **, *** denote 10, 5, 1 per cent levels of significance, respectively.
[b] Reference category: Region controls, the omitted group is the North West; Year controls, the omitted group is 2017; Education controls: the omitted group is that household head never attended school; Tenure controls: the omitted group is that household rents a house; Employment controls: the omitted group is that household head is not working or doing non-profitable works including retired, unemployed, volunteer, incapacitated, homemaker and unwilling to work.

Table 2.4 Tobit Analysis - Marginal effects at the Extensive Margin (M.E.): Household Savings, Financial Assets, NS-Financial Assets and Plans to Send Children to Study Abroad

	Ln(Savings)		Ln(Financial Assets)		Ln(NS-Financial Assets)	
	M.E.	t-stat	M.E.	t-stat	M.E.	t-stat
Study Abroad	0.0149***	3.21	0.0004***	8.07	0.0015***	9.35
Age	0.0096***	4.62	0.0001***	6.63	0.0005***	7.85
Age^2	- 0.0001***	- 4.90	0.0000***	- 6.62	0.0000***	- 7.71
Male	- 0.0049	- 1.23	0.0000	- 0.88	- 0.0001	- 0.78
Risk Attitudes	0.0118**	2.54	0.0005***	8.93	0.0021***	12.28
Self Assessed Health	0.0182***	10.81	0.0002***	10.13	0.0006***	10.40
Primary School	0.1111***	7.98	0.0007***	5.22	0.0017***	4.48
Junior High	0.1740***	12.83	0.0013***	8.03	0.0032***	7.77
Senior High	0.2081***	15.03	0.0017***	9.10	0.0043***	9.56
College/Bachelor	0.2411***	17.17	0.0023***	10.33	0.0063***	11.93
Master/PhD	0.2810***	16.77	0.0030***	10.77	0.0085***	12.01
Self Employed	0.0453***	7.09	0.0007***	9.17	0.0022***	10.33
Employed	0.0089	1.57	- 0.0001*	- 1.69	- 0.0006***	- 4.10
Farming	- 0.0340***	- 4.58	- 0.0003***	- 3.99	- 0.0006***	- 2.99
Rural	- 0.0413***	- 8.13	- 0.0005***	- 8.57	- 0.0012***	- 8.45
Ln(Income)	0.0255***	24.63	0.0003***	11.31	0.0007***	14.06
House No Mortgage	0.0136***	3.16	0.0002***	5.87	0.0006***	5.29
House Mortgage	- 0.0220***	- 3.78	- 0.0001	- 1.17	0.0003**	2.04
No. Workers	0.0048**	2.12	0.0000	0.84	0.0000	0.10
No. Pre School kids	- 0.0189***	- 4.43	- 0.0001***	- 3.00	- 0.0001	- 0.67
No. School Age Kids	- 0.0269***	- 7.70	- 0.0002***	- 6.00	- 0.0003***	- 3.73
No. Continued Study	- 0.0186***	- 4.43	- 0.0001***	- 3.91	- 0.0003***	- 2.74
North East	- 0.0411***	- 5.58	- 0.0003***	- 3.86	- 0.0001	- 0.54
East	0.0133**	2.18	0.0005***	6.60	0.0016***	8.32
North	0.0029	0.42	0.0004***	2.95	0.0007***	3.88
Central	0.0076	1.08	0.0002***	3.08	0.0008***	3.89
South	0.0262***	3.98	0.0002***	3.65	0.0011***	5.53
South West	- 0.0036	- 0.50	- 0.0002***	- 3.07	- 0.0004**	- 2.15
2011 Year	- 0.0653***	- 10.62	- 0.0004***	- 7.00	- 0.0007***	- 4.89
2013 Year	- 0.0530***	- 13.30	- 0.0002***	- 6.93	- 0.0007***	- 7.01
2015 Year	- 0.0153***	- 4.49	0.0000	1.19	0.0001	1.54

[a] *, **, *** denote 10, 5, 1 per cent levels of significance, respectively.
[b] Reference category: Region controls, the omitted group is the North West; Year controls, the omitted group is 2017; Education controls: the omitted group is that household head never attended school; Tenure controls: the omitted group is that household rents a house; Employment controls: the omitted group is that household head is not working or doing non-profitable works including retired, unemployed, volunteer, incapacitated, homemaker and unwilling to work.

Table 2.5 Tobit Analysis - Marginal Effects at the Intensive Margin (M.E.): Household Savings, Financial Assets, NS-Financial Assets and Plans to Send Children to Study Abroad

	Ln(Savings)		Ln(Financial Assets)		Ln(NS-Financial Assets)	
	M.E.	t-stat	M.E.	t-stat	M.E.	t-stat
Study Abroad	0.2165***	3.21	0.4282***	9.39	0.5175***	10.52
Age	0.1396***	4.62	0.1450***	7.52	0.1710***	8.79
Age^2	- 0.0018***	- 4.89	- 0.0017***	- 7.52	- 0.0020***	- 8.60
Male	- 0.0709	- 1.23	- 0.0333	- 0.88	- 0.0305	- 0.78
Risk Attitudes	0.1708**	2.54	0.4955***	11.66	0.7556***	16.68
Self Assessed Health	0.2642***	10.83	0.2547***	15.94	0.2133***	13.33
Primary School	1.6118***	7.94	0.7703***	6.08	0.5857***	4.78
Junior High	2.5250***	12.70	1.4543***	11.79	1.1191***	9.35
Senior High	3.0206***	14.91	1.8594***	14.75	1.5303***	12.46
College/Bachelor	3.4987***	17.08	2.4880***	19.44	2.2370***	17.80
Master/PhD	4.0775***	16.77	3.2572***	20.28	3.0014***	16.88
Self Employed	0.6572***	7.10	0.7709***	12.87	0.7652***	12.81
Employed	0.1295	1.57	- 0.0928*	- 1.69	- 0.2258***	- 4.18
Farming	- 0.4937***	- 4.57	- 0.2862***	- 4.15	- 0.2026***	- 3.04
Rural	- 0.5990***	- 8.17	- 0.5112***	- 11.40	- 0.4207***	- 9.83
Ln(Income)	0.3703***	24.54	0.3071***	32.57	0.2649***	28.97
House No Mortgage	0.1970***	3.16	0.2538***	6.48	0.2200***	5.52
House Mortgage	- 0.3193***	- 3.78	- 0.0653	- 1.18	0.1190**	2.04
No. Workers	0.0694**	2.12	0.0183	0.84	0.0022	0.10
No. Pre School kids	- 0.2747***	- 4.43	- 0.1192***	- 3.10	- 0.0265	- 0.67
No. School Age Kids	- 0.3903***	- 7.72	- 0.2212***	- 6.82	- 0.1228***	- 3.85
No. Continued Study	- 0.2696***	- 4.43	- 0.1627***	- 4.12	- 0.1072***	- 2.78
North East	- 0.5971***	- 5.58	- 0.2787***	- 4.01	- 0.0375	- 0.54
East	0.1926**	2.18	0.4477***	7.63	0.5792***	9.59
North	0.0418	0.42	0.1975***	3.05	0.2651***	4.02
Central	0.1101	1.08	0.2111***	3.19	0.2722***	4.04
South	0.3801***	3.99	0.2469***	3.86	0.3957***	5.98
South West	- 0.0516	- 0.50	- 0.2193***	- 3.17	- 0.1518**	- 2.16
2011 Year	- 0.9479***	- 10.72	- 0.3991***	- 7.63	- 0.2554***	- 4.90
2013 Year	- 0.7688***	- 13.51	- 0.2640***	- 7.37	- 0.2617***	- 7.05
2015 Year	- 0.2214***	- 4.50	0.0388	1.20	0.0531	1.55

[a] *, **, *** denote 10, 5, 1 per cent levels of significance, respectively.
[b] Reference category: Region controls, the omitted group is the North West; Year controls, the omitted group is 2017; Education controls: the omitted group is that household head never attended school; Tenure controls: the omitted group is that household rents a house; Employment controls: the omitted group is that household head is not working or doing non-profitable works including retired, unemployed, volunteer, incapacitated, home-maker and unwilling to work.

Table 2.6 OLS Regression Analysis: Net Wealth and Plans to Send Children to Study Abroad

	Ln(Net Wealth) OLS Regression	
	Coef.	t-stat
Study Abroad	0.2254***	3.26
Age	0.0734***	2.64
Age^2	- 0.0008**	- 2.35
Male	- 0.0978*	- 1.77
Risk Attitudes	0.2099***	3.02
Self Assessed Health	0.4292***	18.86
Primary School	0.0217	0.15
Junior High	0.6447***	4.49
Senior High	1.0338***	6.93
College/Bachelor	1.5352***	10.03
Master/PhD	2.1523***	9.65
Self Employed	1.0144***	11.88
Employed	0.0361	0.47
Farming	- 0.2435***	- 2.61
Rural	- 1.0897***	- 17.84
Ln(Income)	0.2640***	23.19
House No Mortgage	2.3604***	39.65
House Mortgage	1.9354***	22.79
No. Workers	0.1085***	3.55
No. Pre School kids	- 0.0334	- 0.58
No. School Age Kids	- 0.1474***	- 3.33
No. Continued Study	- 0.3211***	- 5.98
North East	- 0.5039***	- 5.37
East	0.7750***	9.86
North	0.5416***	6.16
Central	0.1204	1.33
South	0.5306***	5.97
South West	0.0932	1.04
2011 Year	0.0040	0.05
2013 Year	0.9050***	15.46
2015 Year	0.4633***	8.58
Constant	3.4722***	5.79
$F(31, 32196)$	216.51	
Prob > F	0.0000	
Adj R^2	0.1717	
observations	32,228	

[a] *, **, *** denote 10, 5, 1 per cent levels of significance, respectively.
[b] Reference category: Region controls, the omitted group is the North West; Year controls, the omitted group is 2017; Education controls: the omitted group is that household head never attended school; Tenure controls: the omitted group is that household rents a house; Employment controls: the omitted group is that household head is not working or doing non-profitable works including retired, unemployed, volunteer, incapacitated, homemaker and unwilling to work.

Table 2.7 Lewbel IV Approach: Household Savings, Financial Assets, NS-Financial Assets, Net Wealth (Study Abroad is the Endogenous Variable)

	Ln(Savings)		Ln(Financial Assets)		Ln(NS-Financial Assets)		Ln(Net Wealth)	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Study Abroad	0.5021***	2.72	0.6305***	6.28	0.7239***	6.98	0.5199***	3.18
Age	0.1714***	5.46	0.1453***	8.51	0.1716***	9.74	0.0743***	2.67
Age^2	- 0.0021***	- 5.79	- 0.0017***	- 8.63	- 0.0020***	- 9.65	- 0.0008**	- 2.39
Male	- 0.0736	- 1.18	- 0.0295	- 0.87	- 0.0261	- 0.75	- 0.0937*	- 1.70
Risk Attitudes	0.2256***	2.85	0.4851***	11.27	0.7519***	16.94	0.1940***	2.77
Self Assessed Health	0.3112***	12.10	0.2515***	17.98	0.2104***	14.59	0.4279***	18.80
Primary School	1.2527***	7.55	0.7256***	8.04	0.5274***	5.67	0.0131	0.09
Junior High	2.2186***	13.66	1.3950***	15.80	1.0483***	11.51	0.6379***	4.44
Senior High	2.8539***	16.92	1.7994***	19.61	1.4622***	15.45	1.0200***	6.83
College/Bachelor	3.5431***	20.34	2.4168***	25.51	2.1665***	22.17	1.4986***	9.72
Master/PhD	4.4062***	17.24	3.1640***	22.76	2.9198***	20.36	2.0776***	9.18
Self Employed	0.8199***	8.49	0.7571***	14.42	0.7568***	13.97	1.0065***	11.78
Employed	0.1258	1.46	- 0.0949**	- 2.02	- 0.2312***	- 4.77	0.0432	0.57
Farming	- 0.5071***	- 4.81	- 0.2883***	- 5.03	- 0.2043***	- 3.45	- 0.2422***	- 2.60
Rural	- 0.7078***	- 10.25	- 0.5048***	- 13.45	- 0.4145***	-10.71	- 1.0929***	- 17.89
Ln(Income)	0.4101***	31.62	0.3021***	42.83	0.2615***	35.94	0.2610***	22.73
House No Mortgage	0.2617***	3.89	0.2511***	6.87	0.2194***	5.82	2.3578***	39.60
House Mortgage	- 0.3569***	- 3.72	- 0.0675	- 1.29	0.1204**	2.23	1.9269***	22.66
No. Workers	0.0697**	2.02	0.1184	0.96	0.0016	0.08	0.1092***	3.58
No. Pre SchoolKids	- 0.3172***	- 4.90	- 0.1184***	- 3.36	- 0.0242	- 0.67	- 0.0370	- 0.65
No. SchoolAgeKids	- 0.4479***	- 8.96	- 0.2200***	- 8.10	- 0.1209***	- 4.31	- 0.1504***	- 3.40
No. ContinuedStudy	- 0.3202***	- 5.27	- 0.1620***	- 4.91	- 0.1067***	- 3.13	- 0.3241***	- 6.03
Cons.	- 3.5188***	- 5.20	0.9692***	2.63	- 0.2339	- 0.62	3.4840***	5.81
Region Controls	Yes		Yes		Yes		Yes	
Year Controls	Yes		Yes		Yes		Yes	
Breusch Test: chi2 (1)	22.65		805.81		12.02		13,279.87	
Prob > chi (1)	0.0000		0.0000		0.0000		0.0000	
F (31, 32196)	228.78		397.84		315.55		216.37	
Prob > F	0.0000		0.0000		0.0000		0.0000	
R^2	0.1806		0.2777		0.2344		0.1720	
Root MSE	4. 282		2.329		2.402		3.790	
Under ID test (P Val)	0.0000		0.0000		0.0000		0.0000	
Weak ID test (F Stat)	233.01		233.01		233.01		233.01	
Over ID test (P Val)	0.4129		0.0004		0.0000		0.0749	
Observations (NT)				32,228				

[a]*, **, *** denote 10, 5, 1 per cent levels of significance, respectively.
[b] Reference category: Region controls, the omitted group is the North West; Year controls, the omitted group is 2017; Education controls: the omitted group is that household head never attended school; Tenure controls: the omitted group is that household rents a house; Employment controls: the omitted group is that household head is not working or doing non-profitable works including retired, unemployed, volunteer, incapacitated, homemaker and unwilling to work.

Table 2.8 Lewbel IV Approach: Household Savings, Financial Assets, NS-Financial Assets, Net Wealth (Ln(Income) is the Endogenous Variable)

	Ln(Savings)		Ln(Financial Assets)		Ln(NS-Financial Assets)		Ln(Net Wealth)	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Ln(Income)	0.2415***	13.30	0.1498***	15.10	0.1100***	10.75	0.1192***	7.42
Age	0.1779***	5.65	0.1510***	8.78	0.1774***	10.00	0.0793***	2.85
Age^2	- 0.0022***	- 5.96	- 0.0018***	- 8.86	- 0.0021***	- 9.86	- 0.0008**	- 2.54
Male	- 0.0935	- 1.50	- 0.0480	- 1.41	- 0.0445	- 1.27	- 0.1126**	- 2.04
Risk Attitudes	0.2747***	3.48	0.5316***	12.34	0.7980***	17.96	0.2435***	3.49
Self Assessed Health	0.3334***	12.91	0.2718***	19.26	0.2306***	15.86	0.4474***	19.58
Primary School	1.3098***	7.88	0.7783***	8.57	0.5798***	6.19	0.0658	0.45
Junior High	2.3249***	14.27	1.4922***	16.77	1.1449***	12.48	0.7316***	5.08
Senior High	3.0160***	17.81	1.9479***	21.05	1.6099***	16.88	1.1645***	7.77
College/Bachelor	3.7903***	21.75	2.6452***	27.79	2.3936***	24.39	1.7264***	11.20
Master/PhD	4.7384***	18.68	3.4742***	25.07	3.2281***	22.60	2.3962***	10.68
Self Employed	0.9776***	10.04	0.9009***	16.93	0.8999***	16.40	1.1444***	13.28
Employed	0.3720***	4.20	0.1270***	2.63	- 0.0102	- 0.21	0.2489***	3.18
Farming	- 0.4603***	- 4.35	- 0.2462***	- 4.26	- 0.1623***	- 2.72	- 0.2030**	- 2.17
Rural	- 0.7390***	- 10.68	- 0.5327***	- 14.09	- 0.4423***	- 11.34	- 1.1179***	- 18.25
Study Abroad	0.3884***	4.95	0.4902***	11.44	0.5861***	13.27	0.2798***	4.03
House No Mortgage	0.2881***	4.27	0.2754***	7.47	0.2436***	6.41	2.3816***	39.91
House Mortgage	- 0.2940***	- 3.05	- 0.0095	- 0.18	0.1780***	3.28	1.9844***	23.29
No. Workers	0.1565***	4.44	0.0965***	5.01	0.0797***	4.02	0.1825***	5.86
No. Pre School Kids	- 0.3169***	- 4.88	- 0.1177***	- 3.32	- 0.0235	- 0.64	- 0.0350	- 0.61
No. School Age Kids	- 0.4636***	- 9.26	- 0.2339***	- 8.55	- 0.1348***	- 4.78	- 0.1623***	- 3.66
No. Continued Study	- 0.3469***	- 5.70	- 0.1859***	- 5.59	- 0.1304***	- 3.81	- 0.3454***	- 6.41
Cons.	- 2.3148***	- 3.38	2.0577***	5.50	0.8494**	2.20	4.5011***	7.43
Region Controls	Yes		Yes		Yes		Yes	
Year Controls	Yes		Yes		Yes		Yes	
Breusch Test: chi2 (1)	22.65		805.81		12.02		13,279.87	
Prob $> chi$ (1)	0.0000		0.0000		0.0000		0.0000	
F (31, 32196)	200.69		341.48		275.72		199.94	
Prob $> F$	0.0000		0.0000		0.0000		0.0000	
R^2	0.1763		0.2672		0.2241		0.1683	
Root MSE	4. 294		2.346		2.418		3.799	
Under ID test (P Val)	0.0000		0.0000		0.0000		0.0000	
Weak ID test (F Stat)	1,088.98		1,088.98		1,088.98		1,088.98	
Over ID test (P Val)	0.0013		0.0000		0.0000		0.0000	
Observations (NT)				32,228				

[a] *, **, *** denote 10, 5, 1 per cent levels of significance, respectively.
[b] Reference category: Region controls, the omitted group is the North West; Year controls, the omitted group is 2017; Education controls: the omitted group is that household head never attended school; Tenure controls: the omitted group is that household rents a house; Employment controls: the omitted group is that household head is not working or doing non-profitable works including retired, unemployed, volunteer, incapacitated, homemaker and unwilling to work.

Table 2.9 Censored Regression: Household Savings, Financial Assets, NS-Financial Assets and Uncensored Quantile Regression: Net Wealth, and Plans to Send Children to Study Abroad

Panel A	Ln(Savings) CQR		Ln(Financial Assets) CQR		Ln(NS-Financial Assets) CQR		Ln(Net Wealth) UQR	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Study Abroad (10th)	0.0000	0.00	0.3477***	4.60	0.1913***	2.92	0.1057	1.24
Study Abroad (20th)	0.2435*	1.92	0.4362***	7.47	0.3766***	9.83	0.1802***	4.59
Study Abroad (30th)	0.1629	1.44	0.4124***	7.28	0.4377***	7.76	0.2329***	7.74
Study Abroad (40th)	0.0323	0.37	0.4233***	7.96	0.5614***	8.78	0.2760***	10.11
Study Abroad (50th)	0.0981	1.41	0.4090***	7.96	0.5922***	9.50	0.3285***	12.39
Study Abroad (60th)	0.1569***	2.64	0.4058***	9.29	0.5749***	9.84	0.3567***	14.19
Study Abroad (70th)	0.1860***	3.78	0.4070***	15.41	0.5979***	10.20	0.3801***	14.85
Study Abroad (80th)	0.2471***	4.85	0.4709***	9.33	0.6050***	9.99	0.4158***	15.44
Study Abroad (90th)	0.3132***	6.60	0.5242***	10.19	0.6355***	9.95	0.4527***	13.60
Region Controls	Yes		Yes		Yes		Yes	
Year Controls	Yes		Yes		Yes		Yes	
Observations (NT)				32,228				

Panel B	Ln(Savings) CQR		Ln(Financial Assets) CQR		Ln(NS-Financial Assets) CQR		Ln(Net Wealth) UQR	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Ln(Income) (10th)	0.0000	0.00	0.7653***	20.86	0.5465***	9.70	0.5446***	38.89
Ln(Income) (20th)	0.0450*	1.76	0.5825***	16.24	0.3521***	11.14	0.3797***	58.77
Ln(Income) (30th)	1.0048***	15.91	0.5661***	16.65	0.3313***	13.27	0.3096***	62.49
Ln(Income) (40th)	1.1143***	31.68	0.5478***	15.80	0.3539***	13.53	0.2717***	60.44
Ln(Income) (50th)	0.9541***	25.29	0.5115***	14.62	0.3698***	13.48	0.2263***	51.85
Ln(Income) (60th)	0.7563***	26.71	0.4633***	14.62	0.3627***	12.72	0.1984***	47.92
Ln(Income) (70th)	0.5871***	26.58	0.4089***	14.40	0.3637***	14.40	0.1727***	40.98
Ln(Income) (80th)	0.4456***	16.46	0.3718***	17.11	0.3400***	13.03	0.1536***	34.63
Ln(Income) (90th)	0.3553***	16.86	0.3234***	15.35	0.3218***	19.56	0.1344***	24.52
Region Controls	Yes		Yes		Yes		Yes	
Year Controls	Yes		Yes		Yes		Yes	
Observations (NT)				32,228				

[a] *, **, *** denote 10, 5, 1 per cent levels of significance, respectively.
[b] Reference category: Region controls, the omitted group is the North West; Year controls, the omitted group is 2017; Education controls: the omitted group is that household head never attended school.
[c] The Wald test to check the significance across deciles is not employed in the case of household savings, household financial assets and household non-savings financial assets because the censored quantile regression is a non-linear model but in household net wealth estimation we find that all the adjacent deciles coefficients on $Study\ Abroad$ are statistically equal above the 50th percentile and the effects of the coefficients on $Study\ Abroad$ are different from the median across each decile except the equality between the 60th and the 50th percentile.
[d] In terms of the equality test on household income in the case of household net wealth, we find that all adjacent deciles coefficients on $Ln\ (Income)$ are statistically different across deciles and all effects on $Ln\ (Income)$ are statistically different to the median.

Appendix to Chapter 2

Appendix: Estimations Allowing for the Panel Structure of the Data

Table 2.10 Random Effects Tobit Analysis: Household Savings, Financial Assets, NS-Financial Assets and Plans to Send Children to Study Abroad

	Ln(Savings) RE Tobit		Ln(Financial Assets) RE Tobit		Ln(NS-Financial Assets) RE Tobit	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Study Abroad	0.3580***	3.22	0.4153***	9.62	0.5045***	11.11
Age	0.2184***	4.63	0.1415***	7.72	0.1705***	8.94
Age^2	- 0.0028***	- 4.97	- 0.0017***	- 7.84	- 0.0020***	- 8.86
Male	- 0.1319	- 1.40	- 0.0315	- 0.85	- 0.0313	- 0.81
Risk Attitudes	0.3111***	2.80	0.4710***	10.94	0.7064***	15.56
Self Assessed Health	0.4206***	11.16	0.2316***	16.00	0.1955***	12.86
Primary School	2.5515***	9.14	0.7575***	7.67	0.5862***	5.70
JuniorHigh	3.9820***	14.52	1.4367***	17.77	1.1345***	11.22
Senior High	4.7896***	16.97	1.8509***	18.34	1.5530***	14.80
College/Bachelor	5.6204***	19.53	2.4937***	24.10	2.2796***	21.19
Master/PhD	6.5207***	16.59	3.2522***	21.75	3.0540***	19.67
Self Employed	0.9775***	6.96	0.6993***	12.96	0.7195***	12.70
Employed	0.1950	1.56	- 0.0758	- 1.58	- 0.2028***	- 4.02
Farming	- 0.7806***	- 4.96	- 0.2376***	- 4.02	- 0.1613***	- 2.60
Rural	- 1.0041***	- 9.24	- 0.5485***	- 12.95	- 0.4600***	- 10.51
Ln(Income)	0.5588***	28.55	0.2765***	38.76	0.2451***	32.67
House No Mortgage	0.2999***	3.00	0.2330***	5.97	0.2085***	5.12
House Mortgage	- 0.4729***	- 3.38	- 0.0525	- 0.96	0.1203**	2.10
No. Workers	0.1293**	2.55	0.0263	1.35	0.0111	0.54
No. Pre School Kids	- 0.4530***	- 4.68	- 0.1088***	- 2.91	- 0.0134	- 0.34
No. School Age Kids	- 0.6407***	- 8.31	- 0.2140***	- 7.22	- 0.1187***	- 3.85
No. Continued study	- 0.4272***	- 4.63	- 0.1537***	- 4.33	- 0.0952***	- 2.57
Constant	- 8.7048***	- 8.50	1.3007***	3.29	-0.1253	- 0.30
Region Controls	Yes		Yes		Yes	
Year Controls	Yes		Yes		Yes	
Wald Chi2 (21)	5,263.07		9,451.23		7,640.55	
Prob $> chi2$	0.0000		0.0000		0.0000	
LR: Chibar2 (01)	882.91		1,294.49		892.66	
Prob $\geq$ chibar2	0.0000		0.0000		0.0000	
ρ; Std. Err.	0.2695	0.0093	0.3083	0.0085	0.2554	0.0087
Left censored obs	10,046		983		1,376	
Observations (NT)			32,228			

[a] *, **, *** denote 10, 5, 1 per cent levels of significance, respectively.
[b] Reference category: Region controls, the omitted group is the North West; Year controls, the omitted group is 2017; Education controls: the omitted group is that household head never attended school; Tenure controls: the omitted group is that household rents a house; Employment controls: the omitted group is that household head is not working or doing non-profitable works including retired, unemployed, volunteer, incapacitated, home-maker and unwilling to work.

Table 2.11 Random Effects Tobit Analysis - Marginal Effects at the Extensive Margin (M.E.): Household Savings, Financial Assets, NS-Financial Assets and Plans to Send Children to Study Abroad

	Ln(Savings) RE Tobit		Ln(Financial Assets) RE Tobit		Ln(NS-Financial Assets) RE Tobit	
	M.E.	t-stat	M.E.	t-stat	M.E.	t-stat
Study Abroad	0.0152***	3.22	0.0003***	8.90	0.0013***	10.36
Age	0.0093***	4.63	0.0001***	7.26	0.0004***	8.47
Age^2	- 0.0001***	- 4.97	0.0000***	- 7.36	0.0000***	- 8.41
Male	- 0.0056	- 1.40	0.0000	- 0.85	0.0001	- 0.81
Risk Attitudes	0.0132***	2.80	0.0004***	9.82	0.0019***	13.58
Self Assessed Health	0.0178***	11.15	0.0002***	12.75	0.0005***	11.51
Primary School	0.1081***	9.18	0.0006***	6.85	0.0015***	5.47
Junior High	0.1688***	14.65	0.0012***	11.10	0.0030***	9.92
Senior High	0.2030***	17.12	0.0015***	12.77	0.0041***	12.36
College/Bachelor	0.2382***	19.66	0.0020***	14.92	0.0060***	15.94
Master/PhD	0.2764***	16.62	0.0027***	14.71	0.0080***	15.58
Self Employed	0.0414***	6.96	0.0006***	11.18	0.0019***	11.52
Employed	0.0083	1.56	- 0.0001	- 1.57	- 0.0005***	- 3.98
Farming	- 0.0331***	- 4.97	- 0.0002***	- 3.93	- 0.0004***	- 2.58
Rural	- 0.0426***	- 9.24	- 0.0004***	- 10.89	- 0.0012***	- 9.64
Ln(Income)	0.0237***	28.83	0.0002***	17.10	0.0006***	19.50
House No Mortgage	0.0127***	3.00	0.0002***	5.77	0.0005***	5.04
House Mortgage	- 0.0200***	- 3.38	0.0000	- 0.96	0.0003**	2.10
No. Workers	0.0055**	2.55	0.0000	1.35	0.0000	0.54
No. Pre School kids	- 0.0192***	- 4.68	- 0.0001***	- 2.87	0.0000	- 0.34
No. School Age Kids	- 0.0272***	- 8.31	- 0.0002***	- 6.76	- 0.0003***	- 3.80
No. Continued Study	- 0.0181***	- 4.63	- 0.0001***	- 4.24	- 0.0002**	- 2.56

[a] *, **, *** denote 10, 5, 1 per cent levels of significance, respectively.
[b] Reference category: Tenure controls: the omitted group is household rents a house; Employment controls: the omitted group is that household head is not working or doing non-profitable works including retired, unemployed, volunteer, incapacitated, homemaker and unwilling to work.

Table 2.12 Random Effects Tobit Analysis - Marginal Effects at the Intensive Margin (M.E.): Household Savings, Financial Assets, NS-Financial Assets and Plans to Send Children to Study Abroad

	Ln(Savings) RE Tobit		Ln(Financial Assets) RE Tobit		Ln(NS-Financial Assets) RE Tobit	
	M.E.	t-stat	M.E.	t-stat	M.E.	t-stat
Study Abroad	0.2179***	3.22	0.4133***	9.62	0.4962***	11.11
Age	0.1329***	4.63	0.1408***	7.72	0.1677***	8.94
Age^2	- 0.0017***	- 4.97	- 0.0017***	- 7.84	- 0.0020***	- 8.86
Male	- 0.0803	- 1.40	0.0314	- 0.85	- 0.0307	- 0.81
Risk Attitudes	0.1893***	2.80	0.4687***	10.94	0.6947***	15.56
Self Assessed Health	0.2560***	11.16	0.2304***	16.00	0.1922***	12.86
Primary School	1.5527***	9.14	0.7538***	7.67	0.5765***	5.70
Junior High	2.4232***	14.51	1.4297***	14.78	1.1158***	11.22
Senior High	2.9147***	16.95	1.8419***	18.35	1.5274***	14.80
College/Bachelor	3.4202***	19.50	2.4816***	24.14	2.2420***	21.19
Master/PhD	3.9681***	16.59	3.2363***	21.76	3.0037***	19.67
Self Employed	0.5949***	6.96	0.6959***	12.96	0.7076***	12.70
Employed	0.1186	1.56	- 0.0755	- 1.58	- 0.1995***	- 4.02
Farming	- 0.4750***	- 4.96	0.2364***	- 4.02	- 0.1587***	- 2.60
Rural	- 0.6110***	- 9.24	- 0.5458***	- 12.96	- 0.4524***	- 10.51
Ln(Income)	0.3400***	28.43	0.2752***	38.85	0.2410***	32.67
House No Mortgage	0.1825***	3.00	0.2319***	5.97	0.2051***	5.12
House Mortgage	- 0.2878***	- 3.38	- 0.0522	- 0.96	0.1183**	2.10
No. Workers	0.0787***	2.55	0.0262	1.35	0.0109	0.54
No. Pre School kids	- 0.2757***	- 4.68	- 0.1082**	- 2.91	- 0.0132	- 0.34
No. School Age Kids	- 0.3899***	- 8.31	- 0.2130***	- 7.22	- 0.1167***	- 3.85
No. Continued Study	- 0.2600***	- 4.63	- 0.1529***	- 4.33	- 0.0936**	- 2.57

[a] *, **, *** denote 10, 5, 1 per cent levels of significance, respectively.
[b] Reference category: Tenure controls: the omitted group is household rents a house; Employment controls: the omitted group is that household head is not working or doing non-profitable works including retired, unemployed, volunteer, incapacitated, homemaker and unwilling to work.

Table 2.13 Random Effects Regression Analysis: Net Wealth and Plans to Send Children to Study Abroad

	Ln(Net Wealth) RE Regression	
	Coef.	t-stat
Study Abroad	0.2185***	3.16
Age	0.0779***	2.68
Age^2	- 0.0008**	- 2.40
Male	- 0.0741	- 1.26
Risk Attitudes	0.2126***	3.07
Self Assessed Health	0.4007***	17.33
Primary School	- 0.0097	- 0.06
Junior High	0.6351***	4.14
Senior High	1.0236***	6.43
College/Bachelor	1.5454***	9.46
Master/PhD	2.1218***	8.97
Self Employed	0.9507***	11.02
Employed	0.0545	0.71
Farming	- 0.1234	- 1.31
Rural	- 1.1840***	- 17.76
Ln(Income)	0.2481***	21.87
House No Mortgage	2.2982***	37.03
House Mortgage	1.8887***	21.66
No. Workers	0.1085***	3.48
No. Pre School Kids	- 0.0027	- 0.05
No. School Age Kids	- 0.1310***	- 2.79
No. Continued Study	- 0.3112***	- 5.52
Constant	3.6480***	5.82
Region Controls	Yes	
Year Controls	Yes	
Wald Chi2 (21)	5,624.97	
Prob $> chi2$	0.0000	
ρ	0.2563	
No. Groups	19,977	
observations	32,228	

[a] *, **, *** denote 10, 5, 1 per cent levels of significance, respectively.
[b] Reference category: Region controls, the omitted group is the North West; Year controls, the omitted group is 2017; Education controls: the omitted group is that household head never attended school; Tenure controls: the omitted group is that household rents a house; Employment controls: the omitted group is that household head is not working or doing non-profitable works including retired, unemployed, volunteer, incapacitated, homemaker and unwilling to work.

Table 2.14 Fixed Effects Lewbel IV Approach Model: Household Savings, Financial Assets, NS-Financial Assets, Net Wealth (Study Abroad is the Endogenous Variable)

	Ln(Savings) FE Lewbel		Ln(Financial Assets) FE Lewbel		Ln(NS-Financial Assets) FE Lewbel		Ln(Net Wealth) FE Lewbel	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Study Abroad	0.5119**	12.08	0.7923***	6.94	0.9604***	8.15	0.6296***	3.39
Age	0.1646***	5.44	0.1429***	8.32	0.1696***	9.57	0.0677**	2.42
Age^2	- 0.0020***	- 5.66	- 0.0017***	-8.35	- 0.0020***	- 9.38	- 0.0007**	-2.04
Male	- 0.0747	- 1.51	- 0.0165	-0.48	- 0.0098	-0.28	- 0.0950*	-1.72
Risk Attitudes	0.2377***	3.66	0.4768***	10.99	0.7402***	16.53	0.1927***	2.73
Self Assessed Health	0.3282***	13.70	0.2727***	19.48	0.2351***	16.27	0.4603***	20.21
Primary School	1.2404***	7.86	0.7345***	8.13	0.5626***	6.04	0.0513	0.35
Junior High	2.2083***	14.32	1.4250***	16.15	1.1108***	12.20	0.6810***	4.74
Senior High	2.8687***	18.02	1.8417***	20.08	1.5333***	16.20	1.0778***	7.22
College/Bachelor	3.5621***	22.00	2.4331***	25.62	2.2041***	22.49	1.5480***	10.02
Master/PhD	4.4595***	19.01	3.1930***	22.84	2.9547***	20.48	2.1842***	9.61
Self Employed	0.8615***	10.53	0.7832***	14.83	0.7813***	14.34	1.0531	12.26
Employed	0.1480*	4.54	- 0.0755	-1.60	- 0.2117***	-4.34	0.0770***	1.00
Farming	- 0.5719***	- 5.00	- 0.3549***	-6.17	- 0.2638***	-4.44	- 0.3396***	-3.63
Rural	- 0.7079***	- 10.65	- 0.5030***	-13.33	- 0.4183***	-10.74	- 1.0829***	-17.64
Ln(Income)	0.4137***	4.69	0.3074***	43.32	0.2666***	36.41	0.2695***	23.36
House No Mortgage	0.2297***	3.83	0.2368***	6.44	0.2043***	5.39	2.3145***	38.73
House Mortgage	- 0.3949***	3.40	- 0.1068**	-1.96	0.0847	1.56	1.8884***	22.12
No. Workers	0.0979***	5.31	0.0243	1.29	0.0051	0.26	0.1278***	4.18
No. Pre School Kids	- 0.2324***	- 3.67	- 0.0950***	-2.71	- 0.0129	-0.36	0.0118	0.21
No. School Age Kids	- 0.3710***	- 7.94	- 0.2006***	-6.44	- 0.1105***	-3.99	- 0.1084***	-2.48
No. Continued Study	- 0.2506***	- 4.69	- 0.1465***	-3.95	- 0.1011***	-2.98	- 0.2873***	-5.38
Breusch Test: chi2 (1)	21.37		844.82		15.53		11,176.53	
Prob $>$ chi (1)	0.0000		0.0000		0.0000		0.0000	
F $(22, 32202)$	286.06		518.05		410.51		283.13	
Prob $> F$	0.0000		0.0000		0.0000		0.0000	
R^2	0.1635		0.2608		0.2182		0.1612	
Root MSE	4. 293		2.344		2.419		3.812	
Under ID test (P Val)	0.0000		0.0000		0.0000		0.0000	
Weak ID test (F Stat)	249.29		249.29		249.29		249.29	
Over ID test (P Val)	0.2534		0.0015		0.0019		0.0670	
Observations (NT)				32,228				

[a] *, **, *** denote 10, 5, 1 per cent levels of significance, respectively.
[b] Reference category: Education controls: the omitted group is that household head never attended school; Tenure controls: the omitted group is household rents a house; Employment controls: the omitted group is that household head is not working or doing non-profitable works including retired, unemployed, volunteer, incapacitated, homemaker and unwilling to work.
[c] The Lewbel approach can be used on panel data using the within the transformation of a fixed effects model.

Table 2.15 Fixed Effects Lewbel IV Approach Model: Household Savings, Financial Assets, NS-Financial Assets, Net Wealth (Ln(Income) is the Endogenous Variable)

	Ln(Savings) FE Lewbel		Ln(Financial Assets) FE Lewbel		Ln(NS-Financial Assets) FE Lewbel		Ln(Net Wealth) FE Lewbel	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Ln(Income)	0.2381***	12.08	0.1526***	14.13	0.1117***	10.03	0.1607***	9.20
Age	0.1718***	5.44	0.1487***	8.60	0.1752***	9.83	0.0713**	2.55
Age^2	- 0.0021***	- 5.66	- 0.0018***	- 8.59	- 0.0020***	- 9.60	- 0.0007**	- 2.15
Male	- 0.0943	- 1.51	- 0.0370	- 1.08	- 0.0315	- 0.89	- 0.1120**	- 2.05
Risk Attitudes	0.2893***	3.66	0.5335***	12.31	0.8010***	17.93	0.2418***	3.45
Self Assessed Health	0.3535***	13.70	0.2959***	20.94	0.2586***	17.75	0.4774***	20.88
Primary School	1.3043***	7.86	0.7974***	8.77	0.6279***	6.70	0.1009	0.69
Junior High	2.3234***	14.32	1.5323***	17.23	1.2203***	13.31	0.7612***	5.29
Senior High	3.0425***	18.02	2.0059***	21.68	1.7017***	17.84	1.2023***	8.04
College/Bachelor	3.8240***	22.00	2.6915***	28.25	2.4727***	25.17	1.7523***	11.37
Master/PhD	4.8131***	19.01	3.5589***	25.64	3.3409***	23.35	2.4863***	11.08
Self Employed	1.0295***	10.53	0.9381***	17.50	0.9388***	16.99	1.1677***	13.47
Employed	0.4064***	4.54	0.1504***	3.07	0.0136	0.27	0.2342***	2.96
Farming	- 0.5282***	- 5.00	- 0.3162***	- 5.46	- 0.2250***	- 3.77	- 0.3123***	- 3.33
Rural	- 0.7386***	- 10.65	- 0.5283***	- 13.90	- 0.4430***	- 11.31	- 1.0992***	- 17.89
Study Abroad	0.3688***	4.69	0.4679***	10.86	0.5627***	12.67	0.2381***	3.42
House No Mortgage	0.2584***	3.83	0.2642***	7.14	0.2325***	6.10	2.3355***	39.04
House Mortgage	- 0.3274***	3.40	- 0.0369	- 0.70	0.1530***	2.81	1.9400***	22.71
No. Workers	0.1875***	5.31	0.1037***	5.36	0.0847***	4.24	0.1841***	5.88
No. Pre School Kids	- 0.2363***	- 3.67	- 0.0962***	- 2.73	- 0.0132	- 0.36	0.0129	0.23
No. School Age Kids	- 0.3920***	- 7.94	- 0.2173***	- 8.03	- 0.1267***	- 4.54	- 0.1188***	- 2.72
No. Continued Study	- 0.2831***	- 4.69	- 0.1736***	- 5.24	- 0.1277***	- 3.74	- 0.3051***	- 5.70
Breusch Test: chi2 (1)	21.37		844.82		15.53		11,176.53	
Prob $> chi$ (1)	0.0000		0.0000		0.0000		0.0000	
F (22, 32202)	244.21		434.13		350.31		260.20	
Prob $> F$	0.0000		0.0000		0.0000		0.0000	
R^2	0.1587		0.2509		0.2091		0.1596	
Root MSE	4. 305		2.360		2.433		3.816	
Under ID test (P Val)	0.0000		0.0000		0.0000		0.0000	
Weak ID test (F Stat)	1,145.06		1,145.06		1,145.06		1,145.06	
Over ID test (P Val)	0.0138		0.0000		0.0000		0.0000	
Observations (NT)	32,228							

[a] *, **, *** denote 10, 5, 1 per cent levels of significance, respectively.
[b] Reference category: Education controls: the omitted group is that household head never attended school; Tenure controls: the omitted group is household rents a house; Employment controls: the omitted group is that household head is not working or doing non-profitable works including retired, unemployed, volunteer, incapacitated, homemaker and unwilling to work.
[c] The Lewbel approach can be used on panel data using the within the transformation of a fixed effects model.

Table 2.16 Censored Regression: Household Savings, Financial Assets, NS-Financial Assets and Uncensored Quantile Regression: Net Wealth, and Plans to Send Children to Study Abroad

Panel A	Ln(Savings) CQR		Ln(Financial Assets) CQR		Ln(NS-Financial Assets) CQR		Ln(Net Wealth) QR	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Study Abroad (10th)	0.0000	0.00	0.3939***	6.31	0.2203***	3.14	0.0698	0.54
Study Abroad (20th)	0.0085	0.23	0.4185***	5.97	0.3533***	7.14	0.0612	0.84
Study Abroad (30th)	0.1198	1.23	0.4196***	6.76	0.4318***	8.03	0.0643	1.17
Study Abroad (40th)	0.1006	1.06	0.4048***	8.35	0.5721***	9.59	0.0723*	1.86
Study Abroad (50th)	0.1877**	2.50	0.3953***	8.59	0.5635***	10.78	0.0792**	2.05
Study Abroad (60th)	0.1868***	3.90	0.3413***	9.39	0.5398***	9.31	0.0828**	2.32
Study Abroad (70th)	0.1940***	3.93	0.3080***	8.15	0.5098***	8.66	0.0905**	2.33
Study Abroad (80th)	0.1914***	3.64	0.3199***	7.82	0.4627***	9.53	0.1749***	4.06
Study Abroad (90th)	0.2454***	5.98	0.3664***	7.77	0.4601***	7.64	0.1714***	3.36
Region Controls	Yes		Yes		Yes		Yes	
Year Controls	Yes		Yes		Yes		Yes	
Mundlak Corr.	Yes		Yes		Yes		Yes	
Observations (NT)	32,228							

Panel B	Ln (Savings) CQR		Ln (Financial Assets) CQR		Ln (NS-Financial Assets) CQR		Ln (Net Wealth) QR	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Ln(Income) (10th)	0.0000	0.00	0.8040***	16.22	0.5442***	11.17	0.2294***	5.40
Ln(Income) (20th)	-0.0091	-0.13	0.6721***	18.13	0.4022***	14.40	0.1722***	7.34
Ln(Income) (30th)	1.0967***	11.82	0.6632***	20.90	0.3612***	15.44	0.1447***	8.61
Ln(Income) (40th)	1.3977***	22.82	0.6675***	20.92	0.3755***	15.76	0.1272***	10.37
Ln(Income) (50th)	1.1646***	14.65	0.6252***	24.88	0.4037***	18.04	0.1143***	8.87
Ln(Income) (60th)	0.8239***	18.94	0.6009***	23.21	0.4510***	18.13	0.1043***	9.52
Ln(Income) (70th)	0.6242***	20.99	0.5352***	19.30	0.4911***	18.37	0.0935***	10.89
Ln(Income) (80th)	0.4971***	15.95	0.4469***	16.83	0.4819***	18.16	0.0844***	5.99
Ln(Income) (90th)	0.3648***	13.39	0.3250***	15.77	0.4131***	16.83	0.0744***	4.28
Region Controls	Yes		Yes		Yes		Yes	
Year Controls	Yes		Yes		Yes		Yes	
Mundlak Corr.	Yes		Yes		Yes		Yes	
Observations (NT)	32,228							

[a] *, **, *** denote 10, 5, 1 per cent levels of significance, respectively.
[b] Reference category: Region controls, the omitted group is the North West; Year controls, the omitted group is 2017; Education controls: the omitted group is household head never attended school.
[c] We use the Mundlak correction as the proxy of fixed-effects in the censored quantile regression (CQR) model for panel data, which includes the mean of household head age and the mean of household annual total income because there is no estimator in CQR for panel data.

Chapter 3

Financial Literacy and Risky Asset Holding:

Evidence from Chinese Households

3.1 Introduction

Household risky asset holding has attracted considerable attention among both academics and policymakers since the finance industry is an important component of economic development (Liao et al., 2017). Participation in financial markets involving the holding of risky assets plays an important role in household finances and it is important for households to hold risky assets as part of their portfolios as holding risky assets may help households to avoid shortfalls in retirement income or protect household members' from catastrophic financial shocks – such as the inability to maintain mortgage payments due to the death of a breadwinner (Alzuabi et al., 2021).[47] However, there exists a discord between theory and observed risky asset holding. For example, Haliassos and Bertaut (1995) investigate the explanation for the stockholding puzzle in the U.S., i.e. given the equity premium that has prevailed historically, why do households place such a small proportion of their wealth in stocks. They find that labour income risk is inversely related to holding risky assets. Similarly, Bertaut (1998) provides further explanation for the stockholding puzzle in that households perceive the information required for participation to be costly relative to the benefits received. Factors such as increased risk aversion, income risk, and lower resources can reduce the utility gains from stock market participation. However, the effect of labour income uncertainty has been found to be too small to explain the stockholding puzzle (Heaton and Lucas, 1997; Gollier and Pratt, 1996). Fratantoni (2001) finds that another form of uncertainty associated with homeownership is a promising explanation for the stockholding puzzle.

In China, the stock market has been an important subject for mainstream research over the past decade because of the rapid development of the stock market (Carpenter and Whitelaw, 2017). To be specific, the total capitalization of the Shanghai Stock Exchange and Shenzhen Stock Exchange reached $8,742 billion in 2017, which internationally is only lower than the U.S. New York Stock Exchange and Nasdaq, which reached $32,120 billion in 2017 (Banks around the World, 2017). An increasing number of risky assets have become available to Chinese households due to financial inno-

[47] In Alzuabi et al. (2021), the high-risk assets include directly held stocks; stock mutual funds; risky retirement accounts; risky annuities; trust funds; mortgage bonds; corporate and foreign bonds. Medium-risk assets include safe retirement funds; state and local bonds; tax-free bonds and life insurance policies. Therefore, although high-risk assets are potentially more beneficial they are riskier as well, which means holding high-risk assets may not help households gain sufficient funds but could lead to a loss of wealth.

vation leading to products such as stocks, funds, bonds, financial derivatives, financial wealth-management products, gold and non-RMB assets becoming increasingly available over the past decade in China (Liao et. al, 2017). Nevertheless, a relatively low rate of household risky asset holding is still observed in China. For example, only 8.64% of households hold stocks in 2017 according to the 2017 China Household Finance Survey (CHFS). In contrast, about 13.9% of households in the U.S. report holding stocks in the 2016 U.S. Survey of Consumer Finances (SCF). In a similar vein, approximately 12% of U.K. households held shares between April 2016 and March 2018, as reported in the U.K. Wealth and Assets Survey (WAS), which indicates that risky asset holding is relatively low in China compared to other countries.

This may be due to a lack of financial literacy at the household level in China: A relatively low level of financial literacy amongst household heads has been found in China. Specifically, the proportion of heads of household obtaining a high score of at least six correct answers from a total of eight financial literacy questions in China is 37% in the 2014 China Survey of Consumer Finances (Liao et al., 2017). Furthermore, this proportion is low relative to most OECD countries, as surveyed by Atkinson (2012) in 2010, where close to, or over, half of the heads of household correctly answered at least six questions in, for example, Armenia (46%), the Czech Republic (57%), Estonia (61%), Germany (58%), Ireland (60%), Malaysia (51%), Peru (49%) and the U.K. (53%). The eight financial literacy questions in Liao et al. (2017) and Atkinson (2012) are the same and are designed to test financial knowledge relating to division, the time-value of money, the interest paid on a loan, the calculation of interest plus principal, compound interest, risk and return, inflation and risk diversification.

The level of financial literacy of the head of household has been found to be a key determinant of household financial market participation in both developed countries and China. For example, Van Rooij et al. (2011) find that financial literacy is positively associated with stock market participation using a household-level cross-sectional dataset from the De Nederlandsche Bank's Household Survey (DHS) for the Dutch population in 2005. Similarly, Thomas and Spataro (2018) provide micro-level evidence supporting a positive association between financial literacy and stock market participation using the individual-level dataset from waves 3 and 4 of the Survey of Health, Aging and Retirement in Europe

(SHARE) for nine countries (Austria, Belgium, Denmark, Germany, Italy, France, Switzerland, Sweden, and the Netherlands).

For China, Liao et al. (2017) find a positive association between the level of financial literacy of the household head and household financial market participation using cross-sectional household-level data from the 2014 Survey of Consumer Finances (SCF). Similarly, Zou and Deng (2019) provide evidence that the level of financial literacy of the household head is positively associated with the probability of household financial market participation using a household-level dataset from the 2012 Survey of Consumer Finances (SCF) in Urban China. However, most of the previous research in this area focuses on cross-sectional analysis of the effects of financial literacy, with the effects of financial illiteracy being largely ignored in the existing literature on China.

In addition to financial literacy, there are a number of determinants, which account for risky asset holding. For example, Alzuabi et al. (2021) explore how a household's exposure to background risk affects household portfolio allocation using a household-level cross-sectional dataset from the SCF between 1995 to 2019. In financial markets, individuals simultaneously face risks from various sources, some of which are controllable while others exogenous. These exogenous risks, which make individuals behave more cautiously in their risk-taking, are typically called background risk (Eichner and Wagener, 2012). In other words, background risks are defined as those unavoidable exogenous nonfinancial market risks because they are part of the environment where decisions are taken (see Guiso and Paiella, 2008; Jiang et al., 2010; Fagereng et al., 2016; Brown et al., 2021). In the existing literature, such risks are captured by a range of variables such as risks associated with income, health, business ownership, wealth shocks associated with future inheritance and whether there are multiple earners in the household. For example, Alzuabi et al. (2021) find that the head of household knowing their income for the following year is positively associated with the proportion of high-risk assets held while the households who have no health insurance cover hold a lower level of high-risk assets. Moreover, the households who expect a major financial expense in the next five years or those who anticipate a substantial inheritance, hold a higher proportion of high-risk assets. Similarly, Palia et al. (2014) find an inverse association between background risk and the likelihood of stock market participation. Campbell (2006) stresses the importance of the existence of non-tradable assets (human capital) and illiquid assets (owner-occupied house) in determining a household's asset allocation. In short, these determinants fall into several broad categories: wealth, genetic factors, cognitive ability, financial literacy, trust, the effect of past experience and reinforcement learning, inertia and information in frictions (Badarinza et al., 2016).

This chapter uses the CHFS, as used in the previous chapter, but only includes waves 2013, 2015 and 2017 since there is no information on financial literacy provided in wave 2011. The number of households surveyed in each of the years is as follows: 28,141, 37,289 and 40,011. To explore the robustness of the findings, a range of econometric approaches are employed in this chapter including the Logit model, the Tobit model for both cross-sectional and panel analysis, and the IV Probit model for both cross-sectional and pooled analysis. To be specific, the Logit model is employed to explore the association between financial literacy and household risky asset holding since our main dependent variable is measured by the probability of risky asset holding, which is a binary outcome. In contrast to the existing literature, we split risky assets into high-risk assets and low-risk assets to explore how financial literacy is associated with the likelihood of holding high-risk and low-risk assets. The Tobit estimator is used to investigate how financial literacy impacts the log level of risky assets, which is a censored outcome and in a similar vein, we also model the share of risky assets to total household financial assets.[48] Again, we also investigate the relationship between financial literacy and the amount (and share) of risky assets split into high and low risk categories. Regarding the potential endogeneity of financial literacy, the IV Probit estimator is employed as well. Additionally, we explore the association between household risky asset holding and financial illiteracy.

We find that households, where the head has a higher level of financial literacy, are more likely to hold risky assets based on the cross-sectional Logit analysis, which is in accordance with the existing literature (see, for example, Liao, 2017). In addition, this chapter contributes to the literature by identifying a positive association between financial literacy and household risky asset holding after controlling for unobserved heterogeneity using panel data analysis. In contrast, the existing literature is based upon cross sectional evidence and so potentially suffers from omitted variable bias. The Tobit results

[48] Total household financial assets include both risky assets such as stocks, bonds, funds, financial derivatives (futures, forwards and options), financial wealth-management products (bank financial products, collections of financial brokerages and trusts), non-RMB denominated assets and gold and risk-free assets (such as bank savings and cash).

reveal that financial literacy is positively associated with both the log level of risky assets and the share of risky assets to total household financial assets. Finally, an inverse relationship between financial illiteracy and household risky asset holding is found in this chapter, where financial illiteracy has surprisingly attracted very little attention in the relatively small yet expanding literature on Chinese household risky asset holding.

The rest of the chapter is organized as follows. Section 3.2 reviews the relevant existing research. Section 3.3 describes the data and methodology. Section 3.4 discusses the empirical results and conclusions are presented in Section 3.5.

3.2 Literature Review

This section reviews the relevant literature on financial literacy and household risky asset holding and compares the findings for developed countries and China. Most of the existing research focuses on the relationship between financial literacy and financial market participation (see, for example, Van Rooij et al., 2011; Arrondel et al., 2015; Thomas and Spataro; 2018; Xia et al., 2014; Liao et al., 2017; Zou and Deng, 2019; Pan et al., 2020; Hsiao and Tsai, 2018) and the determinants of household financial market participation (see, for example, Cardak and Wilkins, 2009; Yao and Xu, 2015). The rest of this section discusses each of the above areas in turn focusing firstly on the U.S. and other developed countries and then on China.

3.2.1 The U.S. and other Developed Countries

There has been extensive research on financial literacy in many countries, especially in the U.S. and other developed countries. For example, Lusardi and Mitchell (2008) devised a module on financial literacy for the 2004 U.S. Health and Retirement Study (HRS). The module includes questions related to interest compounding, the effects of inflation, and risk diversification to test basic financial knowledge among older Americans. They find that financial illiteracy is common among the sample aged 50 and over. Specifically, only half of the respondents could answer questions about interest compounding and inflation correctly, and only one third could answer all questions correctly. Similarly, Hogarth and Hilgert (2002) report consistent results based on analysis of the monthly Surveys of Consumers initiated by the Survey Research Center at the University of Michigan, which is a telephone survey covering all age groups. Yoong (2011) uses a monthly individual-level panel dataset from the American Life Panel (ALP) Monthly Surveys in 2003 and employs OLS and Probit methods to estimate the relationship between financial illiteracy related to the stock market and financial market participation. In the ALP Surveys, the respondents are asked basic and advanced questions about financial knowledge. To measure financial illiteracy related to the stock market, an index for a lack of knowledge about stock market investment was constructed through conducting principal component analysis (PCA) on binary indicators for the "Don't know" responses to the advanced questions related to stock market investment. One principal component was retained and the score for this component was taken as an index for stock market investment illiteracy.

The dependent variable in the OLS and Probit regressions is a binary indicator denoting stock ownership. A two-stage least squares (2SLS) estimator is employed as a robustness check in order to mitigate against the potential endogeneity issue of financial illiteracy since linear regression may not lead to correct inferences regarding the causal link between financial illiteracy and stock market participation. The instrumental variables used include: (1) the availability of financial education in high school; (2) the availability of financial education in the workplace; (3) the self-assessed level of economics education; and (4) no knowledge of bond pricing. The fourth one is preferred through testing the validity of the above instruments, specifically by checking for the relevance and exogeneity of the instrument set. The findings are consistent across the OLS, Probit and 2SLS models and suggest that a lack of financial knowledge about stock market investment, as represented by financial illiteracy, has a statistically significant effect, with a one-unit increase in the financial illiteracy index reducing the propensity to hold stocks by 10 percent.

For other developed countries with a mature financial market, the relationship between financial literacy and financial market participation has also drawn attention among academics. For example, Van Rooij et al. (2011) measure financial literacy and study its relationship with stock market participation by using a household-level cross-sectional dataset from De Nederlandsche Bank's Household Survey (DHS) for the Dutch population in 2005. They construct both basic and advanced financial literacy indices by performing factor analysis on basic financial literacy questions related to the interest rate, interest compounding, the influence of inflation, discounting and money illusion. They also perform factor analysis on the advanced questions covering topics such as the difference between stocks and bonds, the function of the stock market, the operation of risk diversification, and the relationship between bond prices and interest rates.

The OLS method and the generalized method of moments (GMM) estimator are employed to analyse the correlation between the dependent variable, stock market participation, which is a binary outcome, and the two types of financial literacy indices. There are several potential problems with OLS regression in this case. First, the indices of financial literacy may be measured with substantial error because many responses may be imprecise and may result from guessing, which may lead to the effect in the OLS specification being biased. On the other hand, experience and participation in the stock market can improve financial literacy, which may lead to the OLS estimates being biased upward, where the magnitude of the coefficient may be overestimated. Therefore, they use information on whether the financial situation of the oldest sibling is "better", "the same", or "worse" than the financial situation of the respondent and treat this information as an instrument in the GMM estimator to solve this potential endogeneity issue of the indices of financial literacy. They find that financial literacy affects financial decision-making. Those with low financial literacy are much less likely to invest in stocks.

Arrondel et al. (2015) obtain similar findings for France, specifically that basic financial literacy is positively associated with stock market participation, but not related to the share of stocks held in financial assets. They use a household-level cross-sectional dataset from the 2011 PATER (PATrimoines et Préférences face au TEmps et au Risque) household survey to explore the relationship between stock market participation and basic financial literacy. They use a Heckman estimator, where the portfolio share invested in stocks (the value of stocks and mutual funds divided by financial wealth) is modelled (i.e. the second stage equation), whilst taking into account the participation decision, defined as a binary outcome, which equals 1 if the respondent holds a positive amount of stocks (i.e. the first stage equation). Financial literacy is measured as an index ranging from 0 to 4, elicited by asking questions on the interest rate, inflation and risk diversification. However, it should be noted that they were not able to find an appropriate instrumental variable (IV) to mitigate against the endogeneity of financial literacy, arising from potential reverse causality (i.e., as stated above, stockholding may improve financial knowledge with learning-by-doing).

Almenberg and Dreber (2015) examine the correlation between the gender gap in stock market participation and financial literacy in Sweden using an individual-level cross-sectional dataset from the 2010 Consumer Survey conducted by the Swedish Financial Survey Authority. The dependent variable, stock market participation, is measured by ownership of stocks and/or indirect participation through ownership of shares in mutual funds, which is a binary outcome. The key explanatory variables, basic and advanced financial literacy, are defined through a set of basic financial literacy questions and advanced financial literacy questions, respectively. The basic financial literacy questions mainly measure the ability to perform basic calculations, i.e. numeracy, and the advanced financial literacy questions focus on financial products and concepts, largely based on Lusardi and Mitchell (2007). Both basic and advanced financial literacy are measured as indices ranging from 0 to 6 and are increasing in the level of financial literacy.

A Probit estimator is used in their analysis because stock market participation is a binary outcome and they focus on the gender differences in financial literacy by including a dummy variable, which equals 1 if the respondent is female. They estimate six Probit models for stock market participation, where the first one only includes the female control and they find that being female is inversely associated with stock market participation. The second regression includes additional personal characteristics (age, educational attainment and personal income) and they find that the female dummy variable becomes almost statistically insignificant and the marginal effect becomes smaller in terms of magnitude. The third one adds the basic financial literacy index, which is statistically significant and positively associated with stock market participation but the female dummy variable becomes statistically insignificant. The fourth one adds both the basic and advanced financial literacy indices, where both financial literacy indices have a statistically significant positive effect on stock market participation. The fifth model adds a risk attitudes index, ranging from 0 to 10 and increasing in risk-tolerance, with both financial literacy indices and the risk attitudes variable attaining statistical significance, where the advanced financial literacy index has a larger impact on stock market participation than the basic financial literacy index. The final regression removes the advanced financial literacy index from the fifth model, where the basic financial literacy index is still statistically significant and positively related to stock market participation, and the female variable is still statistically insignificant. The results above suggest that the estimated gender gap in stock market participation diminishes once basic financial literacy is controlled for. However, the basic and advanced financial indices are treated as exogenous, which means their results may be biased because of the endogeneity issue, as discussed above in the context of other studies.

Similarly, Thomas and Spataro (2018) find that financial literacy is statistically and positively associated with stock market participation using the individual-level dataset from waves 3 and 4 of the Survey of Health, Aging and Retirement in Europe (SHARE) on nine countries (Austria, Belgium, Denmark, Germany, Italy, France, Switzerland, Sweden, and the Netherlands). They mainly focus on wave 4 and match the life history information recorded in wave 3 for those re-interviewed in wave 4. The OLS model and a Probit estimator are employed to explore the effect of financial literacy on stock market participation, which is defined as a binary outcome that takes the value of 1 if the respondent participated in the stock market in 2010. The individual's financial literacy is measured by an index ranging from 1 to 5, increasing in the level of financial literacy based on four financial and numeracy questions: Firstly, "if the chance of getting a disease is 10%, how many people out of 1000 can be expected to get the disease?"; Secondly, "in a sale, a shop is selling all items at half price and before the sale, a sofa costs 300 euros, how much will it cost in the sale?"; Thirdly, "a second-hand car dealer is selling a car for 6,000 euro and this is two-thirds of what it costs new, how much did the car cost new?"; Fourthly, "let's say you have 2,000 euros in a savings account, the account earns 10 percent interest each year, how much would you have in the account at the end of the second year?" If a respondent answers (1) correctly, she/he is then asked (3) and, if she/he answers (3) correctly, she/he is asked (4). Answering (1) correctly leads to a score of 3, answering (3) correctly but (4) incorrectly results in a score of 4, while answering (4) correctly results in a score of 5. In terms of (2), if she/he answers (1) incorrectly, then they will be directed to (2) and, if she/he answers (2) correctly, she/he obtains a score of 2 otherwise he/she attains a score of 1. The marginal effects derived from the Probit model indicate that a unit increase in the financial literacy index increases the probability of participating in the stock market by 2.6%, while an additional year of schooling increases the probability of stock market participation by 0.4%, which indicates the relative importance of financial literacy.

Finally, they use an IV two-stage regression approach as a robustness check since financial literacy is potentially an endogenous regressor. Their IV is based on an index of mathematical ability at age 10, ranging from 1 to 5, where 1 indicates the lowest level of mathematical ability and 5 indicates the highest level based on the question: “how did you perform in Math compared to other children in your class? Did you perform much better, better, about the same, worse or much worse than the average?” The IV estimation results accord with the results from the standard Probit estimation in that financial literacy is positively associated with the probability of participating in the stock market.

In addition to research on the relationship between financial literacy and financial market participation, there exists relevant research on the determinants of risky asset allocation. For example, Cardak and Wilkins (2009) explore the determinants of the risky asset allocation of Australian households using five waves (2001-2005) of household-level panel data from the Household Income and Labour Dynamics in Australia (HILDA) Survey. They employ a Tobit estimator as the dependent variable, the risky asset ratio, is censored at zero and one as many households hold no risky assets and some households only hold risky assets. They find that labour income risk and poor health decrease risky asset holding. Labour income risk is measured in two ways, where the first one is measured by the realized variability of household labour income over the five years of data available and the realized variability is measured by the coefficient of variation of age-adjusted and time-adjusted annual household labour income between 2001 and 2005. The second measure of labour risk is denoted by a dummy variable equal to 1 if there are two or more earners in the household. Poor health is obtained from self-assessments of general health, defined as a dummy variable, which equals 1 if the self-assessed health status of the household head is “poor”. Finally, they collectively interpret the age, education and immigrant status variables as a proxy for financial literacy, which is in stark contrast to the direct measures of financial literacy used in other studies, as described above. Furthermore, it is clear that such variables may capture effects, which are not related to financial literacy.

They find that educational attainment and age are positively associated with the risky asset ratio and non-English speaking background (NESB) immigrant status and poor English skills have negative effects on the risky asset ratio, where the NESB immigrant status is a dummy variable, which equals 1 if the household head was born outside Australia, and poor English skills is a dummy variable, which equals 1 if the first language is not English and their self-assessed ability to speak English is poor.

3.2.2 China

Although there is an established literature on the relationship between financial literacy and household financial market participation, most of the existing studies are based on the context of mature financial markets in developed countries. In contrast, the financial market in China is developing and not yet mature. However, a limited number of studies on financial literacy and household risky asset allocation in China do exist. Most research in this area focuses on how financial literacy affects household financial market participation. For example, Liao et al. (2017) use cross-sectional household-level data from the 2014 Survey Consumer Finances (SCF) to explore the relationship between financial literacy and the risky asset holding behaviour of Chinese urban households. They use two methods to measure financial literacy. The first method is to sum the number of correct answers to financial literacy questions about time deposits, interest rates, inflation, the time-value of money, risk and return, bank functions, knowledge of stocks and so on, to create an index of financial literacy. The second method is to conduct factor analysis on this group of questions and to extract the factor score as the index of financial literacy. Hence, they employ two financial literacy indices and a Probit estimator is used to explore the relationship between financial literacy and risky asset ownership. In addition, a Tobit estimator is employed to explore how financial literacy affects the household risky asset share defined as the ratio of the value of risky assets to the total amount of financial assets in the household portfolio, which is censored at zero.

However, as discussed above, the financial literacy indices may be endogenous: for example, experience in participating in financial markets could enhance participants' financial literacy. Thus, they use the verbal test score as the IV for financial literacy and the information on the verbal test is from the 2010 China Family Panel Study (CFPS), which shares the same sample with the SCF. They argue that the verbal test score reflects basic language skills and implicitly, learning ability, and that it is unlikely to affect one's investment decisions. They find that low financial literacy is a widespread phenomenon in China, which accords with evidence for the U.S. from Hogarth and Hilgert (2002) suggesting that both high school students and adults demonstrate a low level of knowledge of basic economic principles.

The Probit results indicate that the higher is the level of financial literacy the higher is the probability of holding risky assets for a household in China. Specifically, the probability that the household holds risky assets increases by 2.4%, on average, with one additional point of the financial literacy index. Compared with the effect of household income in terms of size, the effect of financial literacy is half that of income, which indicates that financial literacy is less important than income but still has a sizeable effect. This finding is consistent in terms of sign and significance when replaced by the alternative financial literacy index.

In the case of the risky asset share, the results from the Tobit estimation are similar to the results from the Probit model in that the coefficients on the two financial literacy indices are positive and statistically significant. Although the paper presents some interesting results, it should be noted that it focuses on cross-sectional data, which cannot be used to control for unobserved heterogeneity. That is, there may be unobserved effects that could affect both financial literacy and the risky asset holding.

Similarly, Zou and Deng (2019) present relatively recent empirical evidence suggesting that financial literacy increases the probability of household financial market participation using a household-level dataset from the 2012 Survey Consumer Finances (SCF) in Urban China conducted by the financial research center at Tsinghua University in China. A Probit estimator is employed to explore how financial literacy affects household financial market participation, stock market participation, fund market participation and bond market participation. In detail, household financial market participation is a binary outcome, which equals 1 if the household holds risky assets mainly consisting of stocks, funds and bonds. Stock market participation, fund market participation and bond market participation indicate whether the household holds stocks, funds and bonds, respectively, which are denoted by three binary indicators. The key explanatory variable of interest, financial literacy, includes objective and subjective measures. The objective financial literacy measure is an index measured by objective questions regarding China's banking management system, compound interest, diversification of investments, knowledge of stocks, inflation and the exchange rate. The subjective financial literacy measure is based on responses to the following: Do you or your family understand the investment patterns of the following financial products? The financial products include stocks, funds and bonds and options and the responses include: (1) do not understand at all; (2) do not understand well; (3) understand a little; (4) understand fairly well; and, finally, (5) understand very well. The corresponding scores range from 1 to 5 and the scores are added across the three financial products to yield the measure of overall subjective financial literacy.

The results suggest that financial literacy (objective and subjective) is positively associated with household financial market participation. Specifically, they focus on the stock market, fund market and bond market, separately, and employ Probit models for each of these sub-markets. They find that the household is more likely to hold stocks, funds and bonds as the level of financial literacy increases. Specifically, the probability of household financial market participation increases by 12% if financial literacy increases by one sample standard deviation (1.48). Moreover, if financial literacy increases by one standard deviation (1.48), the probabilities of households participating in the stock market, fund market and bond market would increase by 12.57%, 10.21% and 7%, respectively.

The IV Probit estimator is then used to deal with the potential endogeneity of financial literacy. A dummy variable, financial education, which equals one if the respondent thinks that it is necessary for families to receive financial education, is used as the IV. The results suggest that financial literacy only has a positive impact on household fund market participation. There is no evidence of endogeneity in the case of household financial participation, stock market participation, fund market participation and bond market participation because the Wald test is insignificant, which indicates a failure to reject the null hypothesis of exogeneity. Hence, there is limited support for conducting the IV estimation in comparison with the existing literature on countries with a mature financial market (see, for example, Van Rooij et al, 2011), where financial literacy has typically been found to be endogenous. It is interesting to note, however, that whilst Zou and Deng (2019) find no evidence that financial literacy is endogenous in the financial markets they explore, this might reflect differences in the institutional setting. In addition, financial market participation in their study only includes three types of risky assets (stocks, funds and bonds), and omits a number of risky assets such as derivatives (future, option and forward), bank management wealth products as well as gold.

Pan et al. (2020) report evidence consistent with the previous study suggesting that financial literacy is positively associated with the probability of participating in the stock market using household-level cross-sectional data from the 2010 Survey Consumer Finances (SCF) for households in Urban China. They focus on two main explanatory variables, financial literacy and financial advice. Financial literacy is an index measured as the sum of scores from nine questions about six loan types (mortgages, cars, house decoration, education, business and consumer) and three security types (i.e., stocks, funds and bonds). The answers range from 1 to 5, where 1 indicates "knows nothing" and 5 means "knows a lot". Financial advice is denoted by a dummy variable that takes the value of 1 if the household has sought advice from financial institutions and/or professionals. A Probit estimator is used because the dependent variable, stock market participation, is a binary outcome, which equals 1 if the household holds stocks of any kind.

They analyse five different samples: specifically, the first one is the full sample containing all households; the second and third samples are separated by economic expectations, where the second sample of households feel optimistic, while the third sample contains the group of pessimistic households; and, finally, the fourth and fifth samples are split by diversification preference, with the fourth sample comprising households who are willing to invest in assets other than savings and the fifth sample contains the households who are unwilling to do so. It should be noted that splitting the samples based on economic expectations and diversification preference may cause sample selection bias. For example, households who feel optimistic about the economy or those who are willing to invest in assets other than savings are arguably more likely to participate in the stock market, which would bias the relationship between financial literacy and stock market participation estimated from these sub-samples.

In general, the results from all five samples indicate that financial literacy is positively associated with stock market participation, while, interestingly, the financial advice variable is statistically insignificant. Specifically, the magnitude of the effect of financial literacy for the sample of optimistic households (sample 2) is greater than that in the case of the full sample. In addition, the effect of financial literacy on stock market participation for sample 4 (households willing to invest in assets other than savings) is larger in terms of size than that for the full sample estimation.

For robustness, they explore three alternative measures of stock market participation: the ratio of stock value to income; the ratio of stock value to total assets; and the variety of stocks defined as the number of types of stocks held, where the first two are censored outcomes and the third is continuous. For the censored outcomes, they use a Tobit estimator and, for the continuous one, they use OLS and present estimates for the full sample, i.e. all households. The results suggest a positive relationship between financial literacy and the ratio of stock value to income, the ratio of stock value to total assets and the number of types of stocks. Once again, there is no statistically significant effect found for financial advice.

However, their estimation results may suffer from two endogeneity problems. The first one is that there may exist unobserved factors affecting both financial literacy and household stock market participation. The second is that households may invest in bonds and funds rather than stocks because the risks of investing in stocks are higher than bonds and funds and they argue that Chinese households are usually risk-averse. Thus, an index of the frequency of comparing various products before picking one is used as an IV, where the responses are on a 5-point scale, where 1 means "never" and 5 means "always". The IV results indicate that financial literacy remains significant and positively associated with stock market participation, while financial advice remains statistically insignificant, which is consistent with the findings in the standard Probit specification.

In terms of additional robustness checks, they firstly re-define financial advice to whether the households have ever sought advice from not only financial institutions and/or professionals but also from family or friends. The results from the Probit analysis once again reveal that financial literacy attains statistical significance while financial advice does not. Secondly, they omit households that have never compared portfolio options from the estimation sample because these households are arguably overconfident, which may lead to bias in the estimated effect of financial advice. However, the findings remain robust. Thirdly, they measure stock market participation by the ratio of stock value to total assets and the results from the Tobit model remain robust. Fourthly, they measure financial literacy according to knowledge about credit cards, which is defined as an index ranging from 0 to 4, based on whether a household knows about: (1) the fact that borrowing cash on credit cards incurs high fees and interest; (2) the fact that late payments on credit cards lead to penalty interest; (3) risks associated with the use of credit cards; and (4) consequences of defaults on credit card payment. For each question, if a household answers "yes", it scores 1 point, otherwise 0, which means that the minimum value of the financial literacy index is 0 if the household answers "no" to all questions. The Probit results indicate that the alternative financial literacy measure is still statistically significant and positively associated with stock market participation.

For the final robustness check, they explore an alternative measure for financial literacy through the Item Response Theory (IRT) model because there may be gaps between households' actual understanding and their responses to the questions. In other words, households may answer questions correctly by guessing, which means that households with the same scores in financial literacy may actually have different levels of financial literacy. Thus, the IRT model can be used to adjust for this measurement error.[49] Specifically, they first construct a financial literacy index ranging from 9 to 45, which measures how much a household knows about a set of nine loan products and securities and, as discussed above, the answers range from 1 to 5, where 1 indicates "knows nothing" and 5 denotes "knows a lot". Then, the financial literacy index is substituted into the IRT model to estimate the item parameter and the individual ability parameter, i.e. adjusted financial literacy index. After replacing the financial literacy index by the adjusted one, the results remain robust with financial literacy being positively associated with stock market participation.

To summarise, Pan et al. (2020) provide comprehensive evidence of a positive relationship between financial literacy and stock market participation. However, it should be noted that the measure of financial literacy is arguably not objective as the level of financial literacy is measured by the respondent's subjective judgment on a range of relative financial literacy questions. In other words, ideally financial literacy would be measured by whether respondents answer financial literacy questions correctly or not rather than by how much they believe they know about a subject.

Additionally, Xia et al. (2014) examine the influence of financial literacy overconfidence measured by the difference between an individual's subjective and objective financial literacy scores, on stock

[49] IRT is a theory based on the idea that the probability of a correct response to an item is a mathematical function of individual item parameters where the individual parameter is constructed as a single latent trait, e.g. performance on a test item, and the individual latent trait is estimated usually by maximum likelihood.

market participation with a cross-sectional household-level dataset from the 2012 Survey of Consumer Finances in Urban China (SCF).

The financial literacy score is based on six objective and three subjective financial literacy questions. Specifically, the subjective financial literacy score is measured by asking the respondents three questions on stocks, mutual funds and bonds: “How do you rate you or your family’s understanding of the following financial investment types?” The respondents rate their understanding of stocks, mutual funds and bonds on a scale of 1-5, where one indicates “not familiar” and five represents “very familiar”. An index of subjective financial literacy is generated by summing the values across the three questions leading to a range of 3-15. The objective financial literacy score is based on six questions, and is derived by adding together the scores from all six questions. The respondents can attain a score of 1 for each question answered correctly, and thus the measure has a range of 0-6. All six questions are tailored according to the Chinese financial environment. For instance, for the question: “which of the following banks is responsible for managing the financial system?”, the choices include: “(a) Bank of China”; “(b) Industrial and Commercial Bank of China”; “(c) The People’s Bank of China”; “(d) China Construction Bank”; and “(e) I don’t know”. After obtaining the subjective and objective financial literacy scores, the explanatory variable, financial literacy overconfidence, is measured by the difference between the respondents’ subjective and objective financial scores.

Respondents are then classified into four groups in order to construct the financial literacy overconfidence measure: (1) both the subjective and objective scores exceed the mean for the subjective and objective scores in the total sample; (2) the subjective scores are greater than the sample mean of the subjective scores but the objective scores are lower than the sample mean; (3) the objective scores are greater than the sample mean but subjective scores are lower than the sample mean; and (4) both the subjective and objective scores are lower than the sample means. The financial literacy overconfidence measure is a dummy variable, which equals 1 if respondents’ objective scores are lower than the average but they subjectively believe themselves to possess an above average-level of financial literacy. In addition, they create a dummy variable, financial literacy under-confidence, as a control variable, which equals 1 if the respondent has a higher objective financial literacy score than the average and a lower subjective financial score than the mean.

They also generate an alternative measure of financial literacy overconfidence, which is a dummy variable equalling 1 if the respondent did not choose “I don’t know” but answered all six objective financial questions incorrectly. A dummy variable is used to denote stock market participation, which equals 1 if the respondents participated in the stock market, thus the Probit method is employed for estimation. The results indicate that both types of financial literacy overconfidence are positively related to stock market participation. Specifically, the “overconfident” individuals have a 20% higher probability of participating in the stock market, while being under-confident decreases the likelihood of stock market participation by around 10%.

Overall, this paper identifies an interesting additional determinant of stock market participation, with the findings suggesting that financial literacy overconfidence is an important determinant of stock market participation as the effect of overconfidence is greater than that of the objective financial literacy index in terms of size. However, this study does not consider the potential endogeneity issue of financial literacy overconfidence, due to the existence of unobserved factors that may affect both financial literacy overconfidence and stock market participation.

More recently, Hsiao and Tsai (2018) explore the relationship between financial literacy and derivative market participation in Taiwan (China), using cross-sectional individual-level data from the 2011 Literacy Survey conducted by the Financial Supervisory Commission (FSC) of Taiwan (China). Initially, financial literacy is measured by an index ranging from 0 to 8, which is constructed to measure overall financial literacy, with a higher score reflecting a higher level of overall financial literacy. This measure is based on responses to eight financial literacy questions on risk diversification, the basic concepts of investment analysis and portfolio management, knowledge of risk management and derivatives products. The dependent variable, derivatives market participation, is defined as a binary outcome constructed from the question: “have you ever bought derivatives products such as swaps, futures, forwards, options, warrants, credit default swaps, collateralized debt obligation or leverage exchange-traded fund?”

They first use the Logit method to model derivatives market participation for three different specifications, where the first one includes financial literacy and the household’s demographics as the baseline model, the second model adds the household's risk attitudes, information sources, and asset-specific categories of knowledge, and the third model controls for wealth and income. The information sources variable is defined as an index ranging from 0 to 8, measuring the number of information sources from eight possible options including: (1) trade descriptions of financial institutions; (2) books; (3) discussions or interpretations by investment analysts on TV or radio; (4) newspaper or magazine advertisements and operators related to the TV industry; (5) internet and cell phones; (6) conversations with family members and friends; (7) the display section of a financial business office; and (8) school curriculum and handouts. Respondents were allowed to select more than one response and the greater the number of responses selected, the greater their information sources. The asset-specific categories of knowledge variable are based on a question: "which category of information are you interested in?" The answers include: (1) changes in the real estate market; (2) changes in the stock market; (3) interest rate levels; (4) inflation; (5) taxation; and (6) information relating to financial products. Thus, a set of six dummy variables is constructed to capture the asset-specific categories of knowledge.

The results from the Logit estimation show that the effect of overall financial literacy is statistically significant and sizable. For instance, taking the third specification as an example, a unit increase in the level of overall financial literacy increases the probability of derivatives market participation by 25.4%. Compared with the effect of the residential location control, the effect of financial literacy is half the magnitude, which indicates that financial literacy is a less important determinant of derivatives market participation, but it still has a sizeable effect.

For the first robust check, they estimate two subsamples using a Logit estimator, where the first subsample includes respondents with a brokerage account and the second one includes respondents with financial advisors. Although overall financial literacy is found to be positively associated with derivatives market participation in both subsamples, this finding may be biased due to potential sample selection bias. Specifically, the respondents with a brokerage account or a financial advisor are arguably likely to have a relatively high probability of participating in the derivatives market, which is not representative of the population as a whole. The size of the effect of overall financial literacy in the subsample of respondents with financial advisors is greater than that in the subsample of respondents with a brokerage account, which indicates that financial advisors play an important role in household derivatives market participation.

In their second robustness check, Hsiao and Tsai (2018) use the Bivariate Probit approach to model stock market participation and derivatives market participation given that access to derivatives products may be dependent on underlying securities market participation, i.e. stock market participation. The two binary dependent variables, stock market participation and derivatives market participation, relate to whether the respondent has experience in the stock or derivatives market, respectively, which take the value of 1 if the respondent has such experience. The correlation parameter between the probability of stock market participation and the probability of derivatives market participation is statistically significant, indicating that there exist unobserved factors affecting stock or derivatives market participation. Overall financial literacy has a statistically significant positive impact on participation in both markets, which indicates that, after controlling for experience in stock market participation, overall financial literacy is positively associated with derivatives market participation.

The IV Probit technique and a two-stage least squares (2SLS) estimator are then employed to deal with the potential endogeneity issue of financial literacy due to unobserved factors, such as access to derivatives products. Specifically, in China retail customers who do not pass the necessary assessment based on several policy requirements elicited by the local government, such as clear customer rules and risk management rules for business promotion and customer accounts, cannot directly access certain derivatives products. The IV used for overall financial literacy is average monthly expenditure on newspapers, books, magazines and other educational materials relating to the economy, business, trade, finance and accounting. The results from the IV Probit and 2SLS analysis are consistent and indicate that financial literacy is positively associated with the purchase of derivatives. However, the test of exogeneity is not rejected in both specifications, which suggests that there is no endogeneity issue and supports the use of the standard Logit model.

Furthermore, Hisao and Tsai (2018) also explore alternative measures of financial literacy in order to reduce the possibility of measurement error through performing factor analysis on all financial literacy questions related to basic and advanced knowledge on four topics of “money management and savings”, “credit and loan management”, “financial and investment planning” and “insurance and pension planning". The factor analysis supports the construction of two financial literacy indices, with the first index focusing on basic knowledge, and the second relating to advanced financial knowledge. Both financial literacy indices are treated as independent variables for the Logit estimation analysis. The level of advanced financial literacy is found to be positively associated with the probability of derivatives market participation, while conversely, there exists a statistically negative relationship between basic financial literacy and derivatives market participation. The negative effect of basic financial literacy may be explained by two reasons. Firstly, we might expect different effects from basic and advanced financial knowledge since advanced financial knowledge is usually more difficult to access. Secondly, complex derivatives products can create high entry barriers, which means that individuals with basic financial literacy are likely to find it difficult to participate in these markets. Conversely, individuals with advanced financial literacy are more likely to overcome these barriers. The importance of advanced financial literacy in derivatives market participation is demonstrated by the fact that the effect of advanced financial literacy is even greater than that of the income measures in terms of magnitude.

A number of relevant studies exist, which explore the relationship between investment knowledge and securities market participation. For example, Yao and Xu (2015) find that self-assessed investment knowledge is positively associated with Chinese household participation in the securities markets using a household-level cross-sectional dataset from the 2008 Survey of Chinese Consumer Finance and Investor Education (SCCFIE). Self-assessed investment knowledge is captured by the level of understanding of the risk and return associated with various financial products. The measure takes the value of 1 if the household has at least some understanding of the risk and return associated with the various financial products. There are four empirical specifications, where the first one models a binary outcome, which equals 1 if the household held at least one security (i.e. stocks, funds or bonds), and in the second, third and fourth models, the dependent variables are binary outcomes equalling 1 if the household held stocks, funds and bonds, respectively. A Logit estimator is used.

The results indicate that households are more likely to participate in financial securities markets if the household has financial investment knowledge. Specifically, those households who have self-assessed investment knowledge are 6.6 times more likely to hold at least one security as compared to those who have no such knowledge. Moreover, the effect of self-assessed investment knowledge in stock market participation is the highest in terms of magnitude across the four specifications. Specifically, the households, who have at least some understanding of the risk and return associated with various financial products, are 7.6 times more likely to hold stocks as compared to those who have no such knowledge. In contrast, households, who have at least some understanding of the risk and return associated with various financial products, are only 1.8 times more likely to hold bonds compared to those who reported no such understanding. Although the findings indicate that self-assessed investment knowledge plays an important role in participation in securities markets for urban China, it should be acknowledged that this measure arguably does not capture financial literacy. In addition, Chinese rural sectors are not included in their study and it may be the case that rural households behave differently in stock market participation compared with urban households. For example, there is a large gap between urban and rural households in access to financial products (Liao et. al, 2017).

From the review of the studies presented in this section, it seems appropriate to conclude that financial literacy is an important determinant of household financial market participation. Hence, this chapter contributes to further investigating the relationship between financial literacy and participation in household financial markets in China. In detail, we first explore the relationship between financial literacy and the probability of household financial market participation, the value of risky assets held by households and the share of risky assets to total household financial assets, using the 2013, 2015 and 2017 waves of the CHFS. Then, we split risky assets into two types, i.e. high-risk assets and low-risk assets, based on the different risk level for each financial product. This has not been explored in the existing literature for China although some existing studies have found that the magnitude of the effect of financial literacy varies across different types of financial products (see, for example, Zou and Deng, 2019). In addition to financial literacy, following Yoong (2011), we further explore the role of financial illiteracy in risky asset holding, which is largely ignored in the existing literature on China, by focusing on those households who choose “I can’t figure out” or “I don’t know” in the relevant financial literacy questions since financial illiteracy may have different effects to financial literacy measures based on choosing correct or incorrect answers in financial literacy questions (Van Rooij et. al, 2011). Finally, in contrast to most of the existing studies on China, which focus on cross-sectional analysis, cross-sectional analysis and panel data analysis is conducted. The latter is conducted in order to control for unobserved heterogeneity across households, which most of the existing studies on China are unable to control for due to data limitations.

3.3 Data and Methodology

3.3.1 Data

The household-level data employed in this chapter is from the China Household Finance Survey (CHFS), 2013, 2015 and 2017, which is a national survey conducted every two years, starting in 2011.[50] The CHFS includes information about households' demographic characteristics, assets and debt, income and consumption, financial literacy and risky asset holding. In detail, the number of households increases over these years from 28,141 (2013), 37,289 (2015) to 40,011 (2017), respectively. The total sample size increases across each wave as the sampling frame has changed over time in order to ensure the national representativeness of the survey, as discussed in detail in the previous chapter.

First, we focus on waves 2013, 2015 and 2017, as individual cross-sections and then we analyse waves 2013, 2015 and 2017 as an unbalanced panel dataset.[51] For these four samples, we focus on households who provide information on the financial literacy questions, with 1.3%, 8.9%, 35.8% and 14.7% of observations being omitted due to this restriction, respectively.[52] We only include households with a head aged over 20: 0.4%, 0.2%, 0.2% and 0.3% of observations are omitted due to this restriction, respectively. Moreover, after allowing for all missing values on covariates, such as risk attitudes, education attainment, health status, labour market status, marital status and political party membership of the head of household, we have 24,808, 28,212 and 19,718 households (N) in the 2013,

[50] The CHFS was used in the previous chapter. We do not use the 2011 wave in this chapter because this wave does not include information on financial literacy.
[51] In this chapter, we conduct panel analysis since, in contrast to the sample analysed in the previous chapter, we have 29.3% of households responding across all three waves and 31.92% responding across two waves.
[52] The total number of households in the 2017 wave is 40,011 and the number of households surveyed for the first time in 2017 is 13,187. Moreover, in 2017, the financial literacy questions are only asked to newly surveyed households. In addition, we find that there is a large proportion of missing values for the financial literacy questions amongst the households joining the panel in 2017. Specifically, only 1,185 households among the newly surveyed households answered all the financial literacy questions. In other words, only 3% of the total number of households answered all financial literacy questions in 2017. Thus, for those households, who joined the survey prior to 2017, we impute the values for financial literacy based on information from the previous wave and, after imputing the values, 64.2% of the total number of households surveyed in 2017 answered all three financial literacy questions. We discuss the construction of the panel dataset in the context of the financial literacy questions in more detail below.

2015 and 2017 cross-sectional datasets, respectively, and 74,794 observations (NT) in our panel dataset. All monetary variables in the 2015 and 2017 waves are deflated using China's yearly CPI, with the benchmark year 2013 = 100.

There are three main reasons why the CHFS is used for this chapter. Firstly, it contains detailed information on financial literacy and household risky asset holding across urban and rural households. Secondly, in contrast to existing studies for China, the CHFS is a recent dataset and includes almost all provinces of China, and, hence, is representative of the Chinese population. Finally, in addition to cross-sectional analysis, which is undertaken for purposes of comparison with existing studies on China, which have generally used cross-section data, we are able to conduct panel analysis to allow for time-invariant unobserved heterogeneity across households.

We will firstly focus on the relationship between the level of financial literacy of the household head and household risky asset holding as indicated by holding any risky assets including stocks, bonds, funds, financial derivatives (e.g. options, future, forward), financial wealth-management products (e.g. bank financial products, online financial products, collections of financial brokerages and trusts), non-RMB assets (e.g. foreign currency deposits, cash and foreign stocks etc.) and gold. It is possible that the level of financial literacy of the household head has a different impact on different types of risky assets, see, for example, Zou and Deng (2019) who explore how financial literacy affects household financial market participation by splitting the financial market into the stock market, the fund market and the bond market. Thus, we split household risky assets into high risk assets and low risk assets, respectively, based on questions related to risky asset ownership, which are asked to all households: *"Does the family have any stock accounts"; "Does the family have any funds"; "Does the family have any bonds"; "Does the family have any financial derivatives"; "Does the family have any bank wealth management products"; "Does the family have any non-RMB assets";* and *"Does the family have any gold"*. We explore the relationship between financial literacy and the holding of two categories of assets: high risk assets, which include stocks, non-RMB assets, gold and financial derivatives; and low risk assets, which include funds, bonds and financial wealth-management financial products.[53] This taxonomy corresponds closely to that used in Fratantoni (2001) who divides financial assets into risky and safe assets, where risky assets include stocks, equity mutual funds, corporate and municipal bonds, saving bonds and bond funds and safe assets include U.S. Treasury bills and other government bonds.[54]

3.3.2 Cross-sectional Analysis

3.3.2.1 The Logit Model

To examine the determinants of the probability of risky asset holding, we use a Logit estimator as our dependent variable, $Financial\ Asset\ Holding_i$, is defined as a binary outcome, which equals 1 if the households hold any type of risky assets such as stocks, financial derivatives, non-RMB assets, gold, funds, bonds and financial wealth-management products. Thus, the standard Logit estimator in our cross-sectional specification can be expressed as follows:

$$Pr(Risky\ Asset\ Holding_i = 1) = \Lambda(\beta_0 + \beta_1 Financial\ Literacy_i + \beta_2 X_i + \varepsilon_i) \tag{3.1}$$

where there are $i = 1, \dots N$ households, $\Lambda(\cdot)$ is the cumulative probability density function of the logistic distribution, β_0 is the intercept, β_1 captures the relationship between the dependent variable, $Risky\ Asset\ Holding_i$, and the key explanatory variable, $Financial\ Literacy_i$ (defined in detail below). The vector X_i comprises a set of covariates detailed below and ε_i is an error term characterised by the standard logistic distribution. Then following the same approach, the dependent variable is firstly specified as $High\ Risk\ Asset\ Holding_i$, a binary variable, which equals 1 if the household holds any high risk assets such as stocks, financial derivatives, non-RMB assets and gold, and, secondly, as $Low\ Risk\ Asset\ Holding_i$ is a binary variable, which equals 1 if the household holds any low risk assets such as bonds, funds and financial wealth-management products.

Table 3.1 provides full variable definitions and Tables 3.2.A and Table 3.2.B present summary statistics for all dependent variables used in our cross-sectional and unbalanced panel analysis. In 2013,

[53] Following the existing literature, the focus here lies on the relationship between financial literacy and risky assets, which are typically regarded as non-safe assets, where safe assets are defined as bank savings and cash.
[54] In this chapter, bonds were not broken down into more detailed categories because less than 0.01% of households in our sample provide information on whether the household holds Chinese government bonds or Chinese corporate bonds etc.

it can be seen from Table 3.2.A that only 12.9% of households hold risky assets, 9.5% of households hold high risk assets and 6.1% of households hold low risk assets. More households hold risky assets in 2015, specifically, 19.5% of households hold risky assets, 11.8% of households hold high risk assets and 13.1% of households hold low risk assets (see Table 3.2.A). Turning to 2017, we can observe that the proportion of households holding risky assets, high risk assets and low risk assets are still relatively low, namely, 16.1%, 8.5% and 11.9% (see Table 3.2.B). In addition, focusing upon the unbalanced panel in Table 3.2.B we can see that 16.2% of households hold risky assets, 10.0% of households hold high risk assets and 10.3% of households hold low risk assets. This indicates a relatively low risky asset holding rate in China as compared with the U.S., where about 25.1% of households report investing in stocks, bonds or funds in the 2016 U.S. Survey of Consumer Finances (SCF). Although the proportion of risky asset holding increases to some extent in 2015 and 2017, which may be due to the changes in sampling across each wave, it is still evident that the risky asset holding rate in China is relatively low.

3.3.2.2 The Tobit Model

We also explore the relationship between the level of financial literacy of the household head and the log level of risky assets held by the household, defined as $Ln(Risky\ Assets)_i$. Since the log level of risky assets is a censored outcome defined as the natural logarithm of the value of all risky assets held by the household plus one, we implement a Tobit estimator for our cross-sectional analysis. As above, in addition to modelling the log level of risky assets, we split the log level of risky assets into the log level of high risk assets and the log level of low risk assets, as represented by $Ln(High\ Risk\ Assets)$ and $Ln(Low\ Risk\ Assets)$, respectively.

In the cross-sectional Tobit specification, we model the log level of risky assets as follows:

$$Ln(Risky\ Assets)_i = \beta_0 + \beta_1 Financial\ Literacy_i + \beta_2 X_i + \varepsilon_i \tag{3.2}$$

where

$$Ln(Risky\ Assets)_i = Ln(Risky\ Assets)_i^* \quad if\ Ln(Risky\ Assets)_i^* > 0 \tag{3.3}$$

$$Ln(Risky\ Assets)_i = 0 \qquad otherwise \tag{3.4}$$

and the log level of risky assets held by the household is given by $Ln(Risky\ Assets)_i$ such that $i = 1,2,\dots,N$. β_0 is the intercept, and β_1 and β_2 are the coefficients. The key explanatory variable is $Financial\ Literacy_i$ and the vector X_i includes the set of covariates used in the Logit analysis. ε_i is a normally and independently distributed error term with zero mean and constant variance, σ^2.

The distributions of the log level of risky assets, the log level of high risk assets and the log level low risk assets conditional on holding positive amounts of such assets, are shown in Figures 3.1, 3.2 and 3.3 for the cross-sectional datasets for 2013, 2015 and 2017, respectively.

Figure 3.1 shows the distribution of the log level of risky assets for those households with positive amounts of risky assets, i.e. $Ln(Risky\ Assets) > 0$, with the median level of risky assets in 2013, 2015 and 2017 being around ¥35,000 (£3,500), ¥48,342 (£4,834) and ¥46,648 (£4,664), respectively for the sample reporting positive household risky assets. In a similar vein, Figure 3.2 shows the distribution of the log level of high risk assets for those households with positive amounts of high risk assets in 2013, 2015 and 2017, with the median level of high risk assets being lower at around ¥26,000 (£2,600), ¥48,342 (£4,834) and ¥49,354 (£4,935), respectively. Finally, the distribution of the log level of low risk assets is shown in Figure 3.3, where the median level of low risk assets is around ¥40,000 (£4,000), ¥3,9551 (£3955) and ¥46,648 (£4,664) in 2013, 2015 and 2017, respectively.

Additionally, as an alternative dependent variable and robustness check, we also model the share of risky assets held by the household to total household financial assets, denoted by $Share\ of\ Risky\ Assets$ (where total household financial assets include both risky assets such as stocks, bonds, funds, financial derivatives, financial wealth-management products, non-RMB denominated assets and gold, and risk-free assets such as bank savings and cash). The share of high risk assets to total household financial assets held by the household and the share of low risk assets to total household financial assets held by the household are indicated by, $Share\ of\ High\ Risk\ Assets$ and $Share\ of\ Low\ Risk\ Assets$, respectively. Each of these share variables are modelled via a Tobit estimator (as described above)

The distributions of the share of risky assets, the share of high risk assets and the share of low risk assets conditional on holding positive shares of such assets, are shown in Figures 3.4, 3.5 and 3.6 for the cross-sectional datasets for 2013, 2015 and 2017, respectively.

Figure 3.4 shows the distribution of the share of risky assets for those households with positive shares of risky assets, i.e. $Share\ of\ Risky\ Assets > 0$, with the median level of the share of risky assets in 2013, 2015 and 2017 being around 46.16%, 42.49% and 39.37%, respectively, for the sample reporting a positive household share of risky assets to total household financial assets. In a similar vein, Figure 3.5 shows the distribution of the share of high risk assets for those households with positive shares of high risk assets in 2013, 2015 and 2017, with the median level of the share of high risk assets being lower at around 32.61%, 33.33% and 31.90%, respectively. Finally, the distribution of the share of low risk assets is shown in Figure 3.6, where the median level of the share of low risk assets is around 35.71%, 29.25% and 28.92% in 2013, 2015 and 2017, respectively.

3.3.3 Panel Analysis

3.3.3.1 The Fixed Effects Logit Model

In order to control for unobserved heterogeneity across households, we specify an unbalanced panel fixed-effects Logit model, which only includes those households who changed states over two years or three years, to model the probability of holding risky assets as follows:

$$Pr(Risky\ Asset\ Holding_{it} = 1) = \Lambda(\beta_0 + \beta_1 Financial\ Literacy_{it} + \beta_2 X_{it} + \varepsilon_{it}) \tag{3.5}$$

$$\varepsilon_{it} = \mu_i + \eta_{it} \tag{3.6}$$

where the probability of holding any risky assets for household i in time t is given by $Risky\ Asset\ Holding_{it}$, such that $i = 1, 2 \ldots, n$ and $t = 2013, 2015, 2017$. $\Lambda(\cdot)$ is the cumulative probability density function of the logistic distribution, β_0 is the intercept, β_1 captures the relationship between the dependent variable, $Risky\ Asset\ Holding_{it}$, and the key explanatory variable, $Financial\ Literacy_{it}$, and X_{it} includes the set of covariates. ε_{it} is an error term comprising two parts, μ_i and η_{it}, where μ_i represents the household-specific unobserved heterogeneity (i.e. a fixed effect) and η_{it} is an idiosyncratic error term that varies across households and time. We assume that η_{it} is distributed by the standard logistic distribution and μ_i follows a normal distribution with mean zero and variance σ_{μ}^2. Moreover,

μ_i is independent of η_{it} but correlated with X_{it}. The analysis is repeated for high risk and low risk asset holding.

3.3.3.2 The Random Effects Tobit Model

To complement the cross-sectional analysis of the log level of risky assets we also estimate a random effects Tobit unbalanced panel specification for log level of risky assets as follows:

$$Ln(Risky\ Assets)_{it} = \beta_0 + \beta_1 Financial\ Literacy_{it} + \beta_2 X_{it} + \varepsilon_{it} \tag{3.7}$$

where

$$Ln(Risky\ Assets)_{it} = Ln(Risky\ Assets)^*_{it}\ \ if\ Ln(Risky\ Assets)^*_{it} > \ 0 \tag{3.8}$$

$$Ln(Risky\ Assets)_{it} = 0 \qquad otherwise \tag{3.9}$$

$$\varepsilon_{it} = \mu_i + \eta_{it} \tag{3.10}$$

where the log level of risky assets held by the household is given by $Ln(Risky\ Assets)_{it}$, such that $i = 1, 2, \ldots, N$ and $t = 2013, 2015, 2017$. β_0 is the intercept, and β_1 and β_2 are the estimated coefficients. The key explanatory variable is $Financial\ Literacy_{it}$ and matrix X_{it} includes all other covariates as detailed above. Following Mundlak (1978), in order to control for household time invariant effects and to enable the estimated parameters to be considered as an approximation to a standard panel fixed effects estimator, a vector of additional controls including the means of the continuous variables, such as the age of the household head and the mean of total household disposable annual income, is included. ε_{it} is an error term including two parts, μ_i and η_{it}, where μ_i represents household-specific unobserved heterogeneity and η_{it} is a random effect error term that varies across households and time. We assume that η_{it} is independent and identically distributed $N(0, \sigma_\eta^2)$ and μ_i follows a normal distribution with mean zero and variance σ_μ^2 and is independent of η_{it} and X_{it}. The correlation between the error terms of household i at time l and k is a constant given by

$$\rho = corr(\varepsilon_{il}, \varepsilon_{ik}) = \frac{\sigma_\mu^2}{(\sigma_\eta^2 + \sigma_\mu^2)}\ \ l \neq k \tag{3.11}$$

where ρ indicates the proportion of the total unexplained variance in the dependent variable contributed by the panel level variance component. The magnitude of ρ captures the extent of the unobserved intra-household correlation over time, where a low value of ρ indicates little unobservable intra-households correlation. The analysis is repeated for high risk and low risk assets. For a robustness check, we also model the share of risky assets to total household financial assets (where total household financial assets include both risky assets such as stocks, bonds, funds, financial derivatives, financial wealth-management products, non-RMB denominated assets and gold, and risk-free assets such as bank savings and cash), the share of high risk assets to total household financial assets and the share of low risk assets to total household financial assets, respectively, using the random effects Tobit estimator, following the above approach.

3.3.4 IV Probit Analysis

3.3.3.1 Cross-sectional IV Probit Model

As discussed in Section 3.2, it may be the case that $Financial\ Literacy$ is endogenous because experience in financial markets could improve participants' financial literacy, which means that causality may operate in both directions. Therefore, we explore the robustness of our findings using an IV Probit approach. Following Van Rooij et al. (2012), we construct an instrument, $Economics\ Education$, to deal with the potential endogeneity issue related to $Financial\ Literacy$. $Economics\ Education$ is a dummy variable, which equals 1 if the household head received any economics or finance course at school. Although this variable is not available for wave 2017, we can impute the value based on information from waves 2013 and 2015 since the economics education status of the household head is time-invariant as it relates to education acquired at school. In our sample, only about 10% of the heads of household received economic education at school in 2013 and 2015, which may suggest why there is a relatively low level of financial literacy among Chinese households (see, Table 3.2.A).

In the IV cross-sectional analysis, which is conducted separately for waves 2013, 2015 and 2017, we model the following IV Probit specification:

$$Risky\ Asset\ Holding_i^* = \alpha_1 + \gamma_1 Financial\ Literacy_i + \gamma_2 X_i + \omega_i \qquad (3.12)$$

$$Financial\ Literacy_i = \alpha_2 + \Phi_1 Economics\ Education_i + \Phi_2 X_i + \nu_i \qquad (3.13)$$

where $i = 1, 2, \ldots, N$, $Financial\ Literacy_i$ is the endogenous variable, both α_1 and α_2 are intercept terms. γ_1, γ_2, Φ_1 and Φ_2 are the coefficients. ω_i and ν_i are the error terms. The vector X_i includes all other covariates, as detailed above. $Economics\ Education_i$ is the instrumental variable, which is not related to $Risky\ Asset\ Holding_i^*$ and exogenous, i.e. $Cov(Economics\ Education_i,\ \omega_i) = 0$ since this variable measures education at school. Similarly, we also model the $High\ Risk\ Asset\ Holding$ and $Low\ Risk\ Asset\ Holding$ using the IV Probit estimator.

3.3.3.2 Pooled IV Probit Model

In addition to the cross-sectional IV Probit analysis as discussed above, we also conduct pooled analysis, based on an unbalanced panel, to explore the robustness of the findings from the cross-sectional analysis.

For the IV pooled analysis for waves 2013, 2015 and 2017, the specification is defined as follows:

$$Risky\ Asset\ Holding_{it}^* = \alpha_1 + \gamma_1 Financial\ Literacy_{it} + \gamma_2 X_{it} + \omega_{it} \tag{3.14}$$

$$Financial\ Literacy_{it} = \alpha_2 + \Phi_1 Economics\ Education_{it} + \Phi_2 X_{it} + \nu_{it} \tag{3.15}$$

where $i = 1, 2, \ldots, N$ and $t = 2013, 2015, 2017$. $Financial\ Literacy_{it}$ is the endogenous variable, and α_1 and α_2 are the intercepts. γ_1, γ_2, Φ_1 and Φ_2 are the coefficients. ω_{it} and ν_{it} are the error terms and the matrix X_{it} includes all other covariates, as above. $Economics\ Education_{it}$ is the instrumental variable, which is not related to $Risky\ Asset\ Holding_{it}^*$ and exogenous.

3.3.5 The Financial Literacy and Financial Illiteracy Measures

Turning to the key explanatory variable, $Financial\ Literacy$, we measure financial literacy based on three questions tailored according to the Chinese financial environment, which are similar to those devised by Lusardi and Mitchell (2008). Following a previous study for China (Xia et al., 2014), if the respondents answered correctly, a value of 1 is recorded; otherwise, the value is recorded as 0, which means the lowest level of knowledge for the head of household is 0 and the highest value is 3.[55] The questions are as follows:

---

[55] The value recorded as 0 in the financial literacy index includes both incorrect options and those choosing ‘can’t figure out’ in “Interest Rate” and “Inflation” questions and/or choosing ‘don’t know stocks’, ‘don’t know funds’ or ‘don’t know both’ in the “Risk Diversification” question.

1a. Given a 4% interest rate, how much would you have in total after 5 years if you have 100 RMB deposited? (In wave 2013)[56]

Answers: A. Under 120. B. 120. C. over 120. D. cannot figure out

1b. Given a 4% interest rate, how much would you have in total after 1 year if you have 100 RMB deposited? (In waves 2015 and 2017)

Answers: A. Under 104. B. 104. C. over 104. D. cannot figure out

2. With an interest rate of 5% and an inflation rate of 3%, the products you buy with the money you have saved in the bank for one year is:

Answers: A. more than last year. B. the same as the last year. C. Less than last year. D. cannot figure out.

3. Which one do you think is riskier, stocks or funds?

Answers: A. Stocks. B. Funds. C. Don't know stocks. D. Don't know funds. E. Don't know both.

For the cross-sectional analysis, we include the financial literacy index based on the responses to the financial literacy questions in each wave except for wave 2017, since in this case, as discussed above, we impute the value based on the information from the previous wave in the case of households which joined the survey prior to 2017. For the unbalanced panel analysis and pooled IV analysis, to ensure consistency across the waves, we include a time-invariant measure of financial literacy as measured in the first wave that the household responded to the financial literacy questions.[57]

In 2013, the mean of $Financial\ Literacy$ is 0.688, which suggests a low level of financial knowledge for Chinese households (see Table 3.2.A). It appears that financial knowledge at the household-level in China is low, specifically, only 22.72% of households figure out the correct interest rate and only

[56] Although we have a different question about the "Interest Rate" in wave 2013, we can still obtain a consistent financial literacy index, because the two questions about the "Interest Rate" are arguably not different in nature.

[57] Thus, the cross-section analysis captures the relationship between risky asset holding and financial literacy, where both are measured at the same point in time. Our results are unchanged if we use the same approach as in panel analysis for the cross-sectional analysis, i.e. taking the measure of financial literacy at the first time that household responded to these questions. All findings from the cross-section analysis are robust to using this time invariant measure of financial literacy.

15.80% of households correctly understand the effect of inflation, and only one-third of households know that the stocks are riskier than funds in 2013 (see Table 3.3). Moreover, in 2015, 30.66% of households figure out the correct interest rate, 17.41% of households correctly understand the effect of inflation, and 55.11% of households know that the stocks are riskier than funds. Correspondingly, in 2017, 28.84% of households correctly answer the "Interest Rate" question, 16.85% of households correctly answer the "Inflation" question and 51.59% of households correctly answer the "Risk Diversification" question (see Table 3.3). We can see that the financial literacy of the head of household improves across waves, which may be due to the household improving their financial skills and/or new households joining the survey having higher levels of financial literacy.[58] Compared with the U.S., where 77.7% of respondents correctly answer the question about the interest rate, 94.2% of respondents correctly answer the question about the inflation rate and 78.8% of respondents know that stocks are riskier than funds in 2003 (see Yoong, 2011), there exists a large disparity in terms of financial literacy between China and the U.S.

It has been argued that financial illiteracy is distinct from choosing incorrect answers to financial literacy questions (Yoong, 2011). For example, an individual who is not risk-tolerant will be less likely to hold any stocks, if this individual is financially illiterate, than those with some financial knowledge about stocks (Gollier, 2011). However, if this individual erroneously believes that stocks are not risky and have high returns, then he/she will be more likely to participate in the stock market (Yoong, 2011). To explore this further, following Yoong (2011), we construct three variables to explore household financial illiteracy, namely, $CF\ Interest\ Rate$, $CF\ Inflation$ and $DK\ Stock\ Fund$. These are three dummy variables, which equal 1 if the households choose "cannot figure out" for the "Interest Rate" and "Inflation" questions or choose "don't know stocks", "don't know funds" or "don't know both" for the "Risk Diversification" question[59] To explore the robustness of the findings, we replace the main explanatory

[58] Indeed, when we compare the financial literacy scores of households observed in the panel in both 2013 and 2015, we find that 84% of households either attain the same scores in 2013 and 2015 or are just 1 point within the scores for these two years. Such findings suggest that the improvement in overall financial literacy as indicated by the overall changes in the scores is driven by higher financial literacy of the new households joining the panel.

[59] We treated the "correct answers" and "incorrect answers" to financial literacy questions as the omitted category in the financial illiteracy analysis because our focus is on whether the household head does not know about interest rates, inflation and risk diversification rather than the effect of "correct answers" or "incorrect answers". This means we have to separate the option "I don't know/can't figure out" out of the three categories and there is no need to construct variables that capture "I don't know/can't figure out" relative to correct answers or incorrect answers because what we want to know is the effect of "I don't know/ can't figure out" relative to "I know/can figure out" (including "correct answers and incorrect answers").

variable, i.e. $Financial\ Literacy$, by the three financial illiteracy variables entered separately and entered simultaneously in the cross-sectional Logit model specified above.

We can see from Table 3.2.A that a large proportion of households choose “can’t figure out” in the “Interest Rate” and “Inflation” questions or choose “don’t know stocks”, “don’t know funds” or “don’t know both” in the “Risk Diversification” question. Specifically, 50% of households in 2013 choose “can’t figure out” in the “Interest Rate” question, 41.6% of households choose “can’t figure out” in the “Inflation” question and almost 60% of households don’t know stocks, funds or both. This indicates that financial illiteracy is indeed common among Chinese households. The phenomenon of household financial illiteracy increases between 2015 and 2017. To be specific, 44.6% of households choose “can’t figure out” in the “Interest Rate” question, 42.1% of households choose “can’t figure out” in the “Inflation” question and 39.9% of households don’t know stocks or funds in 2015, and in 2017 46.7% of households choose “can’t figure out” in the “Interest Rate” question, about 43.5% of households choose “can’t figure out” in the “Inflation” question and 43.6% of households don’t know stocks or funds (see Table 3.2.B).

3.3.6 Other Explanatory Variables

The matrix X_i contains the control variables generally used in the existing literature on household risky asset ownership (see, for example, Van Rooij et. al, 2011; Liao et. al, 2017; Zou et. al, 2019; Yang et. al, 2019). These variables are defined in Table 3.2. Specifically, $Ln(Income)$ is the natural logarithm of the total amount of disposable annual income of the household plus one.[60] $Ln(Net\ Wealth)$ is the natural logarithm of the total amount of household net worth plus one, which equals total household assets minus total debt if total household assets are equal to or greater than debt, otherwise this variable equals minus one multiplied by the natural logarithm of the modulus of total household assets minus total debt. $No.Children$ is the number of dependent children aged below 16 in the household (after age 16, in China children can choose to work or continue studying at school). $No.Workers$ is the number of workers in the household excluding the household head because we also include the labour market status of the household head, as discussed below. We also control for $No.Aged\ Over\ 60$, which

[60] The CHFS defines household disposable income as: salary net income after tax; net income from agricultural products after-tax; net income from business after-tax; net income from investment after-tax (rent, stock markets; interest from bank deposits, etc.); and net transfer income after-tax (social security, social insurance, annuity, etc.).

is the number of family members aged over 60 in the household excluding the household head (Chinese citizens can choose to retire after age 60).[61]

The risk attitudes of the head of the household have been found to be a key factor for explaining household risky asset holding (see, for example, Liao et al. 2017). Thus, we control for the head of household's risk attitudes, $Risk\ Attitudes$, which is a 5 point index ranging from 0 to 4. This index is increasing in risk tolerance, where 0 denotes a household head unwilling to carry any risk; 1 denotes a household head who prefers projects with slight risk and return; 2 denotes a household head who prefers projects with average risk and return; 3 denotes a household head who prefers projects with slightly high risk and slightly high return; and 4 denotes a household head who prefers projects with high risk and high return.

Health risk has been found to be an important determinant of risky asset holding, specifically, those household heads facing a health problem are less likely to hold risky assets (Rosen and Wu, 2004). Thus, we measure the health of the household head using the survey question: "what do you think of your health status relative to your peers?" The answers include "very poor", "poor", "normal", "good" and "very good". $Self\ Assessed\ Health$ is a 5 point index for the head of household ranging from 0 to 4, where 0 denotes very poor; 1 denotes poor; 2 denotes normal; 3 denotes good and 4 denotes very good. Age is the age of the household head. $Male$ is the gender of the household head, which has been found to play an important role in determining household risky asset holding. For example, Almenberg and Dreber (2015) found that male financial market participation is higher than female financial market participation in Sweden. Following Niu et al. (2020), whether the household head is married, is included as a covariate, denoted by $Married$.

In China, membership of the Communist Party of China distinguishes members from non-members in terms of social status, as the former enjoy explicit or implicit privileges, such as advantages in job hunting and promotions in state-owned enterprises (Yang et al., 2019; Niu et al., 2020).[62] Thus, we

[61] The impact of ethnicity is not controlled for in our analysis because there are 56 ethnic groups in China and approximately 91% of the population is in the Han group, which means we cannot explore the differences between each ethnic group.
[62] Party membership in China is around 6.8% in 2021, but, in our sample this is around 17.7% and one possible explanation is that party members may be more inclined to participate in the survey out of a sense of civic duty.

control for the variable, $Party\ Member$, which is a dummy variable and equals 1 if the household head is a party member. The $Education$ variable is classified into six categories: $No\ Schooling$ (the omitted category) is a dummy variable, which equals 1 if the household head never attended school; $Primary\ School$ is a dummy variable, which equals 1 if the highest educational attainment of the household head is primary school; $Junior\ High$ is a dummy variable, which equals 1 if the highest educational attainment of the household head is junior high school; $Senior\ High$ is a dummy variable, which equals 1 if the highest educational attainment of the household head is senior high school or technical school; $College/Bachelor$ is a dummy variable, which equals 1 if the highest educational attainment of the household head is vocational college or a bachelor degree; and $Master/PhD$ is a dummy variable indicating if the highest educational attainment of the household head is a master's degree or PhD.

Portfolio choice may differ by labour market status. For example, those operating a small business or the self-employed are usually less likely to hold stocks relative to other liquid assets because of the fact that private business ownership tends be focused on somewhat illiquid assets (Heaton and Lucas, 2000). Thus, we control for the labour market status of the head of household: $Employed$ is a dummy variable, which equals 1 if the household head is an employee, i.e. employed by someone else; $Self\ Employed$ is a dummy variable, which equals 1 if the household head is self-employed; $Retired$ is a dummy variable, which equals 1 if the household head is retired; $Not\ Working$ is a dummy variable, which equals 1 if the household head is not working, i.e. the household head is unemployed, incapacitated, a homemaker, a volunteer or unwilling to work; and $Farmer$ (the omitted category) is a dummy variable, which equals 1 if the household head is a farmer.

We control for the whether the household resides in a rural area, as indicated by $Rural$, since financial development is much lower in the rural areas than in the urban areas. The limited access to financial services significantly increases risky asset holding costs for rural households, which means that the urban households can enjoy better financial services than the rural households (Niu et. al, 2020). We also control for region, as represented by $Region$. Specifically, we distinguish between seven regions: $North\ East$, $North$, $East$, $Central$, $South$, $South\ West$, $North\ West$ (the omitted category). In our panel analysis, we control for the year of interview as the data covers three years: 2013 (the omitted category), 2015 and 2017.

Turning to the summary statistics, in wave 2013, we can see from Table 3.2.A that there exists a large disparity in household income between households comparing the minimum value and the maximum value, and the standard deviation of disposable net income. The average age of the head of household in 2013 is over 50, which may indicate that China is facing an aging population. Moreover, we find that most Chinese households are risk-averse because the mean value of $Risk\ Attitudes$ is less than 1 in 2013. In another words, most households are unwilling to carry any risk or prefer projects with slight risk and return, which is in line with the findings in the U.S. (see, for example, Avery and Kennickell, 1991). Turning to the educational attainment of the household head, over 70% of household heads have at least junior high school education, which indicates that the nine-year compulsory education system has been successful in China. However, less than 10% of household heads received economics or finance education during school. Finally, we can see that over 20% of household heads are farmers, which reflects the fact that China is a large agricultural country, as stated in Zeng et al. (2007). We also provide summary statistics for all variables in 2015, 2017 and our unbalanced panel dataset (see Tables 3.1.A and 3.1.B), where similar patterns in the data are evident.

3.4 Results

3.4.1 Cross-sectional Analysis

3.4.1.1 The Logit Analysis

We first analyse the relationship between financial literacy and household risky asset holding as a comparison with the existing literature exploring each wave as a separate cross-section. The results from estimating the three cross-sectional Logit models are shown in Table 3.4, where the estimated coefficients and marginal effects of financial literacy and the other covariates are presented for three dependent variables: $Risky\ Asset\ Holding$, $High\ Risk\ Asset\ Holding$ and $Low\ Risk\ Asset\ Holding$ in 2013. For brevity, as the results for the other covariates for 2015 and 2017 are in line with those for 2013, we only present the estimated coefficients and marginal effects of $Financial\ Literacy$ in 2015 and 2017 in Table 3.4.

From Table 3.4, we can see that the estimated coefficient of $Financial\ literacy$ is positive and statistically significant in the case of the probability of holding risky assets in 2013, which accords with our expectations and the existing literature. Moreover, the marginal effect of financial literacy is 0.0280, which indicates that a one unit increase in the $Financial\ Literacy$ index is associated with a 2.8% increase in the probability that the households hold risky assets. This finding indicates that households, where the head is equipped with a higher level of financial literacy, are more likely to hold risky assets. In comparison with the effect of household annual disposable income, which has drawn attention in the established literature on household risky asset holding (see, e.g., Liao et. al, 2017), it can be seen from Table 3.4 that an increase of one percent in household annual disposable income is associated with a 1.7% increase in the probability of households holding risky assets. Such a finding is not surprising: households with high household income are more likely to hold risky assets. Furthermore, the magnitude of the marginal effect of household income is less than that of financial literacy suggesting that the level of financial literacy is an important determinant of the probability of whether households hold risky assets (which is in line with existing findings for China from Liao et. al, 2017; Zou et. al, 2019).

Turning to $High\ Risk\ Asset\ Holding$ and $Low\ Risk\ Asset\ Holding$ in 2013 (see column 2 and column 3, respectively, in Table 3.4), it can be seen that the estimated coefficient of $Financial\ Literacy$ is positive and attains statistical significance at the 1% level for both outcomes. In terms of the size of the marginal effect of $Financial\ Literacy$, we can see that the effect on the probability of holding high risk assets is greater than that on the probability of holding low risk assets. This is in accordance with the findings of Zou and Deng (2019) that financial literacy has a different effect on different parts of the financial portfolio. Specifically, we find that a one unit increase in the level of financial literacy is associated with an increase in the probability of households holding high risk assets and holding low risk assets of 2.04% and 1.65%, respectively. Once again, taking the effect of household annual disposable income as a comparison, the estimated marginal effects of $Ln(Income)$ on high risk asset holding and low risk asset holding are 1.17% and 1.42%, respectively (see column 2 and column 3 in Table 3.4), which are smaller than the marginal effect of $Financial\ Literacy$ in the case of $High\ Risk\ Asset\ Holding$ and $Low\ Risk\ Asset\ Holding$ in terms of magnitude, respectively. Such a finding is consistent with the results in the case of risky asset holding, which provides further evidence of the importance of the level of financial literacy of the household head for household risky asset holding.

In 2015, see Table 3.4, the estimated coefficient of financial literacy is positive and statistically significant in the case of the three outcomes: $Risky\ Asset\ Holding$, $High\ Risk\ Asset\ Holding$ and

$Low\ Risk\ Asset\ Holding$. Moreover, the marginal effect of $Financial\ Literacy$ in the case of $Risky\ Asset\ Holding$ is 0.0451, which indicates that a one unit increase in the level of financial literacy of the household head is associated with a 4.51% increase in the probability that households hold risky assets. Such a positive effect of financial literacy on the probability of holding risky assets accords with the findings in 2013. Surprisingly, the size of marginal effect of financial literacy on high risk asset holding is smaller than that on low risk asset holding, which is opposite to the findings for 2013. Specifically, a one unit increase in the level of financial literacy is associated with an increase in the probability of holding high risk assets and low risk assets of 2.53% and 3.49%, respectively. Such a change from 2013 to 2015 may be due to a serious stock market crash in 2015: during the stock market crash, the Shanghai Stock Exchange (SSE) Composite Index decreased from 5,178 on 12th June 2015 to 2,638 on 27th January 2016 (The Shanghai Stock Exchange, 2016). This crash may have made people more aware of the risks of the stock market, thereby lowering the magnitude of effect of financial literacy on holding high risk assets in 2015.[63]

Turning to the determinants of the probability of risky asset holding in 2017, we obtain the same findings as in 2013 and 2015 in that $Financial\ Literacy$ is positively associated with the probability of risky asset holding (see Table 3.4). Furthermore, the marginal effect of $Financial\ Literacy$ is 0.0329, which indicates that a one unit increase in the level of financial literacy is associated with a 3.29% increase in the probability of holding risky assets. In addition, when we split the risky assets into high risk assets and low risk assets, the marginal effect of $Financial\ Literacy$ indicates that a one unit increase in financial literacy is associated with an increase in the probability that households hold high risk assets and low risk assets of 1.75% and 2.59%, respectively, which is consistent with the findings for 2015 indicating that the magnitude of the effect of financial literacy on high risk asset holding is smaller than that on low risk asset holding. Such a result suggests that the positive effect of financial literacy is stronger for low risk asset holding than for high risk asset holding.

[63] The estimated marginal effects of household income in the case of *Risky Asset Holding*, *High Risk Asset Holding* and *Low Risk Asset Holding* in 2015 are 0.0212, 0.0167 and 0.0148, respectively. This means that the size of the positive effect of financial literacy is greater than that of household income on the probability that households hold risky assets, high risk assets and low risk assets, which is in accordance with the findings in 2013.

We now briefly turn to the effects of the other covariates in 2013, presented in Table 3.4, where it can be seen that the risk tolerance is positively associated with the probability of holding risky assets, high risk assets and low risk assets, which means that the more risk-tolerant the household head is, the more likely is the household to hold risky assets, high risk assets and low risk assets. Interestingly, we find that the size of the marginal effect of $Risk\ Attitudes$ in the case of $High\ Risk\ Asset\ Holding$ is greater than that in the case of $Low\ Risk\ Asset\ Holding$, which means that the risk attitudes of the household head plays a more important role in determining the probability of holding high risk assets relative to that of low risk assets in 2013. The educational attainment of the head of household has a relatively large effect as compared to the other covariates in terms of size in 2013 (see Table. 3.4). Generally, the higher is the educational attainment of the head of household, the higher is the probability of holding risky assets, high risk assets and low risk assets. For example, the households, where the highest educational attainment of the head of household is primary school, have a 10.80% higher probability of holding risky assets, a 7.7% higher probability of holding high risk assets and an 8.43% higher probability of holding low risk assets in comparison to those heads of household who never attended school. Liao et al. (2017) find a U-shaped age-risky assets profile, which accords with our results that households have an increasing probability of holding risky assets, high risk assets and low risk assets as the age of the head of the household increases but it is a quadratic effect. Regarding the gender of the household head, we find that households are less likely to hold risky assets and low risk assets if the head of household is male. However, no difference is found between male and female headed households in terms of the probability of holding high risk assets.

We find that those households where the head of household is retired are more likely to hold risky assets, high risk assets and low risk assets, which might reflect the accumulation of savings over the lifecycle. Specifically, the estimated marginal effect of $Retired$ is 0.1266 in the case of risky asset holding, which means that households with a retired head have a 12.66% higher probability of holding risky assets than those where the head of household is a farmer (see Table 3.4). Similarly, the marginal effect of $Retired$ in terms of the magnitude is greater than that of other labour market states of the household head, which, as stated above, might reflect the accumulation of savings and also having more time to manage risky assets after retirement. Furthermore, we find that households living in the

$North\ East$ and $South\ West$ regions of China are less likely to hold risky assets, high risk assets and low risk assets, while residing in the $South$ and $East$ regions is positively associated with the probability of holding such assets in 2013, which probably reflects the fact that the $South$ and $East$ regions are more developed economic regions with better financial infrastructure (National Bureau of Statistics of China, 2019). As stated above, for brevity, the covariates for the cross-section models for 2015 and 2017 are not presented as, in general, a consistent pattern of results is found across all models.

3.4.1.2 The Tobit Analysis

We also explore the relationship between financial literacy and the log level of risky assets held by the household. For brevity, we only present the estimated effects of $Financial\ Literacy$ on the log level of risky assets, the log level of high risk assets and the log level of low risk assets. The same controls are included as in the Logit analysis and the pattern of the results remains the same. For financial literacy, the marginal effects are presented at the extensive and intensive margins.[64] It can be seen from Table 3.5 that the marginal effect of $Financial\ Literacy$ at the extensive margin is statistically sig-nificant and positively associated with $Ln(Risky\ Assets)$, $Ln(High\ Risk\ Assets)$, and $Ln(Low\ Risk\ Assets)$ across each wave. Specifically, focusing on wave 2013, a one unit increase in the level of financial literacy is associated with a 2.67% increase in the probability of holding risky assets. Similarly, financial literacy of the head of household is positively associated with the probability of hold-ing high risk assets and low risk assets. In detail, the magnitude of the marginal effect stemming from $Financial\ Literacy$ at the extensive margin is different across the probability of holding high risk assets and the probability of holding low risk assets, which accords with the findings from the Logit specifica-tion. We can observe a similar pattern of results in 2015 and 2017 (see Table 3.5).

Additionally, we shed further light on the effects of financial literacy by exploring marginal effects at the intensive margin as shown in Table 3.5. The estimated effect of $Financial\ Literacy$ in the case of the log level of risky assets is 0.3799 in 2013, which means that among households with a non-zero log level of risky assets, a one unit increase in the level of financial literacy of the head of household is

[64] A marginal effect at the intensive margin relates to the portion of the variation of the explanatory variable that is correlated with the variation of the expected value of the dependent variable conditional on being non-zero, while the marginal effect at the extensive margin relates to the change in the probability that the dependent variable is greater than zero.

associated with an increase in the log level of risky assets of 38%. Similarly, the marginal effect of $Financial\ Literacy$ at the intensive margin is also positive in the case of $Ln(High\ Risk\ Assets)$ and $Ln(Low\ Risk\ Assets)$. We find that financial literacy plays a more important role in determining the log level of low risk assets than the log level of high risk assets since the magnitude of the marginal effect at the intensive margin stemming from $Financial\ Literacy$ is 0.3565 in the case of $Ln(Low\ Risk\ Assets)$, which is higher compared to that in the case of $Ln(High\ Risk\ Assets)$ in 2013. This may reflect the fact that our measure of $Financial\ Literacy$ is based on responses to questions relating to quite basic financial knowledge. In addition, the holding of low risk assets is much more prevalent in our sample than the holding of high risk assets. Similar findings are revealed for 2015 and 2017 (see Table 3.5).

We also analyse the influence of financial literacy on the share of risky assets, the share of high risk assets and the share of low risk assets to total household financial assets and explore the marginal effects at the extensive and intensive margins (total household financial assets include both risky assets such as stocks, bonds, funds, financial derivatives, financial wealth-management products, non-RMB denominated assets and gold, as well as relatively risk-free assets such as bank savings and cash). For 2013, from Table 3.6, it can be seen that the marginal effect of $Financial\ Literacy$ at the extensive margin is 0.0252 in the case of the share of risky assets to total household financial assets. This indicates that a one unit increase in the level of financial literacy of the head of household is associated with a rise in the probability of holding a non-zero share of risky assets of 2.52%. Similarly, we also find that the probability of holding a non-zero share of high risk assets and a non-zero share of low risk assets is positively associated with $Financial\ Literacy$. This positive effect of the level of financial literacy of the household head is larger for the share of low risk assets than for the share of high risk assets in term of size (see Table 3.6). Similar patterns are found for 2015 and 2017.

Turning to the marginal effects of $Financial\ Literacy$ at the intensive margin in 2013, it can be seen from Table 3.6 that among those households with a non-zero share of risky assets, a one unit increase in $Financial\ Literacy$ is associated with an increase in the share of risky assets to total household financial assets of 1.94%. Similarly, a one unit increase in $Financial\ Literacy$ is associated with an increase in the share of high risk assets and the share of low risk assets of 1.29% and 1.58%, respectively. These findings accord with the findings in Liao et al. (2017), which indicate that the level of financial literacy of the head of household is positively associated with the share of risky assets to total household financial assets. Again, similar patterns are also found for 2015 and 2017.

Overall, the results of the Tobit estimation for the log level of risky assets and the share of risky assets to total household financial assets provide evidence that is consistent with the findings from the Logit estimation, as well as the existing literature for China, indicating that the level of financial literacy of the household head is positively associated with the household risky asset holding.

3.4.2 Panel Analysis

3.4.2.1 The Fixed Effects Logit Analysis

In order to control for unobserved heterogeneity across households, the fixed-effects Logit estimator is employed to explore the impact of financial literacy on the probability of risky asset holding using panel data. It is apparent from Table 3.7 that the estimated coefficient of $Financial\ Literacy$ is positive and statistically significant in the case of $Risky\ Asset\ Holding$, $High\ Risk\ Asset\ Holding$ and $Low\ Risk\ Asset\ Holding$ for the unbalanced panel. Specifically, the estimated marginal effect of $Financial\ Literacy$ is 0.0001 in the case of $Financial\ Asset\ Holding$, which indicates that one additional point of the financial literacy index is associated with a 0.01% increase in the probability of household risky asset holding. The results of the fixed-effects estimation approach support the findings from the previous analysis in that the level of financial literacy of the household head is positively associated with household risky asset holding. This result, although relatively small in magnitude, signifies the role of financial literacy once time invariant effects have been accounted for.

3.4.2.2 The Random Effects Tobit Analysis

We model the log level of risky assets using the random effects Tobit estimator with Mundlak controls for the unbalanced panel dataset to complement the cross-sectional Tobit analysis of the log level of risky assets. The results are presented in Table 3.8.

Firstly focusing on the marginal effects of $Financial\ Literacy$ at the extensive margin in the case of modelling the log level of risky assets, it can be observed that a one unit increase in the financial literacy of the household head is associated with a rise in the probability of holding risky assets of 3.23% (see Table 3.8). Similarly, one additional point in the financial literacy index is associated with a higher probability of holding high risk assets and low risk assets.

Turning to the marginal effects at the intensive margin stemming from $Financial\ Literacy$, a one unit increase in the level of financial literacy of the household head is associated with an increase in the log level of risky assets of 41.96% conditional on holding risky assets. Regarding the log level of high risk assets and the log level of low risk assets, financial literacy is more influential in determining the log level of low risk assets than the log level of high risk assets since the size of the marginal effects of $Financial\ Literacy$ at the extensive margin in the case of $Ln(Low\ Risk\ Assets)$ is greater than that in the case of $Ln(High\ Risk\ Assets)$, which, as discussed above, may reflect the nature of the financial literacy measure, which is based on quite basic financial knowledge (see Table 3.8).

The results in Table 3.8 also suggest that the log level of risky assets and the log level of low risk assets held by the households vary over time since ρ is positive, which suggests positive intra-correlation over time, and along with its statistical significance, shows that accounting for the longitudinal element of the data is important.

Finally, we also model the share of risky assets as a proportion of total household financial assets using the random effects Tobit estimator and a similar pattern of results is found based on the analysis of the marginal effects at the extensive and intensive margins. Specifically, the higher the level of financial literacy of the head of the household, the higher is the probability of holding a non-zero share of risky assets and the larger is the share of risky assets held among those who have a non-zero share of risky assets (see Table 3.9). Taking the share of risky assets as an example, the marginal effect at the extensive margin stemming from $Financial\ Literacy$ is 0.0307, which indicates that a one unit increase in the level of financial literacy is associated with a 3.07% increase in the probability of holding a non-zero share of risky assets. Turning to the marginal effects of $Financial\ Literacy$ at the intensive margin, we can see that a one unit increase in the level of financial literacy is associated with a 2.07% increase in the share of risky assets among those households with a non-zero share of risky assets.

To summarise, the findings from the random effects Tobit analysis provide additional evidence indicating that the level of financial literacy of the household head is positively associated with the household risky asset holding.

3.4.3 IV Probit Analysis

3.4.3.1 The Cross-sectional IV Probit Analysis

As discussed above, there may be reverse causality between $Financial\ Literacy$ and the probability of holding risky assets because the households where the head has a higher level of financial literacy are more likely to hold risky assets, and the experience of risky asset holding may improve the level of financial literacy of the household head. Therefore, the explanatory variable, $Financial\ Literacy$, is potentially endogenous and we use the IV Probit estimator to deal with this potential problem. Specifically, in the first stage (see Table 3.10), the estimated coefficient of $Economics\ Education,$ i.e. the instrumental variable, is positive and statistically significant, which is consistent with our expectation that the household head has a higher level of financial literacy if he/she received any economics or finance education at school, i.e. prior to holding risky assets. We assume that $Economics\ Education$ is uncorrelated with the error term in the risky asset holding equation because this information captures the education at school and it is time-invariant. However, as highlighted in Van Rooij et al (2021), this criterion might not be met because the household head's ability may drive literacy, education and risky assets but is unobserved in the risky asset holding regression, which means the IV results should be interpreted with caution. The estimated coefficient of $Financial\ Literacy$ is positive and statistically significant in the case of each of the three outcomes in 2013 (i.e. the second stage).

Taking the IV analysis in 2013 as an example, the estimated marginal effect of $Financial\ Literacy$ on $Risky\ Asset\ Holding$ is 0.0281, which indicates that a one unit increase in the level of financial literacy is associated with an increase in the probability of holding risky assets of 2.81%. Such a positive effect of financial literacy on the probability of holding risky assets is slightly higher than that in terms of magnitude from the cross-sectional Logit regression in 2013, where financial literacy was considered as exogenous. In addition, $Economics\ Education$ is positively associated with $Financial\ Literacy$, which indicates that the heads of household, who received any economics or finance education during school, have, on average, a higher level of financial literacy than those who did not. Similarly, in the case of $High\ Risk\ Asset\ Holding$ and $Low\ Risk\ Asset\ Holding$, a one unit increase in the level of financial literacy is associated with an increase in the probability of holding high risk assets and low risk assets of 2.05% and 1.62%, respectively. In comparison to the cross-sectional Logit analysis in 2013, it can be seen from Tables 3.4 and 3.10 that the magnitude of marginal effect stemming from $Financial\ Literacy$ on the probability of holding high risk assets and low risk assets are close to that estimated in the cross-sectional Logit regressions. Such a finding indicates that the level of financial literacy is still an important determinant of the probability of holding risky assets, high risk assets and low risk assets, both in terms of economic magnitude and statistical significance, even after controlling for endogeneity.

Additionally, the exogeneity Wald test statistic is also statistically significant in the case of $Risky\ Asset\ Holding$, $High\ Risk\ Asset\ Holding$ and $Low\ Risk\ Asset\ Holding$ in 2013, which indicates that endogeneity is an issue in our specifications. Hence, the results from the IV estimation provide valuable further evidence of a positive relationship between financial literacy and the probability of holding risky assets, high risk assets and low risk assets. Similar patterns can also be found in 2015 and 2017 (see Table 3.10).

3.4.3.2 The Pooled IV Probit Analysis

We also apply the IV Probit approach to deal with the potential endogeneity related to our main explanatory variable, i.e. $Financial\ Literacy$, in the context of our pooled dataset and the results are shown in Table 3.11. Similar to the findings from the cross-sectional IV analysis, we find that the estimated coefficient of $Financial\ Literacy$ is positive and statistically significant in the case of $Risky\ Asset\ Holding$, $High\ Risk\ Asset\ Holding$ and $Low\ Risk\ Asset\ Holding$. Specifically, in terms of the marginal effects, we can see from Table 3.11 that a one unit increase in the level of financial literacy is associated with an increase in the probability of holding risky assets of 3.72%. The instrumental variable is positively associated with the endogenous variable, $Financial\ Literacy,$ which is consistent with findings from the cross-sectional IV estimation. We also find the same pattern of results in the case of $High\ Risk\ Asset\ Holding$ and $Low\ Risk\ Asset\ Holding$, thereby further endorsing the robustness of our findings.

3.4.4 Cross-sectional Logit Analysis for Financial illiteracy

Finally, we shed further light on the determinants of the probability that households hold risky assets, following Yoong (2011), by analysing an alternative specification where the main explanatory var-iable is financial illiteracy, which is measured by three dummy variables, i.e. $CF\ Interest\ Rate$, $CF\ Inflation$ and $DK\ Stock\ Fund$. As discussed above in Section 3.3, households who choose "can't figure out" for the "Interest Rate" and "Inflation" questions or "don't know" in the "Risk Diversification" question are arguably different from those who choose incorrect answers for these questions. Thus, we focus on the cross-sectional Logit analysis to explore whether financial illiteracy affects risky asset holding.[65] In detail, we firstly include the three dummy variables in the cross-sectional Logit estimation separately and then include them simultaneously to replace $Financial\ Literacy$ for waves 2013, 2015 and 2017. The results relating to risky asset holding, high risk asset holding and low risk asset holding are presented in Tables 3.12.A, 3.12.B and 3.12.C, respectively.

It is apparent from Table 3.12.A that the estimated coefficients of financial illiteracy, i.e. $CF\ Interest\ Rate$, $CF\ Inflation$ and $DK\ Stock\ Fund$, are negatively associated with the probability of holding risky assets in 2013, 2015 and 2017. The first specification only includes financial illiteracy as measured by those households, where the head chooses "can't figure out" for the "Interest Rate" question, and a one unit increase in $CF\ Interest\ Rate$ is associated with a 4.08% decrease in the probability of holding risky assets in 2013. The second specification only includes $CF\ Inflation$, and the marginal effect is -0.0513, which indicates that a one unit increase in $CF\ Inflation$ is associated with a 5.13% decrease in the probability of holding risky assets in 2013. The third specification includes $DK\ Stock\ Fund$ only and the marginal effect is -0.1137, which indicates that a one unit increase in $DK\ Stock\ Fund$ is associated with an 11.73% decrease in the probability of holding risky assets in 2013. In the fourth specification, we include $CF\ Interest\ Rate$, $CF\ Inflation$ and $DK\ Stock\ Fund$ simultaneously and we find that $DK\ Stock\ Fund$ plays a more important role in determining the probability of holding risky assets since the magnitude of $DK\ Stock\ Fund$ is considerably higher than that of

[65] The analysis has also been conducted for the Tobit models and the pattern of results ties in with the findings discussed in this section.

$CF\ Interest\ Rate$ and $CF\ Inflation$, which is in accordance with our expectations because the explanatory variable, $DK\ Stock\ Fund$, directly relates to knowledge about stocks and funds. These results are in accordance with the findings in Yoong (2011) for the U.S. in that financial illiteracy is inversely associated with the probability of holding risky assets. Moreover, we can see from Table 3.12.A that the marginal effects of all three financial illiteracy variables are greater than that stemming from financial literacy in terms of magnitude in 2013, which may suggest that those who choose "can't figure out" in the "Interest Rate" and "Inflation" questions or "don't know" in the "Risk Diversification" question are arguably different from those who incorrectly answer these financial literacy questions.

In 2015, we can see that the marginal effect of $CF\ Interest\ Rate$ in the first specification is -0.0763, which is greater than that in terms of size for 2013. Similarly, in the second and third specifications, the sizes of the marginal effects of $CF\ Interest\ Rate$ and $DK\ Stock\ Fund$ are also greater than those estimated for 2013, respectively, which indicates that financial illiteracy has a greater effect in 2015 and this may be due to the fact that the stock market crash may have made people more aware of the risks of investing in the stock market, thereby increasing the magnitude of effect of financial illiteracy on holding risky assets in 2015 (see Table 3.12.A). In 2017, the magnitudes of the marginal effects of three financial illiteracy variables in the three separate specifications are all smaller than those in 2015, which also suggests that financial illiteracy has a relatively large effect in 2015.

Turning to the cross-sectional Logit results relating to high risk asset holding and low risk asset holding shown in Table 3.12.B and Table 3.12.C, respectively, a similar pattern of results is found. The estimated coefficients of the three financial illiteracy variables are negative and statistically significant in the case of $High\ Risk\ Asset\ Holding$ and $Low\ Risk\ Asset\ Holding$ in 2013, 2015 and 2017. For brevity, taking the results in 2013 as an example, the marginal effect of $CF\ Interest\ Rate$ in the first specification is -0.0345 in the case of $High\ Risk\ Asset\ Holding$, which is larger than that of $CF\ Interest\ Rate$ in the case of $Low\ Risk\ Asset\ Holding$ (see Table 3.12.C). Similarly, the magnitudes of the marginal effects of $CF\ Inflation$ and $DK\ Stock\ Fund$ in the case of $High\ Risk\ Asset\ Holding$ are also greater than those in the case of $Low\ Risk\ Asset\ Holding$ in 2013. This suggests that financial illiteracy plays a more important role in high risk asset holding, i.e. financial illiteracy has a different effect on different parts of the financial portfolio. Similar patterns can also be found in 2015 and 2017 (see Tables 3.12.B and 3.12.C).

3.5 Conclusion

This chapter has explored the relationship between the level of financial literacy and household risky asset holding employing household-level data from the CFHS (2013, 2015 and 2017). Risky asset holding is captured in three ways, i.e. the probability of holding risky assets, the log level of risky assets, and the share of risky assets in total household financial assets. We also split risky assets into high risk assets and low risk assets based on the different risk-levels of financial products in order to investigate how financial literacy influences the holding of assets characterized by different levels of risk.

We have initially employed the Logit estimator to explore waves 2013, 2015 and 2017 as three cross-sections. The findings suggest that the level of financial literacy of the head of household is positively associated with the probability of holding risky assets, high risk assets and low risk assets. Moreover, we find that the impact of $Financial\ Literacy$ is even greater than that of household income in terms of size in waves 2013 and 2015, and the magnitude of the effect of $Financial\ Literacy$ is only slightly smaller than of household income in 2017. These findings indicate that financial literacy is an important determinant of the probability of holding risky assets.

We have also analysed the effects of financial literacy on the log level of risky assets and the share of risky assets using the Tobit estimator for the three waves 2013, 2015 and 2017. Similarly, we have also split the log level of risky assets into the log level of high risk assets and the log level of low risk assets, and the share of risky assets into the share of high risk assets and the share of low risk assets. The results of the cross-section Tobit specification analysis provide further evidence of the importance of financial literacy of the household head for risky asset holding in terms of both the log level and the share of risky assets.

We have shed further light on the effects of financial literacy on the probability of holding risky assets, high risk assets and low risk assets by applying the fixed-effects Logit estimator to panel data in order to control for unobserved heterogeneity across households over time. Our findings accord with that from the cross-sectional analysis, with $Financial\ Literacy$ found to be positively associated with

$Risky\ Asset\ Holding$, $High\ Risk\ Asset\ Holding$ and $Low\ Risk\ Asset\ Holding.$ In a similar vein, we use the random effects Tobit approach to model the log level of risky assets, the log level of high risk assets, the log level of low risk assets, the share of risky assets, the share of high risk assets and the share of low risk assets. The findings support the positive relationship between the level of financial literacy of the household head and household risky asset holding, as in the case of the cross-section analysis.

The IV Probit approach was then used to deal with the potential endogeneity related to our key explanatory variable of interest, i.e. $Financial\ Literacy$. Our findings remain robust and suggest that the level of financial literacy is positively associated with the probability of holding risky assets, high risk assets and low risk assets and that $Economics\ Education$ has a positive impact on $Financial\ Literacy$. Furthermore, the findings indicate that it is appropriate to instrument the endogenous variable due to the rejection of the null hypothesis of exogeneity, which implies that the non IV results may suffer from an endogeneity problem. Similarly, we have also applied the IV Probit model to the pooled data and the results are consistent with the findings from the cross-sectional IV estimation.

Finally, we have used the cross-sectional Logit model to explore the effect of financial illiteracy, as measured by three dummy variables, i.e. $CF\ Interest$, $CF\ Inflation$ and $DK\ Stock\ Fund$, based on those households who choose “can’t figure out” for the “Interest Rate” and “Inflation” questions, and those choosing “don’t know stocks”, “don’t know bonds” or “don’t both” for the “Risk Diversification” question, on the probability of holding risky assets, high risk assets and low risk assets. We find that financial illiteracy is inversely associated with the probability of holding risky assets, high risk assets and low risk assets, which is in accordance with the findings in Yoong (2011). Such findings suggest that ignorance of basic financial knowledge is inversely associated with the probability of holding risky assets.

In conclusion, based on the descriptive statistics for our measures of the financial literacy of Chinese heads of household in 2013, 2015 and 2017, we find that low levels of financial literacy are common among Chinese households, where a large proportion of household heads are equipped with little financial knowledge, which accords with the findings in Liao et al. (2017). A positive relationship between the level of financial literacy and household risky asset holding has been revealed and the findings have been found to be robust to a range of econometric approaches, namely the Logit model, the

Tobit model and the IV Probit model. In addition, this chapter has identified a negative relationship between financial illiteracy and household risky asset holding, which has been ignored in the existing literature in China.

The evidence from the IV models revealed that households, where the head did not receive any economics or finance education at school, show a lack of financial literacy, and are found to be less likely to hold risky assets. Therefore, such findings suggest that the Chinese Government should pay more attention to economics and finance education for the new generations at school, e.g. adding basic economics or finance courses at high school so that the average household financial literacy and Chinese risky asset holding rate can be improved and, thereby, enhancing financial market growth in the future. Evidence from the U.S. suggests that such interventions may be effective. For example, as discussed in Fox et al. (2005), several wide-ranging financial education plans aimed at school-age children, such as the program conducted by the Jump$tart Coalition, have promoted Children’s financial education effectively in the U.S. In addition, in terms of targeting adults rather than children, the financial regulatory authorities and financial institutions, such as the China Securities Regulatory Commission, could consider providing at least basic financial knowledge for those households who are holding risky assets or are thinking about holding assets in the future. In addition, we find that the issue of relatively low risky asset holding, the stock holding puzzle, is more acute than in the U.S. and other developed countries. So, it is a greater challenge for the Chinese government to overcome than in other developed countries, not only because of the relatively low level of financial literacy but also the huge size of the population.

Finally, in terms of limitations, this study focuses on how financial literacy affects household risky asset holding. One area relates to whether households obtain positive returns from holding such risky assets. This is an area, which is potentially worth exploring because, in general, households hold risky assets in order to obtain positive returns and holding more risky assets may potentially place households in a financially vulnerable situation if they become over-confident in making such investments. Thus, such a topic remains interesting for future research subject to appropriate data availability.

3.6 Figures

Figure 3.1

Distributions of the log level of risky assets in 2013, 2015 and 2017 i.e. $Ln(Riksy\ Assets) > 0$

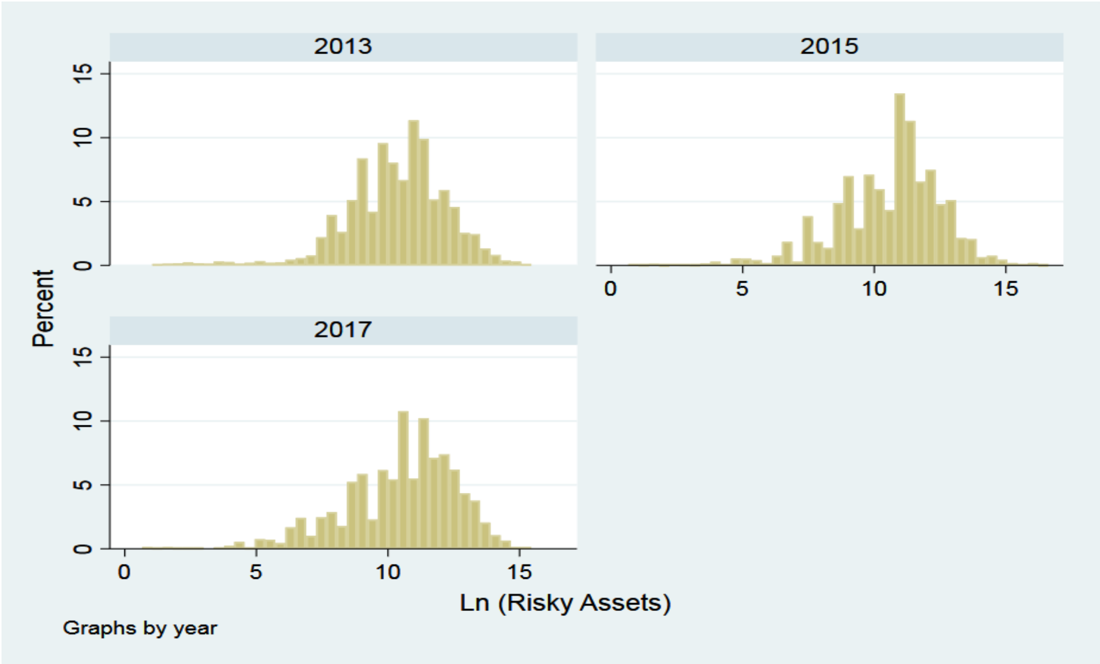


Figure 3.2

Distributions of the log level of high risk assets in 2013, 2015 and 2017, i.e. $Ln(High\ Risk\ Assets) > 0$

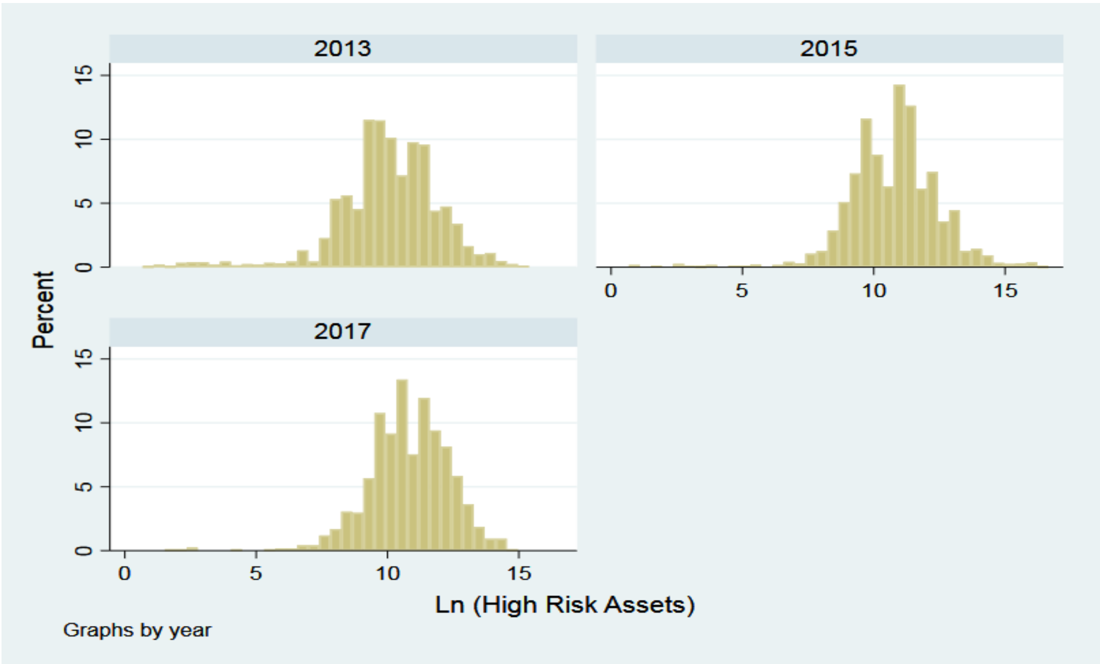


Figure 3.3

Distributions of the log level of low risk assets in 2013, 2015 and 2017, i.e. $Ln(Low\ Risk\ Assets) > 0$

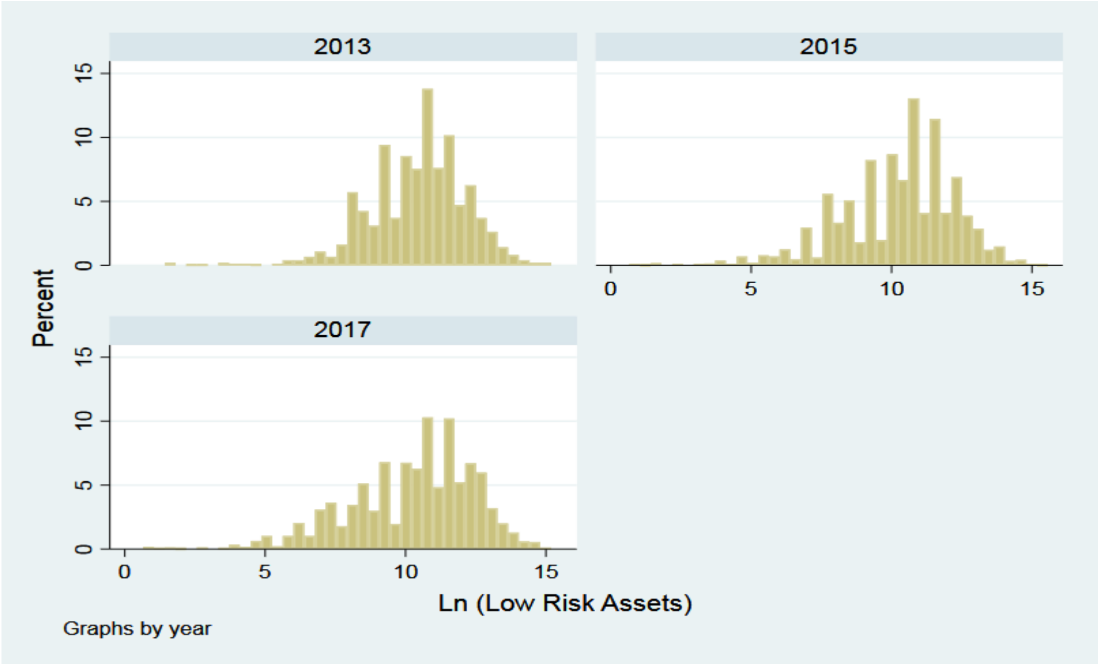


Figure 3.4

Distributions of the share of risky assets in 2013, 2015 and 2017, i.e. $Share\ of\ Risky\ Assets > 0$

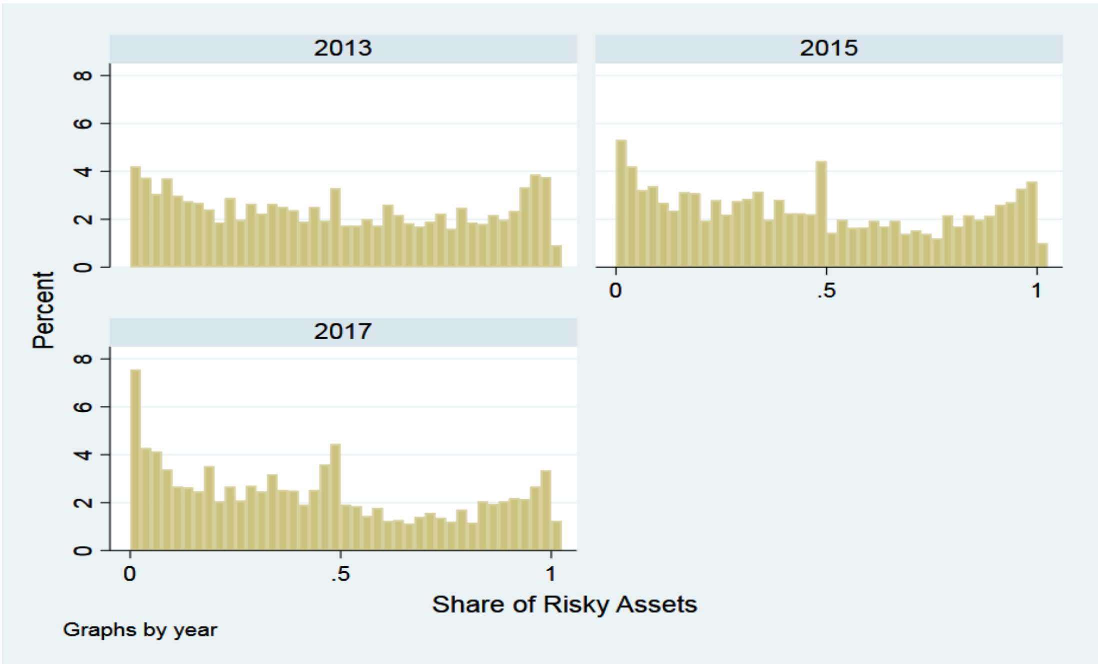


Figure 3.5

Distributions of the share of high risk assets in 2013, 2015 and 2017, i.e. $Share\ of\ High\ Assets > 0$

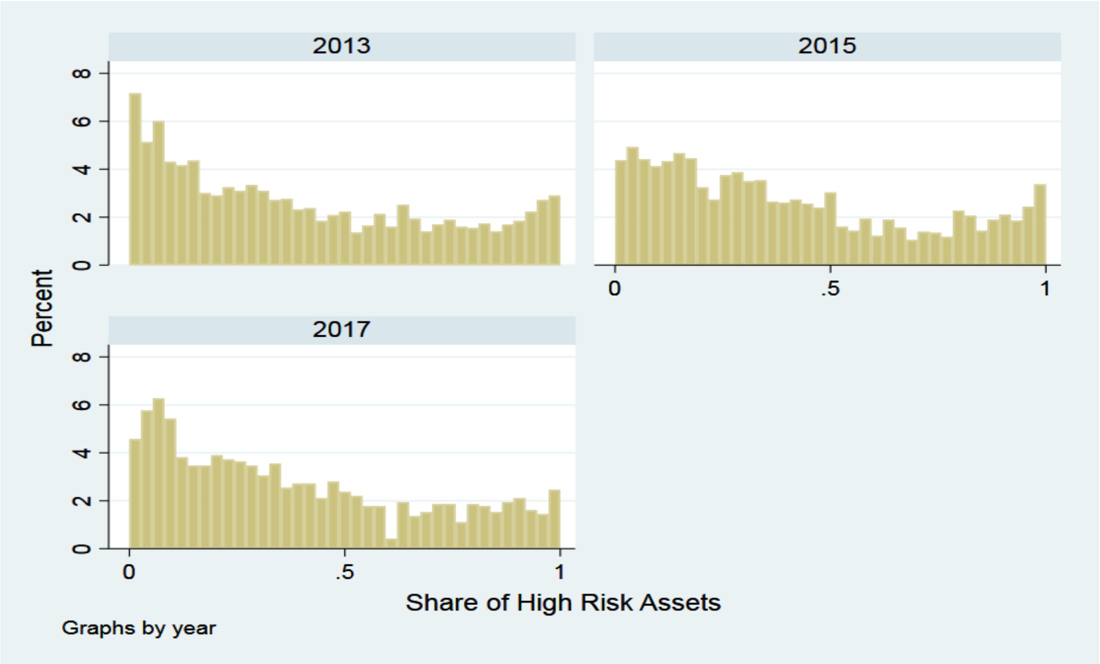
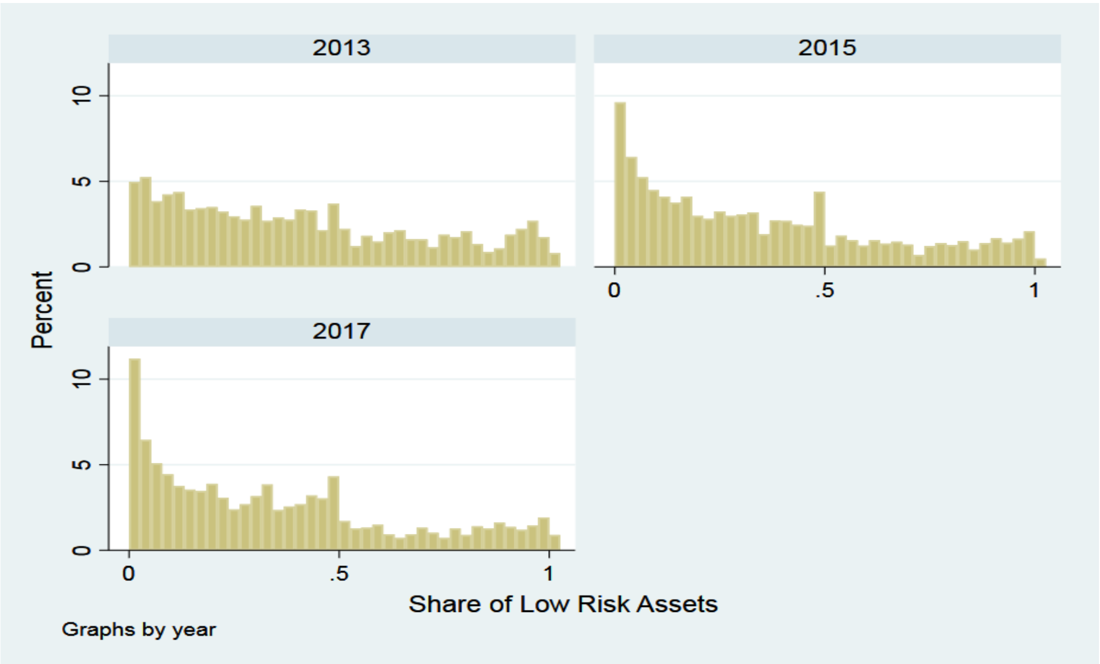


Figure 3.6

Distributions of the share of low risk assets in 2013, 2015 and 2017, i.e. $Share\ of\ Low\ Assets > 0$



3.7 Tables

Table 3.1 Definition of Variables

Variable	Description
Risky Asset Holding	Binary variable (0/1) equals 1 if the household holds any type of the following risky assets: stocks, financial derivatives, non-RMB assets, gold, funds, bonds and financial wealth-management products.
High Risk Asset Holding	Binary variable (0/1) equals 1 if the household holds any high risk assets including stocks, financial derivatives, non-RMB assets and gold.
Low Risk Asset Holding	Binary variable (0/1) equals 1 if the household holds any low risk assets including funds, bonds and bank/internet wealth management products.
Ln(Risky Assets)	The natural logarithm of the value of all risky assets held by the household plus one.
Ln(High Risk Assets)	The natural logarithm of the value of high risk assets held by the household plus one.
Ln(Low Risk Assets)	The natural logarithm of the value of low risk assets held by the household plus one.
Share of Risky Assets	The share of all risky assets held by the household to total household financial assets (total household financial assets include both risky assets such stocks, financial derivatives, non-RMB assets, gold, funds, bonds and financial wealth-management products and risk-free assets such as bank savings and cash).
Share of High Risk Assets	The share of high risk assets held by the household to financial assets.
Share of Low Risk Assets	The share of low risk assets held by the household to financial assets.
Financial Literacy	A 4-point index for the head of household ranging from 0 to 3, which is increasing in the level of financial literacy where 0 denotes the lowest level of financial literacy and 3 denotes the highest level of financial literacy. (see Table 3.3 below for further information)
CF Interest Rate	Dummy variable (0/1) equals 1 if the household head chooses “can’t figure out” for the “Interest Rate” financial literacy question.
CF Inflation	Dummy variable (0/1) equals 1 if the household head chooses “can’t figure out” for the “Inflation” financial literacy question.
DK Stock Fund	Dummy variable (0/1) equals 1 if the household head chooses “don’t know stocks”, “don’t know funds” or “don’t know both” for the “Risk Diversification” financial literacy question.
Ln(Income)	The natural logarithm of the total amount of the household disposable annual income plus one.
Ln(Net Worth)	Household net wealth, defined as the natural logarithm of the total amount of the household net wealth plus one, equals total household assets minus total household debt if assets are equal to and greater than debt, otherwise equals minus one times the natural logarithm of the modulus of total assets minus total debt. Total household assets include agricultural and business assets, land and real estate, vehicles, stocks, financial derivatives, non-RMB assets, gold, funds, bonds and financial wealth-management products, savings and cash etc. Total household debt includes agricultural/business debt, vehicle-purchasing debt, house-purchasing debt, education debt, credit debt and other debt.
No. Children	Number of dependent children aged below 16 in the household.
No. Workers	Number of workers in the household excluding the household head.
No. Aged Over 60	Number of family members who are aged over 60 in the household excluding the household head.
Risk Attitudes	A 5 point index for the head of household ranging from 0 to 4, which is increasing in risk-tolerance, where 0 denotes a household head who is unwilling to carry any risk; 1 denotes a household head who prefers projects with slight risk and return; 2 denotes a household head who prefers projects with average risk and return; 3 denotes a household head who prefers projects with slightly high risk and slightly high return; 4 denotes a household head who prefers projects with high risk and high return.
Self Assessed Health	A 5-point index for the head of household ranging from 0 to 4 where 0 denotes very poor; 1 denotes poor; 2 denotes normal; 3 denotes good and 4 denotes very good.

[a] All monetary variables in the 2015 and 2017 waves are deflated using China’s yearly CPI, with the benchmark year 2013 = 100, year 2015 = 101.4*102 and year 2017 = 101.4*102*102*101.6.

Table 3.1 Definition of Variables (Continued)

Variable	Description
Age	Age of the household head.
Age^2	Age squared of the household head.
Male	Dummy variable (0/1) equals 1 if the household head is male.
Married	Dummy variable (0/1) equals 1 if the household head is married.
Party Member	Dummy variable (0/1) equals 1 if the household head is a party member.
No Schooling (Omitted)	Dummy variable (0/1) equals 1 if the household head never attended school.
Primary School	Dummy variable (0/1) equals 1 if the highest education level of the household head is primary school.
Junior High	Dummy variable (0/1) equals 1 if the highest education level of the household head is junior high school.
Senior High	Dummy variable (0/1) equals 1 if the highest education level of the household head is senior high school or technical school.
College/Bachelor	Dummy variable (0/1) equals 1 if the highest education level of the household head is vocational college or bachelor.
Master/PhD	Dummy variable (0/1) equals 1 if the highest education level of the household head is above bachelor (master or PhD).
Economics Education	Dummy variable (0/1) equals 1 if the household head received any economics or finance education during school.
Employed	Dummy variable (0/1) equals 1 if the household head is an employee, i.e. employed by someone else.
Self Employed	Dummy variable (0/1) equals 1 if the household head is self-employed.
Retired	Dummy variable (0/1) equals 1 if the household head is retired.
Not Working	Dummy variable (0/1) equals 1 if the household head is not working, i.e. the household head is unemployed, incapacitated, a homemaker, a volunteer or unwilling to work.
Farmer (Omitted)	Dummy variable (0/1) equals 1 if the household head is a farmer
Rural	Dummy variable (0/1) equals 1 if the household resides in a rural area, equals 0 if the household resides in an urban area.
North East	Dummy Variable (0/1) equals 1 if the household lives in the northeastern region of China including 3 provinces: Heilongjiang, Jilin, Liaoning.
East	Dummy Variable (0/1) equals 1 if the household lives in the eastern region of China including 7 provinces: Shanghai, Jiangsu, Zhejiang, Anhui, Fujian, Jiangxi, Shandong.
North	Dummy Variable (0/1) equals 1 if the household lives in the northern region of China including 5 provinces: Beijing, Tianjin, Shanxi, Hebei, Neimenggu.
Central	Dummy Variable (0/1) equals 1 if the household lives in the central region of China including 3 provinces: Henan, Hubei, Hunan.
South	Dummy Variable (0/1) equals 1 if the household lives in the southern region of China including 3 provinces: Guangdong, Guangxi, Hainan.
South West	Dummy Variable (0/1) equals 1 if the household lives in the southwestern region of China including 4 provinces: Chongqing, Sichuan, Guizhou, Yunnan.
North West (Omitted)	Dummy Variable (0/1) equals 1 if the household lives in the northwestern region of China including 4 provinces: Shaanxi, Gansu, Qinghai, Ningxia.
2013 Year (Omitted)	Dummy Variable (0/1) equals 1 if the household responded in 2013.
2015 Year	Dummy Variable (0/1) equals 1 if the household responded in 2015.
2017 Year	Dummy Variable (0/1) equals 1 if the household responded in 2017.
NT (Observations)	Total number of observations
N	Total number of households

[a] All monetary variables in the 2015 and 2017 waves are deflated using China's yearly CPI, with the benchmark year 2013 = 100, year 2015 = 101.4*102 and year 2017 = 101.4*102*102*101.6

Table 3.2.A Summary Statistics - All Variables; Cross-section (t = 2013 and 2015)

	Cross-section t = 2013				Cross-section t = 2015			
	Mean	Std. Dev	Min	Max	Mean	Std. Dev	Min	Max
Risky Asset Holding	0.129	0.335	0	1	0.195	0.396	0	1
High Risk Asset Holding	0.095	0.293	0	1	0.118	0.322	0	1
Low Risk Asset Holding	0.061	0.239	0	1	0.131	0.337	0	1
Ln(Risky Assets)	1.226	3.405	0	15.157	1.784	4.044	0	16.224
Ln(High Risk Assets)	0.836	2.831	0	15.157	0.903	3.021	0	16.224
Ln(Low Risk Assets)	0.623	2.506	0	15.054	1.247	3.415	0	15.183
Share of Risky Assets	0.057	0.190	0	1	0.078	0.215	0	1
Share of High Risk Assets	0.033	0.141	0	1	0.034	0.142	0	1
Share of Low Risk Assets	0.024	0.120	0	1	0.044	0.155	0	1
Financial Literacy	0.688	0.821	0	3	1.032	0.922	0	3
CF Interest Rate	0.500	0.500	0	1	0.446	0.497	0	1
CF Inflation	0.416	0.493	0	1	0.421	0.494	0	1
DK Stock Fund	0.597	0.490	0	1	0.399	0.490	0	1
Ln(Income)	10.116	2.074	0	14.914	10.044	2.472	0	14.914
Ln(Net Worth)	11.972	3.245	- 15.244	16.811	11.928	3.915	- 15.136	16.778
No. Children	0.517	0.741	0	4	0.509	0.735	0	4
No. Workers	1.086	1.043	0	6	1.011	0.983	0	6
No. Aged Over 60	0.405	0.580	0	3	0.406	0.590	0	3
Risk Attitudes	0.940	1.206	0	4	0.925	1.172	0	4
Self Assessed Health	1.640	1.196	0	4	2.388	0.936	0	4
Age	51.302	14.178	20	90	52.398	14.043	20	90
Male	0.752	0.432	0	1	0.753	0.431	0	1
Married	0.860	0.347	0	1	0.875	0.331	0	1
Party Member	0.177	0.382	0	1	0.186	0.389	0	1
No Schooling	0.072	0.259	0	1	0.064	0.245	0	1
Primary	0.222	0.415	0	1	0.221	0.415	0	1
Junior	0.329	0.470	0	1	0.331	0.471	0	1
Senior	0.202	0.402	0	1	0.205	0.403	0	1
College/Bachelor	0.165	0.371	0	1	0.168	0.374	0	1
Master/PhD	0.010	0.099	0	1	0.010	0.102	0	1
Economics Education	0.101	0.302	0	1	0.103	0.304	0	1
Employed	0.321	0.467	0	1	0.364	0.481	0	1
Self Employed	0.129	0.335	0	1	0.096	0.295	0	1
Retired	0.173	0.378	0	1	0.091	0.288	0	1
Not Working	0.149	0.356	0	1	0.246	0.430	0	1
Farmer	0.228	0.420	0	1	0.203	0.402	0	1
Rural	0.310	0.463	0	1	0.298	0.457	0	1
North East	0.106	0.307	0	1	0.129	0.335	0	1
East	0.269	0.443	0	1	0.278	0.448	0	1
North	0.175	0.380	0	1	0.159	0.365	0	1
Central	0.131	0.338	0	1	0.110	0.313	0	1
South	0.090	0.287	0	1	0.118	0.322	0	1
South West	0.134	0.340	0	1	0.121	0.326	0	1
North West	0.095	0.293	0	1	0.086	0.280	0	1
Number of Observations		24,808				28,212		

Table 3.2.B Summary Statistics - All Variables; Cross-section & Unbalanced (t = 2017, 2013 & 2015& 2017)

	Cross-section t = 2017				Unbalanced t = 2013 & 2015 & 2017			
	Mean	Std. Dev	Min	Max	Mean	Std. Dev	Min	Max
Risky Asset Holding	0.161	0.368	0	1	0.162	0.369	0	1
High Risk Asset Holding	0.085	0.279	0	1	0.100	0.301	0	1
Low Risk Asset Holding	0.119	0.323	0	1	0.103	0.304	0	1
Ln(Risky Assets)	1.552	3.812	0	15.356	1.523	3.774	0	16.224
Ln(High Risk Assets)	0.646	2.591	0	14.701	0.805	2.838	0	16.224
Ln(Low Risk Assets)	1.208	3.379	0	15.157	1.019	3.132	0	15.391
Share of Risky Assets	0.064	0.195	0	1	0.067	0.201	0	1
Share of High Risk Assets	0.023	0.116	0	1	0.031	0.135	0	1
Share of Low Risk Assets	0.041	0.150	0	1	0.036	0.143	0	1
Financial Literacy	0.973	0.909	0	3	0.892	0.896	0	3
CF Interest Rate	0.467	0.499	0	1	0.472	0.499	0	1
CF Inflation	0.435	0.496	0	1	0.425	0.494	0	1
DK Stock Fund	0.436	0.496	0	1	0.481	0.500	0	1
Ln(Income)	10.490	1.632	0	15.356	10.192	2.140	0	15.391
Ln(Net Worth)	11.432	4.919	- 15.335	17.146	11.810	4.023	- 15.335	17.146
No. Children	0.471	0.742	0	4	0.500	0.739	0	4
No. Workers	0.923	0.941	0	6	1.010	0.994	0	6
No. Aged Over 60	0.430	0.570	0	3	0.414	0.582	0	3
Risk Attitudes	0.865	1.148	0	4	0.907	1.175	0	4
Self Assessed Health	2.350	1.008	0	4	2.128	1.103	0	4
Age	55.770	13.182	20	90	53.044	13.975	20	90
Male	0.791	0.407	0	1	0.763	0.425	0	1
Married	0.879	0.327	0	1	0.871	0.335	0	1
Party Member	0.416	0.493	0	1	0.245	0.430	0	1
No Schooling	0.066	0.249	0	1	0.067	0.250	0	1
Primary	0.250	0.433	0	1	0.229	0.420	0	1
Junior	0.349	0.477	0	1	0.337	0.473	0	1
Senior	0.194	0.395	0	1	0.201	0.401	0	1
College/Bachelor	0.133	0.340	0	1	0.156	0.363	0	1
Master/PhD	0.008	0.089	0	1	0.009	0.096	0	1
Economics Education	0.070	0.255	0	1	0.092	0.290	0	1
Employed	0.338	0.473	0	1	0.341	0.474	0	1
Self Employed	0.103	0.304	0	1	0.109	0.311	0	1
Retired	0.158	0.365	0	1	0.137	0.344	0	1
Not Working	0.188	0.391	0	1	0.200	0.400	0	1
Farmer	0.213	0.409	0	1	0.213	0.409	0	1
Rural	0.359	0.480	0	1	0.318	0.466	0	1
North East	0.133	0.340	0	1	0.121	0.327	0	1
East	0.283	0.450	0	1	0.277	0.448	0	1
North	0.157	0.364	0	1	0.165	0.371	0	1
Central	0.117	0.321	0	1	0.119	0.324	0	1
South	0.101	0.301	0	1	0.103	0.304	0	1
South West	0.125	0.331	0	1	0.125	0.331	0	1
North West	0.084	0.278	0	1	0.089	0.284	0	1
2013 Year					0.332	0.471	0	1
2015 Year					0.393	0.488	0	1
2017 Year					0.275	0.447	0	1
Number of Observations		19,718				74,794		

Table 3.3 Summary Statistics - Financial Literacy; Cross-section (t = 2013, 2015 and 2017)

			All Sample		
Year = 2013					
Response	A	B	C	D	E
"Interest Rate"	12.07%	15.21%	**22.72%**	50.00%	
"Inflation"	**15.80%**	7.56%	35.01%	41.63%	
"Risk Diversification"	**30.30%**	9.96%	0.79%	11.38%	47.57%
Observations			24,808		
Year = 2015					
Response	A	B	C	D	E
"Interest Rate"	8.12%	**30.66%**	16.56%	44.65%	
"Inflation"	**17.41%**	7.76%	32.68%	41.14%	
"Risk Diversification"	**55.11%**	5.00%	0.22%	4.96%	34.71
Observations			28,212		
Year = 2017					
Response	A	B	C	D	E
"Interest Rate"	7.81%	**28.84%**	16.69%	46.66%	
"Inflation"	**16.85%**	7.92%	31.73%	43.50%	
"Risk Diversification"	**51.59%**	4.77%	0.21%	5.43%	38.00%
Observations			19,718		

[a] In wave 2013, regarding the "Interest Rate" question, option "A" refers to "under 120"; option "B" refers to "120"; option "C" refers to "over 120" and option "D" refers to "cannot figure out". In waves 2015 and 2017, regarding the "Interest Rate" question, option "A" refers to "under 104"; option "B" refers to "104"; option "C" refers to "over 104" and option "D" refers to "cannot figure out".
[b] In waves 2013, 2015 and 2017, regarding the "Inflation" question, option "A" refers to "more than last year"; option "B" refers to "the same as last year"; option "C" refers to "less than last year" and option "D" refers to "cannot figure out".
[c] In waves 2013, 2015 and 2017, regarding the "Inflation" question, option "A" refers to "stocks"; option "B" refers to "funds"; option "C" refers to "don't know stocks"; option "D" refers to "don't know funds" and option "E" refers to "don't know both".
[d] Bold text signifies the correct response in the "Interest Rate", "Inflation" and "Risk Diversification" questions.

Table 3.4 The Determinants of the Probability of Risky Asset, High Risk Asset and Low Risk Asset Holding - Cross-sectional Logit Analysis

	Risky Asset Holding			High Risk Asset Holding			Low Risk Asset Holding		
Year = 2013	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat
Financial Literacy	0.3308***	0.0280	12.65	0.2985***	0.0204	10.30	0.3353***	0.0165	9.86
Age	0.1443***	0.0122	13.38	0.1476***	0.0101	11.88	0.1370***	0.0067	9.53
Age^2	- 0.0014***	- 0.0001	- 12.58	- 0.0015***	- 0.0001	- 11.42	- 0.0012***	- 0.0001	- 8.63
Male	- 0.1471***	- 0.0125	- 2.92	- 0.0449	- 0.0031	- 0.80	- 0.3279***	- 0.0161	- 5.09
Married	- 0.0544	- 0.0046	- 0.85	- 0.0595	- 0.0041	- 0.84	- 0.0705	- 0.0034	- 0.84
Party Member	0.0766	0.0065	1.42	- 0.0580	- 0.0040	- 0.96	0.2212***	0.0109	3.28
Risk Attitudes	0.0361***	0.0276	17.35	0.3822***	0.0261	18.65	0.1729***	0.0085	6.92
Self Assessed Health	- 0.0224	- 0.0019	- 1.10	- 0.0396*	- 0.0027	1.74	0.0118	0.0006	0.44
Primary School	1.2747***	0.1080	3.24	1.1292**	0.0770	2.42	1.7184**	0.0843	2.37
Junior High	1.9733***	0.1672	5.11	1.7648***	0.1203	3.87	2.4848***	0.1219	3.48
Senior High	2.4949***	0.2113	6.46	2.3862***	0.1627	5.24	2.8778***	0.1412	4.03
College/Bachelor	2.9912***	0.2534	7.71	2.8478***	0.1942	6.23	3.3348***	0.1636	4.66
Master/PhD	3.2216***	0.2729	7.77	3.0031***	0.2048	5.81	3.5562***	0.1745	4.84
Employed	1.0491***	0.0889	6.67	1.1696***	0.0797	5.01	0.8589***	0.0421	3.80
Self Employed	0.8330***	0.0706	5.10	1.0346***	0.0705	7.97	0.4510*	0.0221	1.89
Retired	1.4947***	0.1266	8.99	1.6810***	0.1146	5.54	1.2027***	0.0590	5.11
Not Working	0.9864***	0.0836	5.92	1.1664***	0.0795	8.58	0.6978***	0.0342	2.89
Ln(Income)	0.2005***	0.0170	10.71	0.1719***	0.0117	11.39	0.2896***	0.0142	9.18
Ln(Net Worth)	0.2303***	0.0195	13.36	0.2226***	0.0152	0.73	0.2106***	0.0103	8.78
No. Children	- 0.0175	- 0.0015	- 0.48	0.0294	0.0020	- 3.75	- 0.0214	- 0.0010	- 0.43
No. Workers	- 0.1004***	- 0.0085	- 3.76	- 0.1134***	- 0.0077	0.06	- 0.1061***	- 0.0052	- 2.92
No. Aged Over 60	0.0011	0.0001	0.03	0.0026	0.0002	- 9.25	0.0231	0.0010	0.44
Rural	- 1.2049***	- 0.1021	- 10.41	- 1.3320***	- 0.0908	- 1.66	- 1.0648***	- 0.0522	- 6.11
North East	- 0.3425***	- 0.0290	- 3.11	- 0.2129*	- 0.0145	5.01	- 0.4480***	- 0.0220	- 3.06
East	0.3178***	0.0269	3.66	0.4999***	0.0341	1.44	- 0.0066	- 0.0003	- 0.06
North	0.1405	0.0119	1.54	0.1519	0.0104	1.13	0.0375	0.0018	0.32
Central	- 0.0108	- 0.0009	- 0.11	0.1307	0.0089	4.22	- 0.2504*	- 0.0123	- 1.85
South	0.2529**	0.0214	2.35	0.5063***	0.0345	- 1.85	- 0.1511	- 0.0074	- 1.04
South West	- 0.2971***	- 0.0252	- 2.76	- 0.2305*	- 0.0157	- 22.96	- 0.3557**	- 0.0175	- 2.52
LR χ^2 (29); p value		5,438.62; p = [0.0000]			4,258.25; p = [0.0000]			2,707.35; p = [0.0000]	
Pseudo R^2		0.2856			0.2733			0.2377	
Observations		24,808			24,808			24,808	
Year = 2015	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat
Financial Literacy	0.4364***	0.0451	19.56	0.3441***	0.0253	13.02	0.3955***	0.0349	16.45
Observations		28,212			28,212			28,212	
Year = 2017	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat
Financial Literacy	0.3513***	0.0329	12.61	0.2946***	0.0175	8.40	0.3255***	0.0259	10.80
Observations		19,718			19,718			19,718	

[a] *, **, *** denote 10, 5, 1 per cent levels of significance, respectively.
[b] Reference category: Education controls: the omitted group is that household head never attended school; Labour market controls: the omitted group is that household head who is a farmer; Region controls, the omitted group is the North West

Table 3.5 The Determinants of the Log Level of Risky Assets, High Risk Assets and Low Risk Assets - Cross-sectional Tobit Analysis

	Ln(Risky Assets)				Ln(High Risk Assets)				Ln(Low Risk Assets)			
Year = 2013	E.M.E	t-stat	I.M.E	t-stat	E.M.E	t-stat	I.M.E	t-stat	E.M.E	t-stat	I.M.E	t-stat
Financial Literacy	0.0267***	12.59	0.3799***	12.38	0.0180***	9.62	0.3061***	9.50	0.0164***	9.89	0.3565***	9.76
LR χ^2 (29); p value	5,221.59; p = [0.0000]				3,784.22; p = [0.0000]				2,774.80; p = [0.0000]			
Pseudo R^2	0.1408				0.1377				0.1326			
Uncensored obs	2,948				2,065				1,486			
Left censored obs	21,860				22,743				23,322			
Observations	24,808				24,808				24,808			
Year = 2015	E.M.E	t-stat	I.M.E	t-stat	E.M.E	t-stat	I.M.E	t-stat	E.M.E	t-stat	I.M.E	t-stat
Financial Literacy	0.0418***	19.82	0.5233***	19.29	0.0180***	10.65	0.3146***	10.53	0.0352***	17.53	0.5039***	17.11
LR χ^2 (29); p value	8,694.40; p = [0.0000]				5,355.30; p = [0.0000]				5,525.66; p = [0.0000]			
Pseudo R^2	0.1543				0.1681				0.1281			
Uncensored obs	4,763				2,375				3,454			
Left censored obs	23,449				25,837				24,758			
Observations	28,212				28,212				28,212			
Year = 2017	E.M.E	t-stat	I.M.E	t-stat	E.M.E	t-stat	I.M.E	t-stat	E.M.E	t-stat	I.M.E	t-stat
Financial Literacy	0.0326***	13.33	0.4236***	13.09	0.0130***	7.27	0.2782***	7.20	0.0270***	11.55	0.3851***	11.39
LR χ^2 (29); p value	5,430.52; p = [0.0000]				2,721.96; p = [0.0000]				4,136.92; p = [0.0000]			
Pseudo R^2	0.1533				0.1630				0.1407			
Uncensored obs	2,923				1,180				2,340			
Left censored obs	16,795				18,538				17,378			
Observations	19,718				19,718				19,718			

[a] *, **, *** denote 10, 5, 1 per cent levels of significance, respectively.
[b] E.M.E. indicates the marginal effects at the extensive margin; I.M.E. indicates the marginal effects at the intensive margin.
[c] All other control variables are included in this analysis.

Table 3.6 The Determinants of the Share of Risky Assets, High Risk Assets and Low Risk Assets - Cross-sectional Tobit Analysis

	Share of Risky Assets				Share of High Risk Assets				Share of Low Risk Assets			
Year = 2013	E.M.E	t-stat	I.M.E	t-stat	E.M.E	t-stat	I.M.E	t-stat	E.M.E	t-stat	I.M.E	t-stat
Financial Literacy	0.0252***	12.07	0.0194***	8.62	0.0159***	8.62	0.0129***	8.52	0.0160***	9.79	0.0158***	9.68
LR χ^2 (29); p value	4,982.08; p = [0.0000]				3,536.70; p = [0.0000]				2,774.80.80; p = [0.0000]			
Pseudo R^2	0.2571				0.2453				0.2309			
Uncensored obs	2,948				2,065				1,486			
Left censored obs	21,860				22,743				23,322			
Observations	24,808				24,808				24,808			
Year = 2015	E.M.E	t-stat	I.M.E	t-stat	E.M.E	t-stat	I.M.E	t-stat	E.M.E	t-stat	I.M.E	t-stat
Financial Literacy	0.0397***	19.14	0.0256***	18.65	0.0180***	10.65	0.3146***	10.53	0.0352***	17.53	0.5039***	17.11
LR χ^2 (29); p value	8,320.41; p = [0.0000]				5,143.48; p = [0.0000]				5,075.92; p = [0.0000]			
Pseudo R^2	0.3054				0.3098				0.2463			
Uncensored obs	4,763				2,375				3,454			
Left censored obs	23,449				25,837				24,758			
Observations	28,212				28,212				28,212			
Year = 2017	E.M.E	t-stat	I.M.E	t-stat	E.M.E	t-stat	I.M.E	t-stat	E.M.E	t-stat	I.M.E	t-stat
Financial Literacy	0.0319***	13.27	0.0206***	13.04	0.0117***	6.67	0.0108***	6.61	0.0274***	11.97	0.0167***	11.81
LR χ^2 (29); p value	5,133.31; p = [0.0000]				2,562.77; p = [0.0000]				3,803.79; p = [0.0000]			
Pseudo R^2	0.2978				0.2856				0.2719			
Uncensored obs	2,923				1,180				2,340			
Left censored obs	16,795				18,538				17,378			
Observations	19,718				19,718				19,718			

[a] *, **, *** denote 10, 5, 1 per cent levels of significance, respectively.
[b] E.M.E. indicates the marginal effects at the extensive margin; I.M.E. indicates the marginal effects at the intensive margin.
[c] All other control variables are included in this analysis.

Table 3.7 The Determinants of the Probability of Risky Asset, High Risk Asset and Low Risk Asset Holding in 2013, 2015 and 2017 - Fixed-effects Logit Analysis

	Risky Asset Holding			High Risk Asset Holding			Low Risk Asset Holding		
Unbalanced	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat
Financial Literacy	0.1740***	0.0001	4.43	0.1896***	0.0117	3.78	0.1938***	0.00002	4.26
Wald χ^2 (24); p value	573.25; p = [0.0000]			253.09; p = [0.0000]			754.30; p = [0.0000]		
Observations (NT)	9,970			5,541			8,945		

[a] *, **, *** denote 10, 5, 1 per cent levels of significance, respectively.
[b] All other control variables are included in this analysis.
[c] The number of observations is different from that in the panel analysis because the fixed-effects Logit estimator is only based on those households who changed their states over time.

Table 3.8 The Determinants of the Log Level of Risky Assets, High Risk Assets and Low Risk Assets in 2013, 2015 and 2017 - Random Effects Tobit Analysis

	Ln(Risky Assets)				Ln(High Risk Assets)				Ln(Low Risk Assets)			
Unbalanced	E.M.E	t-stat	I.M.E	t-stat	E.M.E	t-stat	I.M.E	t-stat	E.M.E	t-stat	I.M.E	t-stat
Financial Literacy	0.0323***	24.48	0.4196***	24.09	0.0149***	14.63	0.2738***	14.57	0.0259***	21.70	0.4105***	21.41
Wald χ^2 (33); p value	7,870.98; p = [0.0000]				3,848.63; p = [0.0000]				5,053.16; p = [0.0000]			
Chibar2 (01); p value	2,088.98; p = [0.0000]				2,426.04; p = [0.0000]				1,239.64; p = [0.0000]			
ρ: Coef.; Std.	0.4678; 0.0095				0.6208; 0.0102				0.4234; 0.0115			
Uncensored obs	10,873				5,735				7,444			
Left censored obs	63,927				69,057				67,350			
Observations	74,794				74,794				74,794			

[a] *, **, *** denote 10, 5, 1 per cent levels of significance, respectively.
[b] E.M.E. indicates the marginal effects at the extensive margin; I.M.E. indicates the marginal effects at the intensive margin.
[c] All other control variables are included in this analysis.

Table 3.9 The Determinants of the Share of Risky Assets, High Risk Assets and Low Risk Assets in 2013, 2015 and 2017 - Random Effects Tobit Analysis

	Share of Risky Assets				Share of High Risk Assets				Share of Low Risk Assets			
Unbalanced	E.M.E	t-stat	I.M.E	t-stat	E.M.E	t-stat	I.M.E	t-stat	E.M.E	t-stat	I.M.E	t-stat
Financial Literacy	0.0307***	23.73	0.0207***	23.37	0.0131***	13.15	0.0111***	13.09	0.0254***	21.75	0.0177***	21.49
Wald χ^2 (33); p value	7,546.90; p = [0.0000]				3,711.30; p = [0.0000]				4,730.05; p = [0.0000]			
Chibar2 (01); p value	2,118.26; p = [0.0000]				2,397.84; p = [0.0000]				1,197.36; p = [0.0000]			
ρ: Coef.; Std.	0.4632; 0.0093				0.6120; 0.0101				0.4084; 0.0113			
Uncensored obs	10,873				5,735				7,444			
Left censored obs	63,927				69,057				67,350			
Observations	74,794				74,794				74,794			

[a] *, **, *** denote 10, 5, 1 per cent levels of significance, respectively.
[b] E.M.E. indicates the marginal effects at the extensive margin; I.M.E. indicates the marginal effects at the intensive margin.
[c] All other control variables are included in this analysis.

Table 3.10 The Determinants of the Probability of Risky Asset, High Risk Asset and Low Risk Asset Holding - Cross-sectional IV Probit Analysis

	Risky Asset Holding			High Risk Asset Holding			Low Risk Asset Holding		
Year = 2013	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat
Financial Literacy	1.0307***	0.0281	14.81	0.9423***	0.0205	10.80	0.9467***	0.0162	10.08
First Stage									
Economics Education	0.2074***		12.12	0.2074***		12.12	0.2074***		12.12
Wald χ^2 (29); p value	8,030.56; p = [0.0000]			5,312.90; p = [0.0000]			3,622.27; p = [0.0000]		
Wald χ^2 (1); p value	58.17; p = [0.0000]			38.80; p = [0.0000]			33.08; p = [0.0000]		
Observations	24,808			24,808			24,808		
Year = 2015	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat
Financial Literacy	1.1090***	0.0450	26.82	1.0183***	0.0255	18.18	0.9234***	0.0351	12.50
First Stage									
Economics Education	0.2131***		12.58	0.2131***		12.58	0.2131***		12.58
Wald χ^2 (29); p value	16,286.46; p = [0.0000]			9,645.46; p = [0.0000]			7,428.72; p = [0.0000]		
Wald χ^2 (1); p value	104.92; p = [0.0000]			73.11; p = [0.0000]			43.39; p = [0.0000]		
Observations	28,212			28,212			28,212		
Year = 2017	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat
Financial Literacy	0.7845***	0.0333	6.01	0.8797***	0.0178	7.82	0.7392***	0.0260	5.27
First Stage									
Economics Education	0.2132***		8.99	0.2132***		8.99	0.2132***		8.99
Wald χ^2 (29); p value	5,245.60; p = [0.0000]			3,917.08; p = [0.0000]			3,836.59; p = [0.0000]		
Wald χ^2 (1); p value	12.72; p = [0.0004]			20.21; p = [0.0000]			10.66; p = [0.0011]		
Observations	19,718			19,718			19,718		

[a] *, **, *** denote 10, 5, 1 per cent levels of significance, respectively.
[b] All other control variables are included in this analysis.

Table 3.11 The Determinants of the Probability of Risky Asset, High Risk Asset and Low Risk Asset Holding in 2013, 2015 and 2017 - Pooled IV Probit Analysis

	Risky Asset Holding			High Risk Asset Holding			Low Risk Asset Holding		
Unbalanced	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat
Financial Literacy	1.0184***	0.0372	27.33	0.9546***	0.0222	21.58	0.8776***	0.0273	16.33
First Stage									
Economics Education	0.2162***		20.23	0.2162***		20.23	0.2162***		20.23
Wald χ^2 (31); p value	30,184.42; p = [0.0000]			19,172.00; p = [0.0000]			15,472.25; p = [0.0000]		
Wald χ^2 (1); p value	171.17; p = [0.0000]			132.43; p = [0.0000]			85.50; p = [0.0000]		
Observations (NT)	74,794			74,794			74,794		

[a] *, **, *** denote 10, 5, 1 per cent levels of significance, respectively.
[b] All other control variables are included in this analysis.

Table 3.12.A The Determinants of the Probability of Risky Asset Holding - Cross-sectional Logit Analysis - Financial Illiteracy

	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat
Year = 2013												
CF Interest Rate	-0.4783	-0.0408	-9.34		--			--		- 0.1532	-0.0125	-2.55
CF Inflation		--		-0.6027	-0.0513	-10.50		--		- 0.2599	-0.0213	-3.84
DK Stock Fund		--			--		-1.3861	-0.1137	-23.95	- 1.3094	-0.1071	-22.22
LR χ^2 (29); p value		5,369.33; p = [0.0000]			5396.38; p = [0.0000]			5937.85; p = [0.0000]			χ^2 (31): 5,978.06; p = [0.0000]	
Pseudo R2		0.2819			0.2833			0.3118			0.3139	
Observations		24,808			24,808			24,808			24,808	
Year = 2015												
CF Interest Rate	-0.7319	-0.0763	-15.91		--			--		- 0.3240	-0.0329	-5.89
CF Inflation		--		-0.7249	-0.0755	-15.87		--		- 0.3566	-0.0362	-6.52
DK Stock Fund		--			--		-1.8542	-0.1900	-23.74	- 1.6939	-0.1719	-21.41
LR χ^2 (29); p value		9,280.82; p = [0.0000]			9,279.77; p = [0.0000]			9817.28; p = [0.0000]			χ^2 (31): 9,979.59; p = [0.0000]	
Pseudo R2		0.3337			0.3337			0.3530			0.3588	
Observations		28,212			28,212			28,212			28,212	
Year = 2017												
CF Interest Rate	-0.5584	-0.0526	-9.97		--			--		-0.3677	-0.0344	-5.46
CF Inflation		--		-0.4225	-0.0400	-7.65		--		-0.0993	-0.0093	-1.49
DK Stock Fund		--			--		-0.9091	-0.0855	-12.55	-0.7891	-0.0738	-10.63
LR χ^2 (29); p value		5,497.96; p = [0.0000]			5,455.01; p = [0.0000]			5,569.94; p = [0.0000]			χ^2 (31): 5,627.09; p = [0.0000]	
Pseudo R2		0.3159			0.3134			0.3200			0.3233	
Observations		19,718			19,718			19,718			19,718	

[a] In this table, the explanatory variable CF Interest Rate is a dummy variable which equals 1 if the household head chooses “can’t figure out” for the “Interest Rate” financial literacy question; CF Inflation is a dummy variable, which equals 1 if the household head chooses “can’t figure out” for the “Inflation” financial literacy question and DK Stock Fund is a dummy variable which equals 1 if the household head chooses “don’t know stock”, “don’t know fund” or “don’t know both” for the “Risk Diversification” financial literacy question.
[b] All other control variables are included in this analysis.

Table 3.12.B The Determinants of the Probability of High Risk Asset Holding - Cross-sectional Logit Analysis - Financial Illiteracy

	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat
Year = 2013												
CF Interest Rate	-0.5036	-0.0345	-8.51		--			--		- 0.1922	-0.0128	-2.80
CF Inflation		--		-0.6085	- 0.0416	-9.08		--		- 0.2461	-0.0163	-3.15
DK Stock Fund		--			--		- 1.4317	-0.0952	-20.50	- 1.3515	-0.0896	-19.06
LR χ^2 (29); p value		4,228.57; p = [0.0000]			4,241.80; p = [0.0000]			4,658.76; p = [0.0000]			χ^2 (31): 4,693.03; p = [0.0000]	
Pseudo R2		0.2714			0.2723			0.2991			0.3013	
Observations		24,808			24,808			24,808			24,808	
Year = 2015												
CF Interest Rate	-0.6600	-0.0488	-11.01		--			--		- 0.3388	-0.0247	-4.81
CF Inflation		--		- 0.5862	- 0.0434	-10.03		--		- 0.2304	-0.0168	-3.35
DK Stock Fund		--			--		- 2.2252	-0.1632	-16.22	- 2.0896	-0.1525	-15.14
LR χ^2 (29); p value		6,861.43; p = [0.0000]			6,838.40; p = [0.0000]			7,204.33; p = [0.0000]			χ^2 (31): 7,274.70; p = [0.0000]	
Pseudo R2		0.3358			0.3347			0.3526			0.3560	
Observations		28,212			28,212			28,212			28,212	
Year = 2017												
CF Interest Rate	-0.4793	-0.0286	-6.22		--			--		- 0.2541	-0.0151	-2.79
CF Inflation		--		- 0.4276	- 0.0256	-5.63		--		- 0.1803	-0.0107	-2.01
DK Stock Fund		--			--		-1.0287	-0.0614	-8.85	- 0.9210	-0.0548	-7.80
LR χ^2 (29); p value		3,561.89; p = [0.0000]			3,554.34; p = [0.0000]			3,615.74; p = [0.0000]			χ^2 (31): 3,640.39; p = [0.0000]	
Pseudo R2		0.3107			0.3101			0.3154			0.3176	
Observations		19,718			19,718			19,718			19,718	

[a] In this table, the explanatory variable CF Interest Rate is a dummy variable which equals 1 if the household head chooses “can’t figure out” for the “Interest Rate” financial literacy question; CF Inflation is a dummy variable, which equals 1 if the household head chooses “can’t figure out” for the “Inflation” financial literacy question and DK Stock Fund is a dummy variable which equals 1 if the household head chooses “don’t know stock”, “don’t know fund” or “don’t know both” for the “Risk Diversification” financial literacy question.
[b] All other control variables are included in this analysis.

Table 3.12.C The Determinants of the Probability of Low Risk Asset Holding - Cross-sectional Logit Analysis - Financial Illiteracy

	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat
Year = 2013												
CF Interest Rate	- 0.4260	-0.0210	- 6.11		--			--		- 0.0948	- 0.0046	- 1.18
CF Inflation		--		- 0.6067	-0.0299	- 7.55		--		- 0.3158	- 0.0153	- 3.40
DK Stock Fund		--			--		- 1.3100	- 0.0636	- 15.64	- 1.2308	- 0.0596	- 14.45
LR χ^2 (29); p value		2,650.28; p = [0.0000]			2,673.57; p = [0.0000]			2,904.43; p = [0.0000]			χ^2 (31): 2,926.71; p = [0.0000]	
Pseudo R2		0.2327			0.2347			0.2550			0.2569	
Observations		24,808			24,808			24,808			24,808	
Year = 2015												
CF Interest Rate	- 0.7220	-0.0642	- 13.54		--			--		- 0.3276	- 0.0288	- 5.24
CF Inflation		--		- 0.7149	-0.0635	- 13.59		--		- 0.3659	- 0.0321	- 5.93
DK Stock Fund		--			--		- 1.6659	- 0.1474	- 18.70	- 1.4941	- 0.1312	- 16.55
LR χ^2 (29); p value		5,538.96; p = [0.0000]			5,540.73; p = [0.0000]			5,830.70; p = [0.0000]			χ^2 (31): 5,962.44; p = [0.0000]	
Pseudo R2		0.2532			0.2532			0.2665			0.2725	
Observations		28,212			28,212			28,212			28,212	
Year = 2017												
CF Interest Rate	- 0.5583	-0.0447	- 8.82		--			--		- 0.3992	- 0.0318	- 5.29
CF Inflation		--		- 0.3831	-0.0308	- 6.22		--		- 0.0544	- 0.0043	- 0.74
DK Stock Fund		--			--		- 0.8719	- 0.0698	- 10.55	- 0.7530	- 0.0600	- 8.89
LR χ^2 (29); p value		3,921.28; p = [0.0000]			3,879.69; p = [0.0000]			3,964.41; p = [0.0000]			χ^2 (31): 4,010.31; p = [0.0000]	
Pseudo R2		0.2729			0.2700			0.2759			0.2791	
Observations		19,718			19,718			19,718			19,718	

[a] In this table, the explanatory variable CF Interest Rate is a dummy variable which equals 1 if the household head chooses “can’t figure out” for the “Interest Rate” financial literacy question; CF Inflation is a dummy variable, which equals 1 if the household head chooses “can’t figure out” for the “Inflation” financial literacy question and DK Stock Fund is a dummy variable which equals 1 if the household head chooses “don’t know stock”, “don’t know fund” or “don’t know both” for the “Risk Diversification” financial literacy question.
[b] All other control variables are included in this analysis.

Chapter 4

Household Debt and Attitudes towards Risk:

Evidence from China

4.1 Introduction

Over the last two decades, there has been a significant increase in the level of household debt in China with an increase from around $517.7 billion in 2007 to around $7,200 billion in 2019.[66] Moreover, the gross domestic product (GDP) in China was $3,550 billion in 2007 and has increased to $14,280 billion in 2019.[67] This means that the proportion of household debt to GDP in China has increased by 35.84%. In contrast, in the U.S., the level of household debt was around $12,000 billion in 2007 and has increased to $13,544 billion in 2019, i.e. a much lower growth rate of 12.87%.[68] In addition, the GDP in the U.S. has increased from $14,452 billion in 2007 to $21,433 billion in 2019, which indicates that the proportion of household debt to GDP has even decreased by 19.84%.[69] In some developed countries, the level of household debt has actually fallen over the last two decades. For example, the level of household debt in the U.K. reached an all-time high of $3,226.6 billion in 2008 and then decreased to $2,482.5 billion in 2019.[70] Although such figures have led to policy-makers being concerned about financial vulnerability and the risk at the household level, there remains a shortage of academic research into the determinants of debt at the household level in China.

Although the debt itself is risk-free, it is an issue for Chinese households due to the extreme poverty in the last century and the traditional culture, which may place huge mental pressure and strain.[71] This is probably why China has the highest household saving rate in the world (OECD (2015). In addition, debt itself as an issue would lower peoples' happiness in China (Liu et al., 2020). Therefore, to some extent, not only does the repayment ability or excessive debt matter for a household but the probability of holding debt and the level of household debt also matter. As stated by Brown et al. (2005), a rise in household debt enables households to better smooth consumption and income to accommodate their various needs at different stages in the life cycle, but it may also place economic and psychological pressure on households. In addition, a higher level of household debt may not only affect the resilience

---

[66] Total household debt is defined as all liabilities of households that require payments of interest or principal by households to the creditors at a fixed date in the future. Debt is calculated as the sum of the following liability categories: loans (primarily mortgage loans and consumer credit) and other accounts payable. The definition of total household debt is the same for China, the U.S. and the U.K.. Data source: https://www.ceicdata.com/en/indicator/china/household-debt.
[67] Data source: https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?locations=CN.
[68] Data source: https://www.ceicdata.com/en/indicator/united-states/household-debt.
[69] Data source: https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?locations=US.
[70] Data source: https://www.ceicdata.com/en/indicator/united-kingdom/household-debt.
[71] Thrift is regarded as a traditional virtue promoted by Confucian ideology in China where children are always educated to be thrifty (Hofstede et al., 2005). Such a traditional cultural may be one reason why Chinese households prefer savings, which also leads Chinese households to be more debt-averse compared to the west (Wang et al., 2001).

of the economy to future shocks but it may also reinforce the existing distribution of wealth, making social and geographic mobility more difficult (Lowe, 2017). Thus, the importance of understanding what affects household debt levels is a key area for policy-making. There are, however, only a small number of empirical studies on household debt in China. For example, Cull et al. (2019) explore the influence of political connections, the social network and household demographic characteristics on formal and informal credit usage using the 2013 China Household Finance Survey (CHFS). The findings suggest that both political connections and social networks are positively associated with formal loans. We review the small existing literature on household debt in China in detail in the next section.

In this chapter, the focus lies on one particular determinant of debt holding and accumulation at the household level, namely attitudes towards risk, which have been empirically identified as an important influence on household debt in the U.S. (Brown et al., 2013). The reason why attitudes towards risk have attracted limited attention in empirical studies on Chinese households may result from the lack of available data and the limited number of studies on household debt at the household and individual level more generally. However, given that debt repayments are usually financed from household income, it is apparent that if there exists uncertainty in household income (due to, for example, redundancy, unemployment, or changes in real wages), then the household head's attitudes towards risk will potentially influence household debt holding, given the distribution of future income and interest rates (Brown et al., 2013). In other words, given the uncertainty surrounding the decision to acquire debt, the level of the household head's risk tolerance is associated with the probability of holding debt and the amount of debt held. For example, a household may want to take out a loan from a bank but the household head may have a low level of risk tolerance, which may reflect the possibility that the household is concerned about being not able to repay loans and thus lowers the probability of borrowing. Therefore, it appears intuitive to predict that the more risk-tolerant an individual is, the higher is the probability that they will hold debt and the higher is the amount of debt held.

In this chapter, we explore the relationship between attitudes towards risk and household debt from an empirical perspective. Specifically, we use the CHFS, as used in the previous chapters, focusing on waves 2011, 2013, 2015 and 2017. The respective number of households surveyed in each of the years is as follows: 8,438, 28,141, 37,289 and 40,011. To explore the robustness of the findings, a range of econometric approaches are employed in this chapter including the Logit model, the Tobit model, and the double hurdle (DH) model. To be specific, the Logit model is initially employed to explore the association between attitudes towards risk and household debt holding since our main dependent variable is measured by holding total household debt, which is a binary outcome.[72] For a robustness check, the Tobit estimator is used to investigate the relationship between risk attitudes and the amount of total household debt, which is generally regarded as a censored outcome in the existing literature (see, for example, Brown and Taylor, 2008). However, in order to allow for the possibility that households first decide whether to hold debt and then decide the level, we also model the level of total household debt using the double hurdle estimator. We further investigate the relationship between risk attitudes and the amount of household debt by splitting debt into the housing and non-housing categories.

Hence, we explore different types of debt holding – housing debt and non-housing debt. Home ownership is generally financed by mortgage debt, especially as households usually acquire this major asset early in their life cycle. Therefore, when house prices rise, households who own a house may expect that their higher wealth allows for greater lifetime consumption, more borrowing and spending (Turk, 2015). Recently, Cloyne et al. (2019) find evidence that house price appreciation induces homeowners to increase borrowing by extracting equity from their home. In China, house prices have been experiencing geometric growth over the past decade, which arguably has led to a substantial increase in the level of residential housing mortgage debt from around \$410 billion in 2007 to around \$3,900 billion in 2018 (The People's Bank of China).[73] Thus, we split total household debt into housing debt and non-housing debt in order to ascertain whether the relationship between debt and risk attitudes differs by type of debt.[74]

We find that households are more likely to hold total household debt if the household head is more tolerant towards risk based on the random effects Logit analysis, which is in accordance with the existing literature for the U.S. (see, for example, Brown et al., 2013). Furthermore, this chapter contributes to the literature on China by identifying a positive association between tolerance towards risk and total

---

[72] Total household debt in this chapter includes housing debt and non-housing debt.
[73] Data source: https://www.ceicdata.com/en/china/loan-consumer-loan/cn-consumer-loan-residential-housing-mortgage-loan.
[74] Housing debt includes mortgages from banks and loans from friends or relatives for housing-purchase/housing-renovation and non-housing debt includes agricultural/business debt, vehicle-purchasing debt, education debt, credit card debt and other debt.

household debt holding after controlling for unobserved heterogeneity. The random effects Tobit results reveal that $Risk\ Attitudes$ is positively associated with the log level of total household debt. Specifically, the marginal effects at the extensive margin indicate that the higher is the level of risk tolerance, the higher is the probability of holding total household debt and the marginal effects at the intensive margin indicate that among those households with a positive amount of total household debt, the higher is the level of risk tolerance, the larger is the amount of total household debt held. In addition, the findings are robust to using the double hurdle approach and suggest that the risk tolerance of the head of household is positively associated with the amount of household debt and that financial illiteracy has a positive impact on the probability of holding total household debt. Finally, a generally monotonic increase in the size of the positive effect of risk tolerance on debt is found in this chapter.

The rest of the chapter is organized as follows. Section 4.2 reviews the relevant existing research. Section 4.3 describes the data and methodology. Section 4.4 discusses the empirical results and conclusions are presented in Section 4.5.

4.2 Literature Review

This section reviews the existing literature on household debt in China and other developed countries. As stated above, there exists a limited number of studies on household debt in China, with one part of the existing literature exploring the determinants of household debt (see, for example, Fan et al., 2017, Han et al., 2019; Cull et al., 2019). Another area of the literature focuses on rural or farm households (see, for example, Xiang et al., 2014; Cui et al., 2017; Sun et al., 2018) and the remaining literature considers debt from an individual-level perspective (see, for example, Liao and Liu, 2012; Xin et al., 2020). The rest of this section discusses these areas in detail, focusing firstly on the U.S. and other developed countries, for purposes of comparison, and then on China.

4.2.1 The U.S. and other Developed Countries

There has been extensive research on household debt and borrowing in the U.S. (see, for example, Han and Li, 2011; Brown et al., 2013; Wildauer, 2016; Coibion et al., 2020) and other developed countries including Italy, Austria, Germany, the Netherlands, France, Switzerland, Belgium, Sweden, Denmark, Spain, Greece, the Czech Republic and Poland (see, for example, Crook and Hochguertel, 2007;

Georgarakos et al., 2014; Altundere, 2014; Cloyne et al., 2019). Han and Li (2011) shed light on the analysis of household borrowing after personal bankruptcy based on a household-level cross-sectional dataset from the U.S. Survey of Consumer Finances (SCF) in 1998, 2001, 2004 and 2007, where the respondents are asked the following questions: “have you (or your spouse) ever filed for bankruptcy?”; And if the respondents answer “yes”, they are then asked “when was the most recent time?” Thus, the main explanatory variable, bankruptcy filing, is a set of dummy variables including ever filed (overall bankruptcy) taking the value of 1 if the head of household (or his/her spouse) has ever filed for bankruptcy and bankruptcy history, i.e. 1 year earlier, 2-5 years earlier, 6-9 years earlier, and more than 9 years earlier, taking the value of 1 if the bankruptcy was filed 1 year earlier, 2-5 years earlier, 6-9 years earlier, and more than 9 years earlier, respectively.

They focus on three types of household debt: credit card debt, first-lien home mortgage, and vehicle loans, which serves to highlight the potential different effects of bankruptcy filing on secured and unsecured loans because credit card debt is unsecured and mortgage and vehicle loans are secured by collateral. Specifically, these types of household debt are measured by binary, continuous and censored outcomes and are modelled using Logit, Ordinary Least Squares (OLS) and Tobit estimators, respectively. Regarding credit card debt, this is defined in a number of ways: (i) a binary outcome taking the value of 1 if the household has a credit card; (ii) a continuous dependent variable capturing the ratio of the credit limit to household normal income; and (iii) the debt amount, which is measured by the ratio of card balances to household normal income and is a censored outcome. Normal income is defined based on the questions: “is your income reported for the previous year usually high or low compared to what you would expect in a ‘normal’ year, or is it normal?” If a respondent answers that it is normal, then the normal income is captured by the income reported for the previous year while if a respondent answers that the income was usually high or low, the SCF then asks, “what would your income have been if it had been a normal year?” For first-lien mortgage, the binary measure equals 1 if the household obtains a mortgage in a given year, and the continuous variable is measured by the loan-to-value (LTV) ratio at the start of the loan. Finally, for vehicle loan, the binary measure takes the value of 1 if the household acquires a vehicle loan in a given year and the continuous outcome is measured by the loan-to-income ratio.

There are two specifications for each regression. For the first specification, the key explanatory variable, ever filed, is included, while, for the second, the four bankruptcy history variables, 1 year earlier, 2-5 years earlier, 6-9 years earlier, and more than 9 years earlier, are included simultaneously. Since the SCF does not have information on when the credit cards were issued, they pool the four cross-sectional datasets. In contrast, the SCF asks retrospectively when mortgage and vehicle loans started, thus they construct pseudo panel data for these two types of secured loans. Specifically, they created a pseudo sample for each of the 5 years prior to the survey. They restricted the sample to within 5 years prior to the survey because they measured borrowing costs using the spread of the vehicle loan interest rate over the yield of 5-year Treasury securities in the year of the vehicle purchase. Thus, from the survey information on the year of loan origination, they inferred whether a mortgage or a vehicle loan was taken out in each of the 5 years. In addition, for each pseudo sample, they included those survey variables that are either time-variant or likely to be time invariant as controls. The time variant variables included the head of household's age and the number of years since the last bankruptcy filing. The potential time-invariant variables included the quintile of normal household income, sex, race, educational attainment, risk aversion, and attitudes toward borrowing various types of debt.

The results from the Logit estimation for credit card loans suggest that the odds ratio estimates indicate that the likelihood of a bankruptcy filer who has a credit card loan, regardless of time, is about 50% less than of that of a non-filer. In addition, the results from the second specification indicate that the probability of a household that filed for bankruptcy 1 year earlier having a credit card loan is 83% less than the probability of a comparable non-filer having a credit card loan. The probability decreases to 46% for those who filed 2-5 years earlier and to 24% for those who filed 6-9 years earlier and for those who filed over 9 years earlier. In the case of the continuous credit debt measure, the ratio of the credit limit to household normal income, the results of the OLS analysis show that bankruptcy has a negative effect on the credit limit to household normal income ratio and the effect remains constant over time except when time since filing is over 9 years. Finally, conditional on having a credit card, filers have a slightly higher debt balance relative to their normal income.

Turning to mortgage debt, the Logit results suggest that the effect of overall bankruptcy (ever filed) is not statistically significant but the effect of bankruptcy history on the likelihood of having a mortgage is negative for filing 1 year earlier and insignificant for those who filed 2-9 years ago, but it turns positive for those who filed more than 9 years earlier. Then the OLS results suggest that, conditional on having obtained a mortgage, those filers who filed 6-9 years earlier have a higher loan-to-value ratio on their mortgages than comparable non-filers. In addition, in the case of vehicle loans, the results from the Logit analysis imply that a current or previous bankruptcy filer is more likely to have a new vehicle loan than a non-filer and the effect is statistically significant over time except for those who filed more than 9 years earlier.

In addition, Han and Li (2011) control for several time-invariant measures, namely the household head's age at the loan origination, race, education and attitudes toward financial risk and debt reported at the time of the survey. This is aimed at mitigating the effect of the timing mismatch issue caused by measurement errors because of the mismatch between the timing of the survey and loan origination. In other words, the dependent and independent variables may not be valued at the same time due to the fact that the SCF is a cross-sectional dataset where household characteristics and financial conditions at the time of the loan application are not directly observable.

In summary, Han and Li (2011) find an inverse relationship between household secured debt and past and/or present personal bankruptcy but, with respect to a potential endogeneity issue, it is important to acknowledge that they do not instrument bankruptcy filing so the estimates may be biased because causality could operate in both directions. Moreover, the measure of credit card loans based on whether the respondent has a credit card is arguably not appropriate because owning a credit card does not directly indicate whether debt, i.e. a loan, has been accrued on the card.

Brown et al. (2013) investigate the correlation between household debt and attitudes towards risk based on a household-level unbalanced panel dataset from the U.S. Panel Study of Income Dynamics (PSID). They focus on household debt, which is available in the PSID for 1984, 1989, 1994, 1999, 2001, 2003, 2005 and 2007, although the information relating to attitudes towards risk is only available in the 1996 PSID and, consequently, attitudes towards risk are treated as being time-invariant. To measure household debt, the head of household is asked the following questions: "Aside from the debts that you have already talked about, like any mortgage on your main home, do you (or anyone in your family) currently have any other debts such as for credit card charges, student loans, medical or legal bills, or on loans from relatives? If you added up all of these debts (for all of your family), about how much would they amount to right now?" Thus, household-level unsecured debt is obtained from the response to this question. Meanwhile, household secured debt is based on the question: "Do you have a mortgage on this property? About how much is the remaining principal on this mortgage?" Total household debt is then constructed by the summation of unsecured and secured debt.

The main explanatory variable of interest, risk attitudes, is based on five questions related to hypothetical gambles with respect to lifetime income such as: "Suppose you had a job that guaranteed you income for life equal to your current total income. And that job was (your/your family's) the only source of income. Then you are given the opportunity to take a new, and equally good, job with a 50-50 chance that it will double your income and spending power. But there is a 50-50 chance that it will cut your income and spending power by a third. Would you take the new job?" The respondents can choose "yes" or "no". The five questions accord with the general approach to classifying individuals in terms of their attitudes towards risk according to their marginal utility income, with the relatively more risk averse individuals characterized by marginal utility of income diminishing at a relatively fast rate. Thus, Brown et al. (2013) assign a range of risk tolerance coefficients to each gamble response category and formulate the categorical sequence of gamble responses into a single cardinal measure of preferences, rather than treating the response as an index because of the assumption that individuals have constant relative risk aversion utility. So, given the gambles above, individuals will accept the risky job when their expected utility is greater than the expected utility of their current/safe job, and true tolerance is log-normally distributed and survey response error is purely random measurement error.

A univariate Tobit model with Mundlak fixed effects is employed to explore how household attitudes towards risk, which are decreasing in risk tolerance, affect the log level of household total debt, the log level of unsecured debt and the log level of secured debt, since these three dependent variables are censored outcomes. The results from the univariate Tobit analysis show that risk attitudes are inversely associated with each type of debt. Specifically, by multiplying the marginal effect of cardinal risk attitudes by the standard deviation, they find that a one standard deviation increase in the cardinal risk attitudes measure reduces unsecured, secured and total debt by about 20.7, 11.4 and 15.3 percentage points, respectively. Compared to the marginal effects of other covariates in terms of magnitude, the effect of risk attitudes on the level of debt is relatively large. For example, a one standard deviation increase in household labour income is associated with an increase in unsecured, secured, and total debt of 16.1%, 39.3% and 13.4%, respectively.

In addition, Brown et al. (2013) shed light on the possible endogeneity of risk attitudes using an instrumental variable, the log expected value of the gamble. In order to test the validity of the instrument, they explore exogeneity by testing whether the residuals from the first stage regression are statistically significant in the debt equation. The residuals from the first stage are found to be statistically insignificant in the outcome equation and the instrument is statistically significant in the risk attitudes equation, endorsing the validity of the instrument. Overall, the results are found to be robust and the negative relationship between risk attitudes and debt is statistically significant.

For robustness, they explore the correlation between risk attitudes and debt by employing quantile regression analysis, which provides a fuller description of the effects of risk attitudes across the entire debt distribution. They focus on those heads of household who hold that particular type of debt due to the truncated nature of the dependent variable. The results of the quantile regression analysis suggest that risk aversion is inversely associated with unsecured, secured, and total debt with the magnitude deceasing monotonically across the distribution (where statistically significant). Moreover, the effect of risk attitudes does not attain statistical significance at the top two deciles of the unsecured debt distribution indicating that, for those who hold the highest levels of unsecured debt, the level of unsecured debt is not influenced by risk attitudes.

Wildauer (2016) explores household debt in the U.S. based on a household-level pooled dataset from the Survey of Consumer Finances (SCF) between 1995 and 2007. The dependent variable is a continuous one defined as the total change in a household's liabilities relative to income. The reason for using the annual change in total outstanding liabilities instead of the total stock of debt is that only the change is directly related to the flows of the current period whereas the stock relies on past decisions, which are not observed. Specifically, based on the information on whether the debt was taken out in the year of the interview or prior to the year of interview, the change in household liabilities is classified into seven cases, where the first four changes are based on the case of the debt taken out in the year of the interview and the remainder are based on the case of the debt taken out prior to the year of interview: (1) the difference between the amount currently outstanding and the amount initially borrowed if the mortgage was used to refinance an earlier credit commitment and taken out in the year of the interview; (2) the amount extracted if the mortgage was used to extract equity from the residence; (3) the amount extracted plus the difference between the amount currently outstanding and the amount initially borrowed if the mortgage was used to extract equity and to refinance an earlier loan and was taken out in the year of the interview; (4) the amount currently outstanding if the household had no prior loan or mortgage; (5) the amount of the principal repaid computed as the difference between the total regular or typical annual payment the household makes and the implicit interest payments based on the reported interest rate for the loan and the current outstanding amount if the mortgage was taken out before the year of interview; (6) zero if the household only pays interest but no principal for a certain period of an interest-only agreement and the mortgage was taken out before the year of interview; and (7) the negative principal repayment (same as situation 5) if the household has fallen behind in payments and the mortgage was taken before the year of interview. Therefore, the total change in household liabilities is the sum of the above changes for the seven debt categories covering the years 1995, 1998, 2001, 2004 and 2007 (the first, second, and third mortgages on the primary residence, mortgages on other residential property, consumer loans, car and vehicle loans, education loans, other loans for property purchase and home improvements).

The explanatory variables include total household income, the current value of the primary residence and other residential real estate minus the value of any potential capital gains on residential real estate purchased in the current period, and the value at the time of purchase of residential real estate obtained in the current period. This is to distinguish between housing assets, which existed before the beginning of the period of the interview and the value of residential real estate purchased in the year of the interview. They also control for those households whose income is lower than in a normal year by means of an indicator variable based on the question: “is your income reported for the previous year unusually high or low compared to what you would expect in a ‘normal’ year, or is it normal?” Moreover, since the aim of the paper is to estimate the correlation between relative income and household borrowing, three measures of peer group income are included. The first two measure the absolute degree of inequality within education-race groups (college/less-than-college and black/white) and are defined as the q^{th} percentile of the education-race group income distribution and the average income above the

q^{th} percentile. The third definition is the cumulative distribution function of income within education-race groups, which corresponds to a head count of households richer than this household.[75] Finally, the outstanding liabilities at the end of the previous period are incorporated as a control since most household borrowing is repaid over a lengthy period of time and household financial wealth is also included.

A pooled OLS estimator with probability weights is employed to estimate the model for three different samples: the borrowing (dependent variable > 0) and non-borrowing/repayment (dependent variable ≤ 0) subsamples and the full sample, where borrowing is defined as an increase in outstanding liabilities held in 1995 and non-borrowing/repayment as a decrease. Weighted estimation is important here because oversampling of households from the upper tail of the income and wealth distributions would lead to them being overrepresented if all observations were implicitly assigned equal weights as in case of un-weighted estimation. The results suggest that households borrow less debt relative to income with rising income levels conditional on being in the borrowing subsample. Specifically, a 1% increase in household income is associated with a 31.4% decrease in household debt relative to income conditional on household debt relative to income being greater than zero. Moreover, higher levels of housing wealth are associated with more borrowing.

In summary, Wildauer (2016) finds evidence of a negative relationship between household income and household debt, but there are a limited number of covariates controlled for in the analysis, with variables such as risk attitudes not included, which have been identified as an important determinant of household debt in the U.S. (see Brown et al., 2013).

In a similar vein, Coibion et al. (2020) investigate the role of local income inequality for the relationship between a household's debt accumulation and their rank in the local income distribution. They find that low-income households borrow relatively less in high-inequality areas than low-income households in low-inequality areas in the U.S. based on two datasets, the New York Federal Reserve Bank Consumer Credit Panel/Equifax (CCP) and the Survey of Consumer Finances (SCF). Specifically, to measure household debt, they focus on the CCP data, which is a quarterly panel of individuals without infor-

[75] The q^{th} in the first two definitions of peer group income indicates the cut-off point including the 80^{th}, 90^{th}, 95^{th} and 99^{th} percentiles of the income distribution.

mation on income. They aggregate individual records into household records using the household identifiers because the CPP database contains information on all individuals with credit files residing in the same household as the individuals residing in the primary sample and the household members are added to the sample based on the mailing address in the existing credit files. They use data from the third quarter of the CCP from 2001-2012 to maximize the match with the SCF survey and they use the third quarter of each subsequent year to generate annual measures of household debt.

Regarding income rank, they impute income for the households in the CPP by using the information on income from the SCF to estimate how household income relates to debt and demographic characteristics available in both the CPP and SCF datasets using the OLS estimator. After obtaining the expected log income (i.e. predicted income), the estimated error terms from the SCF are used to impute the household's income rank in the household's geographic area and the distribution of income in that area. Finally, local inequality is measured by the difference between expected log income at the 90th percentile and expected log income at the 10th percentile in 2001 because it is predetermined relative to subsequent household debt accumulation decisions.

The dependent variable, household debt accumulation, is defined as the change in household debt between 2001 and the relevant year (2002 to 2012) relative to the household's imputed expected income in 2001. The main independent variables are the household's income rank in the 2001 local income distribution, local income inequality in 2001 and their interaction. The results show that the coefficient on a household's rank in the income distribution is negative, which indicates that household debt accumulation (relative to income) is, on average, greater for low-income households. Moreover, the estimated coefficient of inequality level is negative indicating that households living in the more unequal areas within the county accumulated less debt than those in lower inequality areas in the same county. Finally, the estimated coefficient of the interaction between income rank and income inequality is positive, which suggests that debt accumulation is relatively high for high-income households in high-inequality regions as compared to low-inequality regions.

In addition to research on household debt in the U.S., Crook and Hochguertel (2007) compare household debt in the U.S. and three European countries, Spain, Italy and the Netherlands, by exploring the determinants of household debt holding based on four datasets. Specifically, they explore: a household panel dataset for Italy from the Survey of Household Income and Wealth (SHIW) consisting of seven waves from 1991 to 2004; a household-level panel dataset for the Netherlands from the Dutch National Bank Household Survey (DHS) from 1993 to 2006; the first wave (2002) of the Bank of Spain's Survey of Household Finances (EFF); and a cross-sectional dataset for the U.S. from the Survey of Consumer Finances (SCF) from 1992, 1995, 1998 and 2001. The dependent variable, the household debt level, measures the log level of the amount of total debt held by a household and is modelled on a vector of independent variables including age, gender and the educational attainment of the household head, household net worth, and household income.

They explore the determinants of the amount of debt by using a random effects estimator for the Netherlands and Italy. A pooled OLS estimator is used for the U.S. and a standard Tobit estimator is used for Spain where the dependent variable is a censored outcome.[76] The results show that household net worth is inversely associated with the amount of debt in the Netherlands and Italy. Specifically, the estimated coefficients of household net worth are -0.0029 and -0.0033 on the household debt level for the Netherlands and Italy, respectively. The coefficient of household net worth is -0.009 for the U.S. Moreover, the age of the household head has a significant role in all four countries, especially in Italy and the U.S., where the heads of household are less likely to apply for a loan as age increases. However, in the random effects Tobit model, they did not control for the fixed effects using Mundlak corrections, which may bias the estimates.

Similarly, Georgarakos et al. (2014) investigate the association between household debt and social interaction in the Netherlands based on a household-level pooled cross-sectional dataset from the DHS over the period, 2001 to 2008. They measure the dependent variable, household debt, including both collateralized and uncollateralized loans, in two ways: firstly, holding each type of loan, which is a binary outcome; and, secondly, the log level of the outstanding amount of a particular loan type, which is a censored outcome. Social interaction is measured by the perceived average amount of income of the household's peers, where respondents are asked to provide an estimate of the income of their peers.

---

[76] The reason why the dependent variable is censored only for Spain is that the households in the Netherlands, Italy and the U.S. used in the estimations all have positive debt.

They first model the prevalence and amount outstanding for each loan type, collateralized and uncollateralized loans, using a Probit model for holding a loan type and OLS and Tobit models for amounts outstanding. The marginal effect of average peer income on the prevalence of holding a collateralized loan from the Probit analysis is 4%, which means that a one percent increase in average peer income is positively associated with a 4% increase in the probability of having a collateralized loan, which is larger than that on an uncollateralized loan in terms of magnitude, where, for the latter, a one percent increase in the average income of peers is positively associated with the probability of having an uncollateralized loan of 1.6%. Regarding the amount of outstanding loans, the OLS results suggest that a one percent increase in average peer income is positively associated with an increase in the amount of collateralized and uncollateralized loans of 46.6% and 16.2%, respectively, which accords with the results from the Tobit estimation.

Additionally, they perform instrumental variable OLS (IVOLS), instrumental variable Probit (IVProbit) and instrumental variable Tobit (IVTobit) analysis to account for the potential endogeneity of the income of peers with the difference in educational attainment between each respondent and his/her peers as an instrument on the basis that it can influence the respondent's perception of their peers' average income but it is arguably not related to the respondent's household debt. The results from the IV analysis are consistent with the findings from the standard OLS, Probit and Tobit analysis in that the average income of peers is positively associated with the probability of having a collateralized loan and an uncollateralized loan as well as the amount of collateralized and uncollateralized loans after controlling for endogeneity. However, they fail to reject the hypothesis of exogeneity in all IVs, which means that the estimates from the OLS, Probit and Tobit analysis are more efficient compared with their IV counterparts since using the IV approach will needlessly increase the standard errors, i.e. the IV estimators will be inefficient.

Altundere (2014) explores the relationship between social interaction and household debt in Europe using the household-level cross-sectional data for 2006-07 from the Survey of Health, Ageing and Retirement in Europe (SHARE) covering thirteen Europe countries: Austria, Germany, the Netherlands, France, Switzerland, Belgium, Sweden, Denmark, Spain, Italy, Greece, the Czech Republic and Poland. They exploit household debt based on two debt types, mortgage debt and non-mortgage debt, in two ways: the decision to hold debt and the log level of the amount of debt held. The former is a binary outcome taking the value of 1 if the household holds debt and the latter is a censored outcome, which measures the amount of outstanding debt. Thus, Probit and Tobit estimators are used for estimation with the main explanatory variable, social interaction (sociability) being based on the question: “Have you done any of these activities in the last month?” If the household has done at least one of these four activities, namely: (1) voluntary or charity work; (2) attended an educational or training course; (3) gone to a sport, social or other kind of club; and (4) taken part in a political or community-related organization, then the household is classified as “sociable”, otherwise, as “non-sociable”. Hence, the independent variable, sociability, is dummy variable taking the value of 1 if the household is sociable.

The results from the Probit estimation suggest that the marginal effect of sociability on the probability of having a mortgage is 0.0155, which indicates that sociable households are 1.6% more likely to have a mortgage compared to non-sociable households. The Tobit results show that the marginal effect of sociability on the log level of the amount of mortgage debt, conditional on having such debt, is 0.1773, which indicates that sociable households have a 17.73% higher amount of mortgage debt than non-sociable households. Regarding non-mortgage debt, the results from the Probit estimation imply that sociable households have a 2.7% higher probability of incurring non-mortgage debt than non-sociable households. Finally, the Tobit results for non-mortgage debt show that sociable households with non-mortgage debt borrow around 25% more than the non-sociable households.

When people get into financial trouble, they may ask friends for loans or obtain financial advice from their friends, which means sociable households may be more likely to obtain a loan directly or take advice about how to get a loan. However, this line of causality may be reversed because financially troubled households may be more likely to engage in social activities in order to obtain money from their engagements or to meet people for financial advice. Therefore, Altundere (2014) further investigates the possible reverse causality between social interaction and household mortgage and non-mortgage debt because indebted households may engage more frequently in social activities. Specifically, respondents who have taken part in social activities are asked about five possible motivations for the social participation: (1) to meet other people; (2) to contribute something useful; (3) because I am needed; (4) to earn money; and (5) to use my skills or to keep fit. “To meet other people” is the most popular motivation and this motivation may not only indicate a desire to make friends but it also may have the aim to become acquainted with wealthy or financially informed people. So, to eliminate the possible effect of the latter, Altundere (2014) includes an additional control variable that represents the motivation for participation in social activities, i.e. to meet other people. The results suggest that the estimated effects of sociability are virtually unchanged for both mortgage and non-mortgage debt as compared to the baseline estimations.

In general, this study finds a positive relationship between household social interaction and household mortgage and non-mortgage debt in Europe based on Probit and Tobit estimators. Moreover, the possibility of reverse causality between social interaction and household debt has also been considered but it should be acknowledged that they did not apply an IV estimator to solve the potential endogeneity of social interaction caused by reverse causality.

4.2.2 China

Since the level of household debt in China is relatively low compared with developed countries, it is perhaps not surprising that there is a relatively small literature on household debt over the past decades in China. However, household debt is definitely becoming an important component of household finance in China since the total amount of household debt has been increasing rapidly in China recently, which has attracted attention among academics. One example is Fan et al. (2017) who find that the household's social network is positively associated with the amount of informal borrowing for house-purchasing based on a household-level dataset from the second wave (2013) of the China Household Finance Survey (CHFS). The dependent variable, the amount of informal borrowing, is a censored outcome, defined as the log level of the amount of informal borrowing from relatives/friends for purchasing the current residential unit. The key explanatory variable, social network, is measured by three variables, namely: (i) the number of relatives living in the same city; (ii) a dummy variable, which equals 1 if the household head has a local residence permit; and (iii) social net wealth defined as the log of the level of annual social network income minus the log of the level of annual social network expenditure.[77] The empirical analysis includes two steps: (1) testing the effect of informal borrowing on mortgage

[77] Social network income and expenditure are defined as the amount of cash gifts the household received from, or gave to, nonfamily members due to festivals, weddings, funerals, and birthdays, respectively.

borrowing from commercial banks; and (2) testing the effect of informal borrowing on housing demand. However, there are two potential problems in estimating such a model. First, informal and formal borrowing cannot be fully observed due to the censoring problem. Second, informal and mortgage borrowing could be jointly decided by households' latent characteristics that cannot be directly controlled for. Thus, to solve these problems, they apply the conditional-recursive mixed process (CMP) estimator to explore the effect of the social network on informal borrowing.[78] They also explore how informal borrowing affects the probability of having a mortgage/formal loan from a commercial bank when purchasing the current residential unit, which leads to potential omitted variable bias since informal and formal borrowing may be generated with correlated errors.

The results of the CMP analysis suggest a positive relationship between the social network and informal borrowing from friends/relatives for purchasing the current residential unit. To be specific, informal borrowing increases by 2.62% if the household has one more relative living in the same city. When the household head has a local residence permit, household informal borrowing increases by 6.47% and informal borrowing increases by 0.3% if household social net wealth increases by 1%.

In a similar vein, Cull et al. (2019) investigate the influence of political connections, the social network and household demographic characteristics on formal and informal credit usage using a household-level dataset from the 2013 China Household Finance Survey. They first focus on total credit usage measured by a binary outcome, which equals 1 if the household has taken out a loan from banks or other informal sources. Then, they split total credit usage into formal and informal types, where formal credit usage is measured by a binary outcome, which equals 1 if the household has a loan from a bank, while the informal one is a binary outcome, which equals 1 if the household has a loan from other informal sources (i.e. non-bank loans). Regarding the explanatory variables, political connections are measured by a dummy indicator, which equals 1 if any of the household members are a Communist Party member. The social network is captured by the number of siblings that the household head and his/her spouse have. Household demographic characteristics include the age, health status, labour market status, the educational attainment and financial knowledge of the household head, the number

[78] The conditional-recursive mixed process estimator allows for continuous and non-continuous dependent variables in individual equations by using the maximum likelihood approach to estimate equations as a system (Roodman, 2011).

of children in the household, the number of adults of working age in the household and whether the household lives in a rural or urban area.[79]

They apply linear probability models for estimation. The results for the total sample suggest that households are 3.2% more likely to take out a bank loan if anyone in the household is a Communist Party member and that the number of siblings is positively associated with the probability of having loans, bank loans, and non-bank loans of 4.3%, 0.8% and 4.8%, respectively. Moreover, the household has a 9% higher probability of having loans from banks or other sources if the household lives in a rural area. In addition to considering the total sample, they split the households into urban and rural ones because they find that rural and urban residents differ on a number of characteristics, which may affect the source of their loans. For example, rural households tend to have a larger family size and social networks in terms of the number of siblings. The results for the urban sample suggest that the number of dependent children is positively associated with the probability of having loans of 5.2%, which is greater than that in terms of size for the rural sample with 3.4%.

Overall, this paper identifies some interesting determinants of household loans, including formal and informal loans. However, splitting the sample based on households living in a rural or urban area may cause sample selection bias, which is not accounted for in the analysis. For example, households living in rural areas are arguably more likely to borrow from informal sources rather than banks, which may bias the effects of political connections and social networks on informal loans behaviour because the rural subsample is not representative of the population as a whole.

In addition to research on the determinants of household debt from banks, relatives or friends, Han et al. (2019) explore how financial knowledge and risk attitudes affect person to person (P2P) borrowing (via internet loans) using a household-level cross-sectional dataset from the China Survey of Consumer Finances in 2011. A Multinomial Logit estimator is used since the dependent variable, P2P borrowing, is based on the question: “Have you used P2P loans?”; where 1 indicates those who used such loans successfully; 2 indicates those who applied but failed; and 3 refers to those who did not apply. They set

[79] Financial knowledge is measured using the same financial literacy questions about interest rates and inflation from the CHFS as analysed in the previous chapter of this thesis.

those who did not apply as the base outcome category in the Multinomial Logit regression. The independent variable, risk attitudes, is based on the question: "Assuming that a coin is tossed, you will get 2,000 RMB if it comes up heads, you will get nothing if it comes up tails. Supposing you resell such a profit opportunity, what is the lowest amount you would charge for it?"

The second main explanatory variable, financial knowledge, is measured in two parts: familiarity and expertise. Specifically, the familiarity element is captured by two variables, where the first one is about financing: "Have you ever used financing for a specific purpose?" The options include "yes, financed enough money", "only financed some money", or "did not finance any money". They assign this financing variable the values of 1, 2 and 3, corresponding to the three options, respectively. The second familiarity variable is related to their internet borrowing experience: "Have you ever used online products?" Respondents can choose between: "I have applied for it successfully"; "I have applied for it but failed", or "I have never applied for it". This experience variable is also assigned values of 1, 2 and 3 for the above options, respectively. Expertise is captured by three aspects, namely the respondent's industry background, higher education in finance and the level of knowledge of internet financing. First, focusing on the respondent's industry background, a dummy variable is used, which equals 1 if the respondent or his/her spouse works in finance. Second, a dummy variable is used to measure higher education, which equals 1 if the respondent or his/her spouse received higher education in economics or management. Finally, three dummy variables are used to capture the respondent's level of knowledge of internet financing. Specifically, the first one equals 1 if the respondent shops around when selecting a P2P platform, the second one equals 1 if the respondent knows that P2P has financing fees and management fees, and the third dummy variable equals 1 if the respondent realizes that there exists default risk on P2P platforms. In order to reduce the number of variables, factor analysis is used to construct two main variables, the familiarity factor and the expertise factor.

The Multinomial Logit results suggest that an increase in risk aversion is negatively associated with the probability of using P2P loans successfully relative to those who did not apply, while an increase in financial knowledge (captured by familiarity and expertise) is positively associated with the probability of using P2P loans successfully as compared to those who did not apply. A similar pattern can be found for those who applied for a P2P loan but failed relative to those who did not apply.

Additionally, they use the number of elderly individuals (those aged over 60) in the household as an instrumental variable to deal with the potential endogeneity of risk attitudes, based on the argument that due to the needs of supporting the elderly, the respondent's economic burden will increase with the number of elderly individuals in the household thereby making them more risk averse. A 2SLS estimator is used with the results suggesting that the financial knowledge variables (familiarity and expertise) have a significant and positive impact on P2P borrowing. Specifically, a 1% increase in financing familiarity and financing expertise is positively associated with 17.84% and 26.8% increases in the probability of using P2P loans successfully, respectively.

To summarise, they find a negative relationship between risk aversion and the probability of using P2P loans successfully whilst a positive correlation is found between financial knowledge (familiarity and expertise) and P2P borrowing. However, using the number of elderly individuals in the household as a proxy of risk attitudes is arguably not an appropriate way to solve the potential endogeneity problem because, as discussed by Han et al. (2019), the number of elderly individuals is correlated with the risk attitudes. Thus, the unobserved variables in the error term may be correlated with both P2P loans and the number of elderly individuals, which may cause potential endogeneity problems.

In addition, they did not consider the potential endogeneity of financial knowledge in the sense that the experience of using P2P loans may improve the respondent's financial knowledge as well, i.e. the causality may operate in both directions.

Turning to the borrowing behaviour of rural households and farm households, Xiang et al. (2014) examine the influence of non-governmental organizations on formal and informal credit use based on a household-level panel dataset constructed from surveys conducted by the authors from 2006 to 2009, with 749 rural households in total. They first asked farmers whether or not they had received loans in each of the past five years from: the China Foundation for Poverty Alleviation (CFPA) microfinance, which is a non-governmental organization; formal credit institutions such as the Agricultural Bank of China or the Agricultural Development Bank of China and Rural Credit Cooperatives; or from informal networks such as friends and relatives. Thus, the explanatory variable, CFPA microfinance, is a dummy variable, which equals 1 if the household received loans from CFPA microfinance in this year. They also include the lag of CFPA microfinance, which equals 1 if the household received loans from CFPA microfinance in the previous year. A Multinomial Logit estimator is applied because the dependent variable, credit, has four unordered categories - having formal credit only, having informal credit only, having both formal and informal credit, and no credit (the base group). They also apply a fixed-effects Logit model to explore the relationship between CFPA microfinance and the probability of household borrowing from informal and formal sources.

The results from the Multinomial Logit estimation show that the farmer's decision to borrow from CFPA microfinance in the previous year is negatively associated with the probability of borrowing from formal institutions relative to those who have zero loans. Specifically, a farm, which borrowed from CFPA microfinance in the previous year, has a 2% lower probability of borrowing from formal institutions relative to those who have zero loans. Additionally, the results from the fixed-effects Logit model suggest that the farm's decision to borrow from CFPA microfinance in the previous year is negatively associated with the probability of borrowing from formal networks. However, it should be acknowledged that the authors' dataset may not be representative of rural China due to the limited sample size. In addition, the loan amount is not analysed.

In a similar vein, Cui et al. (2017) investigate the determinants of rural household credit using a cross-sectional household-level dataset based on a survey conducted by the authors, which is based upon a relatively small sample size, i.e. 489 observations. Logit and Tobit estimators are used for the estimation, respectively, because there are two dependent variables, where the first one is a binary outcome indicating if the household borrowed from a bank and the second one is a censored outcome defined as the log level of the amount of loans from banks. The explanatory variables include: (i) 'property level' (i.e. total household income, the rural household agricultural land size and the household saving rate); (ii) social relations (i.e. the occupation of the household head and the relationship between the household head and relatives, which is an index, where 1 denotes a poor relationship and 5 denotes a very close relationship); (iii) trading costs (i.e. the perceived complexity of borrowing formalities, which is an index, where 1 denotes cumbersome and 5 denotes simple, and the distance to the market centre from home); and (iv) financial information (i.e. the level of the financial knowledge of the household head, the degree of understanding of the credit policies of financial institutions and the credit guidance received by formal finance institutions). Household characteristics controlled for in the analysis are age, gender, educational attainment, and the health status of the household head, and the number of family members and the number of individuals in full-time education.

The Logit results suggest that the age of the household head is positively associated with the probability of borrowing from banks while total household income is inversely associated with the likelihood of borrowing. Having a household head in poor health is inversely related to the household borrowing from banks. Those household heads with a higher level of financial knowledge are also less likely to borrow, while household heads with a higher degree of understanding of the credit policies of financial institutions and the credit guidance have a higher probability of borrowing. Regarding the amount of the loans from banks, the Tobit results suggest that the households headed by men have a higher amount of loans from banks, while the household saving rate and household income are negatively associated with the amount of loans.

Overall, they explore the determinants of rural household formal borrowing behaviour through both Logit and Tobit analysis but they do not focus on marginal effects, which means they only provide statistical evidence relating to the direction rather than the magnitude of the effects. Additionally, they do not consider the potential endogeneity of determinants such as financial knowledge, which may bias the estimates.

Sun et al. (2018) investigate the correlation between social capital and the ability of farm households to access formal and informal loans, using the second wave of the household-level dataset from the 2013 China Household Finance Survey (CHFS). They first use a Probit estimator to explore how social capital represented by kinship and friendship affects the probability of holding a loan from formal and informal sources. The independent variable, kinship, includes two measures, where the first one is kinship size defined as the number of brothers and sisters in the family and the second one is kinship strength, which is a dummy variable, which equals one if the household members participated in a family sacrifice or tomb-sweeping activity last year.[80] Another key independent variable is friendship defined as the total sum of receipt of and expenditure on gifts related to holidays (such as the Chinese Spring Festival, Mid-Autumn Day), weddings, funerals and birthdays. The Probit results show that both

[80] Family sacrifice and tomb-sweeping in China are types of activity relating to the worship of ancestors.

friendship and kinship size are positively associated with the probability of holding a formal loan. Specifically, one unit increases in friendship and kinship size are associated with a 3.9% and a 2.8% higher probability of holding a formal loan, respectively. Regarding informal loans, a one unit increase in friendship is associated with a 3.6% lower probability of holding such a loan. In contrast, kinship size is positively associated with the probability of holding an informal loan. Then, they use the IV Probit estimator to deal with the potential endogeneity of friendship because there may be causality operating in both directions. In detail, in rural China, a farmer who wants to obtain a formal loan, may give money or a gift to a banker in return for extending the formal loan and/or if a banker is anticipating an important event such as a wedding, a gift in return for a loan may be expected. An instrumental variable is defined as average transport expenditure last year. This is because farm households usually need to visit each other by car, bus or other transportation for maintaining gift-giving and friendship, which means this instrument is related to friendship but not correlated with the farmer's ability to obtain a formal or informal loan at the local village level. The first stage in the IV estimation shows that the instrumental variable, transport expenditure last year, is positively correlated with friendship and is statistically significant, with the Wald test indicating rejection of the null hypothesis of exogeneity. The results in the second stage are in accordance with that from the standard Probit estimation indicating that friendship is positively associated with the probability of holding a formal loan, whilst controlling for endogeneity.

Finally, since they only focus on those households headed by a farmer, the Heckman selection model is used to account for potential sample selection bias. A control variable, the average degree of market liberalization in the province in which the farm is located, is assumed to affect a farmer's probability of holding a loan but not the amount of the loan. The Heckman two-step results show that both friendship and kinship size are positively associated with the probability of holding a formal loan, with a positive and significant Inverse Mills ratio indicating that the Heckman selection approach is appropriate and that the estimates of the amount of loan will be upwardly biased if standard OLS is used.

Overall, this paper provides empirical evidence suggesting that friendship is positively associated with the probability of holding a formal loan but inversely associated with informal loan holding for a household headed by a farmer. However, in the Heckman selection model, the measure of the average degree of market liberalization in the province is not clearly defined and this variable is not available in the CHFS and presumably has been merged in from a different data source. Hence, it is not clear how this covariate is defined and measured, which makes it difficult to assess its appropriateness as an instrument.

From an individual level perspective, Liao and Liu (2012) examine how risk perception and risk preference affect consumers' attitudes towards debt based on an individual level survey conducted by the authors with 347 adults. The dependent variable, attitudes towards debt, is measured by 17 items such as: 'it is a good idea to have something now and pay for it later'. Participants respond to all items using a 5 point scale with endpoints ranging from 1 (strongly disagree) to 5 (strongly agree) and higher scores reflect more positive attitudes towards debt. The explanatory variable, risk perception, is based on five items defined over a 5-point agree/disagree scale, where 1 indicates strongly disagree and 5 indicates strongly agree, with items such as 'credit consumption could cause me mental pain'. Regarding risk preference, this is captured by two variables, namely risk-seeking and risk-aversion. Specifically, the authors combine similar items and apply factor analysis, which is based on 12 items over a 5-point 'agree/disagree' scale as above, where 1 indicates strongly disagree and 5 indicates strongly agree with risk-seeking items such as 'I am more and more convinced that I should take greater financial risks to improve my financial position' and risk-aversion items such as 'if I think an investment will be profitable, I am prepared to borrow money to make this investment'.

They use a hierarchical regression approach to model the consumer's attitudes towards debt employing risk perception, risk-seeking, risk-aversion, educational attainment, personal income and the consumer's debt experience as control variables. The results show that the level of risk perception is positively associated with consumers' attitudes towards debt. In addition, a one-point increase in risk-seeking is positively associated with a 0.3 point increase in attitudes towards consumer debt. However, the potential endogeneity problem is not considered, where the more positive attitudes towards debt are, the more risk-seeking the consumer may be.

Similarly, Xin et al. (2020) explore the determinants of medical debt based on an individual-level cross-sectional dataset with 1,000 patients from the rural New Cooperative Medical Scheme (NCMS) database. A two-part model is used to account for the existence of zeros in medical debt, which invalidates the normality assumption of random errors. The binary dependent variable equals 1 if a person has medical debt, which is modelled by a Logit model in the first part of the model, and the second part models the log level of the amount of medical debt estimated by linear regression, for positive debt values. The independent variables include gender, age, per capita household income, marital status, household size, the quality of hospital, inpatient times, inpatient years, inpatient expenses, NCMS re-imbursable expenses, non-direct medical costs including transportation cost and other costs relating to healthcare, medical assistance and financial assistance from kin.

The results from the two-part model reveal that individuals who received financial assistance from kin are less likely to have medical debt while the amount of reimbursable expenses from the NCMS is inversely associated with the probability of having medical debt. In addition, the results from the second part of the two-part model show that those individuals who received financial assistance from kin have lower levels of medical debt and, the higher are non-direct medical costs, the greater the amount of medical debt the person owes.

To summarize, they examine the impact of personal characteristics, inpatient services and expenses, outpatient services on medical debt but they do not explore the marginal effects stemming from the Logit estimators, which means that these empirical correlations cannot be analysed from a magnitude perspective.

From reviewing the literature, it can be concluded that only a limited number of papers on household debt in China exist. In addition, household head's attitudes towards risk may potentially influence household debt holding because evidence suggests that there exists uncertainty in Chinese household income. For example, Yu and Zhu (2013) find evidence supporting the existence of uncertainty in Chinese household income and that compared with the U.S. households, Chinese household income is much more uncertain.[81] Similarly, Chamon et al. (2013) find that income uncertainty substantially increased from the 1990s to the 2000s in China, which provides further evidence of the existence of household income uncertainty in China. Hence, this chapter contributes to further studying the determinants of

[81] Specifically, income uncertainty is measured by the variances of both transitory and persistent shocks. The transitory and persistent shocks stem from the stochastic component of the residuals, which is obtained from conditioning household income on a set of demographic variables including cubic polynomials of age and dummy variables of education attainment, occupation, sector of employment of the family head, urban-rural status, region of residence and year of survey.

household debt in China and focuses on the relationship between a largely ignored determinant, i.e. risk attitudes, and household debt.

Specifically, we initially explore how the attitudes towards risk of the head of household affect household debt holding and the amount of household debt held based on an unbalanced panel covering the 2011, 2013, 2015 and 2017 waves of the CHFS. Then, we split household debt into two types, i.e. housing debt and non-housing debt. This has not been investigated in the existing literature because most of the existing studies only focus on the determinants of housing debt (see, for example, Fan et al., 2017). In addition, following Cull et al. (2019), we split the total sample into two subsamples according to whether the households reside in rural areas or not. To summarise, we analyse both urban and rural subsamples for total household debt, housing debt, and non-housing debt. Finally, we use the double hurdle model to allow for the two-part process related to holding total household debt: (1) the decision to hold debt; and (2) the decision over the amount of debt held, which has been ignored in the existing literature on China.

4.3 Data and Methodology

4.3.1 Data

The household-level data used in this chapter is from the China Household Finance Survey (CHFS), waves 2011, 2013, 2015 and 2017, which is a national survey conducted every two years, starting in 2011. The CHFS includes information about households' demographic characteristics, assets and debt, income and consumption, and attitudes towards risk. In detail, the number of households increases over these years from 8,438 (2011), 28,141 (2013), 37,289 (2015) to 40,011 (2017). The total sample size increases across each wave as the sampling frame has changed over time in order to ensure the national representativeness of the survey, as discussed in detail in Chapter 2.

We initially investigate waves 2011, 2013, 2015 and 2017 as an unbalanced panel dataset, where we focus on households who provide information on the risk attitudes question with 8.5% of observations being omitted due to this restriction.[82] We only include households with a head aged over 20: 0.2%

[82] In this chapter, for the panel data analysis, we have 5.61% of households responding across all three waves and 47.05% responding across two or three waves.

of observations are omitted due to this restriction. After allowing for missing values on all covariates, such as educational attainment, health status, labour market status, marital status and political party membership of the head of household, we have 49,621 households (N) and 91,354 observations (NT) in our panel dataset. All monetary variables in the 2013, 2015 and 2017 waves are deflated using China's yearly CPI, with the benchmark year 2011 = 100. In addition, we split the total sample into two subsamples according to whether the households reside in rural or urban areas, with the number of observations being 63,378 and 27,976, respectively.

There are three main reasons why the CHFS is used for this chapter. Firstly, it contains detailed information on the household head's attitudes towards risk and household indebtedness across urban and rural households. Secondly, in contrast to the existing studies for China, the CHFS is a relatively recent dataset and includes almost all provinces of China, and, hence, is representative of the Chinese population. Finally, in contrast to the existing studies on household finances in China, we are able to conduct panel analysis to allow for time-invariant unobserved heterogeneity across households.

We firstly focus on the relationship between the risk attitudes of the household head and household debt holding as indicated by holding housing or non-housing debt.[83] We split total household debt into housing debt and non-housing debt based on questions related to debt ownership, which are asked to all households: *'Does the family have any agricultural/business debt'; 'Does the family have any house-purchasing/house-renovation mortgage from banks'; 'Does the family have any house-purchas-ing/house-renovation loans from relatives or friends'; 'Does the family have any vehicle-purchasing debt'; 'Does the family have any debt on children's education'; 'Does the family have any credit card debt';* and *'Does the family have any other debt'*. In additional to exploring total debt holding, we explore the relationship between risk attitudes and the holding of two categories of debt: housing debt, which includes mortgages and any loans from relatives or friends; and non-housing debt, which includes ag-ricultural/business debt, vehicle-purchasing debt, education debt, credit debt and other debt.[84]

[83] We do not focus on the ratio of debt over income because in the survey, household debt is a stock variable, and household income is flow. In addition, there is no information on household debt repayment values in the CHFS.
[84] We are unable to split household debt into formal debt from banks and informal debt from other sources, because information on both formal and informal debt is not available for all debt categories. In addition, we aggregate the components of non-housing debt because less than 3% of households hold a specific category of non-housing debt, e.g., only 2.72% of households hold vehicle-purchasing debt.

4.3.2 The Random Effects Logit Model

In order to explore the relationship between total household debt holding and risk attitudes, we specify a random effects Logit model, which is used to model the probability of holding total household debt as follows:

$$Pr(Total\ Debt\ Holding_{it} = 1) = \Lambda(\beta_0 + \beta_1 Risk\ Attitudes_{it} + \beta_2 X_{it} + \varepsilon_{it}) \tag{4.1}$$

$$\varepsilon_{it} = \mu_i + \eta_{it} \tag{4.2}$$

where the probability of holding any debt for household i in time t is given by $Total\ Debt\ Holding_{it}$, such that $i = 1, 2 \ldots, n$ and $t = 2011, 2013, 2015, 2017$. $\Lambda(\cdot)$ is the cumulative probability density function of the logistic distribution, β_0 is the intercept, β_1 captures the relationship between the dependent variable, $Total\ Debt\ Holding_{it}$, and the key explanatory variable, $Risk\ Attitudes_{it}$, and the matrix X_{it} includes the set of covariates, defined below. Following Mundlak (1978), in order to control for household time invariant effects and to enable the estimated parameters to be considered as an approximation to a standard panel fixed effects estimator, a vector of additional controls including the means of the continuous variables, such as the mean of total household disposable annual income, is included.[85] μ_i represents an independent and identically distributed random effect following a normal distribution with mean zero and variance σ_μ^2. η_{it} is a stochastic error term that varies across households and time. We assume that η_{it} is distributed by the standard logistic distribution. Moreover, μ_i captures household specific unobserved heterogeneity and is uncorrelated with X_{it}. The correlation between the error terms of household i at the time l and k is a constant given by

$$\rho = corr(\varepsilon_{il}, \varepsilon_{ik}) = \frac{\sigma_\mu^2}{(\sigma_\eta^2 + \sigma_\mu^2)} \quad l \neq k \tag{4.3}$$

where ρ indicates the proportion of the total unexplained variance in the dependent variable contributed by the panel level variance component. The magnitude of ρ captures the extent of the unobserved intra-household correlation over time, where a low value of ρ indicates little unobservable intra-

[85] This approach is also employed in the random effects Tobit model and the double hurdle model discussed below.

household correlation (Arulampalam, 1999). The analysis is repeated for holding housing debt and for holding non-housing debt.

Table 4.1 provides full variable definitions and Table 4.2 presents summary statistics for all dependent variables used in the panel analysis. From Table 4.2, we can see that 28.8% of households hold debt, where 14.7% of households have housing debt and 18.3% of households have non-housing debt, respectively.[86] This indicates a relatively low household debt holding rate in China as compared with the U.S., where 76.6% of households report having household debt in the 2019 U.S. Survey of Consumer Finances (SCF). In the urban sample, the proportions holding any debt, housing debt and non-housing debt are 26.2%, 15.1% and 14.8%, respectively, while in the rural sample, these proportions are 34.6%, 13.8% and 26.2%, respectively. Thus, these statistics suggest that rural households are more likely to hold debt and are more likely to have non-housing debt, while urban households have a higher probability of holding housing debt, which may reflect higher property values in urban areas (Wang et al., 2020).

4.3.3 The Random Effects Tobit Model

We also model the log level of total household debt using a random effects Tobit specification as follows:

$$Ln(Total\ Debt)_{it} = \beta_0 + \beta_1 Risk\ Attitudes_{it} + \beta_2 X_{it} + \varepsilon_{it} \tag{4.4}$$

where

$$Ln(Total\ Debt)_{it} = Ln(Total\ Debt)^*_{it}\ \ if\ Ln(Total\ Debt)^*_{it} > 0 \tag{4.5}$$

$$Ln(Total\ Debt)_{it} = 0 \qquad otherwise \tag{4.6}$$

$$\varepsilon_{it} = \mu_i + \eta_{it} \tag{4.7}$$

where the log level of total household debt held by the household is given by $Ln(Total\ Debt)_{it}$, such that $i = 1, 2, \ldots, N$ and $t = 2011, 2013, 2015, 2017$. β_0 is the intercept, and β_1 and β_2 are the estimated coefficients. The key explanatory variable is $Risk\ Attitudes_{it}$ and the matrix X_{it} includes all other

[86] About 4.2% of households have both housing debt and non-housing debt.

covariates, as detailed below in Section 4.3.7. ε_{it} is an error term comprising two parts, μ_i and η_{it}, where μ_i represents household specific unobserved heterogeneity (i.e. a random effect) and η_{it} is a stochastic error term that varies across households and time. We assume that η_{it} is independent and identically distributed $N(0, \sigma_\eta^2)$ and μ_i follows a normal distribution, with mean zero and variance σ_μ^2, and is independent of η_{it} and X_{it}. The correlation between the error terms of household i at the time l and k is a constant given by ρ (as discussed above). The analysis is repeated for housing debt and non-housing debt following the above approach.

Figure 4.1 shows the distribution of the log level of total household debt for those heads of household with positive amounts of total household debt, i.e. $Ln(Total\ Debt) > 0$, with the median level of total household debt being around ¥46,000 (£4,600) for the sample reporting positive total household debt. In a similar vein, Figure 4.2 shows the distribution of the log level of housing debt for those heads of household with positive amounts of housing debt, with the median level of housing debt being around ¥82,700 (£8,270). Finally, the distribution of the log level of non-housing debt is shown in Figure 4.3, where the median level of non-housing debt is around ¥19,500 (£1,950).

Finally, among those urban households with positive amounts of total household debt, the median level of total household debt is around ¥73,478 (£7,347), which is considerably larger than that for rural households with the median level being around ¥25,717 (£2,571) for rural households (see, Table A4.2 in the appendix). Similarly, among those households with positive amounts of housing debt, the median level of housing debt for urban house-holds is around ¥124,773 (£12,477), while for rural households the median level of housing debt is only around ¥28,499 (£2,849), which may reflect higher property values and prices in urban areas. Furthermore, there is only a small difference between urban and rural house-holds regarding the level of non-housing debt. Specifically, the median level of non-housing debt among those urban households with positive amounts of non-housing debt is around ¥26,588 (£2,658) and the median level of non-housing debt among rural households with positive amounts of non-housing debt is ¥17,725 (£1,772).

4.3.4 Additional Robustness Checks

4.3.4.1 The Fixed Effects Logit Model

In order to further explore the robustness of our results to controlling for unobserved heterogeneity across households, we specify a fixed effects Logit model, which only includes those households who changed debt holding states over two years, three years or four years, to model the probability of holding total household debt as follows:

$$Pr(Total\ Debt\ Holding_{it} = 1) = \Lambda(\beta_0 + \beta_1 Risk\ Attitudes_{it} + \beta_2 X_{it} + \mu_i + \varepsilon_{it}) \tag{4.8}$$

where the probability of holding any debt for household i in time t is given by $Total\ Debt\ Holding_{it}$, such that $i = 1, 2 \ldots, n$ and $t = 2011, 2013, 2015, 2017$. $\Lambda(\cdot)$ is the cumulative probability density function of the logistic distribution, β_0 is the intercept, β_1 captures the relationship between the dependent variable, $Total\ Debt\ Holding_{it}$, and the key explanatory variable, $Risk\ Attitudes_{it}$, and the matrix X_{it} includes the set of covariates. μ_i represents a household-specific unobserved fixed effect and is correlated with X_{it}. ε_{it} is an idiosyncratic error term that varies across households and time. We assume that ε_{it} is distributed by the standard logistic distribution. As above, the analysis is repeated for housing debt and non-housing debt holding.

4.3.4.2 The Double Hurdle Model

Households hold a positive amount of debt under the precondition that they first decide to hold debt. In addition, holding zero debt may arise due to two reasons: firstly, they are not willing to hold any household debt; or, secondly, they decide to hold debt but they currently have a zero amount, which means that zero values of household debt can be observed instead of censored. That is to say, observations where household debt is equal to zero are not the result of being unable to observe the distribution below zero. Therefore, to further explore the robustness of our findings, the double hurdle model is used, treating the sample as pooled. Hence, we cluster the households in order to take into account the repeated observations. The double hurdle model is specified as follows:

$$Ln(Total\ Debt)_{it} = s_{it} h_{it}^{*} \tag{4.9}$$

$$s_{it} = \begin{cases} 1 & if\ \gamma_0 + \gamma_1 Risk\ Attitudes_{it} + \gamma_2 No\ Financial\ Knowledge_{it} + \gamma_3 X_{it} + \omega_{it} > 0 \\ 0 & otherwise \end{cases} \tag{4.10}$$

where s_{it} is the probability of holding total household debt, which equals 1 if the dependent variable $Ln(Total\ Debt)_{it}$ is not bounded and 0 otherwise. γ_0 is the intercept and γ_1 is the estimated coefficient of our independent variable of interest, $Risk\ Attitudes_{it}$. The variable used to identify the model is $No\ Financial\ Knowledge_{it}$, the choice of which is discussed further below. γ_3 captures the relationship between the covariates and the probability of holding total household debt. ω_{it} is a standard normal error term.

$$h_{it}^{*} = \alpha_0 + \alpha_1 Risk\ Attitudes_{it} + \alpha_2 X_{it} + \nu_{it} \tag{4.11}$$

where h_{it}^{*} is the continuous latent variable, which can be observed only if $s_{it} = 1$. α_0 is the intercept, α_1 and α_2 are the estimated coefficients. ν_{it} is an error term and has a truncated normal distribution and is uncorrelated with ω_{it}. $Risk\ Attitudes_{it}$ is our main independent variable of interest and the matrix X_{it} includes the set of covariates. In order to identify the model, as stated above, $No\ Financial\ Knowledge$ is included in order to model the probability of holding debt but not the amount of debt held. Conceptually, we argue that, from the demand-side, a financially illiterate household head may borrow because they might not realize the risk associated with taking on debt (Gathergood, 2012). In other words, those households who do not have any financial knowledge may have a higher probability of holding debt. However, from the supply-side, the amount of debt borrowed from banks or friends is not based on whether the household head has any financial knowledge but on affordability in terms of the ability of the household head to repay the debt or, in the case of informal debt, how close the relationship between the household head and the lender is. Thus, $No\ Financial\ Knowledge_{it}$, is used to identify the model, which only determines whether the household holds debt rather than the amount of debt because of the exclusion restriction. We define $No\ Financial\ Knowledge_{it}$ as a dummy variable, which equals 1 if the household head does not have any financial knowledge, i.e. the household head answered all three financial literacy questions incorrectly: (1) given a 4% interest rate, how much would you have in total after 1 year if you have 100 RMB deposited? Answers include: “under 104”, “104”, “over 104” and “cannot figure out”; (2) with an interest rate of 5% and an inflation rate of 3%, the products you buy with the money you have saved in the bank for one year is: “more than last year”, “the same as the last year”, “less than last year” and “cannot figure out”; and (3) which one do you think is more risky, stocks or funds? Answers include: “stocks”, “funds”, “don’t know stocks”, “don’t know funds” and "don't know both". It can be seen from Table 4.2 that nearly 50% of heads of household do not have any financial knowledge and the phenomenon of the lack of financial literacy is more common in rural areas since about 65% of rural heads of household are financially illiterate while the mean value of $No\ Financial\ Knowledge$ for urban households is only 38.6%.

4.3.5 The Risk Attitudes Measure

Turning to the key explanatory variable, $Risk\ Attitudes$, we measure the head of household's attitudes towards risk based on the question: *'in which project below would you want to invest most if you have adequate money?'* The answers include: (1) a project with high risk and high return; (2) a project with slightly high risk and slightly high return; (3) a project with average risk and average return; (4) a project with slight risk and return; and (5) unwilling to carry any risk. Following Hu et al., (2015), we assign a value of 0 to 4 to each of the above five options. Specifically, $Risk\ Attitudes$ is a 5-point index ranging from 0 to 4. This index is increasing in risk-tolerance, where 0 denotes a household head who is unwilling to carry any risk; 1 denotes a household head who prefers projects with slight risk and return; 2 denotes a household head who prefers projects with average risk and return; 3 denotes a household head who prefers projects with slightly high risk and slightly high return; and 4 denotes a household head who prefers projects with high risk and high return. Such a measure of $Risk\ Attitudes$ is the same as that in the U.S. SCF, which has been used extensively in the household finance literature (see, for example, Brown et al. 2011).

In addition, we measure the head of household's attitudes towards risk by including a set of five dummy variables based on the above question rather than an index as a comparison in order to further explore the effect of each specific level of risk attitudes on household debt. Specifically, $No\ Risk\ Return$ equals 1 if the household head is unwilling to carry any risk; $Low\ Risk\ Return$ equals 1 if the household head prefers projects with slight risk and return; $Average\ Risk\ Return$ equals 1 if the household head prefers projects with average risk and return; $Slightly\ High\ Risk\ Return$ equals 1 if the household head prefers projects with slightly high risk and return; and $High\ Risk\ Return$ equals 1 if the household head prefers projects with high risk and return.

From Table 4.2, we can see that the mean value of $Risk\ Attitudes$ is only 0.939, which indicates a low level of risk tolerance among Chinese heads of household. Furthermore, urban heads of household are, on average, more risk tolerant than rural heads of household because the mean value of $Risk\ Attitudes$ for urban households is 1.009, which is greater than that of rural households , i.e. 0.780. In addition, Table 4.2 shows that the mean value of $No\ Risk\ Return$ is 0.518, which means 51.8% of heads of household are unwilling to carry any risk. 18.2% of heads of household prefer projects with low risk and return, 19.6% of heads of household choose projects with average risk and return, and only 4.8% and 5.5% of heads of household prefer projects with slightly high risk and return and high risk and return, respectively. It is not surprising that over half of heads of household are intolerant towards risk, which is in line with the findings in the U.S. (see, for example, Brown et al., 2011).

4.3.6 Other Explanatory Variables

The matrix X_{it} contains the control variables generally used in the existing literature on household debt (see, for example, Brown et al., 2013; Sun et al., 2018 and Han et al., 2019). These variables are defined in Table 4.1 with summary statistics provided in Table 4.2. Specifically, $Ln(Income)$ is the nat-ural logarithm of the total amount of disposable annual income of the household plus one.[87,88] $Ln(Assets)$ is the natural logarithm of the total amount of household assets plus one. It is common in China that households prioritize borrowing money from close friends and relatives because these relations can provide immediate assistance (Cleaver, 2005; Sun et al., 2018). In addition, households usually give gifts to their friends to maintain good relationships. Thus, we control for $Ln(Social\ Network)$, which is the natural logarithm of the total amount of expenditure related to giving to non-family members (plus one) including wedding gifts, funeral money, education, medical treatment, and other donations.[89] Following Sun et al. (2018), we control for another important source households may rely on when they need a loan, i.e. $No.Siblings$, defined as the number of siblings of the household head and his/her spouse, as siblings may help immediately and lend money to households. $No.Children$ is the number of dependent children aged below 16 in the household (after age 16, in China children can choose to work or continue studying at school). $No.Workers$ is the number of workers in the household excluding

[87] The CHFS defines household disposable income as: salary net income after tax; net income from agricultural products after-tax; net income from business after-tax; net income from investment after-tax (rent, stock markets; interest from bank deposits, etc.); and net transfer income after-tax (social security, social insurance, annuity, etc.).
[88] We control for the level of income as we are unable to control for income uncertainty because, although the CHFS provides the opportunity to explore panel data, the CHFS is a relatively short panel and our analysis starts from the first wave. This means that there are no previous time periods to use to construct measures based on past income.
[89] The social network variable may only be relevant in the Chinese context (see, for example, Fan et al., 2017; Cull et al., 2019)

the household head because we also include the labour market status of the household head, as discussed below. We also control for $No. Aged\ Over\ 60$, which is the number of family members aged over 60 in the household excluding the household head (the statutory retirement age in China is 60).

Health has been found to be an important determinant of household debt in China, see, for example, Cui et al. (2017). Thus, we measure the health of the household head using the survey question: 'what do you think of your health status relative to your peers?' The answers include "very poor", "poor", "normal", "good" and "very good". $Self\ Assessed\ Health$ is a 5-point index for the head of household ranging from 0 to 4, where 0 denotes very poor; 1 denotes poor; 2 denotes normal; 3 denotes good and 4 denotes very good. Age is the age of the household head. Following Brown et al. (2013), we also control for $Male$, the gender of the household head, which has been found to play an important role in determining household debt in the U.S. and whether the household head is married, denoted by $Married$.

In China, a member of the Communist Party of China usually has higher personal credibility than a non-member since party membership needs to be qualified and assessed in terms of personal quality and ability, which means a party member may have more access to bank loans (Cull et al., 2019). Thus, we control for the variable, $Party\ Member$, which is a dummy variable and equals 1 if the household head is a party member. The $Education$ variable is classified into six categories: $No\ Schooling$ (the omitted category) is a dummy variable, which equals 1 if the household head never attended school; $Primary\ School$ is a dummy variable, which equals 1 if the highest educational attainment of the household head is primary school; $Junior\ High$ is a dummy variable, which equals 1 if the highest educational attainment of the household head is junior high school; $Senior\ High$ is a dummy, which equals 1 if the highest educational attainment of the household head is senior high school or technical school; $College/Bachelor$ is a dummy variable, which equals 1 if the highest educational attainment of the household head is vocational college or a bachelor degree; and $Master/PhD$ is a dummy variable indicating if the highest educational attainment of the household head is a master's degree or PhD.

The labour market status of the head of household has been found to be an important determinant of household debt (see, e.g. Crook and Hochguertel, 2007; Altundere, 2014). Thus, we control for the labour market status of the head of household as follows: $Employed$ is a dummy variable, which equals 1 if the household head is an employee, i.e. employed by someone else; $Self\ Employed$ is a dummy variable, which equals 1 if the household head is self-employed; $Retired$ is a dummy variable, which equals 1 if the household head is retired; $Not\ Working$ is a dummy variable, which equals 1 if the household head is not working, i.e. the household head is unemployed, incapacitated, a homemaker, a volunteer or unwilling to work; and $Farmer$ (the omitted category) is a dummy variable, which equals 1 if the household head is a farmer.

For the all households sample, we control for whether the household resides in a rural area, as indicated by $Rural$, since the opportunity to access bank loans differs between rural and urban areas. Specifically, urban households have more access to formal loans (see Turvey et al., 2010). We also control for region, as represented by $Region$. Specifically, we distinguish between seven regions: $North\ East$, $North$, $East$, $Central$, $South$, $South\ West$, $North\ West$ (the omitted category, which has the lowest gross regional product (GRP), i.e., it is the most under developed region in China).[90] In our panel analysis, we control for the year of interview as the data covers four years: 2011 (the omitted category), 2013, 2015 and 2017.

Turning to the summary statistics for the all households sample, we can see from Table 4.2 that there exists a large disparity in household income between households by comparing the minimum value and the maximum value, and the standard deviation of disposable net income. The average age of the head of household is over 50, which indicates that China is facing an aging population. Turning to the educational attainment of the household head, the proportion of the households, where the highest educational attainment of the head is above junior high school, is over 70%, which indicates that the nine-year compulsory education system has been successful in China. However, about 50% of household heads lack financial knowledge. Finally, we can see that over 30% of households live in rural areas.

[90] The figures for GRP can be found from the National Bureau of Statistics of China: http://www.stats.gov.cn/tjsj/ndsj/2019/indexeh.htm.

We also provide summary statistics split by urban and rural residence (see Table 4.2).[91] Specifically, the mean value of household income for urban households is higher than that for rural households, which indicates that urban households potentially have a higher ability to repay household debt. The mean value of household social network expenditure for urban households is higher than that for rural households, which probably indicates that urban households have an, on average, broader social networks. It is interesting to find that, compared with rural households, urban households have a smaller number of siblings and dependent children, which probably indicates that the one-child policy is more strongly enforced in urban areas because the government regulates the behaviour of urban households more easily relative to rural households. For example, urban residents working in state-owned enterprises or institutions would lose their jobs if they have more than one child. In contrast, rural households were often allowed to have a second child, particularly if the first child was female because rural households who usually carry out agricultural work had a strong need for a large family and male labour (Zhang, 2017). Finally, we also find that rural areas suffered from a more severe aging problem since the rural households have, on average, more members aged over 60 (Lu and Liu, 2019).

4.4 Results

4.4.1 Random Effects Logit Analysis

The results from estimating the random effects Logit models are shown in Table 4.3, where the estimated coefficients and marginal effects of risk attitudes and the other covariates are presented for the three outcomes: $Total\ Debt\ Holding$, $Housing\ Debt\ Holding$ and $Non\ housing\ Debt\ Holding$ for the all households sample and then split into the urban and rural samples, respectively.

It can be seen from Table 4.3 that the estimated coefficient of $Risk\ Attitudes$ is positive and statistically significant in the case of the probability of holding total household debt for the all households sample, which is in accordance with our expectations. In terms of the magnitude of the marginal effect of risk attitudes, a one unit increase in the $Risk\ Attitudes$ index is associated with a 1.31% increase in the probability of holding total household debt. Such a finding indicates that the households, where the head is more risk tolerant, are more likely to hold total household debt. In comparison with the effect of

[91] Although some differences are apparent in the summary statistics across rural and urban samples, we are unable to control for sample selection bias in the econometric analysis split by region because of the lack of a suitable instrumental variable.

the social network, which has been identified as an important determinant of household debt in the context of China (see, e.g., Sun et al., 2018; Cull et al., 2019), Table 4.3 shows that an increase of one percent in $Ln(Social\ Network)$ is associated with a 0.27% increase in the probability of holding total household debt. Such a finding suggests that those households with a broader social network have a higher probability of holding total household debt, which is common in China, where households generally prioritize borrowing from relatives or friends when needing a loan. In other words, the more expenditure the household gives to non-family members in order to maintain social relationships, the higher is the probability of holding total household debt. Furthermore, the magnitude of the marginal effect of $Ln(Social\ Network)$ is smaller than that of $Risk\ Attitudes$ suggesting that the household head's attitudes towards risk is an important determinant of the probability of whether households hold total household debt (which is in line with existing findings for the U.S. from Brown et al., 2013).

Turning to $Housing\ Debt\ Holding$ and $Non\ housing\ Debt\ Holding$ for the all households sample (see, column 2 and column 3, respectively, in Table 4.3), it can be seen that the estimated coefficient of $Risk\ Attitudes$ is not statistically significant in the case of $Housing\ Debt\ Holding$, but it is positive and attains statistical significance at the 1% level for $Non\ housing\ Debt\ Holding$. This might be due to the fact that the purchase of housing can be regarded as an investment and China's residential property values have steadily risen from 2005 to 2021, which lowers the risk of the investment. Thus, regardless of the risk-tolerance of the household head, households would hold housing debt if they want to purchase a property but cannot afford to buy a house outright.[92] It is important to note that the Chinese Government imposed housing purchase restrictions in 2010, such as raising the down-payment ratio, increasing the mortgage rate and prohibiting mortgages on second home purchases (Cao et al., 2015). Such changes may also affect the riskiness of investment in housing. Turning to non-housing debt, we can see that a one unit increase in the $Risk\ Attitudes$ index is associated with a 1.61% increase in the probability of holding non-housing debt. This is in accordance with the findings of Brown et al. (2013), i.e. attitudes towards risk have different effects on different types of household debt. The positive association found for non-housing debt may reflect the possibility that this type of debt holding is riskier rather than debt holding undertaken to invest in property, especially in the context of rising house prices.

[92] The data relating to residential property in China from 2005 to 2021 can be found at https://fred.stlouisfed.org/series/QCNR628BIS.

Once again, we take the effect of the social network as a comparison; the estimated marginal effect of $Ln(Social\ Network)$ on non-housing debt holding is 0.32% (see column 3 in Table 4.3), which is smaller than that of $Risk\ Attitudes$. Such a finding is consistent with the results in the case of $Total\ Debt\ Holding$, which provides further evidence of the importance of household head's attitudes towards risk for the probability of holding total household debt and non-housing debt.

For the urban sample, see Table 4.3, the estimated coefficient of $Risk\ Attitudes$ is positive and statistically significant in the case of two outcomes: $Total\ Debt\ Holdin$ and $Non\ housing\ Debt\ Holding$. Specifically, the marginal effect of $Risk\ Attitudes$ on the probability of holding total household debt is 0.0102, which means that a one unit increase in the $Risk\ Attitudes$ index is associated with a 1.02% increase in the probability of holding total household debt for urban households. Such a positive effect of risk attitudes on total household debt holding is consistent with the finding for the sample of all households. The household head's attitudes towards risk does not have a statistically significant impact on the probability of holding housing debt for urban households, which may reflect the relatively high and rising value of property in urban areas making such purchases less risky. In contrast, a one unit increase in the $Risk\ Attitudes$ index is associated with a 1.35% increase in the probability of holding non-housing debt, which accords with the findings for the all households sample.

Turning to the rural households (see Table 4.3), we find results, which are consistent with the results for the all households sample and the sample of urban households in that $Risk\ Attitudes$ are positively associated with the probability of holding total household debt and non-housing debt. However, the effect of risk attitudes becomes statistically significant in the case of $Housing\ Debt\ Holding$ in contrast to that of $Housing\ Debt\ Holding$ in the urban sample. Specifically, a one unit increase in the household head's risk tolerance is associated with an increase in the probability of holding total household debt, housing debt and non-housing debt of 1.93%, 0.43% and 2.17%, respectively. Such a finding is not surprising and may be because housing prices in urban areas are much higher than in rural areas (Wang et al., 2020). In accordance with the arguments made above related to the other samples, this arguably means purchasing housing is relatively risky in rural areas and, hence, whether rural households hold housing debt is influenced by the household head's risk tolerance.[93] In addition, it is apparent that the magnitude of the marginal effect of $Risk\ Attitudes$ on the probability of holding total household debt is larger for rural households than that for their urban counterparts (see, column 1 in Table 4.3). This may reflect the fact that urban heads of household have a higher tolerance against the risk associated with debt than rural households. Specifically, the mean value of $Risk\ Attitudes$ for urban households is 1.009 while for rural households it is only 0.780.

In order to explore the effect of specific categories of the household head's attitudes towards risk, we repeat the random effects Logit analysis replacing the $Risk\ Attitudes$ index with the set of risk attitudes dummy variables, i.e. $No\ Risk\ Return$ (the omitted category), $Low\ Risk\ Return,\ Average\ Risk\ Return,\ Slightly\ High\ Risk\ Return$ and $High\ Risk\ Return$. We can see from Table A4.3 in the appendix that, for the all households sample, the marginal effects of the risk attitudes dummy variables are all positive and statistically significant in the case of $Total\ Debt\ Holding$. Specifically, the marginal effects of $Low\ Risk\ Return$, $Average\ Risk\ Return$, $Slightly\ High\ Risk\ Return$ and $High\ Risk\ Return$ are 0.0162, 0.0264, 0.0455 and 0.0501, respectively, which show a monotonic in-crease in the effect on the probability of holding total household debt, as the head of household be-comes more risk tolerant. In other words, the higher the level of the household head's risk tolerance, the higher is the probability that the household holds total household debt, all other things being equal. A similar monotonically increasing effect of the risk attitudes dummy variables in terms of magnitude on the probability of holding non-housing debt is found. However, turning to the probability of holding housing debt, only $Low\ Risk\ Return$ and $Slightly\ High\ Risk\ Return$ attain statistical significance and only at the 10% level, which provides further evidence, as discussed above, that the probability of holding housing debt may not be determined by the household head's attitudes towards risk. Such a finding is even more apparent for urban households, where the household head's risk attitudes do not have any impact on the probability of holding housing debt, which may reflect the higher (and increasing) property

---

[93] The estimated marginal effects of $Ln(Social\ Network)$ on the probability of holding total household debt, housing debt and non-housing debt are 0.0023, -0.0013 and 0.0033, respectively. This means that the size of the marginal effect of risk attitudes is greater than that of the social network on the probability of holding total household debt, housing debt and non-housing debt, which is in support of the importance of the household head's attitudes towards risk for holding household debt.

values in urban areas and the nature of risk associated with this type of investment (Wang et al., 2020). Similar patterns are found for rural households in that the risk attitudes dummy variables are positively associated with the probability of holding total household debt and non-housing debt. In addition, the magnitudes of the marginal effects of the risk attitudes dummy variables are monotonically increasing in risk tolerance with the exception of the size of the marginal effect of $High\ Risk\ Return$ being slightly smaller than that of $Slightly\ High\ Risk\ Return$.

We now briefly turn to the effects of the other covariates for the all households sample, presented in Table 4.3, where it can be seen that households have an increasing probability of holding total household debt, housing debt and non-housing debt as the age of the head of household increases but it is a quadratic effect. This is not in accordance with the findings from Cull et al. (2019), who find that households with an older head are less likely to have a loan. The difference in the findings might stem from the fact that Cull et al. (2019) conduct the analysis only using the 2013 wave of the CHFS and they mainly focus on whether households have held loans previously instead of current outstanding debt. Regarding the gender of the household head, we find that households are more likely to hold total household debt and non-housing debt if the head of the household is male. However, no difference is found between female and male headed households in terms of the probability of holding housing debt. Households are more like to hold housing debt if the household head is married, which is in accordance with the culture in China where a potential mother-in-law is likely to oppose her daughter's marriage if the daughter plans to marry a man without a house (Li and Wu, 2014).

In addition, we find that households are less likely to hold total debt, housing debt and non-housing debt as the household head's self-assessed health status improves. Specifically, a one unit increase in the $Self\ Assessed\ Health$ index is associated with a decrease in the probability of holding total debt, housing debt, and non-housing debt of 3.45%, 1.72% and 2.78%, respectively, which contrasts with the findings from Cui et al., (2017). As noted above, this may be because Cui et al. (2017) focus on whether rural households held loans previously instead of whether they are currently holding an outstanding loan. Educational attainment is negatively associated with the probability of holding total debt, housing debt and non-housing debt except in the case of the effect of having a $\mathrm{Master/PhD}$. For example, the households, where the highest educational attainment of the head of household is primary school, have a 2.90% lower probability of holding total household debt, a 1.88% lower probability of holding housing debt and a 1.43% lower probability of holding non-housing debt in comparison to those heads of household who never attended school. However, the households, where the highest educational attainment of the head of household is above bachelor (master or PhD), have a 3.30% higher probability of holding total household debt and a 2.80% higher probability of holding housing debt. Such findings may be related to the fact that those households where the head has a Master's degree or PhD tend to move to the most developed places in a region such as provincial capitals because these places attract the highly educated and have an advantage in competing for human capital (Liu et al., 2017). However, the house prices in these places are much higher than in other areas, which may lead to a higher probability of holding housing debt.

Furthermore, we find that those households, where the head of household is retired are less likely to hold total household debt, housing debt and non-housing debt, which might reflect that those households where the head of household is retired may have paid off debts or have no incentive to hold debt. In detail, the estimated marginal effect of $Retired$ is -0.1315 in the case of $Total\ Debt\ Holding$, which means that households with a retired head have a 13.15% lower probability of holding total household debt than those where the head of household is a farmer (see Table 4.3). Similarly, the marginal effect of $Retired$ in terms of the magnitude is greater than that of other categories of the labour market status of the household head. The number of dependent children is positively associated with the probability of holding total household debt, housing debt and non-housing debt. This is not surprising given that household expenditure on children has been increasing in China in recent years (Chi and Qian, 2016) and such a phenomenon has been explored in Chapter 2, which may lead to households becoming indebted. We also find that households living in rural areas are more likely to hold total household debt, housing debt and non-housing debt, which may reflect the development of the urbanisation process in China. For example, parents living in rural households may hope that their children move to urban areas with better economic prospects and thus choose to hold debt for education or purchasing houses in urban areas. Furthermore, households living in the $East$, $North$, $Central$ and $South$ regions of China are less likely to hold total household debt, housing debt and non-housing debt in contrast to those living in the $North\ West$, which probably reflects the fact that the $East$, $North$, $Central$ and $South$ regions are more developed economic regions (National Bureau of Statistics of China, 2019). As stated above, for brevity, the full results for the urban and rural samples are not presented as, in general, a consistent pattern of results is found across all models and our focus here lies on the effects of risk attitudes.

The results in Table 4.3 also suggest that those heads of household responding in 2013 and 2015 have a statistically significantly lower likelihood of holding household debt, housing debt and non-housing debt relative to households responding in 2011. In addition, ρ is positive, which suggests positive intra-correlation over time, and along with its statistical significance, indicates that the longitudinal element of the data is important.

4.4.2 Random Effects Tobit Analysis

We also investigate the relationship between risk attitudes and the log level of total household debt held by the household. For brevity, we only present the estimated effects of $Risk\ Attitudes$ on the log level of total household debt, the log level of housing debt and the log level of non-housing debt (see Table 4.4). The same controls are included as in the Logit analysis and the pattern of the results remains the same. For risk attitudes, the marginal effects are presented at the extensive and intensive margins.[94] The marginal effect of $Risk\ Attitudes$ at the extensive margin is statistically significant and positively associated with $Ln(Total\ Debt)$ and $Ln(Non\ housing\ Debt)$. Specifically, for the all households sample, a one unit increase in the household head's risk tolerance is associated with a 1.23% increase and a 1.58% increase in the probability of holding total household debt and non-housing debt, respectively, which accords with the findings from the random effects Logit specification in terms of the magnitude of the marginal effect of the household head's attitudes towards risk on the probability of holding total household debt and non-housing debt. For urban households, we find a similar pattern of results in that the household head's attitudes towards risk are positively associated with the probability of holding total household debt and non-housing debt. In addition, $Risk\ Attitudes$ are positively associated with the probability of holding total household debt, housing debt, and non-housing debt for those households living in rural areas.

---

[94] A marginal effect at the intensive margin relates to the portion of the variation of the explanatory variable that is correlated with the variation of the expected value of the dependent variable conditional on being non-zero, while the marginal effect at the extensive margin relates to the change in the probability that the dependent variable is greater than zero.

Turning to the marginal effects of $Risk\ Attitudes$ at the intensive margin for the all households sample, we can see from Table 4.4 that the marginal effect of $Risk\ Attitudes$ in the case of the log level of total household debt is 0.1293, which means that among households with a non-zero log level of total household debt, a one unit increase in the household head's risk tolerance is associated with a 12.93% increase in the log level of total household debt. In addition, $Risk\ Attitudes$ is positively associated with the log level of non-housing debt among those with a non-zero log level of non-housing debt. Similar patterns of findings are found for the urban sample, while for the rural sample different findings are once again found in that the marginal effect of $Risk\ Attitudes$ at the intensive margin is statistically significant and positive in the case of all three outcomes (see Table 4.4). Furthermore, we find that risk attitudes play a more important role in determining the log level of non-housing debt than the log level of housing debt since the magnitude of the marginal effect at the intensive margin stemming from $Risk\ Attitudes$ is 0.2024 in the case of $Ln(Non\ housing\ Debt)$, which is higher than that in the case of $Ln(Housing\ Debt)$. This may be due to the different motivations behind holding household housing debt and non-housing debt because housing debt differs from non-housing debt. Specifically, a house is not only a place for people to live but is also regarded as the foundation of a household in China, thus the decision to hold housing debt has greater priority relative to other types of debt for a Chinese household. The findings related to ρ are similar to those in the random effects Logit specifications (see Section 4.4.1).

Turning to the set of dummy variables capturing different categories of risk attitudes, the marginal effects of $Low\ Risk\ Return,\ Average\ Risk\ Return,\ Slightly\ High\ Risk\ Return$ and $High\ Risk\ Return$ at the extensive and intensive margins for the all households sample, urban sample and rural sample are presented in Table A4.4 in the appendix. Firstly, focusing on the marginal effects of the risk attitudes dummy variables at the extensive margin, the households, where the head's attitudes towards risk are given by $Low\ Risk\ Return$, have a 1.57% higher probability of holding total household debt relative to those where the head is unwilling to carry any risk. In addition, the households where the head's attitudes towards risk are $High\ Risk\ Return$ have a 4.62% higher probability of holding total household debt relative to those households with a head who is unwilling to carry any risk. In general, the magnitudes of the marginal effects of the risk attitudes dummy variables are monotonically increasing in risk tolerance for the probability of holding total household debt for the all households sample. Turning to the marginal effects of the risk attitudes dummy variables at the intensive margin, we can see from Table A4.4 in the appendix that the higher is the level of the household head's risk tolerance, the greater is the amount of total household debt held by the household. For example, the households where the head's attitudes towards risk are $Low\ Risk\ Return$ have 16.46% more total household debt than those households where the head is unwilling to carry any risk. Similar patterns are found in the case of $Ln(Non\ housing\ Debt)$ for the all households sample, while, in the case of $Ln(Housing\ Debt)$, only $Low\ Risk\ Return$ and $Slightly\ High\ Risk\ Return$ attain statistical significance and only at the 10% level, which is consistent with the findings from the random effects Logit specifications.

For both urban and rural households, we find results, which are similar to those for the all households sample. The marginal effects of the risk attitudes dummy variables are all statistically significant and positive in the case of $Ln(Total\ Debt)$ and $Ln(Non\ housing\ Debt)$ and are monotonically increasing in risk tolerance in terms of magnitude. The findings once again confirm that risk attitudes are important determinants of the amount of total household debt and non-housing debt held by the households and that different levels of risk tolerance have different effects in terms of magnitude.

4.4.3 Additional Robustness Checks

4.4.3.1 Fixed Effects Logit Analysis

For a robustness check, using a fixed effects estimator which only focuses on those households who switch debt status over time (i.e. those always in debt or never in debt are excluded), we also explore the relationship between risk attitudes and the probability of holding total household debt, housing debt and non-housing debt, controlling for unobserved heterogeneity across households. The results are summarised in Table 4.5, where, for brevity, only the effects of $Risk\ Attitudes$ are presented. It is apparent that the effects of $Risk\ Attitudes$ are statistically significant and positive in the case of $Ln(Total\ Debt)$ and $Ln(Non\ housing\ Debt)$ for both the all households sample and the rural sample, which is in accordance with the findings from the random effects Logit specifications. For example, for the all households sample, a one unit increase in the household head's risk tolerance is associated with a 0.79% increase and a 0.46% increase in the probability of holding total household debt and non-housing debt, respectively. However, the estimated effect of $Risk\ Attitudes$ on the probability of holding housing debt is statistically insignificant across all three samples, which is probably due to the fact that the fixed effects Logit estimator excludes those households who maintain states over time in order to control for unobserved heterogeneity and housing debt status is likely to be invariant over time for many households as this type of debt is frequently held over relatively long time periods. Thus, we have different sample sizes than that in the random effects Logit specifications. Nevertheless, the results of the fixed effects Logit estimation provide further evidence of the importance of attitudes towards risk for household debt and this approach has the added benefit of having controlled for unobserved time invariant characteristics.

Repeating the fixed effects Logit analysis for the set of risk attitudes dummy variables rather than the index, the marginal effects of the risk attitudes dummy variables are presented in Table A4.5 in the appendix. Here, we find some different patterns of results as compared to the findings from the fixed effects Logit specifications for the $Risk\ Attitudes$ index. For example, only $High\ Risk\ Return$ attains statistical significance at the 5% level in the case of $Total\ Debt\ Holding$, which means that the households, where the heads' attitudes towards risk are given by $High\ Risk\ Return$, have a 3.80% higher probability of holding total household debt than those households where the head is unwilling to carry any risk for the all households sample. Similar results are found in the case of $Total\ Debt\ Holding$ for rural households in that the households where the heads' attitudes towards risk are given by $High\ Risk\ Return$ have a 7.99% higher probability of holding total household debt than those households where the head is unwilling to carry any risk. Nevertheless, the results of the fixed effects Logit estimation with the risk attitudes dummy variables do provide some evidence that is consistent with the findings from the fixed effects Logit estimation with the $Risk\ Attitudes$ index in terms of indicating that the household head's tolerance to risk is positively associated with the probability of holding total household debt and non-housing debt.

4.4.3.2 Double Hurdle Analysis

As discussed in Section 4.3.5, before households hold any amount of total household debt, they need to make a decision about whether to hold debt. This means that holding total household debt includes two parts: (1) the decision to hold debt, the selection equation; and (2) the amount of debt held, conditional on holding debt. Thus, the double hurdle approach is used to model such a process for an additional robustness check. Specifically, in the selection equation (see Table 4.6), the estimated marginal effect of $No\ Financial\ Knowledge$, i.e. the variable used to identify the model, is positive and statistically significant in the case of all three outcomes for the all households sample, which is in accordance with expectations. Households have a higher probability of holding total household debt, housing debt and non-housing debt if the household heads do not have any financial knowledge. It may be the case that the risk behind taking on debt may be ignored or misunderstood by these financially illiterate households.[95] For example, these households may not consider whether or not they can repay the debt. In order to evaluate the validity of the variable, $No\ Financial\ Knowledge,$ which is used to identity the model, we have tested the exclusion restriction that $No\ Financial\ Knowledge$ is statistically insignificant in the amount of debt specifications for those households with a positive amount of debt, which supports the validity of the exclusion restriction in a statistical sense (see Table 4.6).

The marginal effects of $Risk\ Attitudes$ are 0.1650, 0.0222 and 0.1776 in the case of $Ln(Total\ Debt)$, $Ln(Housing\ Debt)$ and $Ln(Non\ housing\ Debt)$, respectively, for the all households sample. This means that, conditional on holding debt, a one unit increase in the household head's risk tolerance is associated with a 16.50% increase, a 2.22% increase and a 17.76% increase in the log level of total household debt, the log level of housing debt and the log level of non-housing debt held by the households, respectively. This accords with the findings from the random effects Tobit specification. Such findings provide further support that among those households with a positive amount of total household debt, the risk tolerance of the household head is positively associated with the amount of total household debt. Similar patterns of findings are revealed for the rural sample, where $No\ Financial\ Knowledge$ is found to be positively associated with the probability of holding total household debt, housing debt and non-housing debt, and $Risk\ Attitudes$ is positively associated with the log level of total household debt, the log level of housing debt and the log level of non-housing debt (see Table 4.6). For urban households, the marginal effect of $Risk\ Attitudes$ is positively associated with the log level of total household debt and the log level of non-housing debt, which accords with the findings

[95] No Financial Knowledge has not been included in the previous models to enable comparisons across all the models. It should be noted that if No Financial knowledge is included in the previous models, the pattern of the findings remains unchanged.

from the random effects Tobit specifications. Moreover, $No\ Financial\ Knowledge$ is statistically significant in the case of $Ln(Total\ Debt)$ and $Ln(Housing\ Debt)$. Overall, the findings from the double hurdle specifications provide further evidence that the household head's attitudes towards risk play an important role in determining the amount of total household debt, housing debt and non-housing debt held by the households.

Finally, we also conduct the double hurdle analysis with the set of risk attitudes dummy variables and the marginal effects of the risk attitudes dummy variables and $No\ Financial\ Knowledge$ are presented in Table A4.6 in the appendix. It is apparent that the marginal effects of the risk attitudes dummy variables are all statistically significant and positive in the case of $Ln(Total\ Debt)$ and $Ln(Non\ housing\ Debt)$ for all three samples. This accords with the findings from the double hurdle analysis for the $Risk\ Attitudes$ index and provides further evidence that the household head's tolerance to risk is positively associated with the amount of total household debt and non-housing debt held by the household. In addition, the magnitudes of the marginal effects of the risk attitudes dummy variables are monotonically increasing in risk tolerance in the case of $Ln(Total\ Debt)$ and $Ln(Non\ housing\ Debt)$ for urban households. For example, the households where the head's risk attitudes are given by $Low\ Risk\ Return$ have 12.18% more total household debt than those where the head is unwilling to carry any risk and the households have 44.11% more total household debt if the head's risk attitudes are given by $High\ Risk\ Return$. Interesting findings are found for the all households sample and the rural sample in that the marginal effects of the risk attitudes dummy variables are monotonically increasing in risk tolerance in terms of magnitude in the case of $Ln(Total\ Debt)$ and $Ln(Non\ housing\ Debt)$ except for the category, $High\ Risk\ Return$. For example, the size of the marginal effect of $Low\ Risk\ Return$ is 0.2013 and the size of the marginal effect of $Slightly\ High\ Risk\ Return$ increases to 0.6134, while the size of marginal effect of $High\ Risk\ Return$ is slightly lower at 0.6008 in the case of $Ln(Total\ Debt)$. With regard to housing debt, we can see from Table A4.6 in the appendix that not all of the risk attitudes dummy variables are statistically significant for the all households sample and the rural sample. Nevertheless, the results of the double hurdle analysis provide further evidence supporting the importance of risk attitudes in determining the amount of household debt held by households. Finally, Table A4.6 in the appendix also shows the results of the test of the exclusion restriction that $No\ Financial\ Knowledge$ is statistically insignificant in the amount of debt specifications conditional on those households with a positive amount of debt, which supports the validity of the exclusion restriction.

4.5. Conclusion

This chapter has investigated the association between household debt and attitudes towards risk using household-level data from the CHFS (2011, 2013, 2015, 2017). In addition to the all households sample, we have explored urban and rural samples based on whether households live in urban or rural areas in order to explore whether the effect of attitudes towards risk on household debt varies across urban and rural households. Household debt is captured by holding total household debt as well as the amount of total household debt held. In addition, we split total household debt into housing debt and non-housing debt.

We have firstly employed the random effects Logit estimator for the all households sample. The findings suggest that the household head's tolerance of risk is positively associated with the probability of holding total household debt, housing debt and non-housing debt. Furthermore, the effect of risk attitudes is greater than that of proxies for the social network in terms of magnitude in the case of total household debt holding and non-housing debt holding, which indicates that $Risk\ Attitudes$ are an important determinant of the probability of holding household debt. Similarly, the random effects Logit results indicate that $Risk\ Attitudes$ also play an important role in determining total household debt and non-housing debt holding for the urban and rural samples.

We have also employed the random effects Tobit estimator to model the log level of total household debt, the log level of housing debt and the log level of non-housing debt. The results of the random effects Tobit model indicate that the more risk tolerant is the household head the greater is the amount of total household debt and non-housing debt, which provides further evidence supporting the importance of risk attitudes for determining household debt.

To shed further light on the effects of $Risk\ Attitudes$ on the probability of holding total household debt, housing debt and non-housing debt, the fixed effects Logit estimator was used to further explore the robustness of the results stemming from random effects Logit model. The findings accord with that from the random effects Logit analysis, with the risk tolerance of the head of household found to be positively associated with $Total\ Debt\ Holding$ and $Non\ housing\ Debt\ Holding$. Finally, the findings are robust to using the double hurdle approach thereby providing further evidence that the risk tolerance of the head of household is positively associated with the amount of household debt.

In conclusion, our household level data indicates that a large proportion of the heads of Chinese households are unwilling to carry any risk based on the descriptive statistics of our measure of risk attitudes. This observation accords with the findings from U.S. household level surveys (see, for example, Avery and Kennickell, 1991). A positive relationship between the risk tolerance of the heads of Chinese households and household debt has been revealed and the results have been found to be robust to a series of econometric approaches, namely the random effects Logit model, the random effects Tobit model, the fixed effects Logit model and the double hurdle model. Some interesting differences are found across the three types of debt as well as by rural and urban residence.

It is apparent that households characterized by high levels of risk tolerance might be more tolerant of shocks in their financial circumstances and consumption; hence the finding that they are more likely to hold debt and are more inclined to accumulate debt accords with intuition. In contrast, those households, who are less tolerant to risk, are found to have a lower probability of holding debt and are less inclined to accumulate debt. Such findings suggest that the observed debt holding and accumulation partially reflect risk attitudes. If policy-makers are concerned about levels of debt, one might argue that it would be hard to influence an individual's attitudes towards risk and it might be the case that policy intervention in other areas such as improving financial literacy as discussed in the previous chapter might be promising. This might help households understand the potential risks associated with taking on debt and this might be especially important in the case of the risks associated with non-housing debt because, as discussed above, non-housing debt is arguably riskier than housing debt. In addition, given the importance of risk attitudes, further work to enhance understanding of the determinants of risk attitudes is important because it has been found to have an influence on debt holding decisions.

Additionally, the role of risk attitudes has been found to differ across urban and rural households. Specifically, there exists a positive relationship between risk attitudes and the probability of holding housing debt for rural households while such a relationship is not found for urban households, which may reflect the relatively high house prices in urban areas and such findings suggest that policy interventions might be better focused on the effects of high house prices in urban areas such as the increase in the minimum down payment ratio, the cap on the loan-to-value ratio, higher mortgage rates for second homes and restrictions on house-purchasing in the first-tier cities where only those with local *hukou* (household registration) or those who have worked in this city for certain consecutive years, are eligible to purchase one or two houses.

Finally, an additional interesting topic for future research, subject to data availability, concerns whether the effect of risk attitudes varies across different types of non-housing debt. In this chapter, due to the available data, non-housing debt includes agricultural/business debt, vehicle-purchasing debt, education debt, credit card debt and other debt. The effect of risk attitudes may differ across these types of debt because of the different levels of perceived riskiness across debt types. Since the debt itself is risk-free, it is also interesting to explore the effect of risk attitudes on leverage ratios, e.g. the house value to mortgage debt ratio, which may provide a further understanding of household debt behaviour. In addition, it is interesting to explore the effect of risk attitudes on formal and informal debt sources because formal and informal debt are associated with different levels and types of risk. Formal debt can generally be obtained from banks or financial companies while informal debt comes from friends or relatives. In other words, formal sources may assess the household's ability to repay debt before lending and then decide the appropriate amount of debt based on the level of risk the household can undertake while the informal sources may not undertake such an exercise. Thus, household debt from formal sources is arguably less risky than that from informal sources, which may lead to a different relationship between the household head's attitudes towards risk and formal and informal debt.

4.6. Figures

Figure 4.1

Distribution of the log level of total household debt in 2011, 2013, 2015 and 2017 (panel), i.e.

$Ln(Total\ Debt) > 0$

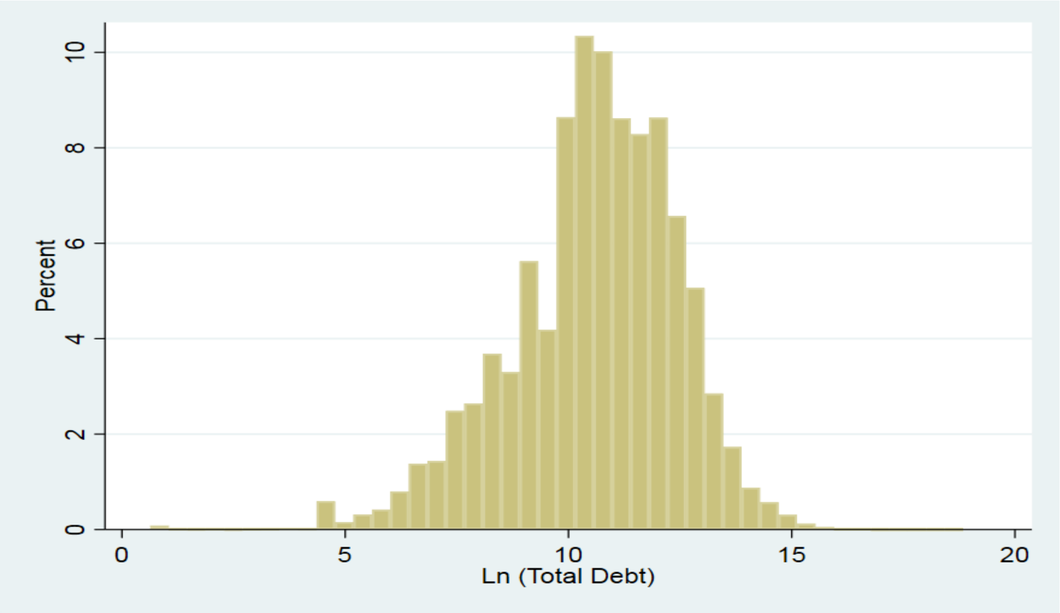


Figure 4.2

Distribution of the log level of housing debt in 2011, 2013, 2015 and 2017 (panel), i.e.

$Ln(Housing\ Debt) > 0$

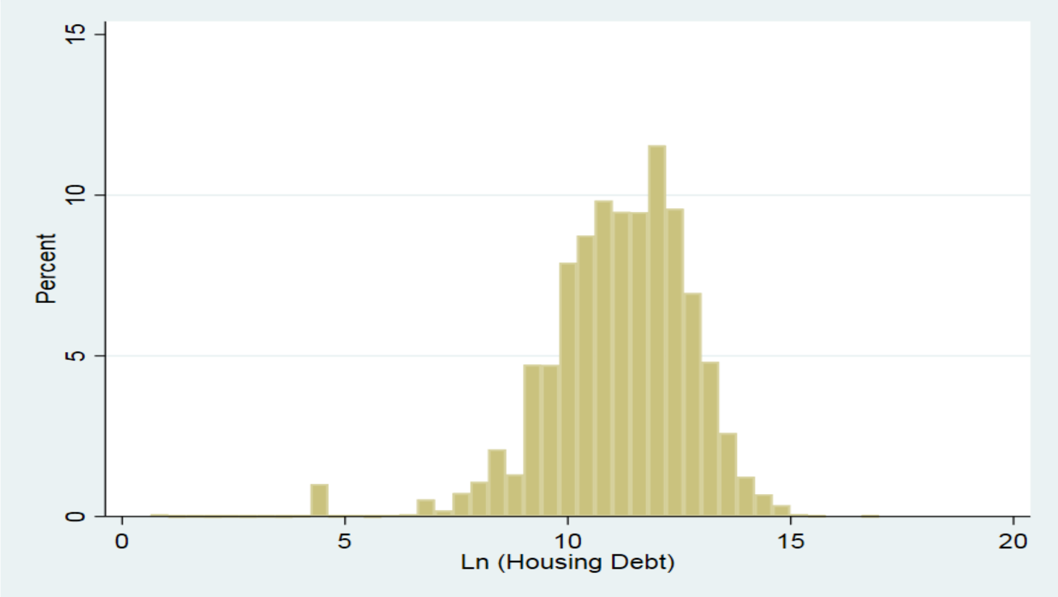
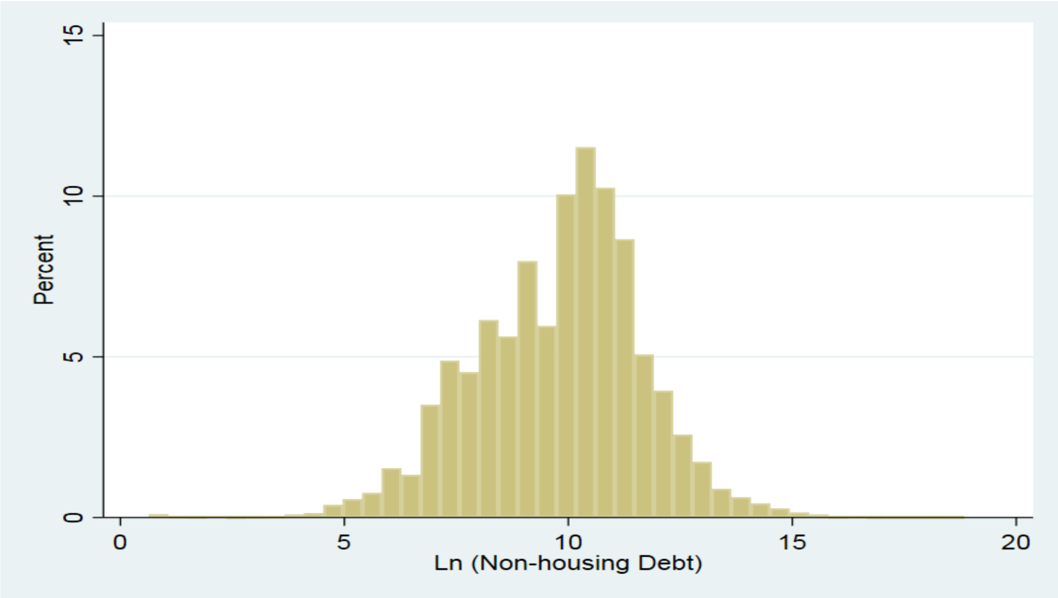


Figure 4.3

Distribution of the log level of non-housing debt in 2011, 2013, 2015 and 2017 (panel), i.e.

$$Ln(Non\ housing\ Debt) > 0$$



4.7. Tables

Table 4.1 Definition of Variables

Variable	Description
Total Debt Holding	Binary variable (0/1) equals 1 if the household holds any housing debt and/or non-housing debt.
Housing Debt Holding	Binary variable (0/1) equals 1 if the household currently holds debt for house-purchasing or house-renovation including mortgages from banks and loans from relatives or friends.
Non housing Debt Holding	Binary variable (0/1) equals 1 if the household currently holds debt for non-housing debt including agricultural/business debt, vehicle-purchasing debt, education debt, credit card debt and other debt.
Ln(Total Debt)	The natural logarithm of the amount of total household debt held by the household plus one. Total household debt includes agricultural/business debt, vehicle-purchasing debt, house-purchasing debt, education debt, credit debt and other debt.
Ln(Housing Debt)	The natural logarithm of the amount of housing debt held by the household plus one.
Ln(Non housing Debt)	The natural logarithm of the amount of non-housing debt held by the household plus one.
Risk Attitudes	A 5 point index for the head of household ranging from 0 to 4, which is increasing in risk-tolerance, where: 0 denotes a household head who is unwilling to carry any risk; 1 denotes a household head who prefers projects with slight risk and return; 2 denotes a household head who prefers projects with average risk and return; 3 denotes a household head who prefers projects with slightly high risk and slightly high return; and 4 denotes a household head who prefers projects with high risk and high return.
No Risk Return (Omitted)	Dummy variable (0/1) equals 1 if the household head is unwilling to carry any risk.
Low Risk Return	Dummy variable (0/1) equals 1 if the household head prefers projects with slight risk and return.
Average Risk Return	Dummy variable (0/1) equals 1 if the household head prefers projects with average risk and return.
Slightly High Risk Return	Dummy variable (0/1) equals 1 if the household head prefers projects with slightly high risk and return.
High Risk Return	Dummy variable (0/1) equals 1 if the household head prefers projects with high risk and return.
Ln(Income)	The natural logarithm of the total amount of the household disposable annual income plus one.
Ln(Assets)	The natural logarithm of the total amount of the household assets plus one. Total household assets include agricultural and business assets, land and real estate, vehicles, stocks, financial derivatives, non-RMB assets, gold, funds, bonds and financial wealth-management products savings and cash etc.
Ln(Social Network)	The natural logarithm of the total amount of expenditure giving to non-family members including wedding gifts, funeral money, education, treatment, donations and others.
No. Siblings	Number of siblings of the household head and his/her spouse.
No. Children	Number of dependent children aged below 16 in the household.
No. Workers	Number of workers in the household excluding the household head.
No. Aged Over 60	Number of family members who are aged over 60 in the household excluding the household head.
Self Assessed Health	A 5 point index for the head of household ranging from 0 to 4, where: 0 denotes very poor; 1 denotes poor; 2 denotes normal; 3 denotes good and 4 denotes very good.
Age	Age of the household head.
Age^2	Age squared of the household head.
Male	Dummy variable (0/1) equals 1 if the household head is male.
Married	Dummy variable (0/1) equals 1 if the household head is married.
Party Member	Dummy variable (0/1) equals 1 if the household head is a party member.

[a] All monetary variables in the 2015 and 2017 waves are deflated using China's yearly CPI, with the benchmark year 2011 = 100, year 2013 = 102*102, year 2015 = 102*102*101.4*102 and year 2017 = 102*102*101.4*102*102*101.6.

Table 4.1 Definition of Variables (Continued)

Variable	Description
No Schooling (Omitted)	Dummy variable (0/1) equals 1 if the household head never attended school.
Primary School	Dummy variable (0/1) equals 1 if the highest education level of the household head is primary school.
Junior High	Dummy variable (0/1) equals 1 if the highest education level of the household head is junior high school.
Senior High	Dummy variable (0/1) equals 1 if the highest education level of the household head is senior high school or technical school.
College/Bachelor	Dummy variable (0/1) equals 1 if the highest education level of the household head is vocational college or bachelor degree.
Master/PhD	Dummy variable (0/1) equals 1 if the highest education level of the household head is above bachelor's degree (master or PhD).
No Financial Knowledge	Dummy variable (0/1) equals 1 if the household head answered incorrectly in all three financial literacy questions about the interest rate, inflation and risk diversification (see Chapter 3 for full details).
Employed	Dummy variable (0/1) equals 1 if the household head is an employee, i.e. employed by someone else.
Self Employed	Dummy variable (0/1) equals 1 if the household head is self-employed.
Retired	Dummy variable (0/1) equals 1 if the household head is retired.
Not Working	Dummy variable (0/1) equals 1 if the household head is not working, i.e. the household head is unemployed, incapacitated, a homemaker, a volunteer or unwilling to work.
Farmer (Omitted)	Dummy variable (0/1) equals 1 if the household head is a farmer
Rural	Dummy variable (0/1) equals 1 if the household resides in a rural area, equals 0 if the household resides in an urban area.
North East	Dummy Variable (0/1) equals 1 if the household lives in the North-eastern region of China including 3 provinces: Heilongjiang, Jilin, Liaoning.
East	Dummy Variable (0/1) equals 1 if the household lives in the Eastern region of China including 7 provinces: Shanghai, Jiangsu, Zhejiang, Anhui, Fujian, Jiangxi, Shandong.
North	Dummy Variable (0/1) equals 1 if the household lives in the Northern region of China including 5 provinces: Beijing, Tianjin, Shanxi, Hebei, Neimenggu.
Central	Dummy Variable (0/1) equals 1 if the household lives in the Central region of China including 3 provinces: Henan, Hubei, Hunan.
South	Dummy Variable (0/1) equals 1 if the household lives in the Southern region of China including 3 provinces: Guangdong, Guangxi, Hainan.
South West	Dummy Variable (0/1) equals 1 if the household lives in the South-western region of China including 4 provinces: Chongqing, Sichuan, Guizhou, Yunnan.
North West (Omitted)	Dummy Variable (0/1) equals 1 if the household lives in the North-western region of China including 4 provinces: Shaanxi, Gansu, Qinghai, Ningxia.
2011 Year (Omitted)	Dummy Variable (0/1) equals 1 if the household responded in 2011.
2013 Year	Dummy Variable (0/1) equals 1 if the household responded in 2013.
2015 Year	Dummy Variable (0/1) equals 1 if the household responded in 2015.
2017 Year	Dummy Variable (0/1) equals 1 if the household responded in 2017.
NT (Observations)	Total number of observations
N	Total number of households

[a] All monetary variables in the 2015 and 2017 waves are deflated using China's yearly CPI, with the benchmark year 2011 = 100, year 2013 = 102*102, year 2015 = 102*102*101.4*102 and year 2017 = 102*102*101.4*102*102*101.6.

Table 4.2 Summary Statistics - All Variables; Panel (t = 2011, 2013, 2015 and 2017)

	All Households				Urban Sample				Rural Sample			
	Mean	Std. Dev	Min	Max	Mean	Std. Dev	Min	Max	Mean	Std. Dev	Min	Max
Total Debt Holding	0.288	0.453	0	1	0.262	0.440	0	1	0.346	0.476	0	1
Housing Debt Holding	0.147	0.354	0	1	0.151	0.358	0	1	0.138	0.345	0	1
Non housing Debt Holdin	0.183	0.386	0	1	0.148	0.355	0	1	0.262	0.440	0	1
Ln(Total Debt)	3.039	4.887	0	18.840	2.865	4.906	0	18.840	3.434	4.822	0	16.276
Ln(Housing Debt)	1.640	3.997	0	16.983	1.743	4.170	0	16.983	1.407	3.565	0	15.011
Ln(Non housing Debt)	1.787	3.864	0	18.840	1.472	3.618	0	18.840	2.501	4.285	0	16.256
Risk Attitudes	0.939	1.182	0	4	1.009	1.190	0	4	0.780	1.148	0	4
No Risk Return	0.518	0.500	0	1	0.483	0.500	0	1	0.598	0.490	0	1
Low Risk Return	0.182	0.386	0	1	0.191	0.393	0	1	0.162	0.368	0	1
Average Risk Return	0.196	0.397	0	1	0.213	0.410	0	1	0.158	0.365	0	1
Slightly High Risk Return	0.048	0.215	0	1	0.058	0.235	0	1	0.026	0.159	0	1
High Risk Return	0.055	0.227	0	1	0.054	0.226	0	1	0.056	0.231	0	1
Ln(Income)	10.453	1.444	0.151	15.391	10.750	1.338	0.151	15.391	9.779	1.450	0.245	15.391
Ln(Assets)	12.551	1.738	0.635	17.096	12.929	1.704	0.635	17.096	11.696	1.491	0.635	17.096
Ln(Social Network)	6.194	3.422	0	14.221	6.509	3.316	0	14.221	5.479	3.548	0	12.608
No. Siblings	3.323	3.402	0	15	3.129	3.255	0	15	3.764	3.676	0	15
No. Children	0.494	0.740	0	4	0.442	0.671	0	4	0.611	0.867	0	4
No. Workers	0.996	0.994	0	6	0.834	0.879	0	6	1.364	1.131	0	6
No. Aged Over 60	0.407	0.579	0	3	0.381	0.568	0	3	0.465	0.599	0	3
Self Assessed Health	2.197	1.087	0	4	2.300	1.052	0	4	1.964	1.130	0	4
Age	52.90	14.152	20	90	51.932	14.894	20	90	55.098	12.021	20	90
Male	0.762	0.426	0	1	0.707	0.455	0	1	0.887	0.317	0	1
Married	0.870	0.337	0	1	0.857	0.350	0	1	0.899	0.301	0	1
Party Member	0.301	0.459	0	1	0.335	0.472	0	1	0.222	0.416	0	1
No Schooling (Omitted)	0.063	0.243	0	1	0.037	0.189	0	1	0.123	0.328	0	1
Primary School	0.222	0.416	0	1	0.150	0.357	0	1	0.386	0.487	0	1
Junior High	0.333	0.471	0	1	0.320	0.466	0	1	0.365	0.481	0	1
Senior High	0.205	0.404	0	1	0.246	0.431	0	1	0.113	0.317	0	1
College/Bachelor	0.165	0.371	0	1	0.232	0.422	0	1	0.014	0.116	0	1
Master/PhD	0.011	0.103	0	1	0.016	0.124	0	1	0.000	0.013	0	1
No Financial Knowledge	0.472	0.499	0	1	0.386	0.487	0	1	0.648	0.478	0	1
Employed	0.355	0.478	0	1	0.420	0.494	0	1	0.208	0.406	0	1
Self Employed	0.110	0.313	0	1	0.128	0.334	0	1	0.070	0.255	0	1
Retired	0.153	0.360	0	1	0.212	0.409	0	1	0.021	0.144	0	1
Not Working	0.177	0.382	0	1	0.183	0.386	0	1	0.164	0.370	0	1
Farmer (Omitted)	0.205	0.403	0	1	0.058	0.234	0	1	0.537	0.499	0	1
Rural	0.306	0.461	0	1								
North East	0.121	0.326	0	1	0.126	0.332	0	1	0.111	0.314	0	1
East	0.286	0.452	0	1	0.300	0.458	0	1	0.255	0.436	0	1
North	0.163	0.369	0	1	0.176	0.381	0	1	0.134	0.340	0	1
Central	0.122	0.327	0	1	0.110	0.313	0	1	0.149	0.356	0	1
South	0.104	0.305	0	1	0.109	0.311	0	1	0.094	0.292	0	1
South West	0.121	0.326	0	1	0.102	0.303	0	1	0.162	0.368	0	1
North West (Omitted)	0.083	0.276	0	1	0.077	0.267	0	1	0.096	0.294	0	1
2011 Year (Omitted)	0.070	0.255	0	1	0.060	0.238	0	1	0.091	0.287	0	1
2013 Year	0.267	0.443	0	1	0.266	0.442	0	1	0.275	0.446	0	1
2015 Year	0.325	0.468	0	1	0.333	0.471	0	1	0.307	0.461	0	1
2017 Year	0.336	0.472	0	1	0.340	0.474	0	1	0.328	0.469	0	1
Number of Observations		91,354				63,378				27,976		

[a] The summary statistics for financial knowledge are based on samples with a different number of observations, namely 79,614 in the all households sample, 53,658 in the urban sample and 25,956 in the rural sample, because those households who did not provide information on financial literacy were omitted.

Table 4.3 The Determinants of the Probability of Total Debt, Housing Debt and Non-housing Debt Holding - Random Effects Logit Analysis

	Total Debt Holding			Housing Debt Holding			Non-housing Debt Holding		
All households	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat
Risk Attitudes	0.0995***	0.0131	10.26	0.0128	0.0011	1.03	0.1474***	0.0161	14.80
Age	0.0773***	0.0102	10.23	0.0557***	0.0046	5.70	0.0693***	0.0076	8.67
Age^2	- 0.0009***	- 0.0001	- 13.99	- 0.0007***	- 0.0001	- 8.85	- 0.0007***	- 0.0001	- 11.67
Male	0.0632**	0.0084	2.13	- 0.0235	- 0.0019	- 0.63	0.1398***	0.0153	4.50
Married	0.0396	0.0052	1.10	0.1428***	0.0117	2.97	- 0.0238	- 0.0026	- 0.63
Party Member	0.0390	0.0052	1.47	0.0178	0.0015	0.53	0.0438	0.0048	1.56
Self Assessed Health	- 0.2606***	- 0.0345	- 22.83	- 0.2103***	- 0.0172	- 14.26	- 0.2538***	- 0.0278	- 21.23
Primary School	- 0.2190***	- 0.0290	- 4.07	- 0.2303***	- 0.0188	- 3.16	- 0.1309**	- 0.0143	- 2.41
Junior High	- 0.3602***	- 0.0477	- 6.56	- 0.4663***	- 0.0381	- 6.31	- 0.2161***	- 0.0236	- 3.89
Senior High	- 0.3689***	- 0.0489	- 6.25	- 0.5545***	- 0.0454	- 7.01	- 0.1672***	- 0.0183	- 2.78
College/Bachelor	- 0.0514	- 0.0068	- 0.80	- 0.1324	- 0.0108	- 1.58	- 0.0596	- 0.0065	- 0.90
Master/PhD	0.2490**	0.0330	2.15	0.3427**	0.0280	2.50	- 0.0297	- 0.0032	- 0.25
Employed	- 0.3298***	- 0.0433	- 8.95	0.2549***	0.0209	5.30	- 0.7223***	- 0.0790	- 19.31
Self Employed	0.0944**	0.0125	2.18	- 0.1825***	- 0.0149	- 3.19	0.2109***	0.0231	4.92
Retired	- 0.9926***	- 0.1315	- 18.81	- 0.6398***	- 0.0523	- 9.06	- 1.1552***	- 0.1264	- 19.50
Not Working	- 0.3544***	- 0.0470	- 8.99	0.0063	0.0005	0.12	- 0.5396***	- 0.0590	- 13.30
Ln(Income)	- 0.0043	- 0.0006	- 0.45	- 0.0570***	- 0.0047	- 4.60	0.0174*	0.0019	1.77
Ln(Assets)	0.1812***	0.0240	21.06	0.5565***	0.0455	40.82	- 0.0473***	- 0.0052	- 5.55
Ln(Social Network)	0.0207***	0.0027	6.27	- 0.0024	- 0.0002	- 0.57	0.0293***	0.0032	8.30
No. Siblings	0.0039	0.0005	0.96	0.0188***	0.0015	3.66	- 0.0097**	- 0.0011	- 2.25
No. Children	0.2075***	0.0275	14.02	0.1861***	0.0152	9.88	0.1593***	0.0174	10.53
No. Workers	0.1473***	0.0195	12.59	0.1522***	0.0124	10.10	0.1232***	0.0135	10.13
No. Aged Over 60	- 0.0766***	- 0.0102	- 3.64	- 0.1205***	- 0.0099	- 4.47	- 0.0315	- 0.0034	- 1.46
Rural	0.5982***	0.0793	18.22	0.3712***	0.0304	8.63	0.6220***	0.0680	18.96
North East	- 0.0712	- 0.0094	- 1.32	- 0.2113***	- 0.0173	- 3.01	- 0.0319	- 0.0035	- 0.60
East	- 0.4940***	- 0.0655	- 10.42	- 0.4801***	- 0.0393	- 7.95	- 0.4290***	- 0.0469	- 9.06
North	- 0.2448***	- 0.0324	- 4.78	- 0.4301***	- 0.0352	- 6.53	- 0.1223**	- 0.0134	- 2.40
Central	- 0.1883***	- 0.0250	- 3.53	- 0.2049***	- 0.0168	- 2.99	- 0.1503***	- 0.0164	- 2.86
South	- 0.3250***	- 0.0431	- 6.00	- 0.1510**	- 0.0124	- 2.22	- 0.4059***	- 0.0444	- 7.43
South West	0.0032	0.0004	0.06	0.3019***	0.0247	4.55	- 0.2253***	- 0.0246	- 4.30
Year 2013	- 0.7673***	- 0.1017	- 18.17	- 0.0314***	- 0.1020	- 23.40	- 0.3511***	- 0.0384	- 7.82
Year 2015	- 0.4555***	- 0.0604	- 10.55	- 0.8189***	- 0.0670	- 15.36	- 0.1281***	- 0.0140	- 2.79
Year 2017	- 0.0018	- 0.0002	- 0.04	- 0.4694***	- 0.0384	- 7.58	0.2439***	0.0267	4.60
ρ; Std Err		0.4013; 0.0077			0.4783; 0.0094			0.3177; 0.0088	
Chibar2 (01); p value		3,114.51; p = [0.0000]			2,491.69; p = [0.0000]			1,479.08; p = [0.0000]	
Wald χ^2 (35); p value		6,240.82; p = [0.0000]			4,289.20; p = [0.0000]			5,200.53; p = [0.0000]	
Observations		91,354			91,354			91,354	
Urban sample	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat
Risk Attitudes	0.0837***	0.0102	6.75	- 0.0114	- 0.0009	- 0.71	0.1412***	0.0135	11.09
ρ; Std Err		0.4269; 0.0099			0.5281; 0.0113			0.3067; 0.0120	
Chibar2 (01); p value		2,029.83; p = [0.0000]			1,967.22; p = [0.0000]			704.59; p = [0.0000]	
Wald χ^2 (34); p value		4,270.84; p = [0.0000]			3,030.11; p = [0.0000]			3,076.62; p = [0.0000]	
Observations		63,378			63,378			63,378	
Rural sample	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat
Risk Attitudes	0.1248***	0.0193	7.90	0.0487***	0.0043	2.45	0.1588***	0.0217	9.74
ρ; Std Err		0.3592; 0.0126			0.3713; 0.0170			0.3424; 0.0135	
Chibar2 (01); p value		983.61; p = [0.0000]			484.79; p = [0.0000]			768.70; p = [0.0000]	
Wald χ^2 (34); p value		1,824.66; p = [0.0000]			1,120.95; p = [0.0000]			1,765.99; p = [0.0000]	
Observations		27,976			27,976			27,976	

[a] *, **, *** denote 10, 5, 1 per cent levels of significance, respectively.
[b] Reference categories: Education controls: the omitted group is that household head never attended school; Labour market controls: the omitted group is that household head who is a farmer; Region controls, the omitted group is the North West; Year controls, the omitted group is the year 2011.

Table 4.4 The Determinants of the Log Level of Total Debt, Housing Debt and Non-housing Debt - Random Effects Tobit Analysis

	Ln(Total Debt)				Ln(Housing Debt)				Ln(Non-housing Debt)			
All households	E.M.E	t-stat	I.M.E	t-stat	E.M.E	t-stat	I.M.E	t-stat	E.M.E	t-stat	I.M.E	t-stat
Risk Attitudes	0.0123***	10.19	0.1293***	10.17	0.0011	1.07	0.0157	1.07	0.0158***	14.87	0.1839***	14.80
ρ; Std Err	0.4021; 0.0067				0.4666; 0.0087				0.3222; 0.0083			
Chibar2 (01); p value	3,585.53; p = [0.0000]				2,598.44; p = [0.0000]				1,607.68; p = [0.0000]			
Wald χ^2 (35); p value	8,380.18; p = [0.0000]				5,343.29; p = [0.0000]				5,733.06; p = [0.0000]			
Uncensored obs	26,288				13,449				16,696			
Left censored obs	65,066				77,905				74,658			
Observations	91,354				91,354				91,354			
Urban sample	E.M.E	t-stat	I.M.E	t-stat	E.M.E	t-stat	I.M.E	t-stat	E.M.E	t-stat	I.M.E	t-stat
Risk Attitudes	0.0096***	6.67	0.1063***	6.66	- 0.0011	- 0.64	- 0.0114	- 0.64	0.0135***	11.22	0.1785***	11.18
ρ; Std Err	0.4240; 0.0085				0.5124; 0.0102				0.3033; 0.0115			
Chibar2 (01); p value	2,335.07; p = [0.0000]				2,169.03; p = [0.0000]				743.52; p = [0.0000]			
Wald χ^2 (34); p value	6,059.79; p = [0.0000]				4,197.10; p = [0.0000]				3,336.43; p = [0.0000]			
Uncensored obs	16,610				9,598				9,363			
Left censored obs	46,768				53,780				54,015			
Observations	63,378				63,378				63,378			
Rural sample	E.M.E	t-stat	I.M.E	t-stat	E.M.E	t-stat	I.M.E	t-stat	E.M.E	t-stat	I.M.E	t-stat
Risk Attitudes	0.0182***	8.05	0.1700***	8.02	0.0041**	2.37	0.0588**	2.37	0.0209***	9.86	0.2024***	9.80
ρ; Std Err	0.3533; 0.0108				0.3599; 0.0163				0.3429; 0.0121			
Chibar2 (01); p value	1,128.96; p = [0.0000]				502.02; p = [0.0000]				867.37; p = [0.0000]			
Wald χ^2 (34); p value	2,277.68; p = [0.0000]				1,191.57; p = [0.0000]				2,030.45; p = [0.0000]			
Uncensored obs	9,678				3,851				7,333			
Left censored obs	18,298				24,125				20,643			
Observations	27,976				27,976				27,976			

[a] *, **, *** denote 10, 5, 1 per cent levels of significance, respectively.
[b] E.M.E. indicates the marginal effects at the extensive margin; I.M.E. indicates the marginal effects at the intensive margin.
[c] All other control variables are included in this analysis.

Table 4.5 The Determinants of the Probability of Total Debt, Housing Debt and Non-housing Debt Holding - Fixed Effects Logit Analysis

	Total Debt Holding			Housing Debt Holding			Non-housing Debt Holding		
All households	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat
Risk Attitudes	0.0378**	0.0079	2.31	- 0.0133	- 0.0001	- 0.62	0.0590***	0.0046	3.33
LR χ^2 (27); p value	909.60; p = [0.0000]			1,375.28; p = [0.0000]			471.67; p = [0.0000]		
Observations	25,092			15,950			20,941		
Urban sample	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat
Risk Attitudes	0.0195	0.0021	0.86	- 0.0319	- 0.0000	- 1.09	0.0453*	0.0033	1.78
LR χ^2 (26); p value	515.16; p = [0.0000]			923.76; p = [0.0000]			197.84; p = [0.0000]		
Observations	14,452			9,678			11,425		
Rural sample	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat
Risk Attitudes	0.0573**	0.0129	2.33	0.0038	0.0004	0.12	0.0753***	0.0107	2.91
LR χ^2 (26); p value	495.19; p = [0.0000]			456.64; p = [0.0000]			465.27; p = [0.0000]		
Observations	10,164			5,974			9,065		

[a] *, **, *** denote 10, 5, 1 per cent levels of significance, respectively.
[b] All other control variables are included in this analysis.

Table 4.6 The Determinants of the Log Level of Total Debt, Housing Debt and Non-housing Debt - Double Hurdle Analysis (Pooled)

	Ln(Total Debt)		Ln(Housing Debt)		Ln(Non-housing Debt)	
All households	M.E.	t-stat	M.E.	t-stat	M.E.	t-stat
Outcome & Selection equation						
Risk Attitudes	0.1650***	10.46	0.0222*	1.70	0.1776***	14.69
Selection equation						
No Financial Knowledge	0.1277***	3.12	0.1026***	3.04	0.0637**	2.01
LR χ^2 (35); p value	6,549.46; p = [0.0000]		6,795.78; p = [0.0000]		2,571.79; p = [0.0000]	
Pseudo R^2	0.0892		0.1201		0.0774	
Exclusion restriction						
H0: No Financial knowledge = 0	p = [0.473]		p = [0.872]		p = [0.166]	
Observations	79,614		79,614		79,614	
Urban sample	M.E.	t-stat	M.E.	t-stat	M.E.	t-stat
Outcome & Selection equation						
Risk Attitudes	0.1246***	6.39	- 0.0061	- 0.36	0.1489***	10.82
Selection equation						
No Financial Knowledge	0.1092**	2.18	0.0936**	2.13	0.0541	1.50
LR χ^2 (34); p value	3,159.74; p = [0.0000]		3,597.50; p = [0.0000]		1,346.57; p = [0.0000]	
Pseudo R^2	0.0886		0.1214		0.0728	
Exclusion restriction						
H0: No Financial knowledge = 0	p = [0.718]		p = [0.527]		p = [0.127]	
Observations	53,658		53,658		53,658	
Rural sample	M.E.	t-stat	M.E.	t-stat	M.E.	t-stat
Outcome & Selection equation						
Risk Attitudes	0.2313***	8.57	0.0604***	3.05	0.2314***	9.81
Selection equation						
No Financial Knowledge	0.1708**	2.42	0.1077**	2.11	0.0944	1.52
LR χ^2 (34); p value	2,124.34; p = [0.0000]		4,095.48; p = [0.0000]		1,896.06; p = [0.0000]	
Pseudo R^2	0.0731		0.0864		0.0738	
Exclusion restriction						
H0: No Financial knowledge = 0	p = [0.402]		p = [0.125]		p = [0.975]	
Observations	25,956		25,956		25,956	

[a] *, **, *** denote 10, 5, 1 per cent levels of significance, respectively.
[b] All other control variables are included in this analysis.
c The null hypothesis H0: No Financial knowledge = 0 is for testing the exclusion restriction where we test the effect of No Financial Knowledge on the amount of debt for those households with positive amount of debt.

Appendix to Chapter 4

Table A4.2 Medians - Outcomes; Panel (t = 2011, 2013, 2015 and 2017)

	All households		Urban		Rural	
	Median		Median		Median	
	All	Ex. Zero	All	Ex. Zero	All	Ex. Zero
Total Debt	0	¥46,000	0	¥73,478	0	¥25,717
Ln(Total Debt)	0	10.735	0	11.205	0	10.155
Housing Debt	0	¥82,700	0	¥124,773	0	¥28,499
Ln(Housing Debt)	0	11.323	0	11.734	0	10.258
Non housing Debt	0	¥19,500	0	¥26,588	0	¥17,725
Ln(Non housing Debt)	0	9.878	0	10.188	0	9.783
Number of Observations	91,354		63,378		27,976	

[a] All figures are in 2011 prices.

Table A4.3 The Determinants of the Probability of Total Debt, Housing Debt and Non-housing Debt Holding - Random Effects Logit Analysis (Risk Attitudes Dummy Variables)

	Total Debt Holding			Housing Debt Holding			Non-housing Debt Holding		
All households	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat
Low Risk Return	0.1226***	0.0162	4.04	0.0728*	0.0059	1.88	0.1105***	0.0121	3.44
Average Risk Return	0.1993***	0.0264	6.63	0.0090	0.0007	0.23	0.3068***	0.0336	9.80
Slightly High Risk Return	0.3436***	0.0455	6.79	0.1160*	0.0095	1.87	0.5080***	0.0556	9.81
High Risk Return	0.3781**	0.0501	8.02	0.0351	0.0029	0.58	0.5435***	0.0595	11.43
ρ; Std Err	0.4012; 0.0077			0.4783; 0.0094			0.3175; 0.0088		
Wald χ^2 (38); p value	6,242.26; p = [0.0000]			4,289.20; p = [0.0000]			5,205.47; p = [0.0000]		
Chibar2 (01); p value	3,112.69; p = [0.0000]			2,490.52; p = [0.0000]			1,476.45; p = [0.0000]		
Observations	91,354			91,354			91,354		
Urban sample	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat
Low Risk Return	0.1029***	0.0125	2.67	0.0230	0.0018	0.46	0.1147***	0.0110	2.79
Average Risk Return	0.1807***	0.0220	4.77	- 0.0268	- 0.0021	- 0.55	0.3123***	0.0299	7.89
Slightly High Risk Return	0.2821***	0.0343	4.77	0.0428	0.0033	0.58	0.4603***	0.0441	7.66
High Risk Return	0.3084***	0.0375	5.12	- 0.0824	- 0.0064	- 1.05	0.5133***	0.0491	8.48
ρ; Std Err	0.4269; 0.0099			0.5281; 0.0113			0.3066; 0.0120		
Wald χ^2 (37); p value	4,270.57; p = [0.0000]			3,030.54; p = [0.0000]			3,078.62; p = [0.0000]		
Chibar2 (01); p value	2,029.95; p = [0.0000]			1,966.68; p = [0.0000]			704.31; p = [0.0000]		
Observations	63,378			63,378			63,378		
Rural sample	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat
Low Risk Return	0.1719***	0.0266	3.40	0.1835***	0.0163	2.91	0.1178**	0.0161	2.22
Average Risk Return	0.2062***	0.0319	4.04	0.0232	0.0021	0.36	0.2814***	0.0384	5.29
Slightly High Risk Return	0.5037***	0.0780	4.65	0.2311*	0.0206	1.73	0.6650***	0.0908	6.06
High Risk Return	0.4956***	0.0767	6.44	0.2277**	0.0203	2.37	0.6118***	0.0836	7.78
ρ; Std Err	0.3588; 0.0126			0.3709; 0.0170			0.3416; 0.0135		
Wald χ^2 (37); p value	1,828.02; p = [0.0000]			1,126.40; p = [0.0000]			1,771.08; p = [0.0000]		
Chibar2 (01); p value	980.40; p = [0.0000]			484.01; p = [0.0000]			764.53; p = [0.0000]		
Observations	27,976			27,976			27,976		

[a] *, **, *** denote 10, 5, 1 per cent levels of significance, respectively.
[b] No Risk Return is a dummy variable, which equals 1 if the household head is unwilling to carry any risk; Low Risk Return is a dummy variable, which equals 1 if the household head prefers project with slight risk and return; Average Risk Return is a dummy variable, which equals 1 if the household head prefers project with average risk and return; Slightly High Risk Return is a dummy variable, which equals 1 if the household head prefers project with slightly high risk and return; High Risk Return is a dummy variable, which equals 1 if the household head prefers project with high risk and return.
[c] Reference category: Risk attitudes dummies controls: the omitted group is that household head who is unwilling to carry any risk.
[d] All other control variables are included in this analysis.

Table A4.4 The Determinants of the Log Level of Total Debt, Housing Debt and Non-housing Debt - Random Effects Tobit Analysis (Risk Attitudes Dummy Variables)

	Ln(Total Debt)				Ln(Housing Debt)				Ln(Non-housing Debt)			
All households	E.M.E	t-stat	I.M.E	t-stat	E.M.E	t-stat	I.M.E	t-stat	E.M.E	t-stat	I.M.E	t-stat
Low Risk Return	0.0157***	4.14	0.1646***	4.14	0.0055*	1.78	0.0810*	1.78	0.0116***	3.40	0.1356***	3.40
Average Risk Return	0.0252***	6.70	0.2638***	6.69	0.0007	0.22	0.0100	0.22	0.0331***	9.89	0.3848***	9.87
Slightly High Risk Return	0.0434***	6.92	0.4545***	6.91	0.0095*	1.92	0.1400*	1.92	0.0544***	9.81	0.6334***	9.80
High Risk Retur	0.0462***	7.87	0.4838***	7.86	0.0029	0.61	0.0434	0.61	0.0581***	11.43	0.6760***	11.40
ρ; Std Err	0.4020; 0.0067				0.4665; 0.0087				0.3219; 0.0083			
Wald χ^2 (35); p value	8,382.12; p = [0.0000]				5,346.74; p = [0.0000]				5,736.82; p = [0.0000]			
Chibar2 (01); p value	3,583.42; p = [0.0000]				2,597.76; p = [0.0000]				1,605.10; p = [0.0000]			
Uncensored obs	26,288				13,449				16,696			
Left censored obs	65,066				77,905				74,658			
Observations	91,354				91,354				91,354			
Urban sample	E.M.E	t-stat	I.M.E	t-stat	E.M.E	t-stat	I.M.E	t-stat	E.M.E	t-stat	I.M.E	t-stat
Low Risk Return	0.0126***	2.83	0.1402***	2.83	0.0016	0.44	0.0244	0.44	0.0107***	2.78	0.1419***	2.78
Average Risk Return	0.0216***	4.92	0.2401***	4.92	- 0.0019	- 0.52	- 0.0287	- 0.52	0.0301***	8.10	0.3997***	8.08
Slightly High Risk Return	0.0332***	4.90	0.3696***	4.90	0.0036	0.65	0.0540	0.65	0.0440***	7.74	0.5843***	7.72
High Risk Retur	0.0341***	4.90	0.3790***	4.90	- 0.0059	- 1.01	- 0.0883	- 1.01	0.0484***	8.46	0.6425***	8.44
ρ; Std Err	0.4241; 0.0085				0.5124; 0.0102				0.3032; 0.0115			
Wald χ^2 (35); p value	6,060.63; p = [0.0000]				4,198.96; p = [0.0000]				3,338.49; p = [0.0000]			
Chibar2 (01); p value	2,335.50; p = [0.0000]				2,068.90; p = [0.0000]				743.08; p = [0.0000]			
Uncensored obs	16,610				9,598				9,363			
Left censored obs	46,768				53,780				54,015			
Observations	63,378				63,378				63,378			

(Continued)

Table A4.4 The Determinants of the Log Level of Total Debt, Housing Debt and Non-housing Debt - Random Effects Tobit Analysis (Risk Attitudes Dummy Variables) (Continued)

	Ln(Total Debt)				Ln(Housing Debt)				Ln(Non-housing Debt)			
Rural sample	E.M.E	t-stat	I.M.E	t-stat	E.M.E	t-stat	I.M.E	t-stat	E.M.E	t-stat	I.M.E	t-stat
Low Risk Return	0.0253***	3.49	0.2369***	3.49	0.0155***	2.79	0.2196***	2.79	0.0154**	2.21	0.1489**	2.21
Average Risk Return	0.0291***	3.96	0.2720***	3.96	0.0016	0.29	0.0232	0.29	0.0361***	5.18	0.3492***	5.18
Slightly High Risk Return	0.0726***	4.74	0.6799***	4.73	0.0197*	1.67	0.2796*	1.67	0.0868***	6.09	0.8401***	6.08
High Risk Retur	0.0732***	6.69	0.6848***	6.67	0.0196**	2.32	0.2790**	2.32	00815***	7.98	0.7892***	7.95
ρ; Std Err	0.3529; 0.0108				0.3596; 0.0163				0.3422; 0.0121			
Wald χ^2 (35); p value	2,281.67; p = [0.0000]				1,196.82; p = [0.0000]				2,035.05; p = [0.0000]			
Chibar2 (01); p value	1,125.02; p = [0.0000]				501.43; p = [0.0000]				862.60; p = [0.0000]			
Uncensored obs	9,678				3,851				7,333			
Left censored obs	18,298				24,125				20,643			
Observations	27,976				27,976				27,976			

[a] *, **, *** denote 10, 5, 1 per cent levels of significance, respectively.
[b] E.M.E. indicates the marginal effects at the extensive margin; I.M.E. indicates the marginal effects at the intensive margin.
[c] No Risk Return is a dummy variable, which equals 1 if the household head is unwilling to carry any risk; Low Risk Return is a dummy variable, which equals 1 if the household head prefers project with slight risk and return; Average Risk Return is a dummy variable, which equals 1 if the household head prefers project with average risk and return; Slightly High Risk Return is a dummy variable, which equals 1 if the household head prefers project with slightly high risk and return; High Risk Return is a dummy variable, which equals 1 if the household head prefers project with high risk and return.
[d] Reference category: Risk attitudes dummies controls: the omitted group is that household head who is unwilling to carry any risk.
[e] All other control variables are included in this analysis.

Table A4.5 The Determinants of the Probability of Total Debt, Housing Debt and Non-housing Debt Holding - Fixed Effects Logit Analysis (Risk Attitudes Dummy Variables)

	Total Debt Holding			Housing Debt Holding			Non-housing Debt Holding		
All households	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat
Low Risk Return	0.0787	0.0164	1.60	0.0035	0.0000	0.05	0.0617	0.0048	1.13
Average Risk Return	0.0612	0.0127	1.23	- 0.0551	- 0.0004	- 0.86	0.1150**	0.0089	2.12
Slightly High Risk Return	0.0834	0.0173	0.98	0.0315	0.0002	0.29	0.1040	0.0081	1.17
High Risk Return	0.1830**	0.0380	2.26	- 0.0576	- 0.0004	- 0.54	0.2761***	0.0214	3.22
LR χ^2 (30); p value	910.99; p = [0.0000]			1,376.32; p = [0.0000]			472.78; p = [0.0000]		
Observations	25,092			15,950			20,941		
Urban sample	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat
Low Risk Return	0.0910	0.0096	1.38	- 0.0250	0.0000	- 0.29	0.1272*	0.0093	1.70
Average Risk Return	0.0651	0.0069	0.98	- 0.0616	- 0.0001	- 0.72	0.1311*	0.0096	1.73
Slightly High Risk Return	0.0752	0.0079	0.72	0.0067	0.0000	0.05	0.1504	0.0110	1.34
High Risk Return	0.0513	0.0054	0.45	- 0.2001	- 0.0002	- 1.36	0.1591	0.0117	1.29
LR χ^2 (30); p value	516.59; p = [0.0000]			924.89; p = [0.0000]			199.34; p = [0.0000]		
Observations	14,452			9,678			11,425		
Rural sample	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat	Coef.	M.E.	t-stat
Low Risk Return	0.0607	0.0138	0.78	0.0483	0.0047	0.48	- 0.0321	- 0.0046	- 0.38
Average Risk Return	0.0406	0.0092	0.52	- 0.0747	- 0.0073	- 0.73	0.0777	0.0113	0.93
Slightly High Risk Return	0.0137	0.0031	0.09	0.0112	0.0011	0.05	- 0.0225	- 0.0033	- 0.14
High Risk Return	0.3524***	0.0799	2.94	0.0943	0.0092	0.59	0.4600***	0.0666	3.65
LR χ^2 (30); p value	498.64; p = [0.0000]			458.08; p = [0.0000]			471.76; p = [0.0000]		
Observations	10,164			5,974			9,065		

[a] *, **, *** denote 10, 5, 1 per cent levels of significance, respectively.
[b] No Risk Return is a dummy variable, which equals 1 if the household head is unwilling to carry any risk; Low Risk Return is a dummy variable, which equals 1 if the household head prefers project with slight risk and return; Average Risk Return is a dummy variable, which equals 1 if the household head prefers project with average risk and return; Slightly High Risk Return is a dummy variable, which equals 1 if the household head prefers project with slightly high risk and return; High Risk Return is a dummy variable, which equals 1 if the household head prefers project with high risk and return.
[c] Reference category: Risk attitudes dummies controls: the omitted group is that household head who is unwilling to carry any risk.
[d] All other control variables are included in this analysis.

Table A4.6 The Determinants of the Log Level of Total Debt, Housing Debt and Non-housing Debt - Pooled Double Hurdle Analysis (Risk Attitudes Dummy Variables)

	Ln(Total Debt)		Ln(Housing Debt)		Ln(Non-housing Debt)	
All households	M.E.	t-stat	M.E.	t-stat	M.E.	t-stat
Outcome & Selection equation						
Low Risk Return	0.2013***	4.15	0.0682*	1.71	0.1484***	3.84
Average Risk Return	0.3355***	6.91	0.0287	0.71	0.3726***	9.82
Slightly High Risk Return	0.6134***	7.33	0.1625**	2.45	0.6506***	10.23
High Risk Return	0.6008***	7.83	0.0588	0.92	0.6349***	10.91
Selection equation						
No Financial knowledge	0.1289***	3.15	0.1035***	3.06	0.0639**	2.26
LR χ^2 (38); p value	6,559.73; p = [0.0000]		6,803.66; p = [0.0000]		2,574.00; p = [0.0000]	
Pseudo R^2	0.0893		0.1202		0.0775	
Exclusion restriction						
H0: No Financial knowledge = 0	p = [0.464]		p = [0.875]		p = [0.183]	
Observations	79,614		79,614		79,614	
Urban sample	M.E.	t-stat	M.E.	t-stat	M.E.	t-stat
Outcome & Selection equation						
Low Risk Return	0.1218**	2.06	- 0.0036	- 0.07	0.1227***	2.77
Average Risk Return	0.2695***	4.60	- 0.0204	- 0.40	0.3333***	7.81
Slightly High Risk Return	0.4375***	4.67	0.0698	0.89	0.5041***	7.67
High Risk Return	0.4411***	4.66	- 0.0739	- 0.89	0.5245***	7.93
Selection equation						
No Financial knowledge	0.1092***	2.54	0.0936***	2.12	0.0553	1.53
LR χ^2 (38); p value	3,163.47; p = [0.0000]		3,599.13; p = [0.0000]		1,348.83; p = [0.0000]	
Pseudo R^2	0.0886		0.1214		0.0729	
Exclusion restriction						
H0: No Financial knowledge = 0	p = [0.711]		p = [0.527]		p = [0.142]	
Observations	53,658		53,658		53,658	
Rural sample	M.E.	t-stat	M.E.	t-stat	M.E.	t-stat
Outcome & Selection equation						
Low Risk Return	0.3485***	4.10	0.2000***	3.21	0.2025***	2.69
Average Risk Return	0.3775***	4.31	0.0463	0.71	0.4043***	5.31
Slightly High Risk Return	1.0578***	5.73	0.2919**	2.13	1.0793***	6.76
High Risk Return	0.8805***	6.69	0.2652***	2.76	0.8591***	7.50
Selection equation						
No Financial knowledge	0.1721***	2.44	0.1108**	2.17	0.0920	1.48
LR χ^2 (38); p value	2,126.30; p = [0.0000]		4,100.94; p = [0.0000]		1,899.82; p = [0.0000]	
Pseudo R^2	0.0733		0.0867		0.0740	
Exclusion restriction						
H0: No Financial knowledge = 0	p = [0.405]		p = [0.118]		p = [0.919]	
Observations	25,956		25,956		25,956	

[a] *, **, *** denote 10, 5, 1 per cent levels of significance, respectively.
[b] No Risk Return is a dummy variable, which equals 1 if the household head is unwilling to carry any risk; Low Risk Return is a dummy variable, which equals 1 if the household head prefers project with slight risk and return; Average Risk Return is a dummy variable, which equals 1 if the household head prefers project with average risk and return; Slightly High Risk Return is a dummy variable, which equals 1 if the household head prefers project with slightly high risk and return; High Risk Return is a dummy variable, which equals 1 if the household head prefers project with high risk and return.
[c] All other control variables are included in this analysis.
[d] The null hypothesis H0: No Financial knowledge = 0 is for testing the exclusion restriction where we test the effect of No Financial Knowledge on the amount of debt for those households with positive amount of debt.

Chapter 5

Conclusion

5.1 Conclusion

This thesis comprises three related, yet independent, empirical studies which have explored important topics in the area of household finance, using household-level data from China. Specifically, Chapter 2 has investigated the role of planning for children's overseas education on household saving behaviour. Chapter 3 has estimated the effect of financial literacy on household risky asset holding. Finally, Chapter 4 has explored the relationship between attitudes towards risk and household debt.

5.1.1 Thesis Summary

5.1.1.1 Summary of Chapter 2

Chapter 2 was motivated by the relatively high saving rate in China, which has attracted much attention among academics (see, for example, Kraay, 2000; Modigliani and Cao, 2004; Meng, 2013). However, an important educational factor has been ignored by the existing literature on household savings in China. Parental investment in their children's human capital, particularly related to overseas education, is potentially a non-negligible determinant of household saving behaviour because the number of Chinese students who choose to study abroad has experienced a rapid increase in recent years and most of these students are financed by their parents. Thus, Chapter 2 used a household-level dataset from the CHFS covering 2011, 2013, 2015 and 2017 to explore the relationship between parental investment in their children's human capital, captured by planning for overseas education, and household saving behaviour, which has received very limited attention in the existing literature. This chapter was primarily based upon pooled cross-sectional analysis, which first examined the effect of planning for overseas education on household savings, household financial assets and household net wealth. Dealing with the potential endogeneity of our key explanatory variable of interest, i.e. planning for overseas education, formed the second focus of the empirical analysis. Finally, the censored quantile regression approach was used to explore the effect of parents planning to send their children to study abroad across the entire distribution of household savings and household financial assets, while uncensored quantile regression approach was used for modelling household net wealth.

Generally, the findings of this chapter supported a positive relationship between parents' planning to have their children educated overseas and household savings, and the findings also revealed that planning for overseas education is positively associated with household financial assets. Furthermore, such a positive effect of planning for overseas education on household savings has been identified after dealing with the potential endogeneity issues discussed in this chapter. In addition, different effects of planning to send children to study abroad on household savings have been found across the whole savings distribution. Specifically, planning to send children to study abroad only has statistically significant effects on household savings at and above the 60^{th} percentile of the household savings distribution.

5.1.1.2 Summary of Chapter 3

Chapter 3 was motivated by the observation that, although a small yet growing number of studies have focused on the influence of financial literacy on household risky asset holding in China, they mainly use cross-sectional data (see, for example, Liao et al., 2017; Zou and Deng, 2019), which cannot be used to control for unobserved heterogeneity. Thus, Chapter 3 used a household-level dataset from the CHFS covering 2013, 2015 and 2017 to explore the relationship between financial literacy and household risky asset holding using panel data. This chapter initially investigated waves 2013, 2015 and 2017 as three separate cross-sections for comparison with the existing literature. Secondly, this chapter investigated how financial literacy affects the probability of holding household risky assets, the amount of risky assets held and the share of risky assets held using panel analysis in order to control for unobserved heterogeneity, which has not been explored in the literature in China. Controlling for unobserved heterogeneity is important in China because household-specific time-invariant heterogeneity may bias the estimates of the effects of financial literacy. Third, this chapter explored the differential effects of financial literacy on high risk assets and low risk assets. An instrumental variable, whether the household head received any economics or finance education during school, was used to deal with the potential endogeneity of financial literacy, which formed the fourth contribution. Finally, this chapter contributed to the existing literature on China by exploring the relationship between financial illiteracy and household risky asset holding.

The results indicated that financial literacy is positively associated with household risky asset holding. Specifically, the positive impact of financial literacy is greater than that of household income in terms of size in waves 2013 and 2015, and the magnitude of the effect of financial literacy is only slightly smaller than that of household income in 2017, which indicates that financial literacy is an important determinant of the probability of holding risky assets. Further evidence of the importance of the role of financial literacy on household risky asset holding has been found once time invariant effects have been accounted for. Furthermore, financial literacy has been found to be positively associated with high risk asset holding and low risk asset holding but the size of the positive effect of financial literacy differs across high risk assets and low risk assets. Specifically, the magnitude of the effect of financial literacy on high risk asset holding is smaller than that on low risk asset holding. Finally, these findings are robust to dealing with the potential endogeneity of financial literacy and the results has also revealed a negative relationship between financial illiteracy and household risky asset holding.

5.1.1.3 Summary of Chapter 4

Chapter 4 was motivated by the significant increase in the level of household debt in China over the last two decades, which has attracted growing attention among academics (see, for example, Fan et al., 2017; Cull et al., 2019). However, the role of risk attitudes has not attracted much attention in this growing literature, although it may be an important determinant of household debt given the existence of uncertainty in Chinese household income and household income usually finances debt repayment. Thus, Chapter 4 investigated the association between risk attitudes and household debt using a household-level dataset from the CHFS (2011, 2013, 2015 and 2017). First, this chapter explored the effect of attitudes towards risk on the probability of holding total debt and the amount of total household debt, and then household debt was split into housing debt and non-housing debt to explore how risk attitudes affect the two types of debt. Second, this chapter split households into urban and rural households in order to explore whether the effect of risk attitudes on household debt differs for urban and rural households. Finally, this chapter investigated the two-part process related to holding total household debt: (1) the decision to hold debt; and (2) the decision over the amount of debt held.

The key findings of this chapter indicated that risk tolerance is positively associated with household debt. Specifically, the more risk-tolerant is the household head, the higher is the probability of holding total household debt and the higher is the amount of total household debt held by the household. The findings reported in this chapter also revealed a positive relationship between risk tolerance and non-housing debt. In addition, we have found differences in the effect of risk tolerance across total household debt, housing debt and non-housing debt by rural and urban households. For example, the magnitude of the marginal effect of risk attitudes on the probability of holding total household debt is larger for rural households than for their urban counterparts, which indicates that risk attitudes play a more important role in determining the probability of holding total household debt for rural households. Finally, the findings are robust to using the double hurdle approach thereby providing further evidence that the risk tolerance of the head of household is positively associated with household debt.

5.1.2 Policy Implications

Important policy implications related to household finance can be drawn from the findings presented in this thesis. From the first empirical chapter, it is apparent that there is a positive link between parents' planning for children's overseas education and household savings, which suggests that the Chinese government should provide more funding or subsidies to relieve the saving pressure for Chinese households so that household savings can be used for other aspects such as housing, raising young children, and insurance against retirement. Specifically, house prices in China have been increasing over time, especially in the large cities such as Beijing, Shanghai and Guangzhou: for example, the average property prices have almost tripled within the last 10 years from 2010 to 2019.[96] In addition, household savings may be used to raise young children in terms of living costs, domestic education and housing or to finance retirement. Moreover, if studying overseas is regarded as beneficial in terms of developing the skills and knowledge of individuals, providing government financial support may ultimately lead to a more productive and skilled workforce in the future thereby enhancing economic growth.

With respect to policy implications from the second empirical chapter, our findings revealed that around 90% of Chinese households, where the head did not receive any economics or finance education at school, revealed a lack of financial literacy, and are found to be less likely to hold risky assets. Therefore, such findings suggest that the Chinese government should pay more attention to economics and finance education for the new generations at school e.g. adding basic economics or finance courses at high school so that the average household financial literacy and Chinese risky asset holding

[96] Data sources: https://www.58.com/fangjiawang.

rate can be improved, thereby enhancing financial market growth in the future. Such a policy intervention may be effective because the evidence from the U.S. suggests that several wide-ranging financial education plans aimed at school-age children have made significant progress, such as the program conducted by the Jump$tart Coalition, have promoted Children's financial education effectively in the U.S. (Fox et al., 2005). In addition, in terms of targeting adults rather than children, the financial regulatory authorities and financial institutions, such as the China Securities Regulatory Commission, could consider providing at least basic financial knowledge for those households who are holding risky assets or are thinking about holding assets in the future.

Finally, the findings of the third empirical chapter are potentially important to policy makers given that risk attitudes are found to be an important determinant of household debt and it is apparent that households characterized by high levels of risk tolerance are more likely to hold debt and are more inclined to accumulate debt. If policy-makers are concerned about levels of debt, one might argue that it would be hard to influence an individual's attitudes towards risk and it might be the case that policy intervention in other areas such as improving financial literacy, as discussed in the previous chapter, might be more promising. This might help households understand the potential risks associated with taking on debt, and this might be especially important in the case of the risks associated with non-housing debt because non-housing debt might have a higher level of risk than housing debt. Specifically, if the households are not able to pay off the outstanding housing debt, the house can be used as a guarantee to lower the risk associated with housing debt. In contrast, non-housing debt does not have such a guarantee.

In addition, we have found some differences in the role of risk attitudes across urban and rural households. To be specific, there exists a positive relationship between risk attitudes and the probability of holding housing debt for rural households while such a relationship is not found for urban households, which may reflect the relatively high housing prices in urban areas. Such findings suggest that policy interventions might be better focused on the effects of high house prices in urban areas such as the increase in the minimum down payment ratio, the cap on the loan-to-value ratio, higher mortgage rates for second homes and restrictions on house-purchasing in the first-tier cities, where only those with local *hukou* (household registration) or those who have worked in this city for certain consecutive years, are eligible to purchase one or two houses.

5.1.3 Shortcomings of the Research and Areas for Future Research

Chapter 2 focused on how planning for overseas education affects household saving behaviour but, due to data limitations, we do not know whether parents actually send their children to study abroad. Thus, it would be interesting to investigate household saving behaviour among those parents who have sent their children to study abroad and those who have not, i.e. employing panel data to compare parents' plans and their actual behaviour. In addition, unfortunately, given the focus on households with children, there is only a small panel element to the data analysed in this chapter. Therefore, we have employed pooled cross-sectional analysis, which does not allow us to control for unobserved heterogeneity. Such issues remain interesting areas for future research subject to data availability.

Chapter 3 focused on the relationship between financial literacy and household risky asset holding. It is potentially worth exploring whether or not households obtain positive returns from holding such risky assets because, in general, households hold risky assets in order to obtain positive returns. In addition, positive returns may lead households to hold more risky assets, which may potentially place households in a financially vulnerable situation if they become over-confident in making such investments. Thus, such a topic remains interesting for future research - again subject to appropriate data availability.

Finally, an additional interesting topic related to Chapter 4 for future research, subject to data availability, concerns whether the effect of risk attitudes varies across different types of non-housing debt, which includes agricultural/business debt, vehicle-purchasing debt, education debt, credit card debt and other debt. Thus, whether the effect of risk attitudes differs across these various types of debt is potentially of interest given the different levels of risk associated with the different types of debt. In addition, it would be interesting to explore whether the effect of risk attitudes differs across formal and informal debt sources because household debt from formal sources is arguably less risky than that from informal sources. Specifically, formal debt can generally be obtained from banks or financial companies, with a specific repayment schedule in terms of repayment amounts and the time period of the loan formally agreed, while informal debt comes from friends or relatives, which may not be characterised by such formal repayment arrangements.

References

Almenberg, J. and Dreber, A., 2015. Gender, stock market participation and financial literacy. *Economics Letters*, *137*, pp.140-142.

Altundere, M., 2014. The relationship between sociability and household debt. *Adam Academy Journal of Social Science*, *4*(2), pp.27-58.

Alzuabi, R., Brown, S., Gray, D., Harris, M.N. and Spencer, C., 2021. Portfolio allocation and borrowing constraints. *The Sheffield Economic Research Paper Series (SERPS), 2021009.*

Arrondel, L., Debbich, M. and Savignac, F., 2015. Stockholding in France: The role of financial literacy and information. *Applied Economics Letters*, *22*(16), pp.1315-1319.

Arulampalam, W., 1999. A note on estimated coefficients in random effects probit models. *Oxford Bulletin of Economics and Statistics*, *61*(4), pp.597-602.

Atkinson, A. and Messy, F.A., 2012. Measuring financial literacy: Results of the OECD/international network on financial education (INFE) pilot study. *OECD Working Papers on Finance, Insurance and Private Pensions No.15.*

Avery, R.B. and Kennickell, A.B., 1991. Household saving in the US. *Review of Income and Wealth*, 37(4), pp.409–432.

Badarinza, C., Campbell, J.Y. and Ramadorai, T., 2016. International comparative household finance. *Annual Review of Economics, 8*, pp.111-144.

Banerjee, A., Meng, X. and Qian, N., 2010. The life cycle model and household savings: Micro evidence from urban China. *National Bureau of Demographic Dividends Revisited*, 21.

Banks around the World, 2017. World's largest stock exchanges. (unpublished). Available at: https://www.relbanks.com/stock-exchanges/largest-stock-exchanges.

Baum, C. and Schaffer, M., 2020. IVREG2H: Stata module to perform instrumental variables estimation using heteroskedasticity-based instruments. *Statistical Software Components, Department of Economics, Boston College.*

Becker, G.S. and Lewis, H.G., 1973. On the Interaction between the Quantity and Quality of Children. *Journal of political Economy, 81*(2), pp.S279-S288.

Bertaut, C.C., 1998. Stockholding behavior of US households: Evidence from the 1983–1989 survey of consumer finances. *Review of Economics and Statistics, 80*(2), pp.263-275.

Brown, S. and Taylor, K., 2008. Household debt and financial assets: Evidence from Germany, Great Britain and the USA. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *171*(3), pp.615-643.

Brown, S., Dietrich, M., Ortiz-Nuñez, A. and Taylor, K., 2011. Self-employment and attitudes towards risk: Timing and unobserved heterogeneity. *Journal of Economic Psychology, 32*(3), pp.425-433.

Brown, S., Garino, G. and Taylor, K., 2013. Household debt and attitudes toward risk. *Review of Income and Wealth*, *59*(2), pp.283-304.

Brown, S., Gray, D., Harris, M.N. and Spencer, C., 2021. Household portfolio allocation, uncertainty, and risk. *Journal of Empirical Finance*, *63*, pp.96-117.

Brown, S., Taylor, K. and Price, S.W., 2005. Debt and distress: Evaluating the psychological cost of credit. *Journal of Economic Psychology*, *26*(5), pp.642-663.

Browning, M. and Lusardi, A., 1996. Household saving: Micro theories and micro facts. *Journal of Economic Literature*, *34*(4), pp.1797-1855.

Campbell, J.Y., 2006. Household finance. *The Journal of Finance, 61*(4), pp.1553-1604.

Cao, J., Huang, B. and Lai, R.N., 2015. On the effectiveness of housing purchase restriction policy in China: A difference in difference approach. *SSRN Working Paper 2584275*.

Cardak, B.A. and Wilkins, R., 2009. The determinants of household risky asset holdings: Australian evidence on background risk and other factors. *Journal of Banking and Finance*, *33*(5), pp.850-860.

Carpenter, J.N. and Whitelaw, R.F., 2017. The development of China's stock market and stakes for the global economy. *Annual Review of Financial Economics, 9,* pp.233-257.

Chamon, M., Liu, K. and Prasad, E., 2013. Income uncertainty and household savings in China. *Journal of Development Economics*, *105*, pp.164-177.

Chamon, M.D. and Prasad, E.S., 2010. Why are saving rates of urban households in China rising? *American Economic Journal: Macroeconomics*, *2*(1), pp.93-130.

Chao, C.C., Laffargue, J.P. and Yu, E., 2011. The Chinese saving puzzle and the life-cycle hypothesis: A revaluation. *China Economic Review*, *22*(1), pp.108-120.

Chen, Y., Li, F. and Qiu, Z., 2013. Housing and saving with finance imperfection. *Annals of Economics and Finance*, *14*(1), pp.207–248.

Chi, W. and Qian, X., 2016. Human capital investment in children: An empirical study of household child education expenditure in China, 2007 and 2011. *China Economic Review*, *37*, pp.52-65.

Chu, T. and Wen, Q., 2017. Can income inequality explain China's saving puzzle? *International Review of Economics and Finance*, *52*, pp.222-235.

Cleaver, F., 2005. The inequality of social capital and the reproduction of chronic poverty. *World Development*, *33*(6), pp.893-906.

Cloyne, J., Huber, K., Ilzetzki, E. and Kleven, H., 2019. The effect of house prices on household borrowing: A new approach. *American Economic Review*, *109*(6), pp.2104-36.

Coibion, O., Gorodnichenko, Y., Kudlyak, M. and Mondragon, J., 2020. Greater inequality and household borrowing: New evidence from household data. *Journal of the European Economic Association*, *18(*6), pp.2922-2971.

Cristadoro, R. and Marconi, D., 2012. Household savings in China. *Journal of Chinese Economic and Business Studies*, *10*(3), pp.275-299.

Crook, J.N. and Hochguertel, S., 2007. US and European household debt and credit constraints. *Tinbergen Institute Discussion Paper No. 2007-087/3*.

Cui, Y., Sun, G., Siddik, M.N.A. and Liu, X., 2017. Analysis on determinants of rural household credit in China. *Journal of Interdisciplinary Mathematics*, *20*(5), pp.1179-1201.

Cull, R., Gan, L., Gao, N. and Xu, L.C., 2019. Dual credit markets and household usage to finance: Evidence from a representative Chinese household survey. *Oxford Bulletin of Economics and Statistics*, *81*(6), pp.1280-1317.

Davey, G., 2005. Chinese students' motivations for studying abroad. *International Journal of Private Education*, *2*, pp.16-21.

Dinardo, J., Fortin, N. and Lemieux, T., 1996. Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica*, *64*(5), pp.1001–1044.

Du, Q. and Wei, S.J., 2013. A theory of the competitive saving motive. *Journal of International Economics*, *91*(2), pp.275-289.

Eichner, T. and Wagener, A., 2012. Tempering effects of (dependent) background risks: A mean-variance analysis of portfolio selection. *Journal of Mathematical Economics, 48*(6), pp.422-430.

Fagereng, A., Guiso, L. and Pistaferri, L., 2016. Back to background risk? *CEPR Discussion Paper No. DP11051.*

Fan, Y., Wu, J. and Yang, Z., 2017. Informal borrowing and home purchase: Evidence from urban China. *Regional Science and Urban Economics*, *67*, pp.108-118.

Feng, J., He, L. and Sato, H., 2011. Public pension and household saving: Evidence from urban China. *Journal of Comparative Economics*, *39*(4), pp.470-485.

Firpo, S., Fortin, N. and Lemieux, T., 2009. Unconditional quantile regressions. *Econometrica*, *77*(3), pp.953-973.

Fox, J., Bartholomae, S. and Lee, J., 2005. Building the case for financial education. *Journal of Consumer Affairs*, *39*(1), pp.195-214.

Fratantoni, M.C., 2001. Homeownership, committed expenditure risk, and the stockholding puzzle. *Oxford Economic Papers, 53*(2), pp.241-259.

Gan, L., Yin, Z., Jia, N., Xu, S., Ma, S., and Zheng, L. (2014). Data you need to know about China. *Springer-Verlag, Berlin.*

Gathergood, J., 2012. Self-control, financial literacy and consumer over-indebtedness. *Journal of economic psychology, 33*(3), pp.590-602.

Ge, S., Yang, D.T. and Zhang, J., 2018. Population policies, demographic structural changes, and the Chinese household saving puzzle. *European Economic Review*, *101*, pp.181-209.

Georgarakos, D., Haliassos, M. and Pasini, G., 2014. Household debt and social interactions. *The Review of Financial Studies*, *27*(5), pp.1404-1433.

Gollier, C., 2011. Does ambiguity aversion reinforce risk aversion? Applications to portfolio choices and asset prices. *Review of Economic Studies, 78*(4), pp.1329-1344.

Gollier, C. and Pratt, J.W., 1996. Risk vulnerability and the tempering effect of background risk. *Econometrica*, *64*, pp.1109-1123.

Gruber, N., 2018. Keeping up with the Zhangs: Relative income and wealth, and household saving behavior. *Journal of Macroeconomics*, *55*, pp.77-95.

Guiso, L. and Paiella, M., 2008. Risk aversion, wealth, and background risk. *Journal of the European Economic association, 6*(6), pp.1109-1150.

Guiso, L. and Sodini, P., 2013. Household finance: An emerging field. *Handbook of the Economics of Finance, 2*, pp. 1397-1532.

Haliassos, M. and Bertaut, C.C., 1995. Why do so few hold stocks? *Economic Journal, 105*(432), pp.1110-1129.

Han, L., Xiao, J.J. and Su, Z., 2019. Financing knowledge, risk attitude and P2P borrowing in China. *International Journal of Consumer Studies*, *43*(2), pp.166-177.

Han, S. and Li, G., 2011. Household borrowing after personal bankruptcy. *Journal of Money, Credit and Banking*, *43*(2/3), pp.491-517.

He, H., Huang, F., Liu, Z. and Zhu, D., 2018. Breaking the “iron rice bowl”: Evidence of precautionary savings from the Chinese state-owned enterprises reform. *Journal of Monetary Economics*, *94*, pp.94-113.

Heaton, J. and Lucas, D., 1997. Market frictions, savings behavior, and portfolio choice. *Macroeconomic Dynamics, 1*(1), pp.76-101.

Heaton, J. and Lucas, D., 2000. Portfolio choice in the presence of background risk. *The Economic Journal*, *110*(460), pp.1-26.

Hofstede, G., Hofstede, G.J. and Minkov, M., 2005. *Cultures and organizations: Software of the mind (Vol. 2).* New York: Mcgraw-hill.

Hogarth, J.M. and Hilgert, M.A., 2002. Financial knowledge, experience and learning preferences: Preliminary results from a new survey on financial literacy. *Consumer Interest Annual*, *48*(1), pp.1-7.

Horioka, C.Y. and Wan, J., 2007. The determinants of household saving in China: A dynamic panel analysis of provincial data. *Journal of Money, Credit and Banking*, *39*(8), pp.2077-2096.

Hsiao, Y.J. and Tsai, W.C., 2018. Financial literacy and participation in the derivatives markets. *Journal of Banking and Finance*, *88*, pp.15-29.

Hu, J., Jiang, M. and Zhang, B., 2015. Social network, financial market participation and asset allocation: Evidence from China. *Xi'an Jiaotong-Liverpool University, Research Institute for Economic Integration Working Paper No. 2015-06.*

Jiang, C., Ma, Y. and An, Y., 2010. An analysis of portfolio selection with background risk. *Journal of Banking & Finance, 34*(12), pp.3055-3060.

Jiang, Y. and Ashley, D., 2013. Mao's children in the new China: Voices from the red guard generation. *Routledge, London*.

Kazarosian, M., 1997. Precautionary savings--a panel study. *Review of Economics and Statistics*, *79*(2), pp.241-247.

Koenker, R. and Bassett Jr, G., 1978. Regression quantiles. *Econometrica*, *46*(1), pp.33-50.

Kong, D. and Dickinson, D., 2016. Investigating the impact of income on savings using a Chinese household level dataset. *Emerging Markets Finance and Trade*, *52*(8), pp.1775-1796.

Kraay, A., 2000. Household saving in China. *The World Bank Economic Review*, *14*(3), pp.545-570.

Lewbel, A., 2012. Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business and Economic Statistics*, *30*(1), pp.67-80.

Lewbel, A., 2018. Identification and estimation using heteroscedasticity without instruments: The binary endogenous regressor case. *Economics Letters*, *165*, pp.10-12.

Li, L. and Wu, X., 2014. Housing price and entrepreneurship in China. *Journal of Comparative Economics*, *42*(2), pp.436-449.

Li, Z. and Feng, S., 2018. Overseas study experience and students' attitudes toward China: Evidence from the Beijing college students panel survey. *Chinese Sociological Review*, *50*(1), pp.27-52.

Liao, J. and Liu, X., 2012. Risk and consumer debt behaviors in China. *Social Behavior and Personality*, *40*(8), pp.1263-1270.

Liao, L., Xiao, J.J., Zhang, W. and Zhou, C., 2017. Financial literacy and risky asset holdings: Evidence from China. *Accounting and Finance*, *57*(5), pp.1383-1415.

Lin, C.C. and Lai, P.S., 2018. An empirical study on the impact of tenure choice on saving for Chinese households. *International Real Estate Review*, *21*(2), pp.275-294.

Liu, S., and Hu, A., 2013. Household savings in China: The Keynesian hypothesis, life-cycle hypothesis, and precautionary savings theory. *The Developing Economies*, *51*(4), 360-387.

Liu, Y., Shen, J., Xu, W. and Wang, G., 2017. From school to university to work: Migration of highly educated youths in China. *The Annals of Regional Science*, *59*(3), pp.651-676.

Liu, Z., Zhong, X., Zhang, T. and Li, W., 2020. Household debt and happiness: Evidence from the China Household Finance Survey. *Applied Economics Letters, 27*(3), pp.199-205.

Lowe, P., 2017. Household debt, housing prices and resilience. *Economic Analysis and Policy*, *55*, pp.124-131.

Lu, J. and Liu, Q., 2019. Four decades of studies on population aging in China. *China Population and Development Studies*, *3*(1), pp.24-36.

Lugauer, S., Ni, J. and Yin, Z., 2019. Chinese household saving and dependent children: Theory and evidence. *China Economic Review, 57*, p.101091

Lusardi, A. and Mitchell, O.S., 2007. Baby boomer retirement security: The roles of planning, financial literacy, and housing wealth. *Journal of Monetary Economics, 54*(1), pp.205-224.

Lusardi, A. and Mitchell, O.S., 2008. Planning and financial literacy: How do women fare? *American Economic Review, 98*(2), pp.413-17.

Mazzarol, T. and Soutar, G.N., 2002. “Push-pull” factors influencing international student destination choice. *International Journal of Educational Management*, *16*(2), pp.82-90.

Meng, X., 2003. Unemployment, consumption smoothing, and precautionary saving in urban China. *Journal of Comparative Economics*, *31*(3), pp.465–485.

Mincer, J., 1958. Investment in human capital and personal income distribution. *Journal of Political Economy*, *66*(4), 281-302.

Ministry of Education of the People’s Republic of China, 2018. The number of students studying abroad and returning to China increased double in 2017. (unpublished).

Ministry of Education of the People’s Republic of China, 2019. Statistics on the situation of Chinese students studying abroad in 2018. (unpublished).

Modigliani, F. and Brumberg, R., 1954. Utility analysis and the consumption function: An interpretation of cross-section data. *Post-keynesian Economics*, *1*, pp.338-436.

Modigliani, F. and Cao, S.L., 2004. The Chinese saving puzzle and the life-cycle hypothesis. *Journal of Economic Literature*, *42*(1), pp.145-170.

Mundlak, Y., 1978. On the pooling of time series and cross section data. *Econometrica*, *46*(1), pp.69-85.

National Bureau of Statistics of China, 2019. Gross regional product data. (unpublished). Available at: https://data.stats.gov.cn/english/=E0103

Niu, G., Wang, Q., Li, H. and Zhou, Y., 2020. Number of brothers, risk sharing, and stock market participation. *Journal of Banking and Finance*, *113*, p.105757.

Nugent, J.B., 1985. The old-age security motive for fertility. *Population and development review, 11*, pp.75-97.

OECD, 2015. Table 16.1 - Household net saving rate: Percentage of household disposable income in national accounts at a glance 2015, *OECD Publishing*, Paris.

Palia, D., Qi, Y. and Wu, Y., 2014. Heterogeneous background risks and portfolio choice: Evidence from micro-level data. *Journal of Money, Credit and Banking, 46*(8), pp.1687-1720.

Pan, X., Wu, W. and Zhang, X., 2020. Is financial advice a cure-all or the icing on the cake for financial literacy? Evidence from financial market participation in China. *International Review of Financial Analysis*, *69*, p.101473.

Pan, Y., 2016. Understanding the rural and urban household saving rise in China. *Regional Science and Urban Economics*, *56*, pp.46-59.

Powell, J.L., 1984. Least absolute deviations estimation for the censored regression model. *Journal of Econometrics*, *25*(3), pp.303-325.

Powell, J.L., 1986. Censored regression quantiles. *Journal of Econometrics*, *32*(1), pp.143-155.

Qian, J.X. and Smyth, R., 2011. Educational expenditure in urban China: Income effects, family characteristics and the demand for domestic and overseas education. *Applied Economics*, *43*(24), pp.3379-3394.

Qian, Y., 1988. Urban and rural household saving in China. *Staff Papers*, *35*(4), pp.592-627.

Roodman, D. 2011. Estimating fully observed recursive mixed-process models with cmp. *Stata Journal*, *11*(2), pp.159-206.

Rosen, H.S. and Wu, S., 2004. Portfolio choice and health status. *Journal of Financial Economics*, *72*(3), pp.457-484.

Shanghai Stock Exchange, 2016. Shanghai Stock Exchange Composite Index in 2015. (unpublished). Available at: http://finance.sina.com.cn/realstock/company/sh000001/nc.shtml

Southwestern University of Finance and Economics, Survey and Research Center for China Household Finance. Household Finance Survey Data Description in 2011 and 2013. (unpublished). (In Chinese)

Starr-McCluer, M., 1996. Health insurance and precautionary savings. *The American Economic Review*, *86*(1), pp.285-295.

Sun, H., Hartarska, V., Zhang, L. and Nadolnyak, D., 2018. The influence of social capital on farm household's borrowing behavior in rural China. *Sustainability*, *10*(12), p.4361.

Thomas, A. and Spataro, L., 2018. Financial Literacy, Human Capital and Stock Market Participation in Europe. *Journal of Family and Economic Issues*, *39*(4), pp.532-550.

Tobin, J., 1958. Estimation of relationships for limited dependent variables. *Econometrica*, *26*(1), pp.24-36.

Turk, R., 2015. Housing price and household debt interactions in Sweden. *IMF Working Paper 15/276.*

Van Rooij, M.C., Lusardi, A. and Alessie, R.J., 2011. Financial literacy and stock market participation. *Journal of Financial Economics*, *101*(2), pp.449-472.

Turvey, C.G., Kong, R. and Huo, X., 2010. Borrowing amongst friends: the economics of informal credit in rural China. *China Agricultural Economic Review, 2*(2), pp.133-147.

Van Rooij, M.C., Lusardi, A. and Alessie, R.J., 2012. Financial literacy, retirement planning and household wealth. *The Economic Journal, 122*(560), pp.449-478.

Wang, C.C., Chan, A.K. and Chen, Z.X., 2001. Segment intenders and non - intenders in China's property market: a hybrid approach. *Journal of Consumer Marketing, 18*(4), pp.319-331.

Wang, X. and Wen, Y., 2012. Housing prices and the high Chinese saving rate puzzle. *China Economic Review*, *23*(2), pp.265-283.

Wang, Y., Li, Y., Huang, Y., Yi, C. and Ren, J., 2020. Housing wealth inequality in China: An urban-rural comparison. *Cities*, *96*, p.102428.

Wei, S.J. and Zhang, X., 2011. The competitive saving motive: Evidence from rising sex ratios and savings rates in China. *Journal of Political Economy*, *119*(3), pp.511-564.

Wildauer, R., 2016. Determinants of US household debt: New evidence from the SCF. *Faculty of Arts and Social Sciences, Kingston University Working Paper 1608.*

World Bank, World Development Indicators, 2017. Gross savings(% of GDP) [World bank national accounts data, and OECD national accounts data files].

Xia, T., Wang, Z. and Li, K., 2014. Financial literacy overconfidence and stock market participation. *Social Indicators Research*, *119*(3), pp.1233-1245.

Xiang, C., Jia, X. and Huang, J., 2014. Microfinance through non-governmental organizations and its effects on formal and informal credit. *China Agricultural Economic Review, 6*(2), pp.182-197.

Xin, Y., Jiang, J., Chen, S., Gong, F. and Xiang, L., 2020. What contributes to medical debt? Evidence from patients in rural China. *BMC Health Services Research*, *20*(1), pp.1-11.

Yang, M., 2007. What attracts mainland Chinese students to Australian higher education. *Studies in Learning, Evaluation, Innovation and Development*, *4*(2), pp.1-12.

Yang, Y., Zhang, C. and Yan, Y., 2019. Does religious faith affect household financial market participation? Evidence from China. *Economic Modelling*, *83*, pp.42-50.

Yao, R. and Xu, Y., 2015. Chinese urban households' security market participation: Does investment knowledge and having a long-term plan help? *Journal of Family and Economic Issues*, *36*(3), pp.328-339.

Yilmazer, T. and Scharff, R.L., 2014. Precautionary savings against health risks: Evidence from the health and retirement study. *Research on Aging*, *36*(2), pp.180-206.

Yin, T., 2011. The “Will” to save in China. *Discussion Papers in Economics and Business*, pp.11-24.

Yoong, J., 2011. Financial illiteracy and stock market participation: Evidence from the RAND American Life Panel. *Financial literacy: Implications for retirement security and the financial marketplace*, *76*.

Yu, J. and Zhu, G., 2013. How uncertain is household income in China. *Economics Letters*, *120*(1), pp.74-78.

Zeng, X., Ma, Y. and Ma, L., 2007. Utilization of straw in biomass energy in China. *Renewable and Sustainable Energy Reviews*, *11*(5), pp.976-987.

Zhang, J., 2017. The evolution of China's one-child policy and its effects on family outcomes. *Journal of Economic Perspectives*, *31*(1), pp.141-160.

Zou, J. and Deng, X., 2019. Financial literacy, housing value and household financial market participation: Evidence from urban China. *China Economic Review*, *55*, pp.52-66.

Zweig, D., Changgui, C. and Rosen, S., 2004. Globalization and transnational human capital: Overseas and returnee scholars to China. *The China Quarterly*, *179*, pp.735-757.