

UNIVERSITY OF BIRMINGHAM

Understanding the Links between Chinese Households' Health Shocks, Financial Literacy, and Financial Decisions

by

LIN TIAN

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

Department of Economics

Birmingham Business School

College of Social Sciences

University of Birmingham

April 2020

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

Though China has experienced substantial development in its financial markets, Chinese households' participation in financial markets is still limited. The reasons for this limited participation and the consequences of financial exclusion are subject to debate. This thesis comprises three empirical researches on household finance and consumption, with financial inclusion at its core. Specifically, we try to unravel the link between health and portfolio choice, the link between financial literacy and online shopping, and the link between financial literacy and borrowing behaviour in the Chinese context.

Using the 2011, 2013 and 2015 waves of the China Health and Retirement Longitudinal Study, we explore the nexus between health shocks and cash holding of middle-aged and elderly residents in rural China. A dynamic analysis framework is exploited to address state dependence, initial conditions, and unobserved heterogeneity. We find robust evidence that, for middle-aged and elderly people living in rural China, health shocks are associated with a higher probability of holding only cash as safe assets, along with a higher proportion of cash in safe assets. This association operates through the rebalancing of financial portfolios. Individuals who live in large households, are eligible for immediate reimbursement of their healthcare expenses, and have access to bank services, are less vulnerable to health shocks.

Using the 2014 wave of the China Family Panel Studies, we estimate how the financial literacy of Chinese urban consumers relates to their online shopping decisions. We find robust evidence that respondents with higher financial literacy show a higher probability of shopping online and spend more online. These effects are stronger for younger respondents, and for respondents with higher levels of income.

Using the 2014 wave of the China Family Panel Studies, we examine how financial literacy relates to Chinese households' use of debt in both formal and informal credit markets. We observe that respondents with higher financial literacy show a lower probability of holding debt, owe a lower level of debt, and are less likely to be financially distressed. We also find that financially literate households prefer formal credit and refrain from informal credit.

To my parents,

I would never be able to thank you enough.

Acknowledgement

Writing a thesis can be an isolating process. It would not be possible without the help and support of these lovely people to whom I would like to express my gratitude.

Firstly, I would like to express my greatest gratitude to my primary supervisor Professor Alessandra Guariglia. I am eternally grateful for her guidance, advice, flexibility, time, and ongoing support throughout the course of the PhD. She was always aware of my progress, seeking to help in a proper manner, cheering me up when I made slow progress. Her professional and rigorous standards and her dedication to research motivate me to pursue excellence and to continue my career in academia. She will always be a source of inspiration for me, both as an academic, and as a human being.

I'm also grateful to my second supervisor Mr. Nicholas Horsewood for his persistent help, indispensable advice and insights, especially in econometrics. His particular point of views regarding research details helped me to see the depth of my research questions. I feel particularly fortunate for having both of my supervisors simultaneously.

I would like to thank Professor Anindya Banerjee and the Department of Economics for the scholarship that supported me to finish this PhD. I would also like to thank Professor David Dickinson, the China Research Cluster at the University of Birmingham, and the Birmingham Business School for offering me the research funding to attend conferences and workshops. The insightful comments I received from these conferences and the wonderful presentations of the participants are instructive and eye-opening.

I would also like to thank the other supportive faculties in the Department of Economics and the Administrator Team in Birmingham Business School.

Finally, my deepest gratitude goes to my parents, Wei Tian and Jing Jin. I owe everything to my parents: this PhD would not have been possible without their selfless support at all levels. Their unconditional love and generosity are limitless.

Table of Contents

Abstract I
Acknowledgement III
Table of Contents IV
List of Figures VII
List of Tables VIII
Acronyms and Abbreviations X
Chapter One: Introduction 1
1.1 Research Background 1
1.2 Thesis Outline 3
1.3 Contributions to the Literature 5
Chapter Two: Health Shocks and Cash Holdings: Evidence from Rural Chinese Households 8
2.1 Introduction 8
2.2 The Chinese Economy 12
2.2.1 Financial inclusion in rural China 12
2.2.2 Public health insurance system in China 13
2.3 Literature Review 15
2.3.1 Measures of health shock 15
2.3.2 Determinants of household portfolio choice 19
2.3.3 Health and household portfolio choice 22
2.3.4 Pathways linking health shocks and portfolio choice 26
2.4 Hypotheses Development 29
2.5 Data and Sample Construction 33
2.6 Methodology 35
2.6.1 Outline of econometric methods 35
2.6.2 Key variables 35
2.6.3 Baseline dynamic models 37
2.6.4 The role of financial wealth 42
2.6.5 The role of reimbursement type, bank accessibility and household size 42
2.7 Empirical Results 44
2.7.1 Descriptive statistics 44
2.7.2 Baseline results and discussion 46
2.7.3 The role of financial wealth 49
2.7.4 Possible pathways: Reimbursement type, bank accessibility, and household size 50

2.8 Robustness Tests 55
2.8.1 Alternative measures of the dependent variables 55
2.8.2 Alternative estimation techniques 57
2.9 Conclusion 59
Tables and figures 62
Appendix 75
Chapter Three: Financial Literacy and Online Shopping Decisions: Evidence from Urban China 85
3.1 Introduction 85
3.2 Literature Review 89
3.2.1 Determinants of online shopping 89
3.2.2 Definition and measurement of financial literacy 91
3.2.3 Financial literacy and financial decisions 97
3.3 General Hypotheses 104
3.4 Data and Sample Construction 105
3.5 Methodology and Specifications 106
3.5.1 Dependent variables: Online shopping 106
3.5.2 Financial literacy 106
3.5.3 Baseline specifications and estimation methodology 108
3.6 Empirical Results 115
3.6.1 Descriptive statistics 115
3.6.2 Main results and discussion 121
3.7 Robustness Tests 128
3.7.1 Control for cognitive abilities 128
3.7.2 Alternative estimation methods 130
3.7.3 Control for peer effects 131
3.8 Extension 133
3.8.1 Heterogenous effects: Online shopping across age groups 133
3.8.2 Heterogenous effects: Online shopping across income/wealth gradients 135
3.9 Conclusion 137
Figures and Tables 139
Appendix 152
Chapter Four: Financial Literacy and Indebtedness: Evidence from China 171
4.1 Introduction 171
4.2 Literature 175
4.2.1 Definition and measurement of financial literacy 175
4.2.2 Financial literacy and borrowing behaviours 177

4.2.3 Other determinants of indebtedness 183
4.2.4 Research gap in the literature 186
4.3 Development of Hypotheses 187
4.4 Dataset and Sample Construction 189
4.5 Methodology and Specifications 189
4.5.1 Dependent variables 189
4.5.2 Financial literacy 192
4.5.3 Baseline specifications 194
4.5.4 Estimation methodology 198
4.6 Main Results and Discussion 201
4.6.1 Summary statistics 201
4.6.2 Financial literacy and indebtedness 205
4.7 Robustness Tests 215
4.7.1 Alternative estimation methods 215
4.7.2 An alternative measure of objective financial literacy 216
4.7.3 An alternative measure of debt level 218
4.7.4 An alternative measure of indebtedness: Financial distress 220
4.8 Conclusion 221
Figures and Tables 223
Appendix 240
Chapter Five: Conclusion 251
5.1 Summary of the Findings 251
5.2 Policy Implications 252
5.3 Limitations and Future Research 253
Reference 255

List of Figures

Chapter Two

Figure 2.1 Bank branches and employers per 1,000 people in China (2008-2017).....62

Chapter Three

Figure 3.1 Online Retail Sales in China: Volume and Growth Rate (2006-2018) 139
Figure 3.2 Online shopping population in China (2013-2018).....139
Figure 3.3 Online shopping decisions by financial knowledge 140

Chapter Four

Figure 4.1 Household debt to GDP ratio: Five countries (2008-2018).....223
Figure 4.2 Household saving, debt, consumption, and housing price: China (2008-2018)...223
Figure 4.3 Dependent variables by financial knowledge test score (0-5).....224
Figure 4.4 Debt position by the level of basic financial knowledge.....227

List of Tables

Chapter Two

Table 2.1 Distribution of health shock indicators for rural households. 63
Table 2.2 Outcome variables by health shocks (N = 17,248) 63
Table 2.3 Descriptive statistics for rural households (N = 17,284) 64
Table 2.4 Baseline model with/without controlling for financial wealth (extensive margin) 65
Table 2.5 Baseline model with/without controlling for financial wealth (intensive margin) 66
Table 2.6 Interactions between health shock and reimbursement type/ bank accessibility/ high household size (extensive margin) 67
Table 2.7 Interactions between health shock and reimbursement type/ bank accessibility/ high household size (intensive margin) 68
Table 2.8 Robust test: Cash/financial assets. Baseline estimates with/without controlling for wealth effect (extensive margin) 69
Table 2.9 Robust test: Cash/financial assets. Baseline estimates with/without controlling for wealth effect (intensive margin) 70
Table 2.10 Robust test: Cash/financial assets. Estimates for interactions (extensive margin) 71
Table 2.11 Robust test: Cash/financial assets. Estimates for interactions (intensive margin). 70
Table 2.12 Robust test: Random-effect OLS, financial wealth (intensive margin) 73
Table 2.13 Robust test: Random-effect OLS, interactions (intensive margin) 74

Chapter Three

Table 3.1 Summary statistics of key variables (N = 2,183) 141
Table 3.2 Key variables by socio-demographics 142
Table 3.3 Basic financial knowledge 143
Table 3.4 Advanced financial knowledge 143
Table 3.5 Subjective assessment of financial knowledge 143
Table 3.6 Distribution of financial knowledge test score (0 - 13) 144
Table 3.7 Online shopping participation: Financial literacy 144
Table 3.8 Online shopping expenditure: Financial literacy 145
Table 3.9 Robustness test 1: Considering the cognitive ability 146
Table 3.10 Robustness test 2: Alternative estimation methods (LMP and Tobit) 147
Table 3.11 Robustness test 3: Controlling for peer effects 148
Table 3.12 Online shopping participation: by age groups 149

Table 3.13 Online shopping expenditure: by age groups 149
Table 3.14 Online shopping participation: by income groups 150
Table 3.15 Online shopping participation: by financial wealth 150
Table 3.16 Online shopping expenditure: by income groups 151
Table 3.17 Online shopping expenditure: by financial wealth 151

Chapter Four

Table 4.1 Summary statistics of key variables (N = 2,237) 228
Table 4.2 Basic financial knowledge 230
Table 4.3 Subjective assessment of financial knowledge 230
Table 4.4 Probability of indebtedness and financial literacy 231
Table 4.5 Debt level and financial literacy 232
Table 4.6 Preference to borrow and financial literacy 233
Table 4.7 Robustness test 1: Alternative estimation methods (LPM and OLS) 234
Table 4.8 Robustness test 2: Probability of holding debt and PRIDIT score for basic financial knowledge 235
Table 4.9 Robustness test 2: Debt level and PRIDIT score for basic financial knowledge 236
Table 4.10 Robustness test 2: Preference to borrow and PRIDIT score for basic financial knowledge 237
Table 4.11 Robustness test 3: Using the debt-to-income ratio as a dependent variable 238
Table 4.12 Robustness test 4: Using financial distress as depenent variable 239

Acronyms and Abbreviations

2SLS	Two-stage-least-square
AFS	Alternative financial service
AHEAD	Study of Assets and Health Dynamics Among the Oldest Old
ALP	RAND American Life Panel
BMI	Body Mass Index
CFPS	China Family Panel Studies
CHARLS	China Health and Retirement Longitudinal Study
CHFS	Chinese Household Finance Survey
CHNS	China Health and Nutrition Survey
CMP	Conditional mixed-process
CNNIC	China Internet Network Information Center
CNY	China Yuan
CRE	Correlated random effect
DHS	DNB Household Survey
DK	Do not know
ELSA	English Longitudinal Study of Ageing
FA	Factor analysis
FE	Fixed effect
FK	Financial knowledge
FRM	Fractional response models
GDP	Gross domestic product
HILDA	Household Income and Labour Dynamics in Australia
HMO	Medicare Health Maintenance organization
HRS	Health and Retirement Study
I.I.D	Independent and identically distributed
ICT	Information and communications technologies
IV	Instrumental variable
LIML	Limited-information maximum likelihood
LMI	Low- and moderate-income
LPM	Linear probability model
MMF	Money market funds
NBS	New Beneficiary Survey
NBSC	National Bureau of Statistics of China

NFCS	National Financial Capability Study
NLSY	National Longitudinal Survey of Youth
NRCMS	New Rural Cooperative Medical Scheme
OECD	Organization for Economic Cooperation and Development
OECD/INFE	Organization for Economic Cooperation and Development International Network on Financial Education
OLS	Ordinary least square
OOP	Out-of-pocket
PBC	People's Bank of China
PCA	Principal component analysis
PP	Percentage points
PRIDIT	Principal component analysis of RIDIT scores
RE	Random effect
SAH	Self-assessed health status
SHIW	Survey of Household Income and Wealth
SRCH	Self-reported change in health
SUR	Seemingly unrelated equations
UEBMI	Urban Employee Basic Medical Insurance
UK	United Kingdom
URBMI	Urban Resident Basic Medical Insurance
US	United States

Chapter One: Introduction

1.1 Research Background

Due to unprecedented financial innovation and credit expansion in the past decades, households have become increasingly active in financial markets. Understanding households' economic and financial behaviours has, therefore, become increasingly important. This research is carried out under the background of global commitment to financial inclusion, population ageing and low levels of financial literacy worldwide. We focus on a representative emerging market, China, which has been experiencing immense transitions in the past decades and made prominent progress in enhancing financial inclusion. The research questions in this thesis are inspired by the following stylised facts about China:

First, although China's financial market has seen prominent development in recent decades, universal financial inclusion still has not been achieved. For instance, in 2017, bank account ownership among people aged over 15 was only 80% in China. Though higher than in most of the developing countries, this figure is still way lower than what is observed in major developed countries, such as the US, UK and Germany, where more than 95% of people have a bank account (Demirguc-Kunt *et al.*, 2018). Within China, sizable gaps are found between urban and rural areas, and across age groups, in terms of access to financial products and services. Without access to formal saving services, less privileged people, such as rural inhabitants and the elderly, often resort to informal means of saving like cash, making them vulnerable to losses. They may also turn to informal credit markets where the risks and hidden costs are much higher.

Second, the opening up of the financial market during the past two decades has improved households' accessibility to various credit and investment opportunities. Yet, many of the financial products are too complicated for financially illiterate investors to master. Meanwhile,

low levels of financial literacy are prevalent globally, and the situation is even more severe in emerging markets, such as China. Prior studies have found that financially illiterate people are less likely to be financially included in terms of using bank accounts and digital banking services. They are also less likely to make sophisticated financial decisions and more likely to make financial mistakes. Therefore, it is crucial to understand how and under what circumstances financial literacy improves households' financial decisions.

Third, China is faced with an ageing problem. The one-child policy and fast development in healthcare has accelerated demographic ageing.[1] In 2017, the Chinese population over 65 accounted for around 10.6% of the total Chinese population, and 22.6% of the world's ageing population (65+) (World Bank, 2017). In this context, health shock have become one of the most critical factors that affect Chinese households' financial status in general and may even bring challenges to financial sustainability.

Considering these three stylised facts in China, we investigate how health shocks and financial literacy are linked to household portfolio choices, and to their borrowing and online shopping decisions.

The rest of this thesis is organised as follows. In the next two sections of this chapter, we summarise the outline of the thesis (Section 1.2) and its contributions (Section 1.3). Chapter Two explores how health shocks are related to households' cash holdings. Chapter Three examines the extent to which financial literacy enhances online shopping. Chapter Four investigates whether financial literacy is linked to households' indebtedness and the choice between formal and informal credit. Chapter Five concludes.

[1] More details about the One-Child policy and population ageing issue can be found in Section 2.1 of Chapter Two.

1.2 Thesis Outline

Studies have found that households' financial portfolios are often under-diversified and mainly made up of low-interest safe assets. Health shocks can further induce a rebalancing in households' financial portfolios towards safe assets. In Chapter Two, we explore the nexus between health shocks and cash holdings of middle-aged and elderly rural residents in China. It is worth mentioning that, for this cohort, almost all of their financial wealth is held in the forms of cash and bank deposits. Our analysis is based on panel data taken from the 2011, 2013 and 2015 waves of the China Health and Retirement Longitudinal Study (CHARLS). Our dataset comprises respondents who participated in at least two consecutive interviews, and ends up with 8,132 observations in each wave. Following the literature, we adopt three binary indicators for health shocks, namely self-reported change in health status, new diagnoses of acute conditions and new diagnoses of minor conditions. We use dynamic binary response models and fractional response models to explore the extensive and intensive margins of cash holdings, respectively. The analysis shows that, for middle-aged and elderly people living in rural China, health shocks are associated with a higher probability of holding only cash in their financial portfolios, along with a higher proportion of their financial portfolios held in cash. Moreover, individuals who are financially worse-off, live alone or with their spouse, are not eligible for immediate reimbursement of their healthcare expenses, and lack access to bank services, are more sensitive to health shocks. These results are robust to alternative measures of the dependent variable and alternative estimation methods. The findings imply that reducing out-of-pocket costs of healthcare, providing immediate reimbursement, and improving access to financial services can efficiently protect households from financial implications following the health shocks. Moreover, since large households provide an essential buffer for health shocks, how to support the one-child generation who lack a family-based safety net against parents' health shocks, is a critical issue to be resolved.

Inspired by the dramatic change in the retail landscape driven by the prosperous e-tailing market, in Chapter Three, we investigate how financial literacy is related to consumption in a digital era, from the perspective of online shopping. China, as the country with the worlds' largest e-tailing market, poses a compelling case for this study. In this chapter, we use the 2014 wave of the China Family Panel Studies (CFPS) to study the association between financial literacy and online shopping decisions. We also investigate the extent to which this association differs across age, income, and wealth groups. In our study, financial literacy is measured in three ways We estimate the extensive margin of online shopping using a set of Probit models and the intensive margins using the Tobit models. We also apply the Instrumental Variable (IV) approach in a conditional mixed-process (CMP) framework (Roodman, 2011), using parental education and community-level financial literacy as instruments for individuals' financial literacy. We find that respondents with higher financial literacy show a higher probability of shopping online and spend more online. These effects are stronger for younger respondents, and respondents with higher levels of income and wealth. The findings from this analysis suggest that policymakers should not only be aware of the benefits of financial literacy in the context of taking up online shopping, but should also understand that such benefits may be more pronounced for the young and the better-off. Furthermore, financial education may be more effective at the early stages of life, and the potential detrimental consequences of online shopping, such as indebtedness, should be considered.

The unprecedented credit expansion in the past decades substantially stimulate consumer spending and, simultaneously, household debt. High ratios of indebtedness are witnessed worldwide. Hence, in Chapter Four, we examine how financial literacy can improve households' debt status and shape households' preference over formal and informal credit. This analysis uses a cross-sectional sample consisting of 2,237 observations from the 2014 wave of the CFPS. We make use of both objective and subjective measures of financial literacy, and

look at households' use of credit, both formal or informal. We also examine households' preferences over formal and informal credit in case they have to borrow in the future. We estimate a set of Probit and Tobit models, and apply an IV approach in both two-stage-least-square (2SLS) and CMP frameworks, using parental education and community-level financial literacy as instruments for individuals' financial literacy. Overall, we find that financially literate households are less likely to be indebted, show a lower level of debt, and are less likely to report financial distress. Moreover, we observe significant evidence that financially literate households are less likely to hold informal credit, but no evidence that financial literacy is linked with the use of formal loans. By contrast, we do find that financially literate households show a strong preference for formal credit and against informal credit in case they have to borrow in the future. These results are robust to alternative measures of financial literacy and indebtedness, and to alternative estimation methods. Our analysis implies that financial literacy is effective in improving debt management and enhancing financial inclusion in terms of using formal credit for the general population.

1.3 Contributions to the Literature

This thesis makes several contributions to the literature.

Most of the existing studies focus on health and risky asset holding in the contexts of developed countries, overlooking households' portfolio choice in the emerging markets where financial markets are under-developed and most of the households' financial wealth is held in the forms of cash and deposits. Hence, the first contribution of Chapter Two is that it studies household portfolio choice from the perspective of cash holdings in the Chinese context for the first time. Second, we examine potential mechanisms lying behind the nexus between health shock and financial portfolio choice, i.e., financial wealth, reimbursement type of the health

insurance, accessibility of banks, and household size. Third, compared to most of the relevant literature that uses static models to identify the relationship between health shocks and household financial choices, this research uses a dynamic framework to consider initial conditions.

Much of the literature on the determinants of online shopping explores supply-side factors, such as product characteristics and the external environment, and demand-side factors, such as personal traits and demographics of the consumers. In Chapter Three, we fill this gap in the literature by focusing on a novel demand-side factor, i.e., financial literacy. Higher financial literacy is documented to be associated with more sophisticated financial decisions and less financial mistakes (see, e.g., Lusardi and Mitchell, 2014). However, very few studies have investigated its link with household consumption, and none has focused on online shopping. Therefore, for the first time, we explore the relationship between financial literacy and online shopping, which may be linked to each other via the adoption of bank accounts/digital payment tools and the motivation to manage consumption. Multiple measures of objective financial literacy and estimation methods have been deployed in this research.

While most of the literature on financial literacy focuses on the asset side of households' balance sheets, such as saving and investments, far fewer studies focus on the liability side, i.e., household indebtedness and borrowing behaviours. Understanding household indebtedness is of immense significance under the background of high ratio of household debt and defaults ever since the global financial crisis in 2008. In the scant literature on financial literacy and indebtedness, most papers focus on the debt level and high-cost credit usage in developed countries. Far less is known about emerging markets where financial markets are far from perfect, and a large proportion of people do not even have access to financial products. Chapter Four fills this gap in the literature by looking at financial literacy and household indebtedness, as well as households' choices between formal and informal credit, in the context of a

representative emerging market. IV approaches are implemented using proper IVs, which allows us to draw a causal inference.

Chapter Two: Health Shocks and Cash Holdings: Evidence from Rural Chinese Households [2]

2.1 Introduction

Due to the unprecedented financial innovation and credit expansion in recent years, households have been provided with a much broader array of financial services and products, which always come with a correspondingly vast array of risks. Therefore, understanding households' financial portfolio choices and identifying the most vulnerable investors are of considerable significance to the stability of the financial system and social welfare. The classic portfolio theories suggest that households should invest in risky assets to take advantage of the equity premium (Tobin, 1958). However, empirical studies have found that a large fraction of people do not invest in any risky assets and hold under-diversified portfolios which concentrate on cash and bank deposits, despite the higher return brought by risky assets (Ivković, Sialm and Weisbenner, 2008; Calvet, Campbell and Sodini, 2009). Furthermore, evidence suggests that the poor are more likely to save informally, keeping all their financial assets in the forms of cash in hand or informal groups (Karlan, Ratan and Zinman, 2014).

An extensive literature has explored the determinants of portfolio choice (e.g., Heaton and Lucas, 2000; Ameriks and Zeldes, 2004; Christelis, Jappelli and Padula, 2010; Bertocchi, Brunetti and Torricelli, 2011; Calvet and Sodini, 2014). In an era when population ageing has become a global concern, background risk stemming from poor health and health shocks has gained popularity amongst academics as a critical determinant of portfolio choice (Gertler and

[2] This chapter was accepted and presented at the Chinese Economist Society (CES) 2015 North America Conference, the Chinese Economic Association (CEA) (UK/Europe) 2016 Annual Conference, and the 2019 meetings of Society of Economics of the HOusehold (SEHO). As a result, this chapter might be found on the conference websites.

Gruber, 2002; Edwards, 2008). Prior studies document that poor health and health shocks are associated with safer portfolio choices (e.g., Rosen and Wu, 2004; Berkowitz and Qiu, 2006; Coile and Milligan, 2009). Moreover, although health insurance is widely available, insurance markets are far from perfect; individuals are unable to insure against all the medical expenditure risk following health shocks (Edwards, 2008; Atella, Brunetti and Maestas, 2012; Goldman and Maestas, 2013). Specifically, health shocks can cause an income loss and an increase in medical expenses in the short run, and impair wealth accumulation in the long run. The effect can be particularly severe for low-income households in developing countries where formal insurance and credit markets are under-developed (Islam and Maitra, 2012). Hence, understanding the financial decisions of people who have experienced health shocks, especially those who lack protection against such shocks, is of great value to individual and societal welfare. However, while most of the relevant studies focus on developed countries (e.g., Berkowitz and Qiu, 2006; Coile and Milligan, 2009; Goldman and Maestas, 2013), little attention has been paid to developing countries. Therefore, this research aims to alleviate the gaps in the literature by exploring health shocks and household portfolio choices in the context of developing countries.

Among all the developing countries, China poses a compelling case when exploring health shocks and portfolio choice. First, China is faced with a severe ageing problem. The one-child policy and fast development in healthcare have accelerated the demographic ageing and raised China's population median age from 19.4 in 1970 to 37.4 in 2013 (Worldmeters, n.d.).[3,4] In 2017, the Chinese population over 65 years old reached 0.16 billion, which account for around

[3] China's one-child policy, introduced in 1979, was part of a birth planning program designed to control the size of its population. It set a limit on the number of children parents could have. The one-child policy was eliminated at the end of 2015, and was replaced by a universal two-child policy that allowed each couple to have two children.

[4] In 2013, the median age for the United States, the United Kingdom, Japan, India and Brazil are 37.4, 40.2, 45.9, 26.5 and 30.3, respectively (World Health Organization, n.d.).

10.6% of the Chinese population, and 22.6% of the world's ageing population (65+) (World Bank, 2017a, b). Second, though the financial market and health insurance system have been developing dramatically in the past decades, China still has not achieved inclusive financial markets and a full-coverage health insurance system.[5] A large fraction of people even do not access healthcare since they cannot afford the high out-of-pocket (OOP) expense (Wang *et al.*, 2018). In such a context, a health shock has become one of the most critical factors that affect Chinese households' financial situation and even society's welfare and financial sustainability.

A few studies examine the relationship between health and risky asset holding in the Chinese context (e.g., Wu, Rong and Xu, 2011; Xie and Sun, 2012). What these studies have ignored is the fact that Chinese households, especially rural residents, hardly invest in financial products except for bank deposits (Sui and Niu, 2018).[6] Most of their financial assets are held in the forms of bank deposits, or cash in hand. Only a small proportion of Chinese households invest in risky assets. Consequently, studies on the risky assets holding of Chinese households have their natural disadvantage in data. Therefore, in this study we evaluate the extent to which health shocks relate to Chinese households' portfolio choices from the perspective of cash holding, and explore the possible mechanism that lies behind the nexus.

The contribution of this study to the literature is fourfold. First, prior studies on health shocks and portfolio choice focus on the developed countries and risky assets holding. To the best of our knowledge, this is the first research that investigates household portfolio choice from the perspective of cash holding. Second, we examine potential mechanisms lying behind the nexus between health shocks and financial portfolio choice, i.e., financial wealth, reimbursement type of the health insurance, accessibility of banks, and household size. Third,

[5] Details will be discussed in Section 2.2.

[6] In the sample of Sui and Niu (2018) drawn from the 2013 wave of the China Household Finance Survey (CHFS), around 15% of urban households own risky financial assets. And only 1% of rural households own risky financial assets.

we deploy three indicators for health shocks, including self-perceived health deterioration, the onset of acute conditions and the onset of minor health conditions, which provides a comprehensive assessment of health events from both objective and subjective perspectives. Fourth, compared with most of the relevant literature that uses static models to identify the relationship between health shocks and household financial choice, this research exploits a dynamic framework to consider initial conditions.

We use dynamic binary response models and fractional response models (FRM) to explore the extensive and intensive margins of cash holdings, respectively. We observe a statistically significant association between health shocks and rural households' cash holdings: the onset of health shocks increases an individual's probability of holding only cash as safe assets as well as the cash ratio. The association between health shocks and cash holding mostly disappears after controlling for the value of financial wealth, which implies that the health effect may operate indirectly: Health shocks can significantly reduce the total stock of financial assets, and cause a rebalance of the financial portfolios. We also find that rural respondents who are not eligible for immediate reimbursement for healthcare expenses, live in communities without banks, or live alone or with the spouse, are more vulnerable to health shocks. Our results are robust to alternative estimation methods and alternative measures of key variables.

The remainder of the Chapter proceeds as follows. In Section 2.2, we provide a brief introduction to China's health insurance system. Section 2.3 reviews the literature on health shocks and household portfolio choice. Section 2.4 develops the hypotheses. Sections 2.5 and 2.6 introduce the data and methodology. Section 2.7 analyses the empirical results. Robustness tests are presented and discussed in Section 2.8, and Section 2.9 concludes the analysis.

2.2 The Chinese Economy

2.2.1 Financial inclusion in rural China

The Chinese economy has undergone a tremendous transition and growth in the last three decades, as a result of the launch of a series of reforming policies in international trade, financial markets, the social security system, and the public healthcare system. In its endeavour to develop financial markets, the government started to commit to financial inclusion from the late 20th century and has been paying particular attention to supporting small- and middle-sized enterprises and rural residents. In 2007, the average number of bank branches in a town/village was lower than 3, and 3,302 towns did not even have a financial institution (The People's Bank of China (PBC), 2013). As of 2014, the number of towns/villages without financial institutions decreased by more than half to 1,570 (PBC, 2015). At the end of 2008, the number of bank branches per thousand people was 0.13, and the number of bank employees per thousand people was 1.96 (PBC, 2009) (see Figure 2.1).[7] By the end of 2017, the number of bank branches and bank employees per thousand people increased to 0.16 and 2.84, respectively, and up to 0.20 million bank branches were in operation in China (PBC, 2018).

Despite these improvements in financial inclusion, the density of bank branches and the number of bank branches per capita resulted in China having a middle ranking in the world in 2014 (89th and 123rd, respectively) (Demirguc-Kunt *et al.*, 2015). A sizable gap in financial inclusion exists between urban and rural China in terms of bank account ownership: As of 2014, around 84% of urban residents aged 15+ have a bank account, while the percentage for rural residents is 74% (Demirguc-Kunt *et al.*, 2015). As of 2017, around 200 million rural adults remain outside the formal financial system (Demirguc-Kunt *et al.*, 2018). Without access to

[7] A large gap in the development of the bank sector exists between rural and urban areas. According to a report published by the government, in 2006, the number of bank branches per thousand people was 0.126 in rural China and 0.2 in urban China (The Central People's Government of the People's Republic of China (Gov China), 2007).

formal saving products, these people often resort to informal means of saving, such as holding cash in hand, which is very vulnerable to theft or devaluation. In an era where digital payment has become prevalent, cash retains a dominant position at the point of transaction for the middle-aged and elderly living in rural areas. According to the 2011, 2013, and 2015 waves of CHARLS, we observe that around 95% of the rural respondents over 45 put all their financial assets in only cash or bank deposits. While all of the rural respondents over 45 hold cash, about 70% of them hold only cash in their financial portfolio, which implies they do not have or use a bank account. The average proportion of cash in their portfolio is as high as 75%.[8]

2.2.2 The public health insurance system in China

The health insurance system in China has been developing rapidly and impressively in terms of the scale and the speed of the coverage expansion. After decades of endeavour, China achieved universal coverage by providing insurance to 95% of the population in 2011 through three major health insurance programmes, namely, the Urban Employee Basic Medical Insurance (UEBMI), the Urban Resident Basic Medical Insurance (URBMI) and the New Rural Cooperative Medical Scheme (NRCMS).

The UEBMI was launched in 1998. It is designed for employees and retirees in both public and private sectors in urban areas, and is compulsory for those in employment. The scheme is financed by premiums from both the employers and employees. The URBMI was launched in 2007, aiming to provide coverage for unemployed urban residents who are not covered by the UEBMI and is voluntary in principle.[9] Extra subsidies from the government are available for

[8] These numbers are calculated based on a pooled sample of 33,162 rural respondents from the 2011, 2013 and 2015 waves of CHARLS. Based on the CHARLS questionnaire, we define financial wealth as including cash, deposits, government bonds, stocks, mutual funds, housing funds, Jizikuan, unpaid salary and other financial investments.

[9] According to the guidelines, people who are eligible for URBMI include young children, primary and secondary school students, and other unemployed urban residents.

low-income families, disabled students, and young children. Financing for the scheme comes from the participants' premiums and government subsidies. The NRCMS, launched in 2003, is designed for rural residents on a voluntary basis, although the farmers are actively encouraged to enrol. The scheme is funded by participants' premiums and subsidies from the local and central governments.

As of 2011, the effective inpatient reimbursement rates of UEBMI, URBMI and NRCMS reached 68%, 48% and 44%, separately, revealing a wide gap in reimbursement across insurance programmes (Yu, 2015). Note that, for all the three healthcare schemes, the central government sets the guidelines, and the local governments determine the premium, level of deductibles, co-payments, and reimbursement caps according to local economic levels.

Since we explore the financial behaviours of rural residents in this study, NRCMS is the insurance scheme of interest. According to Liang, Langenbrunner and World (2013), the NRCMS provided coverage for the inpatient and critical outpatient care of 95% rural residents (equivalent to 62% of the population) in 2011. The NRCMS is operated by local government (usually at the county level) under the broad guidelines set up by the central government. Therefore, the designs and benefits packages vary across the nation. For example, in 16.87% of the rural counties, the NCMS reimburses inpatient services but not outpatient services (Lei and Lin, 2009). In some of the rural areas, the NCMS beneficiaries get immediate reimbursement and pay only the healthcare expenses that exceed the deductibles. While in other rural arears, the beneficiaries have to pay the full amount and get ex-post reimbursement (Lai *et al.*, 2018). Zhong (2011) suggests that immediate reimbursement helps to build up trust and knowledge of the beneficiaries about the insurance. It also lowers the patients' expectation of coverage of healthcare expenses and encourages the price-sensitive patients to seek for treatment.

2.3 Literature Review

2.3.1 Measures of health shocks

Health shocks are often defined as unanticipated adverse health events. The difficulty in measuring health shocks lies in isolating new health information from the acknowledged past and current health status.

2.3.1.1 Subjective measures

Individuals' self-assessed health status (SAH) is widely used in longitudinal analysis. For instance, Lindelow and Wagstaff (2005) examine the effect of health shocks on income, labour market outcomes and medical expenditures using four waves of the China Health and Nutrition Survey (CHNS). The authors define health shocks as the changes in SAH, which is measured by a four-point-scale self-evaluation of health status. Four dummy variables are additionally constructed to distinguish between health changes which are large or small, positive or negative.[10] Similarly, in the research on the effect of health shocks and labour market outcomes, García-Gómez (2011) define those who report "very good" or "good" health in the previous period but report "fair", "bad" or "very bad" in the current period are considered as having experienced a health shock during the period.

However, SAH can be subject to substantial measurement error (Strauss and Thomas, 1998; Gunasekara, Carter and Blakely, 2012). First, the reporting standard for health may be systematically heterogeneous for individuals from different social-economic groups.

[10] Respondents of the CHNS are asked to rate their health status into one of the following categories: excellent, good, fair and poor. The question is asked relative to 'other people in your age'. Lindelow and Wagstaff (2005) further define a jump of two or three SRH ratings as a large health shock and a jump of one SRH ratings as a small health shock.

Consequently, individuals use the response categories given in the self-assessment questions differently (Schultz and Tansel, 1997; Dowd and Todd, 2011). This issue can be partly addressed by including various covariates regarding personal characteristics such as socio-demographics, risk attitude, time preferences, and so on. Second, individuals may misreport SAH out of various incentives. For instance, people may justify their non-participation in the labour market by over-reporting a health condition (Barnay, 2016).

Lindelow and Wagstaff (2005) argue that the problem of measurement error can be mitigated by computing the change in SAH across periods, since time-invariant factors that cause measurement errors in SAH are eliminated by taking the first difference. However, taking the difference cannot adequately eliminate measurement errors, as information on SAH may be lost due to the categorization (Benitez-Silva and Ni, 2008). For instance, individuals may still rank themselves in the poor category though they have experienced an improvement in their "true" health (Gunasekara, Carter and Blakely, 2012), and individuals who ranked themselves the highest level of SAH cannot report subsequent improved health (i.e., the ceiling effect) (Bech *et al.*, 2003; Bowling and Windsor, 2008).

An alternative measure of health shocks based on self-evaluation is the self-reported change in health (SRCH), which is measured by asking the respondents how they perceive their health status compared with some point in the past.[11] A disparity exists between $SRCH$ and the computed SAH changes. For instance, Benitez-Silva and Ni (2008) compare the validity of computed SAH changes and SRHC using a sample collected from the 1992 and 1993 waves of the US Health and Retirement Study (HRS). Their sample consists of 67,705 individuals from 7,700 households that are headed by an individual aged 51 to 61 at the interview. The authors

[11] The $SRCH$ question is commonly framed similarly to the one asked in the Short Form-36 (SF-36) questionnaire (Ware Jr and Gandek, 1998): "How do you rate your health now compared to twelve months ago?" with responses: "Much better", "A little better", "About the same", "A little worse" and "Much worse".

observe that, for 38.63% of the sample, the computed SAH changes in two consecutive waves is inconsistent with the SRCH between the two waves. Such inconsistency between the SRCH and computed SAH changes is also found in the New Zealand population (Gunasekara, Carter and Blakely, 2012).

Nevertheless, prior studies find that SRCH performs better than the computed SAH changes at measuring health dynamics, such as predicting longevity (Benitez-Silva and Ni, 2008) and hospitalization (Gunasekara, Carter and Blakely, 2012). The $SRCH$ is more plausible in terms of eliminating the ceiling effect and information loss from categorization, which are problematic for the computed SAH changes (Bech *et al.*, 2003; Benitez-Silva and Ni, 2008; Bowling and Windsor, 2008).

2.3.1.2 Objective measures

The advantage of self-perceived measures of health is that they integrate complex human judgments on a full array of illness or symptoms of disease (Bressan, Pace and Pelizzon, 2014). However, measurement errors in subjective measures of health shocks can cause substantial bias in estimates. Moreover, SRCH is not flawless at measuring "shocks", since people modify their perception of health status gradually. As a result, it is very likely that people anticipate the deterioration of health and prepare themselves against potential health shocks in advance. Therefore, to overcome the potential endogeneity issue between health and economic consequences, various objective measures are explored.

One strand of literature measures health shock using the onset of health conditions. For instance, Smith (1999, 2004, 2005) uses the new onset of acute diseases (including cancer, heart condition, stroke, and diseases of the lung) and the new onset of minor conditions (including all the other chronic diseases) to measure health shocks. Given their nature, these

acute conditions are arguably less likely to have reporting and justification bias issues compared with minor conditions (which usually refer to a set of chronic diseases) that develop gradually over time (Baker, Stabile and Deri, 2004; Jones, Rice and Zantomio, 2016). Hence, the onset of acute diseases is widely adopted to measure health shocks by later studies (e.g. Berkowitz and Qiu, 2006; Coile and Milligan, 2009; Goldman and Maestas, 2013; Jones, Rice and Zantomio, 2016). One pitfall of it that the onset of acute health conditions may not be totally unexpected. Some of the diseases evolve over a long period and can be anticipated based on available information on the current and past health status, SES and behavioral risk factors. However, the timing of the onset is very likely to be unpredictable (Jones, Rice and Zantomio, 2016), which still makes it a plausible candidate for the indicators of health shocks.

Another widely used objective measure of health shocks is unscheduled and urgent hospitalization. For instance, García-Gómez *et al.* (2013) use administrative data for 1998-2005 from Statistics Netherlands. Individuals between 18 and 64, who experienced unscheduled and urgent hospitalization but had not been admitted in the previous year are defined as having experienced health shocks. The rationale is that the exact timing of an unscheduled and urgent hospitalization is largely unpredicted, which makes this measure more likely to be exogenous to outcome variables. Lindeboom, Llena-Nozal and van der Klaauw (2016) use accidents that cause an unscheduled hospitalization as a measure of health shocks to examine the relationship between health shocks, the onset of disability and employment status. The authors consider all sorts of hospital visits, including outpatient or casualty department visit caused by traffic accidents, workplace accidents, heart attacks, and sport injuries. Wagstaff (2007) further distinguishes between the hospitalization of working members and non-working members in the study regarding the impact of health shocks on the income of Vietnamese households.

The occurrence of accidents is another plausible measure of health shocks since it is very unlikely for one to predict the occurrence of an accident. Dano (2005) measures health shocks using road accidents that cause subsequent hospitalization. The data are collected for Danish residents between 20 and 54 during the period between 1981 to 2000 from the recorded admissions to somatic hospitals, which means less severe injuries which do not require medical treatment or admission are excluded. The rationale is that, though the victims of severe accident may expect the occurrence of the event, it is very unlikely that they know the exact timing of the accidents beforehand. However, due to its nature of rare occurrence, it is hard to collect adequate observations for accidental injury for causality analysis.

Other measures of health shocks include the grip strength and the change in BMI. Huang *et al.* (2013) and Decker and Schmitz (2016) exploit the change in grip strength as an exogenous measure of health shocks. Prior studies document significant associations between low levels of grip strength with the onset of chronic diseases and physical limitations (e.g., Rantanen *et al.*, 2003; Mitra *et al.*, 2016). However, the information on grip strength is costly to collect, hence is not widely provided by the existent longitudinal studies. Wagstaff (2007) defines health shocks as a decline in body mass index (BMI) larger than one standard deviation. However, as the decline in BMI may be actually an improvement in health, it may not correctly measure health adversity.

2.3.2 Determinants of household portfolio choice

Theoretical and empirical studies on household portfolio choice primarily focus on whether households hold risky financial assets and the share of risky assets in their portfolios. Various variables such as age, gender, education, cognitive ability, financial literacy, work and marital status, social interaction, income and wealth have been typically found to play a significant

role in financial portfolio choice (e.g., Rosen and Wu, 2004; Dohmen *et al*., 2005; Christelis, Jappelli and Padula, 2010; Calvet and Sodini, 2014). Considering that cash is the least risky and the most liquid asset among all asset categories, especially in rural areas due to its dominant role in finalising payments, the vast literature on households' risky asset holding still inspires the methodology of our research.

An extensive literature investigates the determinants of asset allocation. Many empirical studies find that elderly individuals are less likely to invest in risky assets, since they are more risk-averse than younger generations (Dohmen *et al*., 2005; Cardak and Wilkins, 2009; Calvet and Sodini, 2014). However, Ameriks and Zeldes (2004) find no evidence that risky asset holding gradually decreases along with age, though there is a tendency that the elderly individuals completely drop out from the stock market. Benzoni, Collin-Dufresne and Goldstein (2007) suggest that portfolio choice actually follows a hump-shape curve along with age, where the young and the old are both less likely to invest in risky assets.

A gender gap exists in financial portfolio choice, mostly through disparity in the risk attitude and financial knowledge between genders: Men are expected to be less risk-averse and more financially literate than women and hence hold riskier financial portfolios (Rosen and Wu, 2004; Dohmen *et al*., 2005; Charness and Gneezy, 2012; Halko, Kaustia and Alanko, 2012). Married males and females are more likely to invest in stocks, since marriage is perceived as a sort of safe asset, especially for females (Bertocchi, Brunetti and Torricelli, 2011; Christiansen, Joensen and Rangvid, 2015).

Individuals with higher educational attainment are more likely to invest in stocks, hold riskier portfolios and obtain higher rate of return. One potential mechanism is that higher education is always linked with higher financial literacy (Campbell, 2006; Calvet, Campbell and Sodini, 2007). A variety of studies have documented that financially literate individuals

are more likely to use bank accounts, save in retirement funds and participate in the stock market (Lusardi and Mitchell, 2011a; van Rooij, Lusardi and Alessie, 2011; Königsheim, Lukas and Nöth, 2017; Grohmann, Klühs and Menkhoff, 2018).

A similar relationship is found between cognitive ability and financial portfolio choice. It is found that individuals who are better at numeracy, memory, and fluency at reading and understanding are more likely to invest in risky assets and hold a larger proportion of risky assets in their portfolios (Christelis, Jappelli and Padula, 2010; Atella, Brunetti and Maestas, 2012).

Social interactions raise the propensity of holding risky assets through peer effects (e.g. Tao, 2006; Brown *et al*., 2008; Abreu and Mendes, 2010). For instance, prior studies have found that social households who frequently interact with their neighbours or attend church are more likely to invest in stocks (Hong, Kubik and Stein, 2004).

Larger households tend to have riskier portfolio choices, since being part of a large household implies plenty of sources of private loans, in-kind assistance, intra-household gifts and transfers, inheritance and bequests, and so on. All of such resources can provide a shield against income or expenditure shocks that lead to a restructuring of portfolio choice towards a safer position (Lee and Kim, 2008; Zhang *et al*., 2014; Cull *et al*., 2016).

Similarly, households with more protective health insurance schemes also have a higher demand for risky assets, since the insurance buffers the effect of labour income shocks and expenditure shocks (Atella, Brunetti and Maestas, 2012; Goldman and Maestas, 2013).

The existence of background risk is also a typical determinant of household portfolio choice. Background risk is defined as the risks arising from the uncertainties in labour income, investment in real estate and private businesses, health shocks, and so on. Investors who carry background risk tend to avoid taking other forms of risk, reduce risky asset holding, and

increase their precautionary saving (e.g. Guiso, Jappelli and Terlizzese, 1996; Heaton and Lucas, 2000a, b).

As a form of background risk, house ownership influences household financial portfolio choice, but the evidence is mixed. For instance, Cocco (2004) and Heaton and Lucas (2000) find that the possession of a house will crowd-out households' risky financial assets holding, since they are unwilling to take more risk on top of house price volatility. However, Cardak and Wilkins (2009) find that the share of risky assets by homeowners is higher than the non-home-owners, and hence interpret homeownership as a substitute for risky assets which improves portfolio diversification. Chetty, Sándor and Szeidl (2017) further split the house's value into property equity wealth and mortgage debt and find the former raises stockholding while the latter reduces stockholding, ceteris paribus.

2.3.3 Health and household portfolio choice

The global ageing issue has raised attention to the role of health risk in household portfolio choices. Prior studies have documented that poor health and health shocks lead to a greater share of safe assets in financial portfolios (e.g., Cardak and Wilkins, 2009; Love and Smith, 2010). These studies predominantly use data from developed countries. Most of the relevant research focuses on health status or health risk and risky asset holding.[12]

As a seminal study in this field, Rosen and Wu (2004) estimate the association between SAH and US households' holding of various financial assets.[13] Using data from 17,822 individuals from four waves (1992-1998) of the HRS, the authors find that having a sick core

[12] Though in some of the research (e.g., Edwards, 2008; Fan and Zhao, 2009; Atella, Brunetti and Maestas, 2012), health status, health risk and health shocks are used interchangeably in interpretation, the three terms are arguably different, and we classify the literature according to the definition used in the specific research.

[13] In Rosen and Wu (2004)'s study, the financial assets are classified into safe assets, bonds, risky assets (stocks, mutual funds, and so on.) and retirement accounts.

member in the household is associated with lower probabilities of holding risky assets and lower proportions of risky assets in the financial portfolios. The main weakness of this research is that the authors use only self-assessment of current health status as the explanatory variable, which is subject to measurement error and impossible to draw a causal inference.

Love and Smith (2009) further examine whether the link between health status and portfolio choice is attributed to causality or unobserved heterogeneity. Their analysis is restricted to a panel constructed on four waves (1998-2004) of the HRS, consisting of 37,479 observations on individuals aged 65 and above in 1998. The authors employ SAH, the number of diagnosed health conditions and OOP expenses as measures of health status. The authors then compare the estimates of health status from the random-effect (RE), correlated-random-effect (CRE), and censored fixed-effect (FE) models. The authors find that, after unobserved heterogeneity is adequately controlled for in the CRE and FE models, the connections between health status and portfolio choice disappear for all the single households and most of the married households. This implies that the difference in households' portfolio choices is mostly driven by unobserved heterogeneity among households.

Fan and Zhao (2009) explore the same research question as Love and Smith (2009) using a panel on 15,188 respondents aged over 60 from two waves (1982 and 1991) of the US New Beneficiary Survey (NBS). Health status is measured by four indicators, including physical functions, chronic conditions, the history of heart attack or stroke, and work-related health limitations. The ordinary least square (OLS) estimates are compared with those derived from the random-effect and fixed-effect models. Slightly different from Love and Smith (2009), the authors find the relationship between health status and portfolio choice highly relies on the measure of health status. While the correlation between chronic conditions/work-related health limitations and the share of risky assets vanishes in the fixed-effect model, it remains significant for physical functions and heart attack or stroke history. This implies a potential

causal effect of physical functions and heart attack or stroke history on household portfolio choice.

Atella, Brunetti and Maestas (2012) investigate whether perceived health status, objective health status and future health risk influence the likelihood of risky asset holdings using data from the 2004 wave of the Survey of Health, Ageing and Retirement in Europe (SHARE).[14] Based on a sample of 11,793 individuals aged over 50 from 5,902 households, the authors find that respondents who live in non-NHS countries and report poor health status or risky health behaviours are, on average, less likely to hold risky assets. Moreover, in the non-NHS countries, highly educated households with lower future health risk are more likely to hold risky assets, while the effect of future health risk is insignificant for low-educated households. One explanation is that the highly educated households accrue higher financial literacy and health literacy, hence are more aware of future health risk and are forward-looking in financial management.

Using data on 18,402 respondents from the first wave of SHARE, Bressan, Pace and Pelizzon (2014) find that the relationship between households' portfolio choice and health highly depends on the measure of health status. More specifically, the authors find that households with poor SAH are less likely to hold risky assets, and hold a lower share of risky assets in their portfolio. By contrast, no significant effect is found when health status is measured by the physical limitations in daily living, levels of depression, and the number of chronic diseases. One potential reason is that SAH captures a wide array of health conditions, either physically or mentally, which makes it more likely to be related to financial decisions.

[14] Atella, Brunetti and Maestas (2012) define future health risk as an increasing function of the average number of risky behaviours, number of diseases diagnosed by a doctor and a decreasing function of grip strength. Risky behaviours include smoking, drinking, and a sedentary lifestyle. The diseases diagnosed by the doctor include high blood pressure, high blood cholesterol and osteoporosis, which are currently asymptomatic but might entail worsening health in the future.

A few studies focus on the effect of health shocks on portfolio choice. For instance, Berkowitz and Qiu (2006) examine the relationship between a new onset of an acute health condition and the composition of the household financial portfolio. Using six waves of the HRS from 1992 to 2002, the authors find that households with core members who experienced acute onsets are less likely to hold risky assets, and have a lower proportion of risky wealth in their portfolios. However, this relationship vanishes after controlling for financial wealth.[15]

Coile and Milligan (2009) investigate whether health shocks are related to the US retirees' holding of a wide range of assets. Two types of health shocks, i.e., the onset of a new acute condition and the diagnosis of new chronic diseases, are explored.[16] Different from most of the previous literature, the authors consider both pre-shock and post-shock periods, as well as the distance to the time when the health shock occurs. They find that the diagnosis of new chronic diseases leads to a lower probability of holding risky assets and the association is more pronounced as time passes. After the health shock occurs, households shift from non-financial assets to financial assets, and from risky assets to safe assets. Furthermore, the effect of health shocks is exaggerated when the respondent has physical or mental incapacity. However, different from Berkowitz and Qiu (2006), the authors find that acute onset does not have any effect on portfolio choice.

Though the literature on the relationship between health and portfolio choice has been growing, most of it uses data from developed countries. Evidence from developing countries in general, and from China in particular, is quite limited.

[15] The authors argue that financial wealth is a potential channel that links health shocks and household portfolio choice. This point is further discussed in Section 2.3.4.

[16] Coile and Milligan (2009) defines heart attack, stroke, or new cancer diagnosis as acute conditions, and high blood pressure, diabetes, lung disease, psychological problems, or arthritis as minor conditions. Apart from these two health shocks, the authors also investigate widowhood shock. We do not discuss widowhood shock here, since, technically, it is not a kind of health shock.

For instance, Wu, Rong and Xu (2011) use data from the 2009 Aordo Survey on Chinese households, which is conducted by a consulting company. The authors observe that households with good health status allocate a higher proportion of their financial assets to risky assets.

Xie and Sun (2012) use the data for two provinces of China collected by the pilot survey of CHARLS to explore the impact of health shocks on the value of financial assets and solely risky assets. In general, the authors observe a significant negative impact of acute health conditions on the risky asset holdings of urban households, which can be explained by the liquidation of financial assets to fund healthcare expenditure. Chronic health shocks have no effect on the asset holding of rural and urban respondents, but lead to a reduction in the consumption of durable goods, and an increase in the investment in real estate for both rural and urban households.

2.3.4 Pathways linking health shocks and portfolio choice

Some scholars attempt to identify the channels that link health and portfolio choices.

In their research on the relationship between health status and portfolio choice in the US context, Rosen and Wu (2004) further test four potential pathways that link health status and household financial portfolio choice, i.e., risk aversion, planning horizon, subjective life expectancy, bequest motives and health insurance. The authors find that even after controlling for the potential channels, the relationship between poor health and safer portfolio choices still holds, which imply none of them mediates the relationship between health and portfolio choices.

Several later studies further investigate these potential channels but obtain different results. For instance, Edwards (2008) analyses a merged panel drawn from the 1993 and 1995 waves of the Study of Assets and Health Dynamics Among the Oldest Old (AHEAD) and the 1998

wave of HRS. Using the IV method, the author observes that self-perceived health risk prompts a safer financial portfolio. After controlling for the self-reported probability of leaving any bequests, the marginal effects of health risk drop by around one-third for single households and even disappear for couples. These results imply that the presence of spouses and bequest motives hedge against the direct association between health risk and financial decisions through informal care arrangements. Using data from the 2002 wave of the Household Income and Labour Dynamics in Australia (HILDA) survey, Cardak and Wilkins (2009) find that, after considering risk and time preferences, the negative relationship between health and risky asset holding becomes insignificant. Wu, Rong and Xu (2011) also find that the nexus between health status and portfolio choice may operate via risk attitude and bequest motives in the Chinese context. However, the authors find no evidence that the health effect operates through medical insurance and planning horizon.

An extensive literature also documents that medical expense risk arising from health shocks generates a shift away from risky assets, and protective health insurance can mitigate the magnitude of the shift (e.g., Elmendorf and Kimball, 2000; Goldman and Maestas, 2013). For instance, Goldman and Maestas (2013) examine the effect of medical expenditure risk on the US households' decisions on risky asset holding, using cross-sectional data drawn from the 1998 and 2000 waves of the HRS. The authors find that individuals enrolled in more protective health insurance schemes have higher demand for risky assets. This relationship holds even after they take into account the endogeneity issue of the enrolment in health insurance schemes.

As mentioned in the last section, Atella, Brunetti and Maestas (2012) investigate whether SAH, objective health status and future health risk affect risky asset holding using data from SHARE. The ten countries explored in this study are classified as NHS countries, which have a national healthcare system, and non-NHS countries. The authors find worse SAH and higher future health risk are associated with safer portfolios, but the relationships are only significant

in non-NHS countries. This implies that households without protective healthcare insurance are more vulnerable to health shocks.

Goldman and Maestas (2013) also find that the sensitiveness of households' portfolio choices to medical expenditure risk varies across wealth gradients; the wealthiest are the least sensitive and those in the lowest wealth group are the most sensitive to medical expenditure risk. Berkowitz and Qiu (2006), instead, examine financial wealth and non-financial wealth separately, and argue that it is financial wealth that primarily mediates the health effect. Using data from the HRS, the authors first observe that a new onset of acute health conditions affects households' financial and non-financial wealth asymmetrically; the acute onset is associated with a significant drop in financial assets but does not affect non-financial assets. This is in line with economic intuitions that financial wealth is more liquid than non-financial wealth. The authors also find that the significant relationship between health shocks and risky asset disappears after controlling for financial and non-financial wealth. These results imply that a health shock causes a reduction in financial wealth with a disproportional reduction in risky and safe assets, and consequently, the restructuring of one's financial portfolio.

2.4 **Hypotheses Development**

Though it is hard to make causal inferences, health shocks have been generally found to associate with a restructuring of the composition of the household portfolio towards a safer position (e.g., Berkowitz and Qiu, 2006; Coile and Milligan, 2009). In rural China, where the financial market is underdeveloped, a vast majority of the financial assets are held in the forms of cash in hand and deposits in banks. Cash (in hand) retains the dominant role as a payment tool, especially in rural areas with a lack of financial institutions.[17] In such contexts, at the onset of health shocks, people tend to withdraw the deposits and other liquid assets, and hold more cash in hand to pay for future potential medical expenses. Hence, we test the following hypothesis.

H1. In rural China, health shocks are associated with a higher probability of holding only cash as safe assets, and a higher proportion of safe assets held in cash.[18]

Individuals who confront health shocks tend to liquidate their assets to fund the medical expenses and smooth consumption. In the context where individuals hold both safe and risky assets in their financial portfolios, the onset of a health shock presumably reduces the financial wealth, and can cause a rebalancing of the financial portfolio towards a safer position (Berkowitz and Qiu, 2006). In rural China, the vast majority of households' financial assets are held in cash and deposits. Despite the rapid development in digital payment in recent years,

[17] According to our analysis of the 2011, 2013 and 2015 waves of CHARLS, only less than 0.05% respondents who reside in rural areas hold financial assets other than cash in hand and deposits.

[18] The financial assets in the main analysis refer to cash, bank deposits, government bonds, stocks, mutual funds and other financial instrument issued by financial institutions. In rural China, households' financial wealth is barely held other than cash in hand and deposits in banks. The situation where an individual holds only cash implies that they do not use a bank account, which may be attributed to either the lack of bank services or that they voluntarily do not use bank services.

cash is still the dominant payment method in most rural areas. Therefore, the onset of a health shock will cause a decrease in total financial wealth, and, consequently, a shift from bank deposits to cash in hand to cope with further possible OOP payments. Therefore, we hypothesize that:

H2. Health shocks affect household portfolio choice through financial wealth.

For rural China, which lacks well-developed infrastructure and financial markets, health shocks can be even disastrous due to inadequate protection provided by the health insurance. The design of the health insurance scheme matters. In our sample, 90% of rural residents are the primary beneficiary of NRCMS. The majority (70%) of the NRCMS beneficiaries have to pay the full medical cost out of their own pockets before they get their reimbursement, while the other beneficiaries get immediate reimbursement and pay only the coinsurance part. Prior studies document that immediate reimbursement significantly encourages the patients to seek outpatient treatment (Zhong, 2011). This implies that health insurance which provides immediate reimbursement is more protective than those which only provide ex-post reimbursement. More specifically, the immediate reimbursement design may work better at relieving the financial burden of poor patients by reducing the current payable cost, the expectation of healthcare expenditures, as well as the time and monetary costs, which arise in the process of ex-post reimbursement. Hence, the eligibility for immediate reimbursement lowers the beneficiaries' precautionary motives to hold large amounts of cash. Therefore, we hypothesize that:

H3. In rural China, the association between health shocks and cash holdings is stronger for individuals who get ex-post reimbursement compared with those who are eligible for immediate reimbursement.

The number of bank branches reflects the development of the local financial market. In many places, digital payments via bank cards or digital payment tools are not yet widely accepted for everyday purchases at local retail stores and markets (The World Bank, 2017). For the rural areas that lack financial institutions and adequate financial education, the residents are unfamiliar with financial services and lack the trust towards financial institutions, which prohibit them from using bank accounts for saving and payment (Karlan, Ratan and Zinman, 2014; Prina, 2015). In a village without bank branches, making transactions or payments via bank payment is almost impossible. Even if the villagers hold bank accounts, they have to commute to the nearest county with bank branches to withdraw their funds, which generates higher time and monetary costs. Therefore, with the onset of health shocks, residents who live in a village without bank branches are more likely to hold a large amount of cash in hand in case of unexpected future medical expenses compared with those who live in a village with bank branches. Therefore, we hypothesize that:

H4. In rural China, the association between health shocks and cash holdings is stronger in villages with no banks.

Apart from protective health insurance and accessible bank services, informal borrowing from relatives can also buffer the effect of health shocks. Previous literature suggests that individuals with a large family have more sources of funding (such as private loans and in-kind

assistance) and share less burden in supporting parents (Zhang *et al.*, 2014; Cull *et al.*, 2016). Therefore, individuals with large families are less motivated to hold a large amount of cash in hand in case of health shocks, but prefer to retain the money in bank accounts (if they have one) for interest income. However, there is also a chance that more household members implies higher financial burden, if one who lives in the rural area holds some extra cash for the sake of other household members. These two effects go opposite directions and may even cancel out. Considering that individuals are in general self-interested, we hypothesize that the former effect may surpass the latter and, hence, that:

H5. In rural China, the association between health shocks and cash holdings is weaker for individuals with large families.

2.5 Data and Sample Construction

This research uses data from the 2011, 2013 and 2015 survey of China Health and Retirement Longitudinal Study (CHARLS). CHARLS is a biennial, nationally representative panel survey of people aged 45 years or older and their spouses from around 450 communities across 28 provinces in China.[19] The survey is designed to be comparable to the Health and Retirement Study (HRS) in the US, the Survey of Health, Ageing and Retirement in Europe (SHARE) and the English Longitudinal Study of Ageing (ELSA) in the UK.

The CHARLS database provides a wide range of information, including socio-demographics, household consumption and income. It also provides detailed information on health status and functioning, pension and insurance and household financial assets, which is of particular interest to this research. The first national wave (wave 2011) of CHARLS additionally provides detailed information on the infrastructure in the community, which allows us to relate the adult life experiences of individuals to the attributes of the places where they live.

The first national survey of CHARLS was launched in 2011, interviewed 17,708 respondents from 10,257 families across 454 counties in 28 provinces. The two follow-ups were launched in 2013 and 2015, respectively. In 2013, 81% of the 2011 individuals were revisited and 2,834 new individuals were added. In 2015, around 97% of the 2013 individuals were revisited and 574 new individuals were added.

After retrieving the data from several modules and merging based on household and individual identifiers, we obtain a preliminary three-period panel which contains 57,416 observations. Since this study focuses on the middle-aged and elderly population (aged 45 and over) who live in the rural area, we first exclude all the urban respondents and those rural

[19] The provinces covered by CHARLS are listed in Appendix 2.2.

respondents who are younger than 45 in any of the waves.[20] This operation leaves us 33,167 observations. We further exclude 492 respondents with negative net household income and cash. As will be discussed in detail in Section 2.6.2.2, the information we need to construct the indicators for health shocks are only available from wave 2013 onwards. Therefore, in this study, we focus on only the respondents who took part in at least two consecutive waves of CHARLS. So far, we obtain a three-period unbalanced panel in which 8,132 individuals participate in all the three waves of the survey, 819 individuals take part in the 2011 and 2013 waves, and 1,588 individuals take part in the 2013 and 2015 waves. Further, to fulfil the prerequisite of the dynamic estimation method proposed by Wooldridge (2005) and Hernández, Jiménez and José Martín (2011) (see details in Section 2.6.3), our sample was transformed into a balanced panel. Finally, we obtain a balanced three-period panel with 8,132 observations in each wave.

Furthermore, all the continuous control variables we used in this study (see Section 2.6.2 for details) are winsorized at the 1% and 99% percentile. All the continuous monetary variables such as assets, consumption and income are converted into real terms using the province-level Consumer Price Index (CPI) in rural areas calculated by the National Bureau of Statistics of China, using 2011 as the base year.

[20] In comparison, we also apply our methods depicted in Section 2.6 on the urban respondents. As expected, we do not find statistically significant associations between health shock and cash holdings in any of the cases. A potential explanation is that, during the time period that our research examines, the urban respondents can choose their way of payment, such as notes, bank card or other digital payment tools, flexibly. As a result, health shocks will not affect their cash (in hand) holding as much, since they do not have to hold their cash in a specific way in case of healthcare expenses following health shocks. This is also the main reason that we focus on the cash holding of rural respondents.

2.6 Methodology

2.6.1 Outline of econometric methods

Compared to prior works, our analysis is performed in a dynamic framework. We exploit three measures for health shocks, either subjective and objective. Two sets of parallel models are estimated to explore the probability that households hold only cash as safe assets and the proportion of their safe assets held in cash, using one health shock indicator at a time. Next, we check whether the relationship between health shocks and household cash holding operates through financial wealth. Then, we include the interactions of health shocks with (1) insurance reimbursement type, (2) bank accessibility and (3) household size each at a time to explore the moderator for the nexus between health shocks and cash holding.

2.6.2 Key variables

2.6.2.1 Dependent variables

Two outcome variables are created. The first outcome variable is a dummy variable equal to 1 if the respondent holds only cash as safe assets, and 0 otherwise:

$$\text{prcash} = \begin{cases} 1, & \text{if the respondent holds only cash as safe assets} \\ 0, & \text{otherwise} \end{cases}$$

Eq.2.1

Then, a continuous outcome variable is constructed to measure the proportion of cash in safe assets:

$$cash\ ratio_{i,t} = \frac{cash_{i,t}}{cash_{i,t}+bank\ deposits_{i,t}} \qquad Eq.2.2$$

2.6.2.2 Key independent variables

2.6.2.2.1 Self-reported health deterioration

Following Benitez-Silva and Ni (2008) and Gunasekara, Carter and Blakely (2012), we adopt SRCH as an indicator of health shock. The relevant question in waves 2013 and 2015 of CHARLS reads: "*Compared with your health when we talked with you in last interview, would you say that your health is better now, about the same, or worse?*" If the respondents answered "Worse", then we suppose he/she suffered from a health shock. If they answered "Better" or "About the same" then we suppose no health shock occurred in the past two years. Accordingly, we construct a dummy variable $SRCH$, which equals 1 if the respondent has experienced health shocks in the past two years, and 0 otherwise.

The advantage of this measure of health shock is twofold. First, it asked about the change in health during the past two years, which avoids the fading out of the effect of the health shock on the economic outcome. Second, respondents are taking their previous condition rather than the condition of the peers as a reference, which partly eliminates judgement bias.

2.6.2.2.2 Onset of health conditions

Following Smith (2005) and Berkowitz and Qiu (2006), we also use the diagnosed onset of health conditions as proxies for health shocks. In the 2013 and 2015 waves of CHARLS, a question is asked: "*Have you been diagnosed with [conditions listed below, read one by one] by a doctor [since R's LAST IW MONTH, YEAR/ in the last two years]*?". Fourteen health

conditions are given in the list.[21] Following the literature (e.g., Smith, 2005; Goldman and Maestas, 2013), we construct a dummy variable $Acute$, which equals 1 if the respondent has been diagnosed with any new onset of cancer, heart condition, stroke, and diseases of the lungs since the last interview, 0 otherwise. We also construct a dummy variable $Minor$, which equals 1 if the respondent has been diagnosed with any other diseases listed in the question since the last interview, and 0 otherwise.

Among all the three measures for health shock, the onset of acute condition is the most plausible one, since it is less likely to have reporting and justification bias issues compared to the minor onsets, and is better at measure "shocks" compared to SRHC.[22]

2.6.3 Baseline dynamic models

2.6.3.1 The probability of holding only cash

Focusing on the extensive margin, we start with a dynamic binary response model of the following type, allowing for state dependence and individual heterogeneity:

$$Y_{i,t} = 1\left(\gamma_1 Y_{i,t-1} + \beta_1 HS_{i,t} + \beta_2 X_{i,t} + \beta_3 Z_i + \varphi_p + \varphi_t + \varphi_i + \varepsilon_{i,t} > 0\right),\ (\mathrm{i} = 1, \ldots, \mathrm{N};\ \mathrm{t} = 1,\ 2) \qquad Eq.2.3$$

[21] The list of health conditions includes: 1. Hypertension, 2. Dyslipidemia, 3. Diabetes or high blood sugar, 4. Cancer or malignant tumor, 5. Chronic lung diseases, such as chronic bronchitis, emphysema, 6. Liver disease, 7. Heart attack, coronary heart disease, angina, congestive heart failure, or other heart problems, 8. Stroke, 9. Kidney disease, 10. Stomach or other digestive disease, 11. Emotional, nervous, or psychiatric problems, 12. Memory-related disease, 13. Arthritis or rheumatism and 14. Asthma.

[22] More discussion can be found in Section 2.3.1.2.

where $Y_{i,t}$ denotes the cash holding of household i at time t. The lagged dependent variable $Y_{i,t-1}$ is included to control for state dependence, so that to reduce the concerns that previous cash holding determines current cash holdings. $HS_{i,t}$ refers to the health shock experienced by respondent i in the past two years as of time t.

Following the literature (e.g., Smith, 2004; Cardak and Wilkins, 2009; Christelis, Jappelli and Padula, 2010; Calvet and Sodini, 2014), we make use of a wide range of standard covariates on the individual and household levels. In *Eq.2.3*, $X_{i,t}$ contains a set of time-varying control variables including age, type of Hukou, marital status, educational attainment, employment status, cognitive ability, engagement in social activities, insurance reimbursement type, household size and household income.[23] $X_{i,t}$ also includes two dummy variables indicating the ownership of a computer and a mobile phone as proxies for the ownership of bank accounts and digital payment accounts. Z_i in *Eq.2.3* is a vector of time-invariant control variables including gender, and community-level factors such as bank accessibility.[24] Definitions of all the control variables are displayed in Appendix 2.1.

φ_p denotes the province fixed effect and φ_t, the time fixed effect. $(\varphi_i + \varepsilon_{i,t})$ is the composite error term, where φ_i represents the time-invariant individual-specific component (i.e. unobserved heterogeneity) and $\varepsilon_{i,t}$ represents an idiosyncratic error term, which is

[23] We measure household size through the answer provided to the question; "In the past week, how many people usually ate meals together in your household (not including guests)?" We use this question because other measures of household size provided in the CHARLS were not consistent across waves.

(ii) In the 2011 and 2013 questionnaires, respondents are asked whether they get reimbursement immediately or have to pay the full cost of the healthcare expenses first and get reimbursed later. Unfortunately, no such question is available in 2015. Since the reimbursement type is subject to the type of health insurance scheme and local legislation, we assume that, for respondents who remained in the same village/community and held the same health insurance schemes, the reimbursement type is the same in 2015 as in 2013. Details on the variable is displayed in Appendix 2.1.

[24] In 2011, CHARLS provides information on the number of bank branches in the community. However, this information is not available in the 2013 and 2015 surveys. According to data from the central bank of China, compared with in 2011, the average number of bank branches per 1,000 people increased by 0.009 unit and 0.015 unit by 2013 and 2015, respectively (see Figure 2.1). Since the magnitude of the increase is not large, we assume that the number of banks in a community remains the same across the waves.

assumed to be independent and identically distributed (i.i.d.). Yet, the composite error terms will be correlated over time due to the individual-specific time-invariant φ_i terms. The γ_1, β_1, β_2, β_3 are parameters to be estimated. Parallel regressions are performed on each of the indicators for the health shocks. For Hypothesis 1 to hold, we expect β_1 to be positive.

This type of dynamic specification has its shortcomings. For instance, it ignores the probability that initial conditions influence the entire path of outcomes (Drakos and Konstantinou, 2013), and the fact that the start of the sample is clearly not the true initial outcome of the process of interest (Heckman, 1982). Moreover, some explanatory variables may be correlated with unobserved individual heterogeneity. As a result, the estimated coefficient and marginal effect might be biased.

To tackle the problem of the initial condition and unobserved heterogeneity, Wooldridge (2005) provides a simple solution for dynamic non-linear models, that is, by parameterizing the individual-specific component of the error term using the within-means of the time-varying control variables in all the period. Some later papers claim that Wooldridge's solution may be over-constrained by treating all the periods of the time-varying explanatory variables equally, despite that the unobserved heterogeneities may rely more on the initial-period than any other periods (e.g., Rabe-Hesketh and Skrondal, 2013). Therefore, Hernandez-Quevedo, Jones and Rice (2008), instead, parameterize the individual-specific effect as a function of the initial value of both the outcome variable and the other time-varying explanatory variables. While Rabe-Hesketh and Skrondal (2013), admitting the influence of explanatory variables from all periods, parameterize the individual-specific effect as a function of the initial values of the dependent variable and explanatory variables, as well as the means of the time-varying explanatory variables excluding the initial values.

Given our short panel, and considering that many control variables barely change across waves, we adopt Hernandez-Quevedo, Jones and Rice (2008)'s method and parameterize the distribution of the time-invariant individual effects φ_i as:

$$\varphi_i = \pi X_{i,1} + \delta Y_{i,1} + \mu_i \qquad \textit{Eq. 2.4}$$

where $X_{i,1}$ is a set of the initial values of the time-varying exogenous variables, and $Y_{i,1}$ is the initial outcome variable. μ_i is an error term which is independent of all the regressors as well as $\varepsilon_{i,t}$.

Hence, the full model for the extensive margin becomes:

$$Y_{i,t} = 1\left(\gamma_1 Y_{i,t-1} + \beta_1 HS_{i,t} + \beta_2 X_{i,t} + \beta_3 Z_i + \delta Y_{i,1} + \pi X_{i,1} + \varphi_p + \varphi_t + \mu_i + \varepsilon_{i,t} > 0\right),$$

$$(\mathrm{i} = 1, \ldots, \mathrm{N}; \mathrm{t} = 1,\ 2) \qquad \textit{Eq.2.5}$$

Eq.2.5 will be estimated using a random-effect Probit model, which allows us to take advantage of the panel nature of our sample and consider unobserved heterogeneity that appears in μ_i.[25] The model will be estimated on each of the health shock indicator one at a time. For Hypothesis 1 to hold, we expect β_1 to be positive.

[25] Since our sample is randomly drawn from a large population, it is appropriate to use a random-effects model (Baltagi, 2008).

In comparison, we will also estimate a set of basic Probit models which does not include the one-period lag of the dependent variable and the initial conditions of the time-varying control variables, all the other variables the same as in *Eq.2.3*.

2.6.3.2 The proportion of cash in hand

In the same spirit, the full dynamic fractional response model (FRM) for the intensive margin takes the following form:

$$E(Y_{i,t}|X,Z,HS) = \Phi\big(\gamma_1 Y_{i,t-1} + \beta_1 HS_{i,t} + \beta_2 X_{i,t} + \beta_3 Z_i + \delta Y_{i,1} + \pi X_{i,1} + \varphi_p + \varphi_t + \mu_i + \varepsilon_{i,t}\big), \quad (\mathrm{i} = 1, \dots, \mathrm{N}; \mathrm{t} = 1,\ 2,\ 3) \qquad \textit{Eq.2.6}$$

where $Y_{i,t} \in [0,1]$ denotes the proportion of safe assets held in cash and is defined by Eq.2.2.

Definitions of the other variables are the same as in Eq.2.3. The model in *Eq.2.6.* is estimated for each of the health shock indicators one at a time. For Hypothesis 1 to hold, we expect β_1 to be positive.

In comparison, we will also estimate a set of basic FRMs which include only the health shocks, the time-varying and time-invariant control variables, as well as the provincial and time fixed effects as explanatory variables.

Following the literature (see Section 2.3.2), we expect that those who are younger, male, married, employed, with a higher level of cognitive ability and educational attainment, eligible for insurance with immediate reimbursement, more engaged in social activities, hold non-agricultural Hukou, live in a larger household, earn higher household income, use computers

and mobile phones and live in communities with banks are less likely to hold only cash, and hold lower levels of cash in their portfolios.

2.6.4 The role of financial wealth

Following Berkowitz and Qiu (2006), we then examine whether the association between health shocks and cash holding operates through financial wealth. To this end, we incorporate the logarithm of financial wealth in the baseline models. Based on the CHARLS questionnaires, we define financial wealth as the sum of cash, deposits, government bonds, stocks, mutual funds, housing funds, Jizikuan, unpaid salary and other financial investments.[26] If the relationship between health shocks and cash holding operates through the rebalancing of financial wealth, then we will observe that the marginal effects of health shocks decrease in magnitude or even disappear after we include financial wealth as an explanatory variable, and Hypothesis 2 will be supported.

2.6.5 The role of reimbursement type, bank accessibility and household size

To verify Hypotheses 3, 4, 5, we augment the baseline model with two interaction terms, $HS_{i,t} * Dummy_{i,t}$ and $HS_{i,t} * (1 - Dummy_{i,t})$, where $Dummy_{i,t}$ denotes one of the three dummies which identify the eligibility for insurance with immediate reimbursement ($Reimbursedum_{i,t}$), bank accessibility ($Bankdum_{i,t}$) and household size ($Hyhhsize_{i,t}$), respectively. Meanwhile, we have to leave out the term $HS_{i,t}$ from the model to get rid of duplication. Definitions for these three dummy variables are displayed in Appendix 2.1.

[26] Jizikuan is a fund that individuals provide to the work unit for the purpose of investment, building apartments, and so on.

Suppose the marginal effect of $HS_{i,t} * Dummy_{i,t}$ is θ_1 and the marginal effect of $HS_{i,t} * (1 - Dummy_{i,t})$ is θ_2. For Hypothesis 3, 4 and 5 to hold, we expect that $\theta_1 < \theta_2$. In other words, the immediate reimbursement of the healthcare expenditures, living in a community with a bank and/or in a large household may help mitigate the impact of health shocks on financial choices.

2.7 Empirical Results

2.7.1 Descriptive statistics

2.7.1.1 Dependent variables

The descriptive statistics for the outcome variables are reported in Panel A of Table 2.1. In the pooled data, we observe that 70% of rural respondents hold only cash in their financial portfolios, and the average share of cash in the safe assets is 74.8%. While the probability of rural respondents holding only cash as safe assets decreases only slightly (by less than one percentage point) from 2013 to 2015, the proportion of safe assets held in cash decreased from 76.4% in 2013 to 73.3% in 2015, which can be attributed to the development of the financial markets during this period.[27]

2.7.1.2 Key independent variables

Panel B of Table 2.1 reports the descriptive statistics for the three health shock indicators. About 46.6% of the respondents in the pooled sample reported a worsening in their health during the past two years. About 7.5% were diagnosed with new acute conditions, and 16.5% were diagnosed with new minor conditions. The much higher occurrence of SRHC may be attributed to its nature in that it captures a broader range of elements besides the physical health conditions, such as lack of vitality which arises from anxiety or depression, (Au and Johnston, 2014).

Table 2.2 reports a breakdown of the dependent variables by each measure of health shock. Panel A reports the statistics of *Prcash* and Panel B reports the statistics of *Cashratio*. For all

[27] These statistics are not reported in the table for brevity.

three health shock indicators, the mean values of *Prcash* and *Cashratio* are higher for respondents who suffered from a health shock, compared to those who did not (Column 1 and 7). The differences in means are statistically significant for self-reported health shocks and the onset of acute conditions, but is insignificant for minor conditions. This is as expected, since the minor conditions are mostly chronic diseases which develop gradually. People who suffers from these conditions may gradually adjust their portfolios to normal positions.

2.7.1.3 Control variables

The descriptive statistics of all the control variables for the pooled data are reported in Table 2.3. The sample has an average age of 61, with half of the respondents being male. About 84% of the sample is married, and 10% are widowed. About 29% of the respondents are illiterate, and 71.2% achieved a medium level of education (high school) or above. Fifty-seven percent of the sample engages in agricultural work, 14.1% of the sample is an employee in a company, and 21.6% does not work. Five percent of the sample run their own business. About 54.1% of respondents are socially active. The average score of cognitive ability is around 10.56 out of a maximum of 21. Around 88.7% of the respondents report ownership of at least one house. Though computers are not prevalent in the rural areas, with computer ownership at only reaching 17.5%, 88.37% of the sample owns mobile phones. The maximum household size reaches up to 9 household members, but the average household size is only 3.15. About 51% of the sample live in a household with more than two members. The average household income is 30,046 in 2011 CNY and the average net financial assets, excluding cash, is 12,720 in 2011 CNY. Among the rural respondents who benefit from a health insurance scheme, 32.69% are eligible for immediate reimbursement. Finally, 8.86% of respondents live in a bank-accessible community.

2.7.2 Baseline results and discussion

2.7.2.1 The extensive margins

Table 2.4 reports the RE Probit estimates for the marginal effect of health shocks on whether the individuals hold only cash in their financial portfolios from the baseline models.[28] Columns (1) to (3) report the estimates for the model which include health shock indicators, the time-varying and time-invariant variables, as well as the province and year dummies as explanatory variables. Columns (4) to (6) report the estimates for the models that further include the one-period lag and the initial condition of the dependent variable, as well as the initial conditions of the time-varying control variables.[29] The same structure applies to Table 2.5.

In Table 2.4, we observe that the marginal effects of $SRCH$ and $Acute$ health conditions are always significantly positive at the 1% level even after considering the initial conditions of the dependent variables and the time-varying explanatory variables (Columns 1, 2, 4 and 5). More specifically, according to the estimates of the full model (see Column 4), individuals who perceive a health deterioration in the past two years are 2.74 percentage points (pp) more likely to hold only cash as safe assets. Those who have been diagnosed with a new acute condition in the past two years are 5.60 pp more likely to hold only cash as safe assets (Column 5). However, the marginal effects of $Minor$ health conditions are never significant. This is not surprising as the minor health conditions in our list generally develop gradually over time. As a result, they may not be totally unanticipated, and their occurrence may not draw a shock as the acute conditions do to households' asset holdings. Therefore, Hypothesis 1 is partly verified.

[28] The estimated marginal effects for the other control variables are not reported for brevity.

[29] The specifications in Columns (7)- (9) further include the logarithm of financial wealth as an explanatory variable to verify Hypothesis 2. This will be discussed in Section 2.7.3.

Additionally, the estimated marginal effect of the one-period lag of the dependent variable (L_prcash) is always positively significant at the 1% level in all the three specifications (Row 1, Column 4 to 6). This implies that state dependence in cash holding exists.

The full estimation results are reported in Table A2.3.1 in Appendix 2.3. The layout of the specifications in Table A2.3.1 resembles Table 2.4. It turns out that the marginal effects of the other control variables are consistent with economic intuition and previous research (e.g., Dohmen et al., 2005; Atella, Brunetti and Maestas, 2012). More specifically, the estimates from the full models (Columns 4 to 6) imply that those who are male, have higher cognitive ability, possess mobile phones, live in small families and in villages where there are banks, are less likely to hold only cash as safe assets. We also find that rural respondents who carry out agricultural work (compared to those who don't work) are less likely to hold only cash as safe assets, presumably because they hold bank accounts and so on for trading agricultural products. Some other control variables which are significant in the static model (Columns 1 to 3), such as age, Hukou dummy, household income, social interactions, house ownership, marital status, education, being employed, are not significant anymore after we introduce the initial condition and one-year lag of cash holding in the models.

2.7.2.2 The intensive margins

Table 2.5 reports the estimates of the marginal effects from a set of dynamic FRMs. The table is organized similarly to Table 2.4. Again, the marginal effects of $SRCH$ and $Acute$ health conditions are always significantly positive at the 1% level and 5% level, respectively, even after considering the initial conditions (Columns 1, 2, 4 and 5). This implies that individuals with $SRCH$ and $Acute$ onset hold larger shares of cash in their safe assets. More specifically, according to the estimates for the full models, the occurrence of $SRCH$ and $Acute$

onset is associated with 2.80 pp (Column 4) and 3.99 pp (Column 6) higher cash ratio, respectively. Therefore, Hypothesis 1 is verified. Again, state dependence is found in the significantly positive marginal effect of $L_cashratio$ for all the full models (Row 1, Columns 4 to 6).

The full estimation results are reported in Table A2.3.2 in Appendix 2.3. The layout of the specifications in Table A2.3.2 resembles Table 2.4. Similar to the extensive case, the estimates from the full models (Columns 4 to 6) imply that those who are male, have higher cognitive ability, possess mobile phones, carry out agricultural work, live in small families, and in communities where there are banks, on average hold higher proportion of their financial assets in the form of cash in hand. Other control variables which are significant in the static model, such as age, Hukou dummy, household income, social interactions, house ownership, marital status, education, being employed, become insignificant after we consider the initial condition and state dependence of cash holding in our models.

Our findings for both the extensive and intensive margins are also in line with prior studies on health shocks and portfolio choice, which focus on risky asset holdings (e.g., Smith, 2005; Coile and Milligan, 2009). These studies suggest that households that experienced health shocks tend to hold safer portfolios. A potential explanation is that the middle-aged and elderly residents in rural China primarily rely on cash to pay for healthcare expenses. For the rural residents who are financially illiterate and unfamiliar with financial services in general, lack of trust towards financial institution is frequently seen (Karlan, Ratan and Zinman, 2014; Prina, 2015). For such people, cash may be regarded as the safest of the safe assets, which stores value and, at the same time, enables transactions and payment. Therefore, they hold a large amount of cash in hand after health shocks, presumably to prepare for potential medical expenses in the future.

2.7.3 The role of financial wealth

Next, we examine whether the relationships between health shocks and cash holding operate through financial wealth. Following Berkowitz and Qiu (2006), we first estimate a specification using the logged financial wealth as the dependent variable. Initial health status, initial financial and non-financial wealth, health shocks and other standard controls are employed as explanatory variables (results displayed in Table A2.4). We observe a significant negative marginal effect of health shocks in the financial wealth equation (Columns 1 and 2). This implies that $SRCH$ and $Acute$ onset are associated with a reduction in financial wealth.

We then go back to the full dynamic baseline models which take into account both state dependence and initial conditions (Eq.2.5 and Eq. 2.6), and further incorporate the logged households' total financial wealth and its one-period lag.

The RE Probit estimates for the marginal effects of the health shocks are reported in Columns (7) to (9) in Table 2.4. We observe that the marginal effects of financial wealth ($Lghfinasset$) are significantly negative at the 1% significance level in all three specifications. This is consistent with previous literature on risky asset holding that suggests the wealthier hold riskier portfolios compared with the less-wealthy (e.g. Guiso and Paiella, 2008). Most importantly, we find that, after controlling for financial wealth, the effects of health shocks either disappear (for $SRCH$, Column 7) or drop by around 40% in magnitude (for $Acute$, Column 8). The $Minor$ onset is still not significant in any of the cases (Column 9). Overall, Hypothesis 2 is partly supported.

The FRM estimates for the marginal effects of health shocks on the proportion of cash are reported in Columns (7) to (9) in Table 2.5. We find that the indicators for health shocks either become insignificant (for $SRCH$, Column 7) or drop by about 40% in magnitude (for $Acute$, though are still significant at the 5% level, Column 8).

These results are in line with Berkowitz and Qiu (2006). The results imply that the occurrence of a health shock causes a decrease in total financial wealth, since the respondents need to pay for medicine or treatment. The health shock also leads to a rebalancing of the financial portfolio which ends up with a higher cash ratio, probably because the rural respondents prefer to hold cash in case of future medical payment. However, after controlling for financial wealth, the effect of health disappears or drops substantially in magnitude. Therefore, Hypothesis 2 is verified.

One point worth noting is that, even after controlling for financial wealth, the effect of the $Acute$ onset, though halves in magnitude, is still significant in both the participation and the proportion equation. One potential reason is that, as discussed in Section 2.3.1, the self-perceived measure of health shocks, by its nature, is not totally unpredicted. Individuals feel the gradual changes in their health status and adjust their financial behaviours accordingly and gradually. In comparison, the $Acute$ onset is less predictable in terms of the timing of the onset. Therefore, the shocks delivered by the $Acute$ onset may be more striking than $SRCH$. Hence, the financial portfolio choices may be more sensitive to the $Acute$ onset.

2.7.4 Possible pathways: Health insurance reimbursement type, bank accessibility, and household size

2.7.4.1 Reimbursement type

To explore the role of reimbursement type in the nexus between health shocks and cash holdings, we augment the full baseline models (Eq.2.5 and Eq.2.6) with interaction terms between health shock and the reimbursement type of the healthcare scheme that the respondent participates in.

Since logged financial wealth (*Lghfinasset*) is always significant at the 1% level, hence an important predictor of households' cash holding, we control for households' financial wealth in all the specifications hereafter.

Columns (1) to (3) in Table 2.6 report the RE Probit estimates of the marginal effects of $SRCH$, $Acute$ onset and $Minor$ onset for the respondents who get ex-post reimbursement (Row 1), and for those who are eligible for immediate reimbursement (Row 3), respectively. After controlling for financial wealth in all the specifications, only the marginal effect of $Acute$ onset is significant and positive for the rural respondents who get ex-post reimbursement (Column 2, Row 1).[30] However, the marginal effect of $Acute$ onset is insignificant for the respondents who receive immediate reimbursement (see Column 2, Row 3). More specifically, rural respondents who experienced $Acute$ onset and get ex-post reimbursement for healthcare expenditure are 4.08 pp more likely to hold only cash as safe assets.

Columns (1) to (3) in Table 2.7 report the FRM estimates of the marginal effects of $SRCH$, $Acute$ onset and $Minor$ onset for the respondents who get ex-post reimbursement (Row 1), and for those who are eligible for immediate reimbursement (Row 3), respectively. Again, we observe that only the $Acute$ onset is significantly positive among the three indicators for health shocks, and only for the rural respondents who get ex-post reimbursement (see Column 2, Row 1). More specifically, the cash ratio is 3.42 pp higher for individuals with $Acute$ onset and get ex-post reimbursement.

The estimates for the extensive and intensive margins are consistent with previous literature (e.g., Atella, Brunetti and Maestas, 2012; Goldman and Maestas, 2013), which suggest that individuals with more protective health insurance schemes are less vulnerable to health shocks.

[30] In a set of results not reported, we exclude financial wealth from the specifications and find that $SRCH$ significantly positive for rural respondents who get ex-post reimbursement, but insignificant for those who get immediate reimbursement.

The results also suggest a way to improve the effectiveness of health insurance schemes, that is, by providing immediate reimbursement rather than ex-post reimbursement to the enrollees. So far, Hypothesis 3 is verified.

2.7.4.2 Bank accessibility

We then incorporate interaction terms between the health shock and bank accessibility in the full baseline model (Eq.2.5 and Eq.2.6). Columns (4) to (6) in Table 2.6 report the RE Probit estimates of the marginal effects of health shocks for the respondents who live in a community without banks (Row 5), and for those who live in a community with at least one bank branch (Row 7), respectively. We observe that only the marginal effect of $Acute$ onset for the respondents who live in a village without bank branches is significant and positive. More specifically, the rural respondents who experienced $Acute$ onset and live in a community without a bank are 3.59 pp (at the 5% level) more likely to hold only cash in their financial portfolios (Column 5, Row 5). For those who live in a community with at least one bank branch, the marginal effect of $Acute$ onset is positive but insignificant.

Columns (4) to (6) in Table 2.7 report the FRM estimates of the marginal effects of health shocks. Again, among the three indicators for health shocks, only the marginal effect of $Acute$ onset is significant and positive for those living in a village without bank branches. More specifically, the $Acute$ onset is associated with a cash ratio which is 2.80 pp (3.74%) higher than average for the rural respondents who reside in a village without a bank (Column 5, Row 5). Therefore, Hypothesis 4 is verified.

The results above imply that the rural respondents who live in a village with bank branches are less likely to be affected by health shocks in terms of cash holdings. A potential reason is that the presence of banks in a village reflects the financial development in this area. In a village

with a bank, digital payment via bank account is possible and withdrawing funds from the bank is convenient. Hence, there is no need for individuals to hold a large amount of cash in hand, which is neither safe nor profit-generating.

2.7.4.3 Large households

To examine how the nexus between health shocks and cash holding varies for large and small households, we generate a new dummy variable, $Hyhhsize$, to identify large households. This dummy variable equals 1 if the household has more than two members, and 0 otherwise. We then augment the baseline models (Eq.2.5 and Eq.2.6) with the interactions between $hyhhsize$ and health shock indicators.

Columns (7) to (9) of Table 2.6 reports the RE Probit estimates for the marginal effect of health shocks for the respondents who live alone or with spouse (Row 9), and for those who live with a larger family (Row 11), respectively. We observe that the marginal effect for $Acute$ onset is positively significant at the 5% level (Column 8, Table 2.6). More specifically, individuals who suffered from a new onset of acute conditions and live alone or with the spouse are 4.88 pp more likely to hold only cash as safe assets.

Columns (7) to (9) of Table 2.7 report the FRM estimates for the marginal effect of health shocks for the respondents who live alone or with the spouse (Row 9), and for those who live with a larger family (Row 11), respectively. We observe that the marginal effects for all three health shock indicators are significantly positive at the 10% significance level for the respondents who live alone or with the spouse (Columns 7 to 9, Row 9). More specifically, the cash ratio for individuals characterized by $SRCH$, $Acute$ onset and $Minor$ onset and who live in a small household is 1.57 pp, 3.20 pp, and 2.06 pp higher than average, respectively. We also find that the marginal effects of $Minor$ onset are significantly negative at the 10%

significance level for those who live in large families (Column 9, Row 11). Therefore, Hypothesis 5 is verified.

These findings suggest that, overall, a large family can work as a cushion for the financial distress brought by health shocks, which is in line with Zhang *et al.* (2014) and Cull *et al.* (2016). In fact, this association can be decomposed into two effects in opposite directions. On the one hand, in case of the onset of health shocks, having a large household implies a broader source for informal loans and a way of sharing the financial burden. Therefore, those rural residents who do not have large households have to hold a relatively large amount of cash in hand in case of future healthcare expenses following the health shocks. On the other hand, it is also possible that large household, to some extent, brings more financial burden. In such a situation, one may hold extra cash in hand to be prepared to support the other members in case they confront shocks. But our analysis reveals that, even though large household may cause more financial burden for an individual, this effect is offset or even exceeded by the positive role that a large household performs as a way of burden sharing in case of health shock.

Overall, our results reveal that the three mechanisms that we examine in this section can mitigate the associations between health shock and cash holdings. The intuition behind is that these three mechanisms basically represent the resources of funding. When individuals confront health shocks, they can either withdraw money from the bank, get reimbursement through the health insurance scheme or borrow from the other household members, to pull through. As a result, the access to these resources can mitigate the effect of health shocks on one's financial asset positions.

2.8 Robustness Tests

2.8.1 Alternative measures of the dependent variables

First, we use alternative measures of the dependent variables. Following a variety of studies on health and risk asset holding (e.g., Rosen and Wu, 2004; Fan and Zhao, 2009), we look at cash holding out of total financial assets. For the extensive margin, the new measure becomes

$$\text{prcash}_{\text{fin}} = \begin{cases} 1, & \text{if the respondent holds only cash as financial assets} \\ 0, & \text{otherwise} \end{cases} \qquad \textit{Eq.2.7}$$

where the financial assets are as defined in Section 2.6.4.

Similarly, the dependent variable for the intensive margin of cash holding becomes

$$cash\ ratio_{i,t} = \frac{cash_{i,t}}{financial\ assets_{i,t}} \qquad \textit{Eq.2.8}$$

Table 2.8 reports the RE Probit estimates for the marginal effect of health shocks. The specifications in Columns (1) to (3) control for time-varying and time-invariant control variables, the one-period lag of dependent variable, the initial values of the dependent and time-varying explanatory variables, initial health status and the year and province fixed effects. The specifications in Columns (4) to (6) further control for financial wealth. Similar to the baseline results, we find that the marginal effects of $SRCH$ and $Acute$ onset are significantly positive, though the latter is only significant at the 10% level (Column 1 and 2). After controlling for

financial wealth, the marginal effect of $Acute$ onset is still significant (Column 5), while the marginal effect of $SRCH$ becomes insignificant (Column 4).

The FRM estimates for the intensive margin are reported in Table 2.9, with a similar structure to Table 2.8. We observe that individuals with $SRCH$ hold higher proportions of cash in their financial portfolio (significant at the 1% level, Column 1), while $Acute$ and $Minor$ onsets are not significant at all (Column 2 and 3). After controlling for financial wealth, $SRCH$ becomes insignificant (Column 4).

These findings imply that a health shock is associated with a higher probability of holding cash as the only financial assets, and with a higher ratio of cash among total financial assets. Moreover, these relationships operate through total financial wealth. Hence, Hypotheses 1 and 2 are supported.

The results for the specifications with the interactions for the extensive margin are reported in Table 2.10. The findings are similar to the baseline results. We find that, for the respondents who get ex-post reimbursement, the marginal effect of $Acute$ onset is significantly positive at the 5% level (Column 2), which implies that an immediate reimbursement design of the health insurance can mitigate the effect of health shocks on household financial portfolios. For the respondents who live in a village without bank branches, the marginal effect of $Acute$ onset is significantly positive, though at the 10% level (Column 5), implying that the finances of people who live in a village without bank are more vulnerable to health shocks. For the respondents who live alone or with the spouse, the marginal effect of $Acute$ onset is significantly positive at the 5% level (Column 8). We also find that the marginal effect of $Minor$ onset is significantly negative at the 10% level for those who live with large families (Column 9). These findings imply that respondents who live alone or with the spouse are more sensitive to health shocks compared with those who live with a large family. By contrast, we find no significant

marginal effect of any health shock indicators for the respondents who get immediate reimbursement, live in villages with bank branches or live in large families.

Table 2.11 reports the marginal effect for the intensive margin. For brevity, we observe significant evidence that, for respondents who get ex-post reimbursement, only the marginal effects of $Acute$ onset are significant at the 5% level (Column 5). For respondents who live alone or with the spouse, only the marginal effects of $Acute$ onset are significantly positive at the 5% level (Column 8). We also observe a significantly negative marginal effect of $Minor$ for the respondents living with large families. However, none of the health shock indicators are significant for the respondents who get immediate reimbursement or live with large families. Moreover, we find no evidence that the associations between health shocks and cash holdings differ by the accessibility of bank services.[31]

In a sum, our findings in this section are similar to the baseline results, except for the role of bank accessibility in the intensive margins. These findings support Hypotheses 3 and 5, and partly support Hypothesis 4.

2.8.2 Alternative estimation techniques

The dynamic FRM cannot fully account for the panel nature of our data though we have controlled time and province fixed effects and initial values. Therefore, we estimate a set of random-effect models for the intensive margin of cash holdings. The dependent variable is defined by Eq.2.2. The estimates for Eq.2.6 are presented in Table 2.12. Compared with the FRM estimates, the random-effect estimates decrease in magnitude, yet are still marginally significant. More specifically, we observe that the occurrence of $SRHC$ or $Acute$ is associated

[31] However, if we exclude financial wealth from the set of explanatory variables, both SRHC and Acute are significantly positive for the households who live in a village with at least one bank branch. These results are not reported for brevity.

with a 2.54 pp (Column 1) or 2.94 pp (Column 2) higher proportion of cash in the safe assets, respectively. Again, after controlling for financial assets, the marginal effect of *SRHC* becomes insignificant, and the marginal effect of $Acute$ onset drops in magnitude (Columns 4 and 5). Therefore, Hypotheses 1 and 2 are once again verified

The estimates for the interactions in the equation of the cash ratio are reported in Table 2.13. We observe that for the respondents who get ex-post reimbursement, only the marginal effect of $Acute$ is significant and positive (Column 2). This finding is similar to the baseline results, implying that the rural respondents who get ex-post reimbursement are more sensitive to the $Acute$ onsets. However, we do not find any significant marginal effects of the health shock indicators for those who get immediate reimbursement. Moreover, we find significantly positive marginal effects of *SRHC* for those who live alone or with the spouse (Column 7). We also observe a significantly negative effect of $Minor$ onset for the respondents who live in large families (Column 9). These findings imply that rural respondents who live alone or with the spouse, on average, hold higher levels of cash after health shocks compared with the respondents who live with large families, even after considering the prior cash levels. Nevertheless, none of the health shock indicators show significant marginal effects in the case of bank accessibility. Therefore, Hypotheses 3 and 5 are verified.

2.9 Conclusion

A growing literature has uncovered that poor health status and health shocks increase an individual's risk aversion (Decker and Schmitz, 2016) and are associated with a lower share of risky assets in households' portfolios (Rosen and Wu, 2004; Coile and Milligan, 2009; Atella, Brunetti and Maestas, 2012). However, most of the studies in this field are based on data from developed countries, and all of them focus on risky assets holdings.

This study examines the relationship between health shocks and cash holdings for middle-aged and elderly people living in rural China, based on data from the 2011, 2013 and 2015 waves of CHARLS. We also examine several smoothing mechanisms for the effects of health shocks, i.e. the presence of a bank in the village, the reimbursement type of medical insurance and household size. We observe a higher probability of holding only cash as safe assets and a higher proportion of safe assets held in cash for the individuals who report deterioration in health status or are diagnosed with acute onset during the past two years. These relationships are stronger for those who were diagnosed with new onset of acute conditions, as the marginal effects associated with $Acute$ onset are always larger in magnitude and significance compared with $SRCH$. Furthermore, we find that these relationships are mediated by a rebalancing in total financial wealth, and are mitigated by the immediate reimbursement design of the health insurance, the availability of bank services, and a large family. These findings are robust to using alternative measures of the dependent variable and alternative estimation methods.

It is also worth noting that the situation where the respondent holds only cash as safe assets implies that they do not use a bank account. This can be attributed to either that they do not have access to bank accounts or that they voluntarily do not use it. Since, in all the specifications, we have controlled for the exogenous bank service accessibility, the ownership of computers and mobile phones as proxies for the ownership of bank account, we are trying

to shed light on the relationship between health shocks and voluntary exclusion from bank services of the middle-aged and elderly rural habitants.

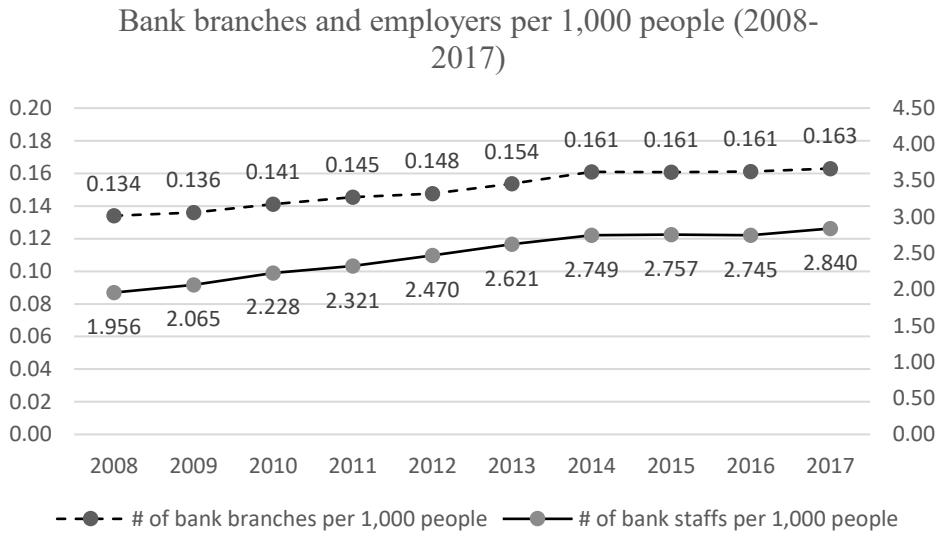
The findings of our analysis have strong policy implications. A large amount of cash in hand rather than being invested in bank deposits and other more profit-generating financial instruments will impede the households' wealth accumulation in the long run. It may also harm their ageing wellbeing. As revealed by this research, reducing OOP costs of healthcare and providing immediate reimbursement can efficiently protect rural households from health shocks. Improving financial inclusion, such as increasing the number of bank branches can also help. Additionally, our results show that large households provide an essential buffer against health shocks. In the light of this finding, how to support the one-child generation who lack a family-based safety net against parental health shocks, is an issue both urgent and necessary to solve and calls for a supportive social security system and policies.

Finally, the results of our analysis should be interpreted with caution. First, though we have applied econometric methods to control for the initial condition issue and state dependence, we cannot draw causal inference. A possible endogeneity issue between health shocks and cash holding needs to be addressed. Second, due to the limitation of the data, we cannot identify the exact timing of the health shock, and the way that the patient paid for the medicine and treatment. If the interview took place right after the health shock, there might either be a drop in cash ratio if the patient paid by cash, or an increase in cash ratio if the patient paid by card. This leads to difficulty in interpretation.

Future research can search for sophisticated instruments for health shocks to identify potential causality between health shocks and cash holdings. Moreover, we could extend our research by using more recent waves of the CHARLS database (when they become available) to investigate the dynamics of the relationship in a longer time span. Future research could also

look at the transitions in other areas of households' economic behaviours following a health shock, such as education, and property investments and borrowing decisions, to obtain a full picture of household decisions in the presence of health shocks.

Tables and figures



Source: The People's Bank of China

Figure 2.1 Bank branches and employers per 1,000 people in China (2008-2017)

Table 2.1 Distribution of health shock indicators for rural households.

Variable	Mean	Std. Dev.	Min	Max	Obs
Panel A					
Prcash	0.6982	0.4591	0	1	17,284
Cashratio	0.7483	0.4042	0	1	17,284
Panel B					
SRCH	0.4661	0.4989	0	1	17,284
Acute	0.0753	0.2639	0	1	17,284
Minor	0.1647	0.3709	0	1	17,284

Notes: Detailed definitions are displayed in Appendix 2.1.

Table 2.2 Outcome variables by health shocks (N = 17,248)

Health shocks	Indicator = 0					Indicator = 1					Pr(|t|>|t|)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	
Panel A: Prcash	Mean	Std. Dev.	Min	Max	Obs	Mean	Std. Dev.	Min	Max	Obs	
SRCH	0.668	0.471	0	1	4,604	0.732	0.443	0	1	4,038	0
Acute	0.692	0.462	0	1	7,993	0.774	0.419	0	1	649	0
Minor	0.696	0.460	0	1	7,220	0.712	0.453	0	1	1,442	0.738
Panel B: Cashratio	Mean	Std. Dev.	Min	Max	Obs	Mean	Std. Dev.	Min	Max	Obs	
SRCH	0.721	0.418	0	1	4,604	0.78	0.387	0	1	4,038	0
Acute	0.743	0.407	0	1	7,993	0.809	0.369	0	1	649	0
Minor	0.746	0.405	0	1	7,220	0.759	0.400	0	1	1,442	0.938

Notes: This table reports the statistics of Prcash and Cashratio by health shock indicators, i.e., *SRCH*, *Acute* and *Minor*. Columns (1)-(5) report the statistics for the respondents who have not experienced a specific type of health shock, and Columns (6)-(10) report the statistics for the respondents who experienced the specific type of health shock. Detailed definitions are displayed in Appendix 2.1.

Table 2.3 Descriptive statistics for rural households (N = 17,284)

Variable	Mean	Std. Dev.	Min	Max
Age	60.9776	8.5523	45	101
Gender	0.5059	0.5000	0	1
Hukou (non-agricultural)	0.0496	0.2172	0	1
Marital status				
Married	0.8399	0.3667	0	1
Married (living separate)	0.0462	0.2100	0	1
Widowed	0.1032	0.3042	0	1
Not married	0.0106	0.1025	0	1
Education				
Low	0.2876	0.4526	0	1
Medium	0.6341	0.4817	0	1
High	0.0783	0.2687	0	1
Work status				
Agriculture	0.5677	0.4954	0	1
Employed	0.1414	0.3485	0	1
Self-employed	0.0513	0.2207	0	1
Unpaid housework	0.0234	0.1512	0	1
Not working	0.2162	0.4116	0	1
Social	0.5407	0.4984	0	1
Cognition	10.5596	4.1541	0	21
House ownership	0.8870	0.3167	0	1
Computer	0.1748	0.3799	0	1
Mobilephone	0.8837	0.3206	0	1
Lghhitot	8.4353	2.6494	0	12.2758
Hhsize	3.1461	1.6454	1	9
Hyhhsize	0.5116	0.4999	0	1
Lghfinasset	3.9690	4.7141		12.5671
Hfinasset	12,720.74	32,852.32	1	286,945
Reimburse_imm	0.3269	0.4691	0	1
Bankdum	0.0886	0.2842	0	1

Notes: Detailed definitions are displayed in Appendix 2.1.

Table 2.4 Baseline model with/without controlling for financial wealth (extensive margin)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
L_prcash				0.0698**	0.0691**	0.0657*	0.0047	0.0040	0.0036
				(0.0336)	(0.0337)	(0.0340)	(0.0120)	(0.0122)	(0.0121)
Ini_Prcash				0.1359***	0.1370***	0.1430***	0.0140	0.0138	0.0141
				(0.0269)	(0.0270)	(0.0272)	(0.0146)	(0.0148)	(0.0147)
Lghfinasset							-0.0478***	-0.0478***	-0.0480***
							(0.0007)	(0.0007)	(0.0007)
SRCH	0.0296***			0.0274**			0.0034		
	(0.0086)			(0.0115)			(0.0074)		
Acute		0.0482***			0.0560***			0.0329**	
		(0.0164)			(0.0216)			(0.0139)	
Minor			0.0107			0.0060			-0.0028
			(0.0112)			(0.0147)			(0.0098)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Initial X	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Observations	13,399	13,385	13,385	11,895	11,768	11,768	11,895	11,768	11,768

Notes: RE Probit estimates for marginal effects are reported. The dependent variable is the probability of holding only cash as safe assets. A full set of contemporary time-varying and time-invariant variables are controlled but relevant estimates are not reported for brevity. See Appendix 2.1 for complete definitions of all variables and Table A2.3.1 for the full table including the estimates for all the variables. Robust standard errors are clustered at household level and reported in the parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 2.5 Baseline model with/without controlling for financial wealth (intensive margin)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
L_cashratio				0.2030***	0.2050***	0.2020***	0.0359***	0.0352***	0.0346***
				(0.0167)	(0.0169)	(0.0168)	(0.0100)	(0.0100)	(0.0100)
Ini_Cashratio				0.0423**	0.0416**	0.0474***	0.0085	0.0073	0.0088
				(0.0179)	(0.0180)	(0.0178)	(0.0138)	(0.0139)	(0.0139)
Lghfinasset							-0.0509***	-0.0510***	-0.0510***
							(0.0009)	(0.0009)	(0.0009)
SRCH	0.0320***			0.0280***			0.0067		
	(0.00771)			(0.00970)			(0.0064)		
Acute		0.0369**			0.0399**			0.0230*	
		(0.0146)			(0.0189)			(0.0128)	
Minor			0.0051			0.0010			-0.0037
			(0.0100)			(0.0125)			(0.0088)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Initial X	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Observations	13,399	13,385	13,385	11,895	11,768	11,768	11,895	11,768	11,768

Notes: Estimates for marginal effects are reported for the fractional response models which have controlled for the year and provincial fixed effects. The dependent variable is the proportion of safe assets held in cash. A full set of contemporary time-varying and time-invariant variables are controlled but relevant estimates are not reported for brevity. See Appendix 2.1 for complete definitions of all variables and Table A2.3.2 for the full table including the estimates for all the variables. Robust standard errors are clustered at household and reported in the parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 2.6 Interactions between health shocks and reimbursement type/ bank accessibility/ high household size (extensive margin)

Dep. Var: Prcash	Random Effects Probit								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Health shocks	SRCH	Acute	Minor	SRCH	Acute	Minor	SRCH	Acute	Minor
Health_Shock * (1- Reimbursedum)	0.0061	0.0408**	0.0037						
	(0.0090)	(0.0180)	(0.0118)						
Health_Shock * Reimbursedum	-0.0063	0.0242	-0.0109						
	(0.0132)	(0.0243)	(0.0170)						
Health_Shock * (1- Bankdum)				0.0040	0.0359**	-0.0004			
				(0.0079)	(0.0153)	(0.0105)			
Health_Shock * Bankdum				-0.0179	0.0304	-0.0086			
				(0.0226)	(0.0484)	(0.0283)			
Health_Shock * (1-Hyhhsize)							0.0122	0.0488**	0.0192
							(0.0104)	(0.0221)	(0.0143)
Health_Shock * Hyhhsize							-0.0089	0.0234	-0.0195
							(0.0102)	(0.0194)	(0.0135)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lagged Dep.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Initial Dep.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Initial X	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lg(financial asset)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11,895	11,768	11,768	11,895	11,768	11,768	11,895	11,768	11,768

Notes: RE Probit estimates for marginal effects are reported. The dependent variable is the probability of holding only cash as safe assets. A full set of control variables is included. See Appendix 2.1 for complete definitions of all variables. Robust standard errors are clustered at household level and reported in the parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 2.7 Interactions between health shocks and reimbursement type/ bank accessibility/ high household size (intensive margin)

Dep. Var: Cash/safe asset	Fractional Response Model								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Health shocks	SRCH	Acute	Minor	SRCH	Acute	Minor	SRCH	Acute	Minor
Health_Shock * (1- Reimbursedum)	0.0083	0.0342**	-0.0017						
	(0.0078)	(0.0167)	(0.0104)						
Health_Shock * Reimbursedum	0.0010	-0.0022	-0.0058						
	(0.0115)	(0.0211)	(0.0141)						
Health_Shock * (1- Bankdum)				0.0083	0.0280**	-0.0024			
				(0.0068)	(0.0140)	(0.0094)			
Health_Shock * Bankdum				-0.0151	0.0366	-0.0093			
				(0.0186)	(0.0415)	(0.0231)			
Health_Shock * (1-Hyhhsize)							0.0157*	0.0320*	0.0206*
							(0.0088)	(0.0193)	(0.0118)
Health_Shock * Hyhhsize							-0.0038	0.0130	-0.0233*
							(0.0088)	(0.0173)	(0.0126)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lagged Dep.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Initial Dep.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Initial X	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lg(financial asset)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11,895	11,768	11,768	11,895	11,768	11,768	11,895	11,768	11,768

Notes: Estimates for marginal effects are reported for the fractional response models. The dependent variable is the probability of holding only cash as safe assets. A full set of control variable is included. See Appendix 2.1 for complete definitions of all variables. Robust standard errors are clustered at household level and reported in the parentheses. $***$ $p < 0.01$, $**$ $p < 0.05$, $*$ $p < 0.1$

Table 2.8 Robust test: Cash/financial assets. Baseline estimates with/without controlling for wealth (extensive margin)

Dep. Var: Prcash_fin	Random Effects Probit					
	(1)	(2)	(3)	(4)	(5)	(6)
L_prcash_fin	0.0619*	0.0621*	0.0603*	-0.0101	-0.0095	-0.0102
	(0.0324)	(0.0328)	(0.0329)	(0.0097)	(0.0098)	(0.0098)
Prcash_fin_ini	0.1500***	0.1500***	0.1550***	0.0177	0.0166	0.0178
	(0.0297)	(0.0300)	(0.0302)	(0.0130)	(0.0130)	(0.0132)
Lghfinasset				-0.0427***	-0.0428***	-0.0430***
				(0.0011)	(0.0011)	(0.0011)
SRCH	0.0318**			-0.0007		
	(0.0127)			(0.0065)		
Acute		0.0417*			0.0255**	
		(0.0235)			(0.0118)	
Minor			0.0008			-0.0093
			(0.0163)			(0.0087)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Initial X	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11,895	11,768	11,768	11,895	11,768	11,768

Notes: RE Probit estimates for marginal effects are reported. The dependent variable is the probability of holding only cash as safe assets. A full set of control variables is included. See Appendix 2.1 for complete definitions of all variables. Robust standard errors are clustered at the household level and reported in the parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1

Table 2.9 Robust test: Cash/financial assets. Baseline estimates with/without controlling for wealth (intensive margin)

Dep. Var: Cash/Financial assets	Random Effects Probit					
	(1)	(2)	(3)	(4)	(5)	(6)
L_cashratio_fin	0.1990***	0.2010***	0.1980***	0.0190**	0.0194**	0.0187**
	(0.0163)	(0.0163)	(0.0163)	(0.0088)	(0.0088)	(0.0089)
Cashratio_fin _ini	0.0437**	0.0427**	0.0480***	0.0025	0.0025	0.0035
	(0.0172)	(0.0173)	(0.0172)	(0.0131)	(0.0131)	(0.0132)
Lghfinasset				0.0021**	0.0021**	0.0022**
				(0.0010)	(0.0010)	(0.0010)
SRCH	0.0278***			0.0035		
	(0.0099)			(0.0059)		
Acute		0.0259			0.0195	
		(0.0187)			(0.0120)	
Minor			-0.0014			-0.0062
			(0.0127)			(0.0082)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Initial X	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11,895	11,768	11,768	11,895	11,768	11,768

Notes: Estimates for marginal effects are reported for the fractional response models. The dependent variable is the proportion of financial assets held in cash. A full set of control variables is included. See Appendix 2.1 for complete definitions of all variables. Robust standard errors are clustered at the household level and reported in the parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 2.10 Robust test: Cash/financial assets. Estimates for interactions (extensive margin)

Dep. Var: Prcash_Fin	Random Effects Probit								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Health shocks	SRCH	Acute	Minor	SRCH	Acute	Minor	SRCH	Acute	Minor
Health_Shock * (1- Reimbursedum)	0.00410	0.0370**	-0.0109						
	(0.00822)	(0.0162)	(0.0106)						
Health_Shock * Reimbursedum	-0.0104	0.00733	-0.00733						
	(0.0122)	(0.0205)	(0.0159)						
Health_Shock * (1- Bankdum)				0.000997	0.0256*	-0.0104			
				(0.00717)	(0.0134)	(0.00941)			
Health_Shock * Bankdum				-0.0190	0.0485	-0.00312			
				(0.0220)	(0.0497)	(0.0274)			
Health_Shock * (1-Hyhhsize)							0.0119	0.0460**	0.00465
							(0.00972)	(0.0200)	(0.0131)
Health_Shock * Hyhhsize							-0.0136	0.0116	-0.0227*
							(0.00928)	(0.0170)	(0.0120)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lagged Dep.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Initial Dep.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Initial X	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lg(financial asset)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11,895	11,768	11,768	11,895	11,768	11,768	11,895	11,768	11,768

Notes: RE Probit estimates for marginal effects are reported. The dependent variable is the probability of holding only cash as safe assets. A full set of control variables is included. See Appendix 2.1 for complete definitions of all variables. Robust standard errors are clustered at the household level and reported in the parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1

Table 2.11 Robust test: Cash/financial assets. Estimates for interactions (intensive margin)

Dep. Var: Cash/financial asset	Fractional Response Model								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Health shocks	SRCH	Acute	Minor	SRCH	Acute	Minor	SRCH	Acute	Minor
Health_Shock * (1- Reimbursedum)	0.00536	0.0342**	-0.0100						
	(0.00714)	(0.0158)	(0.00968)						
Health_Shock * Reimbursedum	-0.00156	-0.0111	0.000778						
	(0.0109)	(0.0181)	(0.0135)						
Health_Shock * (1- Bankdum)				0.00517	0.0162	-0.00680			
				(0.00627)	(0.0129)	(0.00868)			
Health_Shock * Bankdum				-0.0163	0.0556	-0.00363			
				(0.0181)	(0.0435)	(0.0219)			
Health_Shock * (1-Hyhhsize)							0.0130	0.0370**	0.00991
							(0.00834)	(0.0187)	(0.0116)
Health_Shock * Hyhhsize							-0.00673	0.00470	-0.0204*
							(0.00807)	(0.0165)	(0.0111)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lagged Dep.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Initial Dep.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Initial X	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lg(financial asset)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11,895	11,768	11,768	11,895	11,768	11,768	11,895	11,768	11,768

Notes: Estimates for marginal effects are reported for the fractional response models. The dependent variable is the probability of holding only cash as safe assets. A full set of control variables is included. See Appendix 2.1 for complete definitions of all variables. Robust standard errors are clustered at the household level and reported in the parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 2.12 Robust test: Random-effect OLS, financial wealth (intensive margin)

Dep. Var: Cash/Safe assets	Random Effects OLS					
	(1)	(2)	(3)	(4)	(5)	(6)
L_cashratio	0.134***	0.138***	0.134***	0.0187	0.0206	0.0180
	(0.0216)	(0.0218)	(0.0217)	(0.0138)	(0.0139)	(0.0140)
Cashratio_ini	0.133***	0.129***	0.138***	0.0540**	0.0508**	0.0545**
	(0.0231)	(0.0232)	(0.0231)	(0.0214)	(0.0215)	(0.0215)
Lghfinasset				-0.0700***	-0.0699***	-0.0700***
				(0.000936)	(0.000939)	(0.000938)
SRCH	0.0254***			0.00724		
	(0.00975)			(0.00683)		
Acute		0.0294*			0.0220*	
		(0.0165)			(0.0127)	
Minor			0.00523			-0.00205
			(0.0126)			(0.00903)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Initial X	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11,895	11,768	11,768	11,895	11,768	11,768

Notes: Random-effect OLS estimates for marginal effects are reported. The dependent variable is the proportion of financial assets held in cash. A full set of control variables is included. See Appendix 2.1 for complete definitions of all variables. Robust standard errors are clustered at the household level and reported in the parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 2.13 Robust test: Random-effect OLS, interactions (intensive margin)

Dep. Var: Cash/financial asset	Fractional Response Model								
Health shocks	SRCH	Acute	Minor	SRCH	Acute	Minor	SRCH	Acute	Minor
Health_Shock * (1- Reimbursedum)	0.0080	0.0306*	0.0003						
	(0.0084)	(0.0158)	(0.0114)						
Health_Shock * Reimbursedum	0.0057	0.0042	-0.0061						
	(0.0116)	(0.0207)	(0.0143)						
Health_Shock * (1- Bankdum)				0.0094	0.0207	-0.0012			
				(0.0072)	(0.0132)	(0.0094)			
Health_Shock * Bankdum				-0.0158	0.0369	-0.0111			
				(0.0218)	(0.0458)	(0.0321)			
Health_Shock * (1-Hyhhsize)							0.0177*	0.0274	0.0182
							(0.0092)	(0.0167)	(0.0122)
Health_Shock * Hyhhsize							-0.0038	0.0157	-0.0220*
							(0.0098)	(0.0189)	(0.0130)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lagged Dep.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Initial Dep.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Initial X	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lg(financial asset)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11,895	11,768	11,768	11,895	11,768	11,768	11,895	11,768	11,768

Notes: Random-effect OLS estimates for marginal effects are reported. The dependent variable is the probability of holding only cash as safe assets. A full set of control variables is included. See Appendix 2.1 for complete definitions of all variables. Robust standard errors are clustered at the household level and reported in the parentheses.
$***$ $p < 0.01$, $**$ $p < 0.05$, $*$ $p < 0.1$

Appendix

Appendix 2.1 Definitions of the variables

Table A3.1 Definitions of the variables

Variable	Definition
Dependent variables	
Prcash	Dummy equals 1 if the respondent holds only cash as safe assets, 0 otherwise.
Cashratio	The proportion of safe assets held in cash. Cashratio = cash / (cash + bank deposits)
Health shock indicators	
SRCH	Self-perceived health deterioration since two years ago (when the respondent took the last interview). Dummy equals 1 if the respondent reports worse health status since last interview, 0 otherwise.
Acute	Dummy equal to 1 if respondent *i* has been diagnosed with an onset of acute conditions in the past two years (since last interview), 0 otherwise. Acute conditions include Cancer or malignant tumor, Chronic lung diseases, Heart attack and Stroke.
Minor	Dummy equal to 1 if respondent *i* has been diagnosed with an onset of chronic conditions in the past two years (since last interview), 0 otherwise. Chronic conditions include Hypertension, Dyslipidemia, Diabetes or high blood sugar, Liver disease, Kidney disease, Emotional, nervous, or psychiatric problems, Memory-related disease, Arthritis or rheumatism and Asthma.
Age	The age of the respondent when taking the interview.
Gender	The gender of the respondent.
Hukoudum	Dummy equals 1 if the respondent holds agriculture hukou, and equals 0 if the respondent holds non-agriculture hukou or unified hukou[32], 0 otherwise.
Hhsize	The number of household members who usually ate meals together (not including guests).
Lghfinasset	The logarithm of real financial assets of the respondent and spouse adjusted using the CPI. We take 2011 as the base year.
Lghhitot	The logarithm of real household income (of the respondent, the spouse and other household members), adjusted using CPI.
Mobilephone	Dummy equals 1 if the respondent possesses a mobile phone, 0 otherwise.
Computer	Dummy equals 1 if the respondent possesses a computer, 0 otherwise.
Marital status (with married as the reference category)	
Married_apart	Dummy equals 1 if the respondent reports being married but not living together temporarily, 0 otherwise.
Widowed	Dummy equals 1 if the respondent reports being widowed, 0 otherwise.
Not_married	Dummy equals 1 if the respondent reports divorced/separated/never married, 0 otherwise.
Education (with illiterate as the reference category)	
Edulow	Dummy equals 1 if the respondent is illiterate, 0 otherwise.
Edumed	Dummy equals 1 if the respondent achieved an education level at middle school and below, 0 otherwise.
Eduhigh	Dummy equals 1 if the highest education level achieved is high school and above, 0 otherwise.
Work status (with currently not working as the reference category)	

[32] Unified residence Hukou is a kind of residence permission as replacement for agricultural Hukou in the area where agricultural Hukou is abandoned after the residence registration system reformation.

Work_agri	Dummy equals 1 if the respondent is doing agricultural work, 0 otherwise.
Work_emp	Dummy equals 1 if the respondent is doing employed work, 0 otherwise.
Work_slfemp	Dummy equals 1 if the respondent is doing self-employed work, 0 otherwise.
Work_unpaid	Dummy equals 1 if the respondent is doing unpaid family business, 0 otherwise.
Bankdum	Dummy equals 1 if the respondent lives in a community with at least one bank branch, 0 otherwise.
Reimbursedum	Dummy equals 1 if the respondent participates in a health insurance scheme which provides reimbursement immediately, 0 if the respondent joins a health insurance scheme in which he/she pays first and gets reimbursed later.
Hyhhsize	Dummy equals 1 if the respondent lives in a household with more than two members, 0 if the respondent lives by his/her own or with the spouse.

Appendix 2.2 Chinese provincial units

In the 2013 wave of CHARLS, respondents come from 28 province units (ordered alphabetically):

Anhui	Henan	Shandong
Beijing	Hubei	Shanghai
Chongqing	Hunan	Shanxi
Fujian	Inner Mongolia	Sichuan
Gansu	Jiangsu	Tianjin
Guangdong	Jiangxi	Xinjiang
Guangxi	Jilin	Yunnan
Guizhou	Liaoning	Zhejiang
Hebei	Qinghai	
Heilongjiang	Shaanxi	

Appendix 2.3 Baseline model (full table)

Table A2.3.1 Baseline model (extensive margins)

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
L_prcash				0.0698**	0.0691**	0.0657*	0.0047	0.0040	0.0036
				(0.0336)	(0.0337)	(0.0340)	(0.0120)	(0.0122)	(0.0121)
Ini_Prcash				0.1359***	0.1370***	0.1430***	0.0140	0.0138	0.0141
				(0.0269)	(0.0270)	(0.0272)	(0.0146)	(0.0148)	(0.0147)
Lghfinasset							-0.0478***	-0.0478***	-0.0480***
							(0.0007)	(0.0007)	(0.0007)
SRCH	0.0296***			0.0274**			0.0034		
	(0.0086)			(0.0115)			(0.0074)		
Acute		0.0482***			0.0560***			0.0329**	
		(0.0164)			(0.0216)			(0.0139)	
Minor			0.0107			0.0060			-0.0028
			(0.0112)			(0.0147)			(0.0098)
Age	-0.0282***	-0.0284***	-0.0281***	-0.0538	-0.0401	-0.0546	-0.0269	-0.0232	-0.0299
	(0.0075)	(0.0075)	(0.0075)	(0.0441)	(0.0439)	(0.0444)	(0.0252)	(0.0264)	(0.0250)
Gender	-0.0394***	-0.0388***	-0.0395***	-0.0371***	-0.0400***	-0.0394***	-0.0576***	-0.0580***	-0.0582***
	(0.0085)	(0.0085)	(0.0085)	(0.0116)	(0.0117)	(0.0116)	(0.0087)	(0.0088)	(0.0087)
Hukoudum	-0.1020***	-0.1030***	-0.1030***	-0.0333	-0.0360	-0.0360	0.0327	0.0336	0.0335
	(0.0214)	(0.0215)	(0.0214)	(0.0431)	(0.0433)	(0.0433)	(0.0309)	(0.0311)	(0.0306)
Hhsize	0.0093***	0.0092***	0.0092***	0.0010**	0.0098**	0.0104**	-0.0034	-0.0034	-0.0033
	(0.0032)	(0.0032)	(0.0032)	(0.0048)	(0.0048)	(0.0048)	(0.0029)	(0.0029)	(0.0029)
Lghhitot	-0.0062***	-0.0061***	-0.0061***	-0.0031	-0.0025	-0.0027	0.0043***	0.0045***	0.0044***
	(0.0021)	(0.0021)	(0.0021)	(0.0027)	(0.0027)	(0.0027)	(0.0016)	(0.0016)	(0.0016)
Social	-0.0169*	-0.0179**	-0.0172*	-0.0113	-0.0118	-0.0120	0.0067	0.0060	0.0069
	(0.0090)	(0.0090)	(0.0090)	(0.0121)	(0.0122)	(0.0122)	(0.0079)	(0.0080)	(0.0080)
Cognition	-0.0114***	-0.0114***	-0.0115***	-0.0118***	-0.0119***	-0.0120***	-0.0024*	-0.0024*	-0.00251**
	(0.0014)	(0.0014)	(0.0014)	(0.0020)	(0.0020)	(0.0020)	(0.0013)	(0.0013)	(0.0013)
Hsown	-0.0316*	-0.0319*	-0.0320*	-0.0350	-0.0362*	-0.0336	-0.0077	-0.0076	-0.0068

	(0.0163)	(0.0163)	(0.0163)	(0.0216)	(0.0217)	(0.0216)	(0.0124)	(0.0124)	(0.0124)
Computer	-0.0334**	-0.0349**	-0.0351**	-0.0256	-0.0294	-0.0289	0.0160	0.0143	0.0161
	(0.0144)	(0.0144)	(0.0144)	(0.0190)	(0.0191)	(0.0191)	(0.0120)	(0.0122)	(0.0120)
Mobilephone	-0.0422**	-0.0412**	-0.0412**	-0.0489**	-0.0509**	-0.0506**	-0.0153	-0.0150	-0.0173
	(0.0167)	(0.0167)	(0.0167)	(0.0222)	(0.0223)	(0.0223)	(0.0129)	(0.0129)	(0.0130)
Married_apart	-0.0076	-0.0074	-0.0079	-0.0136	-0.0139	-0.0103	-0.0494**	-0.0492**	-0.0491**
	(0.0200)	(0.0200)	(0.0200)	(0.0282)	(0.0282)	(0.0282)	(0.0200)	(0.0201)	(0.0203)
Widowed	0.0526***	0.0508***	0.0516***	-0.0030	-0.0082	-0.0060	-0.1110***	-0.1130***	-0.114***
	(0.0171)	(0.0171)	(0.0171)	(0.0387)	(0.0394)	(0.0391)	(0.0161)	(0.0165)	(0.0162)
Not_married	0.0137	0.0104	0.0121	0.0238	0.0392	0.0381	-0.0393	-0.0328	-0.0356
	(0.0396)	(0.0395)	(0.0396)	(0.0879)	(0.0896)	(0.0897)	(0.0360)	(0.0384)	(0.0369)
Edumed	-0.0320**	-0.0320**	-0.0317**	-0.0173	-0.0239	-0.0188	-0.0106	-0.0173	-0.0105
	(0.0128)	(0.0128)	(0.0129)	(0.0292)	(0.0292)	(0.0291)	(0.0207)	(0.0207)	(0.0205)
Eduhigh	-0.0811***	-0.0834***	-0.0828***	-0.0729	-0.0771	-0.0750	-0.0020	-0.0081	-0.0027
	(0.0214)	(0.0214)	(0.0214)	(0.0619)	(0.0621)	(0.0622)	(0.0345)	(0.0352)	(0.0346)
Work_agri	-0.0332***	-0.0334***	-0.0345***	-0.0287*	-0.0285*	-0.0312*	-0.0037	-0.0034	-0.0048
	(0.0122)	(0.0122)	(0.0122)	(0.0170)	(0.0170)	(0.0170)	(0.0106)	(0.0107)	(0.0107)
Work_emp	-0.0489***	-0.0496***	-0.0514***	-0.0191	-0.0204	-0.0206	0.0145	0.0123	0.0130
	(0.0162)	(0.0163)	(0.0162)	(0.0223)	(0.0225)	(0.0225)	(0.0159)	(0.0160)	(0.0160)
Work_slfemp	-0.0332	-0.0342	-0.0349	-0.0311	-0.0311	-0.0336	0.0093	0.0098	0.0069
	(0.0222)	(0.0223)	(0.0222)	(0.0300)	(0.0301)	(0.0300)	(0.0202)	(0.0203)	(0.0201)
Work_unpaid	-0.0370	-0.0338	-0.0348	-0.0509	-0.0504	-0.0534	-0.0008	0.0007	-0.0022
	(0.0301)	(0.0304)	(0.0304)	(0.0403)	(0.0408)	(0.0411)	(0.0262)	(0.0264)	(0.0266)
Bankdum	-0.1070***	-0.1080***	-0.1070***	-0.0565**	-0.0599**	-0.0600**	-0.0211	-0.0229*	-0.0198
	(0.0223)	(0.0224)	(0.0223)	(0.0272)	(0.0274)	(0.0272)	(0.0135)	(0.0137)	(0.0133)
Reimbursedum	-0.0117	-0.0125	-0.0125	-0.0078	-0.0074	-0.0096	-0.0003	-0.0002	-0.0021
	(0.0122)	(0.0123)	(0.0123)	(0.0153)	(0.0154)	(0.0154)	(0.0091)	(0.0092)	(0.0091)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Initial X	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes

Observations	13,399	13,385	13,385	11,895	11,768	11,768	11,895	11,768	11,768

Notes: RE Probit estimates for marginal effects are reported. The dependent variable is the probability of holding only cash as safe assets. A full set of contemporary time-varying and time-invariant variables are controlled but relevant estimates are not reported for brevity. See Appendix 2.1 for complete definitions of all the variables. Robust standard errors are clustered at household level and reported in the parentheses. $***$ $p < 0.01$, $**$ $p < 0.05$, $*$ $p < 0.1$

Table A2.3.2 Baseline model (intensive margins)

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
L_cashratio				0.2030***	0.2050***	0.2020***	0.0359***	0.0352***	0.0346***
				(0.0167)	(0.0169)	(0.0168)	(0.0100)	(0.0100)	(0.0100)
Ini_Cashratio				0.0423**	0.0416**	0.0474***	0.0085	0.0073	0.0088
				(0.0179)	(0.0180)	(0.0178)	(0.0138)	(0.0139)	(0.0139)
Lghfinasset							-0.0509***	-0.0510***	-0.0510***
							(0.0009)	(0.0009)	(0.0009)
SRCH	0.0320***			0.0280***			0.0067		
	(0.00771)			(0.00970)			(0.0064)		
Acute		0.0369**			0.0399**			0.0230*	
		(0.0146)			(0.0189)			(0.0128)	
Minor			0.0051			0.0010			-0.0037
			(0.0100)			(0.0125)			(0.0088)
Age	-0.0257***	-0.0260***	-0.0258***	-0.0434	-0.0341	-0.0445	-0.0238	-0.0243	-0.0288
	(0.0067)	(0.0067)	(0.0067)	(0.0364)	(0.0372)	(0.0368)	(0.0222)	(0.0240)	(0.0217)
Gender	-0.0363***	-0.0361***	-0.0365***	-0.0337***	-0.0359***	-0.0357***	-0.0531***	-0.0536***	-0.0542***
	(0.0076)	(0.0076)	(0.0076)	(0.0096)	(0.0097)	(0.0096)	(0.0074)	(0.0074)	(0.0074)
Hukoudum	-0.0916***	-0.0922***	-0.0917***	-0.0218	-0.0249	-0.0251	0.0369	0.0370	0.0367
	(0.0182)	(0.0183)	(0.0183)	(0.0351)	(0.0356)	(0.0357)	(0.0269)	(0.0271)	(0.0269)
Hhsize	0.0116***	0.0116***	0.0116***	0.0095**	0.0095**	0.0102**	-0.0028	-0.0027	-0.0024
	(0.0029)	(0.0029)	(0.0029)	(0.0040)	(0.0041)	(0.0041)	(0.0025)	(0.0025)	(0.0025)
Lghhitot	-0.0079***	-0.0079***	-0.0079***	-0.0041*	-0.0036	-0.0037	0.0038***	0.0039***	0.0040***
	(0.0019)	(0.0019)	(0.0019)	(0.0023)	(0.0023)	(0.0023)	(0.0013)	(0.0013)	(0.0013)
Social	-0.0168**	-0.0175**	-0.0171**	-0.0150	-0.0156	-0.0166	0.0021	0.0015	0.0013
	(0.0080)	(0.0080)	(0.0080)	(0.0102)	(0.0103)	(0.0103)	(0.0066)	(0.0067)	(0.0066)
Cognition	-0.0101***	-0.0102***	-0.0103***	-0.0103***	-0.0103***	-0.0105***	-0.0028**	-0.0028**	-0.0029***
	(0.0012)	(0.0012)	(0.0012)	(0.0017)	(0.0017)	(0.0017)	(0.0011)	(0.0011)	(0.0011)
Hsown	-0.0299**	-0.0304**	-0.0305**	-0.0301	-0.0308*	-0.0287	-0.0041	-0.0032	-0.0030
	(0.0146)	(0.0146)	(0.0146)	(0.0185)	(0.0186)	(0.0185)	(0.0111)	(0.0112)	(0.0111)

Computer	-0.0244*	-0.0261**	-0.0262**	-0.0157	-0.0186	-0.0180	0.0197**	0.0183*	0.0202**
	(0.0127)	(0.0128)	(0.0128)	(0.0158)	(0.0159)	(0.0159)	(0.0100)	(0.0101)	(0.0099)
Mobilephone	-0.0319**	-0.0305**	-0.0306**	-0.0340*	-0.0350*	-0.0349*	-0.0063	-0.0057	-0.0078
	(0.0149)	(0.0150)	(0.0150)	(0.0186)	(0.0188)	(0.0188)	(0.0105)	(0.0106)	(0.0106)
Married_apart	0.0013	0.0016	0.0013	-0.0074	-0.0075	-0.0036	-0.0381**	-0.0372**	-0.0364**
	(0.0177)	(0.0178)	(0.0177)	(0.0234)	(0.0234)	(0.0233)	(0.0158)	(0.0159)	(0.0159)
Widowed	0.0587***	0.0576***	0.0583***	0.0008	-0.0037	-0.0014	-0.1090***	-0.1090***	-0.1110***
	(0.0153)	(0.0153)	(0.0153)	(0.0321)	(0.0330)	(0.0326)	(0.0140)	(0.0148)	(0.0145)
Not_married	0.0180	0.0148	0.0163	0.0114	0.0227	0.0220	-0.0188	-0.0127	-0.0164
	(0.0355)	(0.0355)	(0.0356)	(0.0785)	(0.0798)	(0.0805)	(0.0400)	(0.0425)	(0.0417)
Edumed	-0.0295**	-0.0293**	-0.0290**	-0.0170	-0.0219	-0.0181	-0.0168	-0.0214	-0.0165
	(0.0116)	(0.0116)	(0.0117)	(0.0243)	(0.0243)	(0.0242)	(0.0179)	(0.0178)	(0.0176)
Eduhigh	-0.0727***	-0.0745***	-0.0740***	-0.0671	-0.0708	-0.0701	-0.0090	-0.0134	-0.0105
	(0.0188)	(0.0189)	(0.0189)	(0.0498)	(0.0499)	(0.0501)	(0.0259)	(0.0264)	(0.0261)
Work_agri	-0.0383***	-0.0388***	-0.0397***	-0.0220	-0.0215	-0.0230	0.0002	0.0004	0.0006
	(0.0109)	(0.0109)	(0.0109)	(0.0144)	(0.0145)	(0.0145)	(0.0093)	(0.0094)	(0.0093)
Work_emp	-0.0521***	-0.0539***	-0.0554***	-0.0159	-0.0180	-0.0169	0.0157	0.0133	0.0149
	(0.0144)	(0.0144)	(0.0144)	(0.0190)	(0.0191)	(0.0191)	(0.0131)	(0.0131)	(0.0131)
Work_slfemp	-0.0379*	-0.0389**	-0.0399**	-0.0192	-0.0184	-0.0206	0.0160	0.0161	0.0144
	(0.0195)	(0.0196)	(0.0196)	(0.0249)	(0.0250)	(0.0250)	(0.0164)	(0.0165)	(0.0163)
Work_unpaid	-0.0363	-0.0354	-0.0364	-0.0317	-0.0308	-0.0314	0.0112	0.0123	0.0116
	(0.0274)	(0.0277)	(0.0278)	(0.0348)	(0.0351)	(0.0354)	(0.0210)	(0.0211)	(0.0214)
Bankdum	-0.1080***	-0.1090***	-0.1080***	-0.0667***	-0.0694***	-0.0701***	-0.0355***	-0.0367***	-0.0342***
	(0.0205)	(0.0206)	(0.0205)	(0.0229)	(0.0230)	(0.0227)	(0.0102)	(0.0103)	(0.0103)
Reimbursedum	-0.0063	-0.0072	-0.0072	-0.0025	-0.0032	-0.0048	0.0056	0.0047	0.0026
	(0.0108)	(0.0109)	(0.0109)	(0.0123)	(0.0124)	(0.0124)	(0.0073)	(0.0074)	(0.0073)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Initial X	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Observations	13,399	13,385	13,385	11,895	11,768	11,768	11,895	11,768	11,768

Notes: Estimates for marginal effects are reported for the fractional response models which have controlled for the year and provincial fixed effects. The dependent variable is the proportion of safe assets held in cash. See Appendix 2.1 for complete definitions of all variables. Robust standard errors are clustered at household and reported in the parentheses. $***$ $p < 0.01$, $**$ $p < 0.05$, $*$ $p < 0.1$

Appendix 2.4: Financial wealth equation

Table A2.4 Estimates for the financial wealth equation

Dep. Var: Lghfinasset	(1)	(2)	(3)
Slfhealthchange	-0.1320***		
	(0.0413)		
Newseveredum		-0.1642**	
		(0.0739)	
Newmildum			-0.0445
			(0.0544)
Selfhealth_ini	-0.0915***	-0.1009***	-0.1040***
	(0.0294)	(0.0294)	(0.0293)
Lghfinasset_ini	0.3360***	0.3376***	0.3378***
	(0.0151)	(0.0151)	(0.0151)
Age	0.3170*	0.3171*	0.3182*
	(0.177)	(0.178)	(0.177)
Gender	0.3050***	0.3022***	0.3050***
	(0.0453)	(0.0454)	(0.0454)
Hukou	0.0694	0.0698	0.0648
	(0.1801)	(0.179)	(0.180)
Hhsize	-0.0172	-0.0171	-0.0169
	(0.0162)	(0.0162)	(0.0162)
Social activity	0.1379***	0.1409***	0.1387***
	(0.0434)	(0.0435)	(0.0435)
Cognitive ability	0.0444***	0.0446***	0.0449***
	(0.0072)	(0.0072)	(0.0072)
Married_apart	-0.0802	-0.0836	-0.0835
	(0.1069)	(0.1070)	(0.1070)
Widowed	0.3909***	0.3958***	0.3962***
	(0.1197)	(0.1192)	(0.1201)
Not_married	-0.1976	-0.1832	-0.1941
	(0.2812)	(0.2833)	(0.2824)
Edumed	0.1151	0.1150	0.1101
	(0.1034)	(0.1031)	(0.1032)
Eduhigh	0.4383*	0.4247*	0.4258*
	(0.2511)	(0.2518)	(0.2510)
Work_agri	0.0738	0.0753	0.0789
	(0.0575)	(0.0575)	(0.0575)
Work_emp	0.0701	0.0728	0.0795
	(0.0836)	(0.0835)	(0.0834)
Work_slfemp	0.1889*	0.1937*	0.1959*
	(0.1121)	(0.1130)	(0.1129)
Work_unpaid	0.2621*	0.2620*	0.2662*
	(0.1376)	(0.1387)	(0.1389)
Lg(household income)	0.0563***	0.0565***	0.0563***
	(0.0090)	(0.0090)	(0.0090)
Observations	11,895	11,768	11,768

Notes: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Chapter Three: Financial Literacy and Online Shopping Decisions: Evidence from Urban China [33]

3.1 Introduction

The retail landscape has significantly changed in recent years. The flourishing penetration of smartphones, internet and 3G+ networks, along with the development in mobile payments, e-banking and logistics, have expedited and sustained the growth of global electronic retailing (e-tailing hereafter) markets worldwide. According to a report published by eMarketer (2018), in 2017, global e-tailing sales reached $2.304 trillion, a 24.8% increase over the previous year, and make up 10.2% of global total retail sales. Consumers are attracted to online shopping because of convenience, increasing ease of use, and time and cost savings. Specifically, shopping online enables them to easily obtain critical knowledge about specific products and undertake comparisons among a wide range of goods and brands (Alba *et al.*, 1997; Häubl and Trifts, 2000; Grewal, Iyer and Levy, 2004).

Among all countries, China stands out as a special case. Having witnessed an above-average growth during the last few years, China's e-tailing market has surpassed the United States (US) since 2013. In 2018, it accounted for almost half of the worldwide total e-tailing sales (eMarketer, 2019). Underpinning this growth is the world's largest online population. According to a series of reports of China Internet Network Information Centre (CNNIC), the overall internet penetration rate in China has grown from 45.8% in 2013 to 59.6% in 2018, with the ratio in urban area (74.6%) almost twice as high as in rural area (38.4%) in 2018 (CNNIC, 2015, 2019). Among all the internet users, around 74% had made a purchase online

[33] This chapter was accepted by the 6th Young Finance Scholars (YFS) Conference and the SOAS Centre for Global Finance Annual Conference 2019. As a result, this chapter can be found on the conference websites.

in 2018. Moreover, the share of digital buyers in the total population has increased from 21.8% in 2013 to 43.1% in 2018 (reaching a total of 610 million) (see Figures 3.1 and 3.2). As a result, the Chinese e-tailing market has become a vital driver of consumption and, hence, gross domestic product (GDP).[34] Yet, the increased ease of access and the advantages of online shopping do not automatically translate themselves into effective use: There is still a notable share of Chinese internet users who have never shopped online (Figure 3.2).

A large literature has explored the determinants of online shopping behaviour, focusing on both the supply side (e.g. internet penetration, logistics infrastructure, product characteristics external environment and so on; Li and Zhang, 2002; Sohn, 2017) and the demand side (e.g. personal traits and demographics; Limayem, Khalifa and Frini, 2000; Bosnjak, Galesic and Tuten, 2007). Most of the research focuses on developed countries, and few have investigated online shopping behaviour from an economics perspective.

One factor that may be important but has been ignored in the literature on online shopping is financial literacy, which refers to one's understanding and application of financial knowledge in making financial decisions (Remund, 2010). A large strand of literature has documented that financial literacy enhances retirement planning, savings, wealth accumulation, and stock market participation (e.g. van Rooij, Lusardi and Alessie, 2011; Clark, Morrill and Allen, 2012; Xiao, Chen and Chen, 2014; Lusardi, Michaud and Mitchell, 2017; Bianchi, 2018; Hsiao and Tsai, 2018). A few papers also explore the influence of financial literacy on consumption (Dinkova, Kalwij and Alessie, 2018; Jappelli and Padula, 2017). Yet, to the best of our knowledge, the association between financial literacy and online shopping, which is an

[34] According to the National Bureau of Statistics of China (NBSC) (2019), in 2018, the e-tailing sales in China totaled above $1.3 trillion, accounting for 23.64% of the total retail sales and 10.0% of GDP. The growth in consumption expenditures contributed 76.2% to the GDP growth in 2018.

important part of daily consumption, has never been explored. This research fills this gap in the literature by looking at how financial literacy relates to online shopping decisions in China.

Online shopping decisions may be linked with financial literacy as they require the knowledge and capability of understanding and use of online payment tools. In the Chinese context, the online shopping process is mostly finalised trough digital payment options (Alipay and WeChat Wallets) or debit and credit cards.[35] And as documented by prior studies, financial literacy enhances the use of bank accounts, online banking and digital payment tools (Servon and Kaestner, 2008; Königsheim, Lukas and Nöth, 2017; Grohmann, Klühs and Menkhoff, 2018). Furthermore, financially literate individuals are more sophisticated at expenditure management, such as assessing the affordability of products and expenditures, planning and monitoring the household budget, and so on (e.g. Cohen and Nelson, 2011; Agarwalla *et al.*, 2015). Since online shopping provides an alternative shopping channel which is time and cost-saving, consumers who are more financially literate may be more motivated to adopt online shopping to fulfil the need for expenditure management.

In this research, we use the 2014 wave of the China Family Panel Studies (CFPS) to study the association between financial literacy and online shopping decisions. We also investigate the extent to which this association differs across age, income, and wealth groups. Our results are based on a set of Probit and Tobit models, as well as an IV approach applied in a conditional mixed-process (CMP) framework (Roodman, 2011), using parental education and community-level financial literacy as instruments for individuals' financial literacy. We find that respondents with higher financial literacy show a higher probability of shopping online and

[35] Alipay is an affiliate of the Chinese ecommerce group Alibaba, and WeChat Pay is owned by the messaging and social media powerhouse Tencent. Both cooperate with money market funds (MMF) to attract users' savings and spare change. The leftover balance automatically goes to an MMF, which outperforms bank deposits in terms of both performance and liquidity, and can be withdrawn by consumers whenever it is required. The account owners can also easily invest their leftover balance in other financial instruments.

spend more online. These effects are stronger for respondents who are young or have higher levels of income and wealth.

The remainder of the chapter is organized as follows. Section 3.2 summarizes the literature on the determinants of online shopping, the definition of financial literacy and the effects of financial literacy on households' financial decisions. Section 3.3 develops our hypotheses. Section 3.4 describes the dataset. Section 3.5 describes our baseline specifications, variables of interest, and estimation methodology. The empirical results are reported and discussed in Section 3.6. Section 3.7 performs a number of robustness checks. Section 3.8 performs a few extensions, and Section 3.9 concludes this study.

3.2 Literature Review

3.2.1 Determinants of online shopping

An extensive literature examines the determinants of consumers' participation in online shopping. However, most of the literature focuses on factors such as the external environment, personal characteristics (such as attitudes towards online shopping and self-efficacy), experiences, product characteristics, internet access and so on (e.g. Perea y Monsuwé, Dellaert and De Ruyter, 2004; Bosnjak, Galesic and Tuten, 2007; Hui and Wan, 2007; Pérez-Hernández and Sánchez-Mangas, 2011). Only a small strand of literature has focused on the role of socio-demographic characteristics in online shopping decisions.

Among these, a number of studies have found age to be a significant predictor of online shopping. Kuoppamäki, Taipale and Wilska (2017) find that in Finland, those aged over 60 are less likely to carry out mobile shopping. By contrast, Hernández, Jiménez and José Martín (2011) assert that, though older people may have difficulties with information technology (IT) skills to perform the first purchase, they may not diverge from the younger users once the initial barriers to online shopping have been overcome.

A gender gap in online shopping decisions exists, which may originate from the gender gap in attitudes towards online shopping (Rodgers and Harris, 2003). Dittmar, Long and Meek (2004) and Dai *et al.* (2019) suggest that, while the female consumers demonstrate a more positive attitude toward offline shopping compared to online shopping, male consumers show no difference in their attitudes towards the two shopping means. In general, online shopping is more appealing to men than to women (Hasan, 2010; Dai *et al.,* 2019). However, Räsänen and Koiranen (2016) find no significant gender difference in the online shopping participation of Finnish senior citizens. Hernández, Jiménez and José Martín (2011) assert that once the consumers become experienced e-shoppers, gender difference becomes less significant.

Some studies observe a positive relationship between education and participation in and expenditure on online shopping, even after income is controlled for. For instance, Räsänen and Koiranen (2016), Kuoppamäki, Taipale and Wilska (2017) and Pérez-Hernández and Sánchez-Mangas (2011) find that those with secondary education or higher are more likely to shop online. However, some other studies do not find a significant relationship between education and online shopping decisions (e.g., Mahmood, Bagchi and Ford, 2004; Zhou, Dai and Zhang, 2007). Zhou, Dai and Zhang (2007) claim that high education is not necessary for carrying out online shopping since it is not a complicated task.

Household structure and family context typically affect online shopping decisions. Pérez-Hernández and Sánchez-Mangas (2011) find that, in the Spanish context, those who live alone are more likely to shop online, and those who live in a household with more dependent children are less likely to shop online. However, focusing on mobile shopping, specifically, Kuoppamäki, Taipale and Wilska (2017) observe that Finnish individuals who live with a spouse or have dependent children are more likely to carry out mobile shopping.

Online shopping has always been associated with carrying various types of risks as compared to traditional offline shopping (Ko *et al*., 2004), such as the safety of credit cards and privacy-related concerns (Brashear *et al.*, 2009). Existent literature has revealed that online shoppers are generally more risk-tolerant (Trocchia and Janda, 2000; Doolin *et al.*, 2005; Brashear *et al.*, 2009).

Other factors that influence online shopping decisions include working status and income. It is found that workers are more likely to shop online compared with the unemployed (Pérez-Hernández and Sánchez-Mangas, 2011). Also, online shopping is more frequent in higher-income quartiles (Susskind, 2004; Räsänen and Koiranen, 2016).

3.2.2 Definition and measurement of financial literacy

3.2.2.1 Definition of financial literacy

Financial literacy is a concept of multiple facets. By reviewing the literature on financial literacy since 2000, Remund (2010) points out that, at its broadest definition, financial literacy comprises not only the knowledge of financial concepts but also the ability to apply the knowledge in financial planning.[36] Another definition is proposed by the OECD/INFE (Organization for Economic Cooperation and Development International Network on Financial Education), who defines financial literacy as a "combination of awareness, knowledge, skill, attitude and behavior necessary to make sound financial decisions and ultimately achieve individual financial wellbeing" (Atkinson and Messy, 2012, P.5). Supporting questionnaires on financial knowledge, financial behaviours and financial attitudes are designed by OECD/INFE, and a measure of financial literacy is proposed, covering financial knowledge, financial attitude and financial behaviour (Atkinson and Messy, 2012).

In most of the empirical studies, financial knowledge is used to proxy financial literacy, and the terms financial knowledge and financial literacy are often used interchangeably (see, e.g., Behrman *et al.*, 2012; Alessie, Van Rooij and Lusardi, 2011; Atkinson and Messy, 2011; Lusardi and Mitchell, 2011a). Only a few studies that examine the patterns of financial literacy provide a holistic analysis of financial literacy by looking at the financial attitudes and financial behaviour aspects as well (e.g., Agarwalla *et al.*, 2015; OECD, 2016). The terms financial literacy and financial knowledge will be used interchangeably hereafter.

[36] The exact wording of Remund (2010, P.278)'s definition of financial literacy is: 'Financial literacy is a measure of the degree to which one understands key financial concepts and possesses the ability and confidence to manage personal finances through appropriate, short-term decision-making and sound, long-range financial planning, while mindful of life events and changing economic conditions.'

3.2.2.2 Measures of financial literacy

3.2.2.2.1 Test scores: the "Big Three" questions and extensions

Three financial literacy questions (the Big Three questions, hereafter) were included in the 2004 wave of the Health and Retirement Study (HRS), with the aim of linking individuals' financial literacy with their financial behaviours.[37] The Big Three questions ask about concepts on interest compounding, inflation and risk diversification. The first two concepts are relatively basic but crucial to financial numeracy. The third concept is more complicated and requires a basic understanding of stocks and mutual funds. The different levels of difficulty aim at differentiating financial sophistication using a limited number of questions. The Big Three questions have been widely used in surveys and studies conducted in many countries, such as Australia, Germany, Italy, Sweden, Switzerland, the Netherlands, Japan and so on (Lusardi and Mitchell, 2014).

One of the obstacles to measuring financial literacy is the lack of a comprehensive set of questions on all the aspects of financial literacy (Huston, 2010). Therefore, more concepts related to competitive saving and investment decisions are considered in later studies, to capture as many dimensions as possible of financial literacy (e.g., Lusardi, Mitchell and Curto, 2014; OECD, 2016)[38]. For instance, the DNB Household Survey (DHS) includes 16 financial knowledge questions in their questionnaire. van Rooij, Lusardi and Alessie (2011) classify the sixteen questions into basic and advanced questions according to the results of factor analysis.

[37] The HRS is a nationally representative longitudinal dataset of Americans over the age of 50. A module on planning and financial literacy was commissioned to the HRS 2004 by Lusardi and Mitchell (2011a), including 3 questions on compound interest, inflation and risk diversification concepts. The exact wording of the 3 questions is shown in Appendix 3.2.1.

[38] In the notable series of comparative analysis on worldwide financial literacy conducted by the OECD/INFE (OECD, 2016), 7 questions covering the time-value of money, interest paid on loans, interest compounding, risk-return nexus, inflation and risk diversification are asked. Lusardi, Mitchell and Curto (2014) analyse ten financial knowledge questions on capital markets, risk diversification, knowledge of fees as well as numeracy in the HRS 2008.

The basic set includes five questions on numeracy, compound interest, inflation, time value and the purchasing power of money. The advanced set includes eight questions on much more complex topics ranging from the features and working of various financial products, risk diversification, basic asset pricing and the working of the stock market. In a series of OECD/INFE reports on global financial literacy (e.g. OECD, 2016), seven questions covering the time-value of money, interest paid on loans, interest compounding, risk-return nexus, inflation and risk diversification are asked.

Another challenge in measuring financial literacy lies in the combination of all sorts of information in constructing an integrated measure. A typical practice is to count the number or calculate the fraction of the financial knowledge questions which the respondent answers correctly (Lusardi, 2008; Anderson, Baker and Robinson, 2017; Hsiao and Tsai, 2018). Such a measure is simple to calculate and simple to interpret. Another plausible practice is to perform a factor analysis on the answers to the questions, since researchers conjecture that similar patterns of the answers are driven by common factors (e.g. van Rooij, Lusardi and Alessie, 2011; Deuflhard, Georgarakos and Inderst, 2018). Factor analysis can extract these meaningful common factors and alleviate the potential problem of multicollinearity. This method is particularly effective coping with a long list of financial questions. However, a disadvantage of both counting the correct answers and the factor analysis index (FA index) is that, in both cases, the financial knowledge questions are equally weighted, despite the fact that some questions are more difficult than others.

To tackle this problem, a two-step weighted scoring technique, the principal component analysis of RIDIT scores (PRIDIT), was introduced to the financial literacy literature (Behrman *et al.*, 2012; Lusardi, Mitchell and Curto, 2014).[39] In the first step, each question is given a

[39] The PRIDIT approach was initially developed by Brockett *et al.* (2002) to handle binary or categorical indicators that proxy for a difficult-to-observe dependent variable. Similar to "probit" and "logit", "ridit"

RIDIT score according to how difficult it is for the respondents to answer it correctly. More specifically, a higher credit is applied for the correct answer to a question that more respondents answer incorrectly, and a higher penalty for the incorrect answer to a question that more respondents answer correctly. Next, a principal component analysis is applied to the RIDIT scores calculated in the first stage. By principal component analysis, a RIDIT score (hence the corresponding answer to the question) will be given a higher weight in building up the final financial literacy index if it is less correlated with the combination of the other RIDIT scores.

3.2.2.2.2 Subjective financial knowledge

Some questionnaires augment the financial knowledge questions with a question on the self-perception of financial knowledge. A significantly positive association, though not quantitatively large, is found between perceived financial literacy and the actual test score in several databases (e.g. Lusardi and Mitchell, 2007a; Bursztyn *et al.*, 2012). This fact has generated two strands of literature.

The first strand focuses on the misperception in financial literacy. Over-confidence in financial literacy is widely witnessed in several countries such as the US, Germany, and the Netherlands (Lusardi and Mitchell, 2014; Anderson, Baker and Robinson, 2017). For instance, Lusardi (2011) finds that, among the 38 percent of US respondents who score themselves 6 or 7 (7 is the highest) in perceived level of financial knowledge, only half were able to do calculations involving interest rates and inflation.[40] One exception is Japan, where respondents rated themselves lower compared with the actual test score (Sekita, 2011). Compared to the

represents a type of transformation based on the observed distribution of a response variable for a specified sample (Bross, 1958).

[40] In the National Financial Capability Study (NFCS), the respondents are asked to self-assess their level of financial knowledge on a scale from 1 to 7, where 1 indicates very low and 7 very high financial knowledge.

young, the elderly are less aware of their low level of financial literacy and rate themselves high (Lusardi and Mitchell, 2011b), which raises the concern that the most vulnerable cohorts are the least aware of their shortfall in financial literacy and are at high risk of loss caused by poor financial decisions. Women score lower on financial knowledge test compared to men but are less likely to overestimate their financial literacy (Anderson, Baker and Robinson, 2017).

The second strand of literature examines how such misperceptions affect financial decisions. For example, Anderson, Baker and Robinson (2017) explore the extent to which misperception of financial literacy relates to household precautionary savings and retirement planning. 5,814 randomly selected LinkedIn respondents were asked the Big Five questions as well as their self-assessed levels of financial literacy in 2014.[41] The authors find that those who perceive themselves financially literate, rather than those who are actually financially literate, are more likely to accrue precautionary savings and plan for retirement. However, it is very likely that the relationship actually goes the other way around, i.e. that the holding of precautional saving and a retirement plan makes people believe that they make such decisions out of financial sophistication.

Allgood and Walstad (2016) suggest that both actual and perceived financial knowledge contribute to sophisticated financial behaviours. The authors examine the financial behaviours in terms of credit cards, investments, loans, insurance, and financial advice of 28,146 US respondents from the 2009 wave of the NFCS. The sample is divided into four groups according to different combinations of high/low perceived and actual financial literacy. The authors find that actual and perceived financial literacy may have independent explanatory powers to one's financial behaviours. Compared with those who have low perceived and low

[41] The 2009 NFCS in the US augments the Big Three questions with two additional financial literacy questions on mortgage interest and bond pricing. These five questions are labled as the Big Five questions in later studies.

actual financial literacy, those with either high actual or high perceived financial knowledge are more likely to make sophisticated financial decisions.[42]

3.2.2.3 Patterns of financial literacy

Many works have examined the patterns of financial knowledge in the general population; similar patterns are found worldwide (Lusardi and Mitchell, 2011b; OECD, 2017).

Specifically, financial illiteracy is pervasive among the general population for all age groups. Lusardi and Mitchell (2011a) find that in the 2004 survey of the HRS, Americans aged 50+ faced widespread shortfalls in financial knowledge. Only one-third of the respondents understand all the concepts of compound interest, inflation and risk diversification. Ten percent of the sample understands none of the concepts. The observed level of financial knowledge is strikingly low, considering that these respondents are supposed to be familiar with inflation and risk diversification: They had experienced at least two periods of high inflation and had witnessed numerous economic and stock market shocks (Lusardi and Mitchell, 2014). The financial literacy for young Americans is also not positive: only 27% of the young respondents (aged 23 to 28 in 2008) in the National Longitudinal Survey of Youth (NLSY) correctly answered all of the Big Three questions (Lusardi, Mitchell and Curto, 2010).

Certain demographic groups demonstrate particular acute underperformance (Alessie, Van Rooij and Lusardi, 2011; Hanson and Olson, 2018). For instance, compared to women, men, in general, perform better at the basic and advanced financial literacy questions and are less likely to answer "do not know (DK)" (e.g. Alessie, Van Rooij and Lusardi, 2011; Klapper,

[42] More specifically, Allgood and Walstad (2016) find that those with high actual or perceived financial knowledge are more likely to participate in the stock market, set up and actively manage their retirement accounts, own a residential house, compare different offers for mortgage and auto loans, hold health/life/auto insurances, and seek advice from financial professionals. They are also less likely to have costly credit card behaviour or pay the mortgage late.

Lusardi and Van Oudheusden, 2015). Financial literacy is generally higher for people with higher educational attainment (e.g. Lusardi and Mitchell, 2011a). However, even the elite cohort with high levels of education and high positions are not financially literate as expected (Anderson, Baker and Robinson, 2017). A wide gap in financial literacy is also witnessed across employment status (e.g. Lusardi and Mitchell, 2011b). For instance, in the Dutch context, financial literacy is higher for the self-employed than for employees, and the non-employed demonstrate the lowest financial literacy (Alessie, Van Rooij and Lusardi, 2011).

3.2.3 Financial literacy and financial decisions

The neoclassical consumption model posits that individuals choose consumption and saving at each point of life so as to maximizes their discounted lifetime expected utility. The underlying assumption of this model is that individuals understand numeracy and financial concepts, and are able to project expected future incomes (Lusardi and Mitchell, 2007b). However, empirical evidence shows that households are, on average, not sufficiently financially literate to make sound financial decisions (e.g. Hung, Parker and Yoong, 2009; Almenberg and Dreber, 2015). A large number of studies examine whether financial literacy influences financial decision making. Most of them focus on household saving, retirement planning and stock market participation, but still shed light on our analysis. These studies are surveyed below.

3.2.3.1 Retirement planning and saving

Bernheim, Garrett and Maki (2001) are among the first to uncover the link between financial knowledge and economic behaviours. The authors examine the effect of high school financial

education on saving behaviour based on a cross-sectional sample of 2,000 US individuals aged 30-49 in 1995. The authors find that financial education in high school significantly increases the respondents' saving rate and wealth, probably because the financial education familiarizes them with financial matters and removes psychological barriers to financial decision making. High retirement saving is also observed for the American employees who have joined financial education programmes provided by the companies (Bernheim and Garrett, 2003).

Some researches assert that financially literate individuals accumulate higher retirement wealth through proper retirement planning. For instance, Lusardi and Mitchell (2007a) analyse the retirement planning of a sample of 811 respondents from the RAND American Life Panel (ALP). The authors exploit both objective and subjective measures of financial knowledge, instrumented by the financial training offered by the school and the employers. They find that financially literate employees are more likely to plan for retirement. Alessie, Van Rooij and Lusardi (2011) examine a DHS panel of 1,665 respondents who were over 25 when being interviewed in 2005 and 2010 (before and after the financial crisis). Using the financial situation of siblings and parents as IVs for financial literacy, their empirical results suggest that respondents with a higher level of financial literacy think more about retirement and save more.

A few studies focus on precautionary saving. The consequences of the absence of precautionary saving is potentially disastrous. Without adequate liquid assets, people may turn to other alternatives such as short-term credit, friends and family and potentially even public subsidy programs, and may fall in indebtedness (Chase, Gjertson and Collins, 2011). Lusardi, Schneider and Tufano (2011) explore households' financial fragility using data from the 2009 TNS Global Economic Crisis survey. Those who cannot raise $2,000 funds in 30 days are considered to be financially fragile. Financial fragility is frequently observed, especially for individuals with low levels of education and no financial education. The results show that financially literate individuals are more likely to accumulate precautionary savings. Anderson,

Baker and Robinson (2017) focus on the same topic based on a sample of 5,814 LinkedIn users collected in 2014. The authors examine the misperception of financial literacy and compare the predictive powers of actual and perceived financial literacy. They find that though both perceived and actual financial literacy facilitate precautionary savings, the former is more effective.

3.2.3.2 Stock market participation and portfolio choice

Another strand of literature finds that higher financial literacy is associated with a higher propensity to invest in stocks and other financial instruments (e.g., Abreu and Mendes, 2010; van Rooij, Lusardi and Alessie, 2011; Hsiao and Tsai, 2018).

Abreu and Mendes (2010) study the impact of financial literacy on portfolio diversification using survey data on 1,268 investors disclosed by the Portuguese Securities Commission. The authors employ three distinct aspects of financial literacy, including their score for the financial knowledge test, their educational attainment and their sources of (financial) information. They find that financially literate individuals are more likely to participate in the stock market, and tend to hold more diversified portfolios. One limitation of their research is that education may not be appropriate to proxy financial literacy, since the former reflects one's attainment in a general sense, but the latter focuses on specific knowledge about the working of the financial market and financial products.

van Rooij, Lusardi and Alessie (2011) use a sample of over 2,000 representative Dutch households from the 2005 DHS to uncover the role of financial literacy in explaining the "stock-holding puzzle" that many households do not invest in stocks even though stocks can bring a higher rate of return (Campbell, 2006). The authors construct FA indices for both basic and advanced financial literacy. They also apply an IV approach, using the exposure to

economic education prior to job market entry as an instrument for advanced financial literacy. The analysis shows that low levels of financial literacy may be an explanation for the stock-holding puzzle: Households with higher financial literacy are more likely to invest in stocks.

Bianchi (2018) uses administrative data from a financial institution on 511 French clients at a monthly frequency from Sep 2002 to April 2011. Both subjective and objective measures of financial knowledge are estimated. Portfolio choice is assessed from three dimensions: how risky is the portfolio, how frequencly the investors rebalance the portfolio and in what direction they rebalance the portfolio in terms of risky asset share. Their estimates reveal that, in general, more sophisticated households earn higher portfolio returns by actively rebalancing their portfolios and shifting towards funds which provide higher returns. However, the lack of consideration of endogeneity makes it impossible to draw a causal inference.

3.2.3.3 Financial inclusion

One of the prerequisites for individuals to accumulate retirement saving and to invest in the stock market is that they use bank accounts. Prior studies also investigate how and to what extent does financial literacy enhance the take-up of bank accounts, and digital financial services is also a topic of debate.

Based on a sample of 243 US urban residents, Servon and Kaestner (2008) investigate whether training on financial literacy, access to information and communication technologies (ICT) and training on how to use the Internet can help low- and moderate-income (LMI) individuals to adopt online banking. The authors suggest that skills in technology motivate LMI individuals to learn about financial literacy, and hence help banks to expand into currently untapped markets. In other words, financial literacy, together with skills in technology, increase the intention of urban LMI individuals to carry out online banking.

Grohmann, Klühs and Menkhoff (2018) explore how financial literacy improves the demand for financial inclusion at the country level, and how it works together with the supply-side factors, such as the financial infrastructure. Using financial education as an instrument for financial literacy, the authors find that, regardless of the measures of financial inclusion being used, financial literacy enhances financial inclusion at all levels of financial infrastructure. However, it works with financial infrastructure in different manners depending on the definition of financial inclusion. More specifically, the authors find that, if financial inclusion refers to taking up bank accounts, the positive marginal effect of financial literacy on financial inclusion is largest in countries with low levels of financial infrastructure. Under such circumstances, financial literacy works as a substitute for financial infrastructure. When it comes to the use of financial services such as saving in formal financial institution, financial literacy is more effective in countries with higher levels of financial infrastructure. Under such circumstances, financial literacy works as a necessary complement to financial infrastructure in enhancing financial inclusion.

Lyons, Grable and Zeng (2019) examine the role of financial literacy on financial inclusion in terms of the demand for loans. By analyzing a sample of 23,676 individuals from the 2013 wave of the Chinese Household Finance Survey (CHFS), the authors find that the financially included households who are more financially literate are more likely to borrow from formal sources, such as banks. However, the financially excluded households are less likely to benefit from financial literacy and are more likely to borrow from relatives, friends or informal institutions. Similar to Grohmann, Klühs and Menkhoff (2018), this research implies that financial literacy can only improve financial inclusion (in terms of formal borrowing) when a sufficient level of financial infrastructure is in place. However, before reaching such a specific level of financial infrastructure, increasing the supply of financial services, may be the only way to approach financial inclusion.

Financial literacy may also prompt the use of digital financial services. Königsheim, Lukas and Nöth (2017) examine how self-perceived financial knowledge affects the use of digital financial services. Based on a sample of 1,752 German respondents who hold retail bank accounts, the authors find that respondents with higher perceived levels of financial knowledge are more likely to adopt digital financial services, such as online investment advice, credit, and payment services.

3.2.3.4 Consumption

The literature on financial literacy reveals a glaring lack of attention to household consumption behaviours. One of the notable exceptions is Jappelli and Padula (2017). The authors posit that, in a life-cycle consumption model, financially literate investors expect higher returns on assets, and consequently, reduce current consumption. Their empirical analysis is performed on a two-wave panel with 8,743 observations taken from the 2006 to 2010 waves of the Survey of Household Income and Wealth (SHIW) on Italian households. Using the major that the respondent studied and parental educational background as instruments for financial literacy, the authors observe a positive relationship between financial literacy and consumption growth, but only for the high-wealth cohort.[43] A potential explanation is that it is less likely for wealthy investors to be restricted by borrowing constraints, and it is easier for them to smooth out consumption intertemporally.

Dinkova, Kalwij and Alessie (2018) obtain different results using a Dutch panel. Their dataset consists of three waves of information for 948 households. In contrast to Jappelli and Padula (2017), they find no evidence for a link between financial knowledge and consumption

[43] Specifically, the IVs include whether the parents have a college degree and whether the respondent has a degree in economics.

growth. However, the authors find that higher financial knowledge is associated with higher household consumption on public use.[44] A gender gap is also found in both the level of financial knowledge and the way it contributes to household consumption. The financial knowledge of the husband is found to dominate the wife's in terms of its effect on consumption. However, as the wife's financial knowledge increases, the household's consumption on public use tends to decrease.

[44] The household consumption on public use equals the total household consumption less the consumption on personal use. More specifically, the total reported household consumption consists of mortgage/rent, utilities, transport, insurances, debt/loans, food (excluding restaurants), and alimony payments, which are considered as public use consumption, and expenditure on buying clothes, out-of-pocket medical expenses, gifts, leisure activities, tobacco and alcohol, considered as personal use consumption.

3.3 General Hypotheses

Compared with traditional offline shopping, online shopping requires higher financial knowledge and capability of the consumers due to the involvement of digital payment in the process. According to the literature, financial literacy induces sophisticated financial behaviours and enhances the take-up of bank accounts and digital financial services (e.g., Servon and Kaestner, 2008; Agarwalla *et al.*, 2015; Königsheim, Lukas and Nöth, 2017; Grohmann, Klühs and Menkhoff, 2018). In the Chinese context, digital payment tools, such as Alipay and Wechat Wallet, perform a critical role in personal financial management, since the left-over balance in the payment tools can be easily invested in open-ended funds. Consumers with high financial literacy are more informed about the benefits of holding bank cards or digital payment tools. Hence, the financially literate consumers are more motivated to acquire and use bank cards or digital payment tools, therefore fulfilling the premises of online shopping. This leads to the following hypothesis:

H1: Consumers with higher financial literacy are more likely to participate in online shopping.

Moreover, financially literate individuals are found to be more sophisticated at expenditure management. For instance, Agarwalla *et al.* (2015) find that the financially literate young people are more likely to set up long-term financial plans, assess the affordability of products and expenditures, and monitor household budgets. Since online shopping provides an alternative shopping channel which is time and cost-saving, consumers who use digital payments and are more financially literate may make more use of online shopping. Therefore, we also propose the following hypothesis:

H2: Consumers with higher financial literacy spend more online.

3.4 Data and Sample Construction

We use the 2014 wave of the CFPS, which is a nationally representative longitudinal survey of Chinese communities, households, and individuals, funded by the Chinese government through Peking University. The 2014 survey contains information on 32,307 respondents from around 12,000 families from 25 provinces who reside in either rural or urban areas. Information on socio-demographics, household composition, household consumption and wealth, as well as individuals' online shopping decisions and financial literacy, are collected.

After retrieving the data from several modules and merging them based on household and individual identifiers, we obtain a preliminary cross-sectional dataset which contains 32,307 respondents from around 12,000 families. In the CFPS, only the family member who is most familiar with the family's financial situation is asked the financial literacy questions. Therefore, the 13 financial knowledge questions are only asked to 3,380 respondents, 200 of which reside in rural areas, and 3,180 of which reside in urban areas. To rule out the possibility that the exclusion from online shopping might be caused by the inaccessibility of internet and bank services, we focus only on urban respondents. As typical in survey data, respondents may sometimes decline to answer a question or skip a question. We treat such cases by dropping the observation from the sample. Moreover, since income, wealth and consumption are sensitive to outliers, we winsorize the top and bottom 1% of these variables before taking logarithms.[45] After we exclude the missing values for all the variables in the baseline specification, the final sample consists of 2,183 urban respondents.

[45] We use household income, wealth and consumption as control variables in our specifications. See Section 3.5.3 for more details about the control variables.

3.5 Methodology and Specifications

3.5.1 Dependent variables: Online shopping

The CFPS 2014 includes a question on each respondent's individual online shopping expenses in the "Mobile Phone and Internet" module. The question asks:

In the past 12 months, how much did you spend on online shopping (including paying online)?

_____yuan (1…10,000,000)

To explore whether the financially literate are more likely to shop online and whether they spend more on online shopping, two outcome variables are created. The expenditure on online shopping (*Exp_online*) is constructed based on the answers to the question above. The participation in online shopping is denoted by a dummy variable (*D_online*), which equals 1 if *Exp_online* is positive, and 0 otherwise.

3.5.2 Financial literacy

3.5.2.1 Objective financial knowledge

The thirteen financial knowledge questions included in the CFPS take the forms of either multiple choice questions or true-or-false questions. Both types of questions include the option for respondents to answer with "don't know" or "refuse". The CFPS questions resemble those in the DHS and OECD reports (Atkinson and Messy, 2016), excluding some of the most difficult concepts such as asset pricing, and including a few easier questions which are unique to the Chinese context, i.e., the annual interest rate of one-year fixed deposits in the bank and the name of the bank that is responsible for making and carrying out monetary policies. The thirteen questions are classified into five basic and eight advanced questions.

The five basic financial questions in the CFPS ask about the working of simple interest and compound interest (Question 1, 2 and 3), the concept of the real interest rate (Question 4) and the time value of money (Question 5). These concepts are essential for transactions, financial planning and everyday financial decision-making (van Rooij, Lusardi and Alessie, 2011). As such, they are closely relevant to consumption (Dinkova, Kalwij and Alessie, 2018; Jappelli and Padula, 2017). The eight advanced questions cover topics such as the risk-return nexus, risk diversification, working of the financial market, and the risk profile of various financial products, which lie at the basis of investment and portfolio choice. The exact wording of these questions is displayed in Appendix 3.2.2. Responses to these questions are discussed in Section 3.6.1.2.

Based on the answers to the financial knowledge questions, we construct three sets of indicators for objective financial knowledge.

The first set of indicators counts the number of correct answers to basic and advanced questions, respectively. Therefore, the summary score for basic financial knowledge, *Fk_score_b*, ranges from 0 to 5, and the summary score for advanced financial knowledge, *Fk_score_a*, ranges from 0 to 8.

Then, following van Rooij, Lusardi and Alessie (2011), we perform factor analysis using the principal-component method on the responses to the five basic financial knowledge questions and the eight advanced financial knowledge questions, respectively. Two indices are created for basic and advanced financial knowledge, namely, *FA_Indx_b* and *FA_Indx_a*, respectively. These two indices extract the most important common factors underlying the answers to the basic/advanced questions given by the respondents. Details on the derivation of the FA indices are reported in Appendix 3.3.1.

Finally, following Lusardi, Mitchell and Curto (2014), we take advantage of the two-step PRIDIT approach proposed by Behrman *et al.* (2012). The PRIDIT approach comprises two steps. In the first step, a RIDIT score is given to each question according to how difficult it is for the respondents to answer. The second step performs a principal component analysis on the RIDIT scores. The calculation of the RIDIT scores for each question is depicted in Appendix 3.3.2.

3.5.2.2 Subjective financial knowledge

The financial literacy module of the questionnaire complements standard questions with a question on self-perception, which reads: "*In your opinion, your financial knowledge is, compared to the people of your age: 1. Way above the average level 2. Above the average level 3. About average level 4. Below the average level 5. Way below the average level*". Based on this question, we construct a score for subjective financial literacy, which ranges from 1, indicating way below the average, to 5, indicating way above the average.

All the indicators for financial literacy, either objective or subjective, are normalized so as to range from 0 to 1. This operation aims to eliminate the influence of scale and allow comparison of the predictive powers of different sets of indicators for financial literacy.

3.5.3 Baseline specifications and estimation methodology

3.5.3.1 Baseline specifications

We start by estimating whether the respondent shops online (extensive margin). We then look at the determinants of the intensive margin, i.e. the expenditure on online shopping. The baseline model for the extensive margin takes the following form:

$$Y_i = 1\big(\beta_1 FK_{b_i} + \beta_2 FK_{a_i} + \gamma X_i + Location\ dummies\ + u_i > 0\big), (i = 1, \dots, N)$$

Eq. 3.1

where Y_i is coded as 1 if the respondent i had shopped online in 2013, and 0 otherwise. FK_{b_i} denotes respondent i's basic financial knowledge, and FK_{a_i} denotes respondent i's advanced financial knowledge. In the case of subjective financial knowledge, we include solely self-perceived financial knowledge instead of the objective basic and advanced financial knowledge measures.

Regressions are estimated on the objective indicators (i.e. summary score, FA indices, PRIDIT indices) and the subjective indicator for financial literacy each at a time. X_i denotes a set of control variables. Following the literature on online shopping and consumption decisions (e.g. Hernández, Jiménez and José Martín, 2011; van Rooij, Lusardi and Alessie, 2011; Lian and Yen, 2014; Dinkova, Kalwij and Alessie, 2018), we consider:

(i) Individual-level control variables including age, gender, marital status, duration of education, working status, self-reported health shocks and whether the respondent is risk-reverse. We also add a variable that measures how sophisticated the respondent's attitude towards spending and saving is.[46]Additionally, we include a discrete variable to measure the frequency that the respondents communicate via internet.

(ii) Household-level control variables including number of dependent children and dependent elderly, household net annual income (in logarithm), household financial and non-financial wealth (both in logarithm).

[46] A respondent is considered to hold less sophisticated attitude towards spending and saving if he/she (1) considers spending to be more satisfactory than saving, (2) is present bias, or (3) believes that money should be spent whenever one has it. Details of the definition is displayed in Appendix 3.1.

(iii) Two community level variables which measure the development of facility and infrastructure in the community.[47]

To account for the variation in online shopping habits and internet penetration rate across regions, we include two location dummies to identify whether the household resides in the western area or central area, with the coastal area as the reference group.[48] Finally, u_i denotes the error term.

For each set of financial literacy measures, we estimate three models which differ in the explanatory variables included. In the first model, we include only financial literacy as the explanatory variable(s). In the second model, we additionally control for individual-level variables. In the third model, we include a full set of control variables, including individual- and household-level variables. Location dummies are included in all the specifications.

For Hypothesis 1 to hold, we expect to observe positive estimates for β_1 and β_2.

The baseline model for the intensive margin incorporate the same set of control variables. Since we only observe the non-negative online expenditures, the baseline model for the intensive margin takes the form:

$$Y_i = \text{Max}\,(0, \beta_0 + \beta_1 FK_{b_i} + \beta_2 FK_{a_i} + \gamma X_i + Location\ dummies + u_i)$$

$$(i = 1, \dots, N),$$

Eq. 3.2

[47] These two variables are constructed based on a set of dummy variables using factor analysis. These dummy variables identify whether the community possess facilities such as convenience stores, schools and so on, and whether the community has access to public services such as electricity, transportation and so on. The lists of the facilities and public services can be found in Appendix 3.1.

[48] The definitions for all the control variables are displayed in Appendix 3.1.

where Y_i denotes respondent i's expenditure on online shopping expressed in logarithm. All other variables are the same as in Eq. 3.2. For Hypothesis 2 to hold, we expect to observe positive estimates for β_1 and β_2.

3.5.3.2 Discussion of selected control variables

In this section, we highlight some control variables which are included in our models but not widely used in the literature on online shopping decisions.

In the Chinese context, shopping information from peers or from social media shapes consumers' preferences to a great extent. As of 2014, around 20% of Weibo users had been shopping online via Weibo (Statista, 2015).[49] Therefore, we include a variable for the frequency of using the internet to communicate, whose values range from 0 (never) to 7 (every day).

As suggested by Servon and Kaestner (2008), financial literacy is more effective in promoting the use of online banking in the presence of computer skills. Therefore, we include a variable on the households' ownership of a computer to proxy for the computer skills of the respondent. Even though a respondent is not sufficiently skilled to shop online, he/she can still get help from other household members if they have a computer at home.

Following Grohmann, Klühs and Menkhoff (2018) and Lyons, Grable and Zeng (2019), the link between financial literacy and financial inclusion may also depend on the development of infrastructure. Therefore, we also incorporate two variables measuring the development of infrastructure and facility at the community level. Detail items considered for infrastructure is demonstrated in Appendix 3.1.

[49] Weibo is one of the most popular social media platforms in China.

3.5.3.3 Estimation methodology

We estimate the extensive margin models using Probit models and the intensive margin ones using Tobit models.[50] However, the Probit and Tobit estimates may be biased, since financial literacy may be subject to an endogeneity problem. First, financial literacy may be measured with substantial error, since many responses are imprecise and may result from guessing, especially those to complex questions (van Rooij, Lusardi and Alessie, 2011). Second, the baseline estimates may be biased due to unobserved time-invariant individual characteristics such as innate ability and intelligence (Behrman *et al.*, 2012). Third, selection bias can exist if those who wish to shop online are motivated to invest in financial knowledge to achieve this. Moreover, if people who attempt to shop online improve their financial literacy in the process, then shopping online would be influencing financial literacy rather than the other way around. As a result, reverse causality may occur and make the estimates biased (Hilgert, Hogarth and Beverly, 2003). Failing to control for these issues may lead to substantial bias and inconsistency in the estimates (Hyslop and Imbens, 2001). To remedy the endogeneity problems and to disentangle causality from correlation, we therefore also use an IV approach in estimation.

Following Jappelli and Padula (2017), the first IV we select for the respondent's financial knowledge is the mother's duration of education.[51] Parental educational attainment may make a valid instrument because, on the one hand, empirical evidence has demonstrated a strong and monotonic relationship between financial literacy scores and parental education (Lusardi, 2009,

[50] Similar results were obtained for the models explaining the amount spent online when using a Tobit estimator instead of OLS (see Section 3.7.2).

[51] We also considered the fathers' duration of education as instrument for the respondents' financial literacy. However, it failed most of the instrument validity tests. Therefore, we do not report the results. Details for the tests of instrument validity of all our reported specifications appear in Appendix 3.6, and are discussed in Section 3.6.2.

P.263; Jappelli and Padula, 2017). On the other hand, it is very likely that parental educational attainment is related to personal online shopping decisions only via its effect on individuals' financial literacy.

Following Bucher-Koenen and Lusardi (2011) and Calcagno and Monticone (2015), the second IV we use is the average level of financial literacy in the community, excluding the respondent. On the one hand, the community level of financial literacy is likely to increase individual financial knowledge through social interactions. On the other hand, financial literacy of the others is beyond the control of the respondents. For concerns on the representativeness of the community average, we only consider the communities which have at least five respondents (the *Com_5* sample, with 1,771 observations), and those with at least ten respondents (the *Com_10* sample, with 1,353 observations) as a robustness check.

Several existing papers takes basic financial literacy as exogenous since the authors believes that the basic financial knowledge questions only capture one's cognitive ability such as numerical ability (e.g. van Rooij, Lusardi and Alessie, 2011). While we treat both basic and advanced financial literacy as endogenous. This is because that the basic financial knowledge questions ask about concepts on interest rate and the time value of money, which is closely related to economic decisions and improves along with the experiences. Additionally, the basic and advanced financial literacy in our sample are significantly correlated at the 1% level, with a correlation coefficient at 0.53. Furthermore, different from most of the previous studies that calculate IV estimators using two-stage least squares (2SLS) method, we adopt the conditional mixed process (CMP) modeling framework for the IV approach, using both IVs simultaneously. CMP is, at its core, a seemingly unrelated equations (SUR) method, which allows for cross-equation correlations and produces more efficient estimates than those derived from single-

equations.[52] The CMP framework can exploit limited-information maximum likelihood (LIML) estimation, which beats 2SLS estimation in case of weak IV for small samples (Roodman, 2011). Moreover, a CMP framework allows us to capture the binary or censored nature of the dependent variables and produces more efficient estimates (Roodman, 2011).

[52] We estimate the model using the Stata 15.0 command -cmp- developed by Roodman (2011). Our estimation results under the CMP framework prove the existence of cross-equation correlations.

3.6 **Empirical Results**

3.6.1 Descriptive statistics

Table 3.1 displays the key information on the 2,183 respondents relevant to this research. Table 3.2 further summarizes the respondents' online shopping and financial literacy profiles by several socio-demographic variables.

3.6.1.1 Dependent variables

Table 3.1 displays the key information on the 2,183 respondents relevant to this research. The first four columns report the statistics for the full sample. Column (5) and (6) report the statistics for online shoppers and non-online shoppers, respectively. As shown in the first panel of Table 3.1, 17% of the respondents in our sample have shopped online in the past two years, and the average annual spending on online shopping is around 447 CNY for the full sample. The average annual spending on online shopping is 2,651 CNY for the online shoppers, which on average accounts for 10% of the households' consumption expenditure per capita (26,785 CNY, the 4th panel). Table 3.2 provides a breakdown of online shopping behaviours by some socio-demographic variables. It shows that, in general, respondents who are female, younger, have a higher education level, a higher family income, and a higher family wealth are all more likely to shop online and spend more on online shopping.

3.6.1.2 Financial literacy

3.6.1.2.1 Objective financial literacy

- Basic financial knowledge questions

Table 3.3 reports the distribution of the answers to each basic financial knowledge question. Less than 56% of the respondents in our sample respond correctly to the question of the annual interest rate of a one-year fixed deposit. This percentage is surprisingly low, given that China is well-known for its high saving rate. The 2nd and 3rd basic questions in the CFPS are identical and hence comparable to those in the OECD (OECD, 2016) and the DHS questionnaires.[53] Compared with the OECD population, urban Chinese residents show a rather low level of basic financial knowledge. For example, 65% of the respondents across the 17 OECD economies (OECD, 2016), 79% of Hong Kong respondents (OECD, 2016) and 90.8% of Dutch respondents (van Rooij, Lusardi and Alessie, 2011) correctly calculated the simple interest. Yet, only 44.58% of the respondents in our sample answered the questions correctly. This figure puts China within the lowest third of the 30 countries sampled in the OECD (2016) report.[54] The proportion is even lower for rural CFPS respondents.[55] While 37% of respondents across 17 OECD countries were able to answer both questions 2 and 3 correctly (OECD, 2016), the ratio is only 26.33% in our urban sample.

The 4th and 5th basic questions in the CFPS are almost identical to those in the DHS survey. While 82.6% of the DHS sample understood the concept of inflation and 8.5% answered "DK", less than 60% of our sample answered the inflation question correctly, and around 21% declared "DK". The gap is narrower for the question on the time value of money; 72.3% of the DHS sample answered correctly, and only 4.3% declared "DK", while almost 69% of our

[53] Note that mainland China is not included in the OECD (2016) report. Yet, the report does include data for Hong Kong residents, which allows us to compare our sample with a cohort from similar backgrounds.

[54] The low percentage for the CFPS urban residents may be due to the fact that we treat the "DK" responses as incorrect answers. If we treated them as missing, the percentage of correct answer would increase to 56.7%, still much lower than the OECD average (65%).

[55] Among the 91 rural respondents remaining in the dataset before the step when we restrict the sample to urban respondents, only 35.25% calculate the simple interest correctly and 34.53% answered "DK". Even considering all rural respondents who answered financial knowledge questions in the raw data (191 in total) who answered the financial knowledge questions, 33.51% answered correctly and 41.88% answered "DK".

sample answered correctly, but a rather high proportion (15.95%) of our sample did not know the answer.

Finally, less than 15% of the respondents in our sample answered all five basic questions correctly, and 10.72% of them failed to give correct answers to any of the five basic financial knowledge questions. The performance is much worse compared with the DHS sample in which 40.2% correctly answered all the five questions, and only 2.3% failed to answer any of the questions correctly. In summary, respondents in our sample on average correctly answered 2.72 out of the 5 basic questions (see the second panel in Table 3.1). This is much lower than the average score for the DHS sample of van Rooij, Lusardi and Alessie (2011), which is 3.94/5.[56]

- Advanced financial knowledge questions

Table 3.4 summarizes the answers to the advanced questions.[57] Compared with the basic questions, the percentage of respondents who correctly answered each advanced question is much lower, and a sizable proportion of them answered "DK".

The question on which the respondents scored the highest was that on the risk-return nexus (Question 6): 86.40% of respondents provided the correct answer. Surprisingly, while 68.73% of our sample were aware of the highest risk of investing in stocks compared with bank savings, national debts and capital funds (Question 9), only 34% of respondents understood the benefit of diversification provided by mutual funds (Question 7).[58] It is also surprising that only slightly more than 30% of respondents knew that the People's Bank of China is the central

[56] The sample size of van Rooij, Lusardi and Alessie (2011)'s DHS sample is 1,580.

[57] Some of the advanced questions in the CFPS, i.e. questions 7, 9, 10, 11, 13, are identical to those asked in the DHS questionnaire, and hence comparable to the DHS sample of van Rooij, Lusardi and Alessie (2011).

[58] A similar disparity is found in the DHS sample of van Rooij, Lusardi and Alessie (2011). While 68.5% of respondents in their sample were aware that stocks display the highest fluctuations over time, only 48% were aware that a stock mutual fund provides safer returns than a company stock.

bank of China (Question 8). Less than 16% of respondents knew that the ownership of a share of stock means they are the stockholder of the company, and 42% of respondents answered "DK" to the relevant question (Question 10). The percentage of correct answers for Question 10 is much lower than that in the DHS sample, where 62.2% of the respondents understood the meaning of stock holding, and only 11% answers "DK" (van Rooij, Lusardi and Alessie, 2011). Even fewer respondents (11.3%) in our sample understood how the capital funds work, and more than half of them did not know the answer (Question 11), in contrast to the 66.7% correct rate and the 21.7% "DK" responses for the DHS sample (van Rooij, Lusardi and Alessie, 2011). Moreover, less than 30% of our sample understood that the bank's financial products can also suffer from loss (Question 12). Finally, less than 30% of respondents knew the function of the stock market (Question 13), while the percentage for the DHS sample was as high as 67% (van Rooij, Lusardi and Alessie, 2011).

Panel B of Table 3.4 shows that only 1.05% of the respondents were able to answer all the advanced financial knowledge questions correctly. To summarize, the average score for the advanced financial knowledge questions is 2.95/8 (see Table 3.1).

Overall, less than 20% of the respondents answered more than 75% of the 13 financial knowledge questions correctly, a criterion for being financially literate according to the OECD reports.

Next, we evaluate the extent of heterogeneity in financial knowledge across socio-demographic groups (see Table 3.2). We observe a significant decline, though not huge in magnitude, in basic and advanced financial knowledge as people age. This may be due to the contrasting effects of financial literacy and ability declining with age, and of financial knowledge accumulating through experience (Chen and Volpe, 1998; Alessie, Van Rooij and Lusardi, 2011). In line with the literature (e.g. Bucher-Koenen *et al.*, 2017), a gender gap exists,

and women consistently score lower than men in both basic and advanced questions. Additionally, consistent with various studies (e.g., see, Lusardi and Mitchell, 2014; Atkinson and Messy, 2016), we find both basic and advanced financial knowledge increases with educational attainment. However, even for those with at least a college degree, the average basic and advanced financial knowledge scores are respectively 3.73/5 and 4.44/8. We also find that people who have higher family income and wealth perform better in the financial knowledge questions.[59]

We also present the distribution of the FA indices and PRIDIT indices across gender, age groups and educational attainment, to confirm the validity of these indices (see Appendix 3.4). The FA and PRIDIT indices demonstrate similar patterns as the summary scores, i.e. a large gender difference (Panel A), a downward trend as age increases (Panel B), and upwards trend as educational achievement increases (Panel C).

3.6.1.2.2 Subjective financial literacy

The first two rows in Table 3.5 report the distribution of self-perceived level of financial knowledge. Around 40% of the respondents in our sample assess their level of financial knowledge as average. More than half of the respondents assess their financial knowledge as below average. Only around 6% assess their financial knowledge as above average. The self-perceived score of financial knowledge is positively correlated with the overall financial test score, with a Pearson's pairwise correlation of 0.319 (significant at the 5% level).

[59] To check the statistical significance of the trend, we group the number of correct answers to the thirteen financial knowledge questions in five quintiles for basic and advanced financial knowledge, respectively. Pearson Chi-square test are conducted to test the relationship between financial literacy quintiles and age, gender, income, financial and non-financial wealth and education. The p-values of the Pearson Chi-square test are zero for both basic and advanced financial knowledge paired with all the six variables, implying statistically significant trends. Results are not reported for brevity, but are available upon request.

We then plot individuals' online shopping decisions across the objective and subjective levels of financial knowledge in Figure 3.3. We observe that individuals' participation in and expenditure on online shopping increase strongly with both actual and perceived level of financial knowledge. These positive unconditional relationships between financial literacy and online shopping provide preliminary evidence in support of our hypotheses.

3.6.1.2.3 Misperception of financial literacy

The third row in Table 3.5 reports the mean of the financial knowledge test score for each level of subjective financial literacy. In general, the average test score is higher for the respondents who perceive higher levels of financial literacy.

We then summarize the distribution of the overall financial knowledge test score in Table 3.6. The median number of questions that the respondents answer correctly is 6. We consider the test scores that are no more than 1 (the 10% cut-off point of the distribution) to be "Way below average", and those above 10 (the 90% cut-off point of the distribution) to be "Way above average". Additionally, answering 2 or 3 questions (the 25% cut-off point of the distribution) correctly are considered to be "Below average", answering 4 to 7 questions correctly means the respondent is about average, and answering 8, 9 or 10 questions correctly is considered as "Above average". By a rough comparison between this distribution (in Table 3.6) with the average test score corresponding to each level of subjective financial literacy (Row 3, Table 3.5), we find a possible systematic misperception of financial literacy. More specifically, those who report being about or below average tend to underestimate their financial literacy (3.88 vs 1, 5.398 vs 3, 6.526 vs 6), and those who report being above average tend to overestimate their financial literacy.

3.6.1.3 Control variables

The third to fifth panels of Table 3.1 report the statistics for the individual-, household- and community-level control variables. Comparing the statistics for online shoppers and non-online shoppers, we find that the online shoppers, on average, are more likely to be female, younger, employed or self-employed (including engaging in agricultural working), less risk-averse, have higher educational attainment, have not experienced health shocks, communicate more frequently online, hold a more sophisticated attitude towards spending and saving. The online shoppers are also more likely to live in a larger household with more dependent children and fewer dependent elderly, earn a higher income, and hold more financial and non-financial wealth.[60]

The pairwise correlations among the variables are reported in Appendix 3.5. The correlations are generally low (usually lower than 0.3) and we do not expect multicollinearity to be problematic.

3.6.2 Main results and discussion

3.6.2.1 Financial literacy and online-shopping participation

As laid out in the methodology section, we seek to examine the associations between the level of financial literacy and the likelihood for a consumer to shop online. The dependent variable is a binary variable indicating whether a consumer had shopped online in 2013. The

[60] In this paragraph, we only mention the variables which are significantly different (at the 1% level) between online shoppers and non-online shoppers. Marital status and computer ownership do not differ significantly between online shoppers and non-online shoppers. One potential explanation for the insignificant role of computer ownership is that most of online purchases are carried out on mobile devices (CNNIC, 2015).

main independent variables are either of the normalized financial literacy measures, i.e., summary scores, FA indices, PRIDIT indices and subjective scores.

We start with three Probit models to estimate the relationship between financial literacy and whether the respondent carries out online shopping. Table 3.7 is a short version of the results table, which reports the estimated marginal effects for each set of measures of financial knowledge (FK, for brevity in the sections hereafter) in separate panels. The first three columns of Table 3.7 report Probit estimates for the specifications which include (i) financial literacy only, (ii) financial literacy and individual-level control variables, and (iii) financial literacy and the full set of control variables, respectively. Location dummies indicating whether the respondent resides in the western or central area are controlled for in all the specifications. Bootstrap estimates of standard errors are computed with 200 replications in each case.[61] The marginal effects of all the explanatory variables are reported in Appendix 3.8.

Compared with prior research on financial literacy and consumption, which uses an overall score for financial literacy (Dinkova, Kalwij and Alessie, 2018; Jappelli and Padula, 2017), the separation of basic and advanced financial literacy allows us to analyse the effect of each of them. According to the results in Table 3.7, basic and advanced financial literacy affect online shopping participation differently. While basic FK attracts an insignificant marginal effect when controlling for the full set of control variables, the marginal effect associated with advanced FK is significant and positive at the 1% level regardless of the way it is measured. This implies that, though basic FK on interest, inflation and risk diversification may affect consumption in general, advanced FK matters more for consumers' online shopping decisions. As shown in Column (3), the marginal effects of the normalized indicators for advanced FK in the full baseline specification are all around 0.15. More specifically, one additional correct

[61] We also estimate the specifications using robust estimators clustering at the community level. The results are similar to the bootstrap estimators.

answer to the advanced FK questions is associated with a 1.884 pp higher probability of shopping online, ceteris paribus (Panel 1, Column (3)).[62] As for the subjective indicator, respondents who report one higher point on their perceived level of FK show a 1.278 pp higher likelihood to shop online (see Column (3), Panel 4). The findings from the baseline specification support our first hypothesis, according to which respondents with higher financial literacy, especially higher objective financial literacy, are more likely to shop online.

Next, we exploit the IV approach to tackle the potential endogeneity issue of our financial literacy variables, using mother's duration of education and the community level of financial literacy as instrument variables for one's financial literacy. The results for the validity tests of the IVs are reported and discussed in Table A.3.7.1.

Columns (4) and (5) of Table 3.7 report the IV estimates for the *Com_5* and *Com_10* samples, respectively. The marginal effects associated with basic FK are insignificant no matter what measure we use, while the marginal effects associated with advanced FK remain positive and statistically significant, regardless of the measure or the sample used. The marginal effects associated with the subjective indicator becomes less significant in the *Com_5* sample (10% level) and insignificant in the *Com_10* sample. In summary, the IV estimates reveal a significantly strong relationship between advanced financial literacy and online shopping participation, and show weak evidence that respondents with higher subjective financial literacy are more likely to shop online. Hence, our first hypothesis is further verified by the IV approach.

[62] One additional correct answer to the advanced financial knowledge test score corresponds to a 0.125 (= 1/8, where 8 is the range of the original summary score for advanced financial knowledge) increase in the normalized score for advanced financial knowledge, which in turn corresponds to a 0.1507 * 0.125 * 100 = 1.884 pp higher chance to participate in online shopping. Similarly, one higher point of perceived level of financial literacy is equivalent to a 0.2 (=1/5, where 5 is the range of the original score for subjective financial knowledge) increase in the normalized score for subjective financial knowledge. Therefore, one higher point of perceived financial literacy corresponds to a 0.0639 * 0.2 * 100 = 1.278 pp higher probability to shop online.

As discussed in Section 3.5.2, basic FK refers to the basic concepts that are vital to saving and consumption decisions, while advanced FK focuses on more complicated concepts on various financial instruments and investment. In light of prior studies that document a positive correlation between financial literacy and carrying out online banking and digital payments (Cohen and Nelson, 2011; Grohmann, Klühs and Menkhoff, 2018), the mechanism may go in such a way that advanced FK increases individuals' demand for bank cards and digital payment accounts, through which they can carry out investment and wealth management. The holding of these payment tools, in turn, fulfils their demand for shopping online, which is more effective and efficient compared to offline shopping. This is consistent with prior studies that document a positive correlation between financial literacy and sophisticated consumption management (Agarwalla *et al.*, 2015).

In both the Probit and IV estimations, the marginal effects associated with subjective FK are always considerably lower than the marginal effects associated with advanced FK, or even insignificant, regardless of how they are measured. This implies that it may be what the respondents actually know, rather than what the respondents perceive they know, that motivates them to shop online. To verify this inference, we also estimate specifications which include basic, advanced, and subjective FK simultaneously (see Appendix 3.7). In all the cases, advanced FK is significant at the 1% level, and subjective FK is always insignificant. One potential reason for the insignificant role of self-perceived FK is that people may overstate or understate their true levels of financial literacy, so that the subjective financial literacy may be meaningless. As a matter of fact, we do find that in the Chinese context, respondents systematically misperceive their levels of financial literacy.[63]

[63] See discussion in Section 3.6.1.2.3.

The Probit and IV estimates of the marginal effects for a full set of explanatory variables are reported in Appendix 3.8. The sign, magnitude and significance of the IV estimates are in general unchanged compared with the Probit estimates. In line with previous literature (e.g. Pérez-Hernández and Sánchez-Mangas, 2011; Räsänen and Koiranen, 2016; Kuoppamäki, Taipale and Wilska, 2017; Dai *et al.,* 2019), our estimates reveal that, in general, consumers who are female, more educated, aged below 36, who communicate online more frequently and with higher family income and wealth are more likely to shop online. It is worth mentioning that the marginal effects of these factors on the propensity to shop online are much lower than the marginal effect of advanced financial literacy. For example, though we find people with higher levels of education are more likely to shop online, the magnitude of the association is only one-eighth of that between advanced financial knowledge and the propensity to shop online. Similarly, the marginal effect of household income, another vital determinant of consumption behaviours, is slightly higher than that of educational attainment, but only counts for one-sixth of the marginal effect of advanced financial knowledge.

3.6.2.2 Financial literacy and online-shopping expenditure

The Tobit estimates for the expenditure on online shopping are reported in columns (1) to (3) of Table 3.8. The dependent variable is the logarithm of annual individual online shopping expenditure.[64] Again, when the full set of control variables are included, the marginal effect of basic FK becomes insignificant. The coefficients associated with advanced FK, regardless of the way it is measured, are always positive and significant at the 1% level even after controlling for a full set of covariates (see Column 3, Panel 1). More specifically, one additional correct

[64] We also consider scaling the expenditure on online shopping by annual personal income and household net income, respectively. The findings are in line with what we find from the baseline estimates in Section 3.6.2.2.

answer to the advanced FK questions is associated with a 1.74% higher expenditure on online shopping, ceteris paribus.[65] However, though the marginal effects of self-perceived financial knowledge are significant when individual level control variables are included, it becomes insignificant in the model that further incorporate the household level control variables (Column 3, Panel 4). Therefore, the baseline estimates support our Hypothesis 2, according to which individuals with higher levels of advanced financial literacy tend to spend more on online shopping.

Columns (4) and (5) in Table 3.8 report the IV estimates for the *Com_5* and *Com_10* samples, respectively. The results for the validity of the instruments used in the IV approach are reported in Table A3.6.2. Again, basic FK hardly has any effect on the expenditure on online shopping, while advanced FK always shows a significant and positive coefficient. The IV estimates for the subjective score are significant at the 5% level for the *Com_5* sample and insignificant for the *Com_10* sample. Therefore, the IV estimates provide some further support to Hypothesis 2. Similar to the case of the extensive margin, the Tobit and IV estimates reflect a higher predictive power of objective financial literacy compared to self-perceived FK in terms of online expenditure.

The Tobit and IV estimates of the marginal effects for a full set of explanatory variables are reported in Appendix 3.9. The sign, magnitude and significance of the IV estimates barely change compared with the Tobit estimates. In summary, consumers who are female, spend more years in education, are aged below 36 years, are risk seeking, communicate online more frequently, earn higher family income, hold more wealth and live in western area (compared with those living in Coastal area) spend more on online shopping. Again, the association

[65] One additional correct answer in the advanced financial knowledge test corresponds to a 0.125 (= 1/8) increase in the normalized summary score for advanced financial knowledge. Since the dependent variable is the logarithm transformed expenditure on online shopping, one additional correct answer therefore leads to a 1.74% (≈ exp(0.1377 * 0.125) -1) higher online expenditure.

between educational attainment and online shopping expenditure is much lower compared to financial literacy, with the estimates about one-eighths of the estimates for advanced financial literacy.

Overall, the analysis in this section reveals that financially literate individuals are more likely to shop online and tend to spend more on online shopping. Therefore, Hypotheses 1 and 2 are verified. However, this relationship is primarily driven by advanced financial literacy: We observe no evidence that basic financial literacy or self-perceived financial literacy enhance online shopping. The results also endorse our approach of treating basic and advanced financial literacy separately, as we find no significant role of basic financial literacy in determining one's online shopping decisions.

The relationship between advanced financial literacy and online shopping may operate via two mechanisms. First, in China, most online shopping is finalized by online payment using bank cards or digital payment tools, rather than by cash-on-delivery. Financial literacy increases individuals' demand for bank accounts and digital payment accounts (Königsheim, Lukas and Nöth, 2017; Grohmann, Klühs and Menkhoff, 2018), which fulfils the prerequisite of online shopping. Second, as documented in prior studies, financial literate individuals are more likely to monitor income and expenditure, set up long-run financial plans and consider the affordability of goods before purchasing (Agarwalla *et al.*, 2015). Therefore, financially literate individuals may prefer online shopping to offline shopping, since the former is more time- and cost-saving (Perea y Monsuwé, Dellaert and De Ruyter, 2004).

3.7 Robustness Tests

3.7.1 Control for cognitive abilities

One topic of debate is whether financial literacy measures actual financial knowledge or simply ability and cognition (e.g. van Rooij, Lusardi and Alessie, 2011). Evidence suggests positive relationships between financial literacy and cognitive ability and between cognitive ability and financial behaviours (e.g. Lusardi, Mitchell and Curto, 2010; van Rooij, Lusardi and Alessie, 2011; Gustman, Steinmeier and Tabatabai, 2012). There are also studies arguing that, for example, the relationship between education and financial participation is mediated by cognitive ability rather than financial literacy (Cole, Paulson and Shastry, 2012). Therefore, to disentangle the effects of knowledge from talents and abilities, we additionally introduce cognitive ability into our specifications.

The CFPS 2014 provides information on the respondents' performance in word and math tests. The scores for the word test and math test count the number of correct answers and range from 0 to 34 and 0 to 24, respectively. We find strong correlations between cognitive abilities and financial knowledge; the pairwise correlations between FK scores and the scores for cognitive abilities are at least 0.4 and significant at the 1% level.

Table 3.9 reports the simple estimates in Panel A and the IV estimates in Panel B. For brevity, we do not report the marginal effects of the other covariates.

Focusing on the participation equation, we find that the Probit estimates and the IV estimates are slightly lower than the baseline estimates reported in Table 3.7. More specifically, we find that one more correct answer to the advanced FK questions is associated with a 1.809 pp (= 0.1447/8) higher probability of shopping online (Column 1, Panel A), only slightly lower than the corresponding baseline results. Also, one higher point of self-perceived financial literacy is associated with a 1.294 pp (= 0.0647/5) higher probability to shop online. In none of the cases

is basic financial literacy significant. The positive relationships between advanced/subjective financial literacy and online shopping participation are also supported by the IV estimates. Hence, we provide further evidence for Hypothesis 1 that financial literacy enhances participation in online shopping.

For the expenditure equation, the estimated marginal effect of financial literacy after controlling for cognitive ability is also slightly lower than corresponding baseline estimates. More specifically, one more correct answer to the advanced FK is associated with 1.60% higher expenditure on online shopping (Column 5, Panel A).[66] The significantly positive relationship between advanced/subjective financial literacy and online shopping expenditure is also proved by the IV estimates. Therefore, we provide further evidence for Hypothesis 2.

One point worth noting is that cognitive ability (the score for the word test) is only significant in the extensive margin specifications on subjective financial literacy (Column 4). The marginal effects of the word test score are also significant in all the intensive margin specifications regardless the way we measure financial literacy. However, the magnitude of the effect of cognitive ability is much lower than that of financial literacy. In comparison, when we exclude financial literacy from the specification, the score for the word test is always significantly correlated with online shopping participation and expenditure even after controlling for the full set of control variables, and the magnitude of its marginal effects is much higher. Given the strong pairwise correlation between financial knowledge and cognitive ability, the significant association between cognitive ability and online shopping behaviours may attribute to the significant association between financial literacy and online shopping behaviours. In other words, though cognitive ability enhances online shopping, the effect is mediated by financial literacy.

[66] 1.60%≈ exp(0.1269 * 0.125) -1). See Footnote 65 for detailed explanation for the calculation.

3.7.2 Alternative estimation methods

To check if our findings are sensitive to the selection of estimation techniques, we additionally estimate the relationship between financial literacy and whether one shops online using a linear probability model (LPM). Considering the abnormally large marginal effects of financial literacy in the expenditure equation, we also estimate the expenditure equation using a Tobit model, since around 83% of our sample never shopped online. The estimates for financial literacy from the LPM and Tobit models are reported in Table 3.10. For brevity, we do not report the marginal effects of the other covariates.

As expected, the LPM estimates for advanced financial literacy in the participation equation is significantly positive, regardless of the measures being used, but larger in magnitude compared with the baseline results. More specifically, the LPM estimate indicates that one more correct answer to the advanced FK questions is associated with a 2.60 pp (= 0.2079/8) higher chance to shop online (Column 1, Panel A). However, we find no evidence that subjective financial literacy prompts individuals to shop online. These results are further supported by the IV estimates reported in Columns (1) to (4) in Panel B. Therefore, our Hypothesis 1 is further supported by using alternative estimation methods.

The OLS estimates are reported in Columns (5) to (8) in panel A. We find significant associations between advanced financial literacy and online shopping expenditure, but no significant associations are observed between basic/subjective financial literacy and the online shopping expenditure. Specifically, one more correct answer to the advanced FK questions is associated with 23.45% higher expenditure on online shopping.[67] This percentage is much

[67] The percentage of 23.45% is calculated by exp(1.6854 * 0.125) -1). See Footnote 652 for detailed explanation for the calculation.

higher than the baseline estimates, probably because that the majority of the sample had never shopped online. The positive relationship between advanced financial literacy and online shopping is further supported by the IV estimates (Column 5 to 8, Panel B). Therefore, our results are robust to alternative estimation methods, and Hypothesis 2 is verified.

3.7.3 Control for peer effects

Peer effect is a nonnegligible factor in consumers' online shopping decisions (Smith, Menon and Sivakumar, 2005; Thaichon, 2017). Individuals may make their purchasing decisions considering the decisions and feedback of other individuals that are close in some dimension (Moretti, 2011). Therefore, to capture potential peer effects in online shopping decisions, we augment the Probit model with a variable *D_peer*. *D_peer* measures the proportion of online shoppers (excluding the respondent him-/herself) among a group of people who are of similar age (+/- 10 years) and residing in the same community. We augment the OLS model with a variable *Lg_exp_peer*, which measures the average expenditure on online shopping of the other respondents who are of similar age (+/- 10 years) and reside in the same community. To ensure the two variables capturing peer effects are valid, we focus on the communities with at least 30 respondents being interviewed in CFPS 2014. This operation leaves us with a sample of 995 observations.

Table 3.11 reports the estimates after controlling for peer effects. Probit estimates for the participation equation after controlling for peer effects are reported in Columns (1) to (4) in Panel A. We find that peer effects may not influence whether the individual shop online, since the marginal effects of *D_peer* are never significant in any of the cases. Meanwhile, the marginal effects of advanced financial literacy are always significant in the participation equations, and are slightly higher than the baseline results. Subjective financial literacy is also

significantly positive at the 5% level. We also report the IV estimates for the participation equations in the first four columns in Panel B. In this case, *D_peer* becomes weakly significant in the three models using objective measures of financial literacy (Columns 1 to 3, Panel B), and becomes significantly positive at the 5% level in the model using the subjective measure of financial literacy (Column 4, Panel B). Meanwhile, advanced and subjective financial literacy are still significant, and the magnitude of their marginal effects is higher than the peer effect. These results imply that peer effects itself cannot fully explain people's participation in online shopping. Even after we control for peer effects, we still find that individuals with higher financial literacy are more likely to shop online. Hence, Hypothesis 1 is verified.

Tobit estimates for the expenditure equation after controlling for peer effects are reported in Columns 5 to 8 in Panel A. We find that the marginal effects of *Lg_exp_peer* are significantly positive in all the specifications, implying part of the expenditure in online shopping can be explained by peer effects. Meanwhile, the marginal effects of advanced financial literacy are still significantly positive at the 1% level regardless of the measure we use (Column 5 to 7, Panel A). However, no significant effects are observed for basic and subjective financial literacy. According to the estimates using the IV approach, the marginal effects of *Lg_exp_peer* are still significant in all the specifications. Nevertheless, the positive relationships between advanced financial literacy and online shopping expenditure still hold (Columns 5 to 8, Panel B). It is also worth mentioning that the marginal effects of advanced financial literacy are always much larger than the peer effects. Therefore, we conclude that after controlling for peer effects, financial literacy is still positively related to expenditure on online shopping, which is consistent with our Hypothesis 2.

Overall, peer effects indeed affect individuals' online shopping decisions, but only marginally. After controlling for peer effects, financial literacy is still positively associated with the participation in and expenditure on online shopping.

3.8 Extensions

In this section, we perform several extensions to study how socio-demographics and financial literacy interact to affect households' demand for online shopping.

3.8.1 Heterogenous effects: Online shopping across age groups

The association between financial literacy and online shopping could depend on the consumers' position in the life cycle. Servon and Kaestner (2008) assert that, in the context of online banking, appropriate technological knowledge and skills have to be in place for financial education to effectively prompt the use of online banking of urban LMI individuals. Since it is more difficult for older individuals to learn technological skills such as computer skills (Czaja *et al.*, 1989; Hernández, Jiménez and José Martín, 2011), financial literacy may affect the economic decisions of the young and the old in different ways. Additionally, financially literate individuals are more likely to plan for retirement and save more upon retirement (e.g. van Rooij, Lusardi and Alessie, 2012). Therefore, it is very likely that those who are more financially literate and are close to retirement are more concerned about retirement saving, and, as a result, are not interested in online shopping.

In light of these findings, we expect the link between financial literacy and online shopping to be weaker for older respondents. To test whether this is the case, we extend the full baseline model for the participation in online shopping by including the interaction terms between financial literacy and four age groups, i.e., 18-35, 36-45, 46-60 and above 60.

Table 3.12 reports the Probit estimates for the marginal effects associated with financial literacy for different age groups.[68] The other control variables are the same as in the baseline

[68] Since the simple baseline estimates and the IV estimates do not differ much, in Section 3.8, we only report the baseline estimates for the specifications with interactions.

models, but are not reported for brevity. Similar to the baseline results, in none of the cases is basic FK significant. However, advanced FK is always significantly positive for the respondents below 60, and the marginal effects associated with advanced FK decrease in magnitude as age increases (Columns 2, 4 and 6). These results prove our hypothesis, that the relationship between financial literacy and online shopping participation gets weaker as age increases. They are in line with the findings of Servon and Kaestner (2008), who find that financial literacy can be more effective with adequate technology skills.

Next, we explore whether the relationship between financial literacy and online shopping expenditure differs between age groups as well. Table 3.13 reports the Tobit estimates of the marginal effects associated with financial literacy for different age groups. For all the age groups, basic FK and subjective FK shows no significant association with online shopping expenditure. The marginal effects of advanced FK are always significantly positive; they are the highest for those between 36 and 45, second highest for those between 20 and 35, third highest for those between 46 and 60, and the lowest for whose over 60 (Columns 2, 4 and 6).

To conclude, our analysis in this section shows that financial literacy enhances online shopping participation and expenditure for young and middle-aged consumers, but the magnitude of the effect declines with age. Financial literacy does not prompt the engagement of elderly consumers (60+) in online shopping, probably because of the difficulties they encounter in the process, which impedes the effectiveness of financial literacy. Financial literacy does not enhance the online expenditure of elder consumers as much as the younger generations, presumably because repsondents over 45 restrict their consumption to prepare for retirement or to slow down the loss of wealth after retirement.

3.8.2 Heterogenous effects: Online shopping across income/wealth gradients

Lyons *et al.* (2006) and Servon and Kaestner (2008) assert that it is more difficult for financial literacy to prompt sophisticated financial behaviours for LMI individuals. Therefore, we also investigate whether the relationship between financial literacy and online shopping differs across income gradients. To this end, we initially incorporate in the full baseline model for the participation in online shopping interaction terms between financial literacy and dummy variables indicating whether the respondent belongs to low, moderate, and high-income groups.[69]

Table 3.14 reports the marginal effects associated with financial literacy for different income groups estimated by Probit models. We can see that, with the exception of the summary score and the FA index for basic FK, which are marginally significant for the low-income group, basic FK and subjective FK are hardly significant in any of the specifications. By contrast, advanced FK is significant and positive for the moderate- and high-income groups (Columns 2, 4, and 6).

Table 3.15 reports the marginal effects associated with financial literacy for different financial wealth groups estimated by Probit models. Similar to the case for income groups, the summary score and PRIDIT index for basic FK demonstrate a weak positive relationship with online shopping participation for the low-wealth cohort. So does the subjective FK, while advanced FK is positively correlated with online shopping participation, especially for the high-wealth cohort.

Next, we explore whether the relationship between financial literacy and online shopping expenditure differs across income and wealth gradients as well. Table 3.16 reports the Tobit

[69] These groups are created by dividing the income distribution evenly into three parts. In this section we also obtained differentiating households by financial wealth.

estimates of the marginal effects associated with different types of financial literacy for different income groups. In general, there is no significant association between FK and online expenditure for poor consumers. However, we find advanced FK is significant for moderate- or high-income consumers, and the magnitude is higher for the latter.

Table 3.17 reports the Tobit estimates of the marginal effects associated with different types of financial literacy for different financial wealth groups. We observe weak evidence (significant at the 10% level) that respondents with higher level of basic financial literacy are more likely to shop online, but such relationship only holds for the lowest financial wealth group. We also find that advanced financial literacy is significantly positive (at the 5% and 1% level) for all the financial wealth groups, and is the highest for the high-wealth group.

Overall, our analysis in this section shows that financial literacy enhances online shopping participation and expenditure for consumers with moderate- or high-income, and the effect is stronger for the high-income cohort. Financial literacy does not affect online shopping participation and expenditure for the low-income consumers, probably because these consumers are less exposed to new technology and therefore lack the chance to learn and obtain digital skills (Scheerder, van Deursen and van Dijk, 2017), which impair the effectiveness of financial literacy in enhancing online shopping.

3.9 **Conclusion**

The worldwide shortfalls in financial literacy have raised grave concerns among policymakers, especially in the wake of the global financial crisis and population ageing problem. Efforts are underway to identify the consequences of low financial literacy and to explore efficient ways to improve financial literacy through financial education.

Using data from the 2014 wave of the CFPS, this chapter investigates the effect of financial literacy on online shopping decisions of Chinese urban residents. Objective and subjective measures of financial literacy are constructed. The empirical results reveal that individuals with higher advanced or subjective financial literacy are more likely to shop online and spend more online. However, contrary to some previous research that finds subjective financial literacy to be at least as important as objective financial literacy in determining specific sophisticated financial behaviours (e.g. Allgood and Walstad, 2016; Anderson, Baker and Robinson, 2017), we find the objective measure to be more predictive in the context of online shopping. One potential explanation is that the misperception of financial literacy in our sample makes the subjective measure of financial literacy less reliable. By looking at subsamples, we find that the strength of the link between financial literacy and online shopping basically declines with age, and is strongest for wealthier consumers.

Our empirical results have significant policy implications. Policymakers should not only be aware of the fact that financial literacy can enhance online shopping in general, but also that financial education is more effective at the early stage of life. Financial education programmes targeted at younger people would, therefore, be more effective. Also, financial literacy may only work for individuals with moderate- or high-income. Other issues have to be addressed before financial literacy take effect for the poor consumers, such as technology training.

Furthermore, it is hard to tell whether online shopping improves wellbeing or not. On the one hand, online shopping can efficiently reduce the time spent on shopping. The time saved by engaging in online shopping can be devoted to wellbeing-enhancing activities, such as exercising or socializing (Castellacci and Tveito, 2018). On the other hand, online shopping may be detrimental to wellbeing if people get addicted to it, spend more than before and become indebted (Wang, Malthouse and Krishnamurthi, 2015). To enjoy the former consequence and avoid the latter, financial education should also consider content that can protect the online shoppers from overspending and indebtedness.

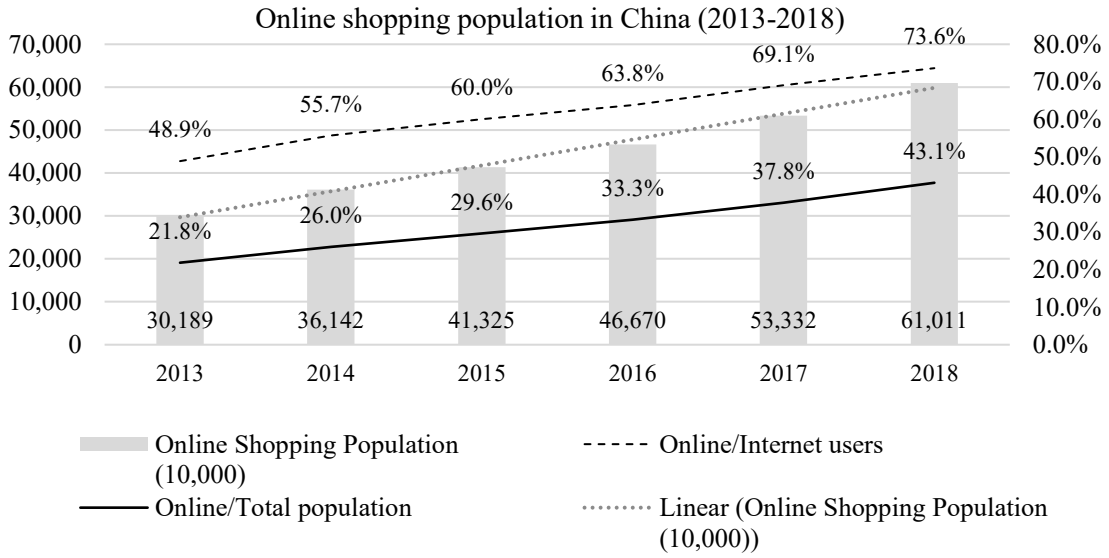
Several limitations exist. First, due to the limitation of the data, we cannot verify whether the relationship between financial literacy and online shopping decisions operates through the take-up of bank cards or digital payment tools. Second, our analysis is restricted to the respondents who answered the financial knowledge questions in the 2014 wave of the CFPS, leading to a loss of observations. Therefore, our final sample may not be representative of the whole population. Future research can further investigate whether and to what extent are the online shoppers indebted, and how financial literacy can work to alleviate any negative consequences.

Figures and Tables



Source: CNNIC (2016); NBSC (2019)

Figure 3.1 Online Retail Sales in China: Volume and Growth Rate (2006-2018)



Source: NBSC (China Statistical Yearbook), CNNIC E-commerce Market Report (2015).

Figure 3.2 Online shopping population in China (2013-2018)

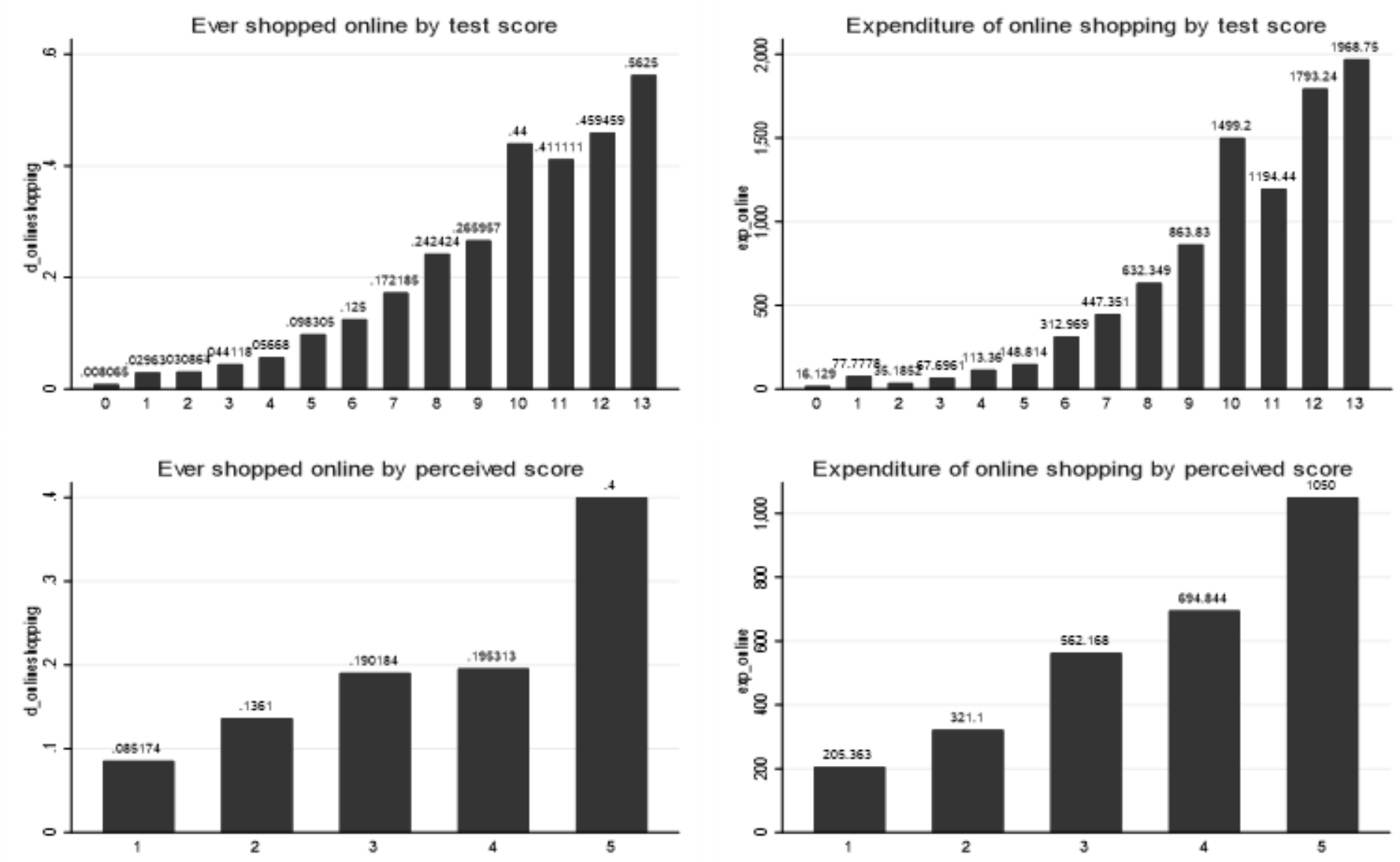


Figure 3.3 Online shopping decisions by financial knowledge

Table 3.1 Summary statistics of key variables (N = 2,183)

Variable	(1) Mean	(2) S.E.	(3) Min	(4) Max	(5) D_online=1 N=368	(6) D_online=0 N=1,815
Dependent Variable						
Exp _online	447	1471	0	9000	2,651	0
D _online	0.17	0.37	0	1	1	0
Financial Knowledge						
Original Score						
Fk_score	5.93	2.97	0	13	8.16	5.48
Fk_score_b	2.83	1.55	0	5	3.68	2.66
Fk_score_a	3.10	1.83	0	8	4.48	2.82
Subj_fk	2.41	0.79	1	5	2.63	2.37
Normalized Score						
Fk_score_b	0.46	0.23	0	1	0.63	0.42
Fk_score_a	0.57	0.31	0	1	0.73	0.53
Subj_fk	0.39	0.23	0	1	0.56	0.35
Individual Characteristics						
Male	0.43	0.50	0	1	0.35	0.45
Edu_y	10.09	4.18	0	19	13.51	9.40
Married	0.84	0.37	0	1	0.83	0.84
Age(36-45)	0.22	0.42	0	1	0.39	0.19
Age(45-60)	0.39	0.49	0	1	0.23	0.42
Age(60+)	0.29	0.45	0	1	0.03	0.34
Not-working	0.46	0.50	0	1	0.15	0.52
Risk_Averse	0.67	0.47	0	1	0.40	0.73
Health_Shock	0.34	0.47	0	1	0.24	0.35
Online_Comm	1.43	2.37	0	7	2.62	1.19
Financial attd	3.35	1.01	1	5	3.53	3.31
Household Characteristics						
Family_Size	3.25	1.40	1	13	3.41	3.22
N_depchildren	0.87	0.70	0	5	1.07	0.83
N_depelderly	0.67	0.84	0	3	0.30	0.74
Own_computer	0.01	0.10	0	1	0.00	0.01
Net HH Income	63,575	46,495	500	219,300	85,639	59,101
Fin_asset	75,749	122,331	0	500,000	125,533	65,655
Non_fin_asset	663,701	871,193	1,520	3,803,848	954,939	604,651
PCE_percapita	21,065	18,712	1,590	263,524	26,785	19,882
Community-level						
Facility1	0.06	0.20	-0.67	0.56	0.06	0.06
Facility2	0.05	0.41	-0.95	0.93	0.04	0.05

Notes: Definitions of the variables are reported in Appendix 3.1.

Table 3.2 Key variables by socio-demographics

	Household income			
	Low	Moderate	High	Diff
D_online	0.07	0.12	0.24	*
Exp_online	94.19	259.87	711.89	*
Fk_b	2.07	2.60	3.28	*
Fk_a	2.42	2.82	3.55	*
Subj_fk	2.21	2.38	2.51	*
Obs	387	751	1,045	

	Gender		
	Female	Male	Diff
D_online	0.19	0.14	*
Exp_online	511.93	362.13	*
Fk_b	2.68	3.03	*
Fk_a	2.85	3.43	*
Subj_fk	2.34	2.51	*
Obs	1,235	948	

	Financial asset			
	Low	Moderate	High	Diff
D_online	0.10	0.12	0.24	*
Exp_online	225.11	212.24	698.30	*
Fk_b	2.15	2.62	3.41	*
Fk_a	2.47	2.85	3.65	*
Subj_fk	2.29	2.27	2.55	*
Obs	748	401	1,034	

	Age				
	20-35	36-45	46-60	60+	Diff
D_online	0.62	0.30	0.10	0.02	*
Exp_online	1910.14	669.59	273.32	22.14	*
Fk_b	3.44	3.17	2.74	2.50	*
Fk_a	4.10	3.57	3.01	2.52	*
Subj_fk	2.59	2.38	2.38	2.42	Insig
Obs	210	488	855	630	

	Non-financial asset			
	Low	Moderate	High	Diff
D_online	0.07	0.13	0.23	*
Exp_online	119.73	263.98	686.67	*
Fk_b	2.29	2.66	3.17	*
Fk_a	2.53	2.88	3.47	*
Subj_fk	2.16	2.42	2.53	*
Obs	517	545	1,121	

	Education					
	Illiterate	Primary	Junior	Senior	Collegeabove	Diff
D_online	0.0041	0.02	0.08	0.18	0.46	*
Exp_online	0.83	31.46	88.01	465.62	1457.60	*
Fk_b	1.69	2.12	2.70	3.09	3.76	*
Fk_a	1.62	2.33	2.83	3.32	4.50	*
Subj_fk	2.19	2.38	2.33	2.43	2.66	Insig
Obs	242	267	662	571	441	

Notes: * indicates that the difference in the key variables (online shopping and financial knowledge) across socio-demographic groups is significant at the 1% or 5% significance level. Insig indicates the specific key variable doesn't differ significantly across socio-demographic groups.

Table 3.3 Basic financial knowledge

Panel A: Percentages of answers (N = 2,183)

	Correct	Incorrect	DK
1. Annual interest rate of one-year fixed deposit	55.88	8.92	35.20
2. One-year interest and principal	44.58	34.03	21.39
3. Interest compounding	50.36	26.37	23.27
Both 2 & 3	26.33	31.39	-
4. Purchasing power (inflation)	59.82	19.38	20.80
5. Time value of money	68.94	15.11	15.95

Panel B: Distribution of correct answers (N = 2,183)

Correct	None	1	2	3	4	5
Percentage	10.72	13.06	15.78	21.56	24.15	14.73

Notes: Panel A reports the percentages of respondents who answered correctly, incorrectly and "do not know". Panel B reports the distribution of the number of correct answers.

Table 3.4 Advanced financial knowledge

Panel A: Percentages of answers (N = 2,183)

	Correct	Incorrect	DK
6. In general, investments with high returns have high risks.	86.40	5.98	7.62
7. In general, the risk of investing in one-share stock is smaller than that of investing in stock fund.	33.95	25.36	40.69
8. Which of the following bank is responsible for making and carrying out monetary policies?	32.57	38.09	29.34
9. Which of the following investment has the highest risk?	68.73	12.14	19.13
10. If you purchase stocks of some company, that means?	15.86	42.49	41.65
11. Which one of the following is correct in terms of describing capital fund?	11.30	33.15	55.55
12. Which of the following statement is correct about a bank's financial products?	29.13	35.00	35.87
13. Which of the following statements correctly describes the core function of stock market?	27.42	22.56	50.02

Panel B: Distribution of correct answers (N = 2,183)

Correct	None	1	2	3	4	5	6	7	8
Percentage	6.36	14.27	20.76	21.05	17.08	9.84	6.57	3.01	1.05

Notes: Panel A reports the percentages of respondents who answered correctly, incorrectly and "do not know". Panel B reports the distribution of the number of correct answers.

Table 3.5 Subjective assessment of financial knowledge

Subjective Score (N = 2,183)

Subjective Score	1(way below)	2	3	4	5 (way above)	DK
Percentage	11.97	41.52	39.64	5.23	0.63	1.00
Test score (0-13)	3.880	5.398	6.526	7.551	8.857	-

Table 3.6 Distribution of financial knowledge test score (0 - 13)

FK_Score (N = 2,183)

Percentage	10%(way below)	25%	50%	75%	90%(way above)
FK_Score	1	3	6	8	10

Table 3.7 Online shopping participation

VARIABLES	(1) Probit	(2) Probit	(3) Probit	(4) IV_Com_5	(5) IV_Com_10
Panel 1: Summary score					
Basic FK	0.1532***	0.0342	0.0184	-0.0491	-0.0377
	(0.0274)	(0.0247)	(0.0250)	(0.0545)	(0.0588)
Advanced FK	0.3692***	0.1775***	0.1507***	0.2539***	0.2074**
	(0.0335)	(0.0300)	(0.0299)	(0.0831)	(0.1029)
Panel 2: FA indices					
Basic FK	0.1545***	0.0343	0.0177	-0.0479	-0.0345
	(0.0275)	(0.0248)	(0.0251)	(0.0552)	(0.0593)
Advanced FK	0.3695***	0.1769***	0.1509***	0.2489***	0.2005**
	(0.0336)	(0.0299)	(0.0298)	(0.0808)	(0.1004)
Panel 3: PRIDIT indices					
Basic FK	0.1488***	0.0304	0.0163	-0.0466	-0.0392
	(0.0245)	(0.0230)	(0.0231)	(0.0518)	(0.0562)
Advanced FK	0.3633***	0.1697***	0.1411***	0.2859***	0.2184*
	(0.0358)	(0.0336)	(0.0332)	(0.0962)	(0.1139)
Panel 4: Subjective score					
Subj_FK	0.1941***	0.0836***	0.0639**	0.1758*	0.1096
	(0.0375)	(0.0314)	(0.0315)	(0.0975)	(0.1141)
Individual control	No	Yes	Yes	Yes	Yes
Household control	No	No	Yes	Yes	Yes
Location Fixed Effect	Yes	Yes	Yes	Yes	Yes
Observations	2,183	2,183	2,183	1,771	1,353

Notes: Columns (1) to (3) report Probit estimates of Eq.3.1. Columns (4) and (5) report IV estimates of the same equation for the *Com_5* and *Com_10* samples, respectively. The indicators for financial literacy are normalized. Definitions of variables are displayed in Appendix 3.1. Bootstrap estimates of standard errors are computed with 200 replications and reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 3.8 Online shopping expenditure

VARIABLES	(1) Tobit	(2) Tobit	(3) Tobit	(4) IV_Com_5	(5) IV_Com_10
Panel 1: Summary score					
Basic FK	0.1608***	0.0372	0.0085	-2.5255	-2.2074
	(0.0258)	(0.0252)	(0.0291)	(2.3365)	(2.5500)
Advanced FK	0.4417***	0.1806***	0.1377***	7.7279**	3.5589
	(0.0301)	(0.0296)	(0.0317)	(3.4311)	(4.2803)
Panel 2: FA indices					
Basic FK	0.1613***	0.0376	0.0083	-2.5179	-2.2681
	(0.0258)	(0.0253)	(0.0293)	(2.3828)	(2.5998)
Advanced FK	0.4415***	0.1790***	0.1376***	7.5236**	3.4309
	(0.0300)	(0.0293)	(0.0315)	(3.3479)	(4.1777)
Panel 3: PRIDIT indices					
Basic FK	0.1606***	0.0336	0.0035	-2.3453	-1.9335
	(0.0229)	(0.0239)	(0.0273)	(2.2019)	(2.4182)
Advanced FK	0.4373***	0.1713***	0.1186***	8.6315**	4.4933
	(0.0321)	(0.0327)	(0.0350)	(3.9217)	(4.7953)
Panel 4: Subjective score					
Subj_FK	0.2689***	0.0901***	0.0535	7.6313*	5.7677
	(0.0364)	(0.0292)	(0.0359)	(4.2264)	(4.2513)
Individual control	No	Yes	Yes	Yes	Yes
Household control	No	No	Yes	Yes	Yes
Location Fixed Effect	Yes	Yes	Yes	Yes	Yes
Observations	2,183	2,183	2,183	1,771	1,353

Notes: Columns (1) to (3) report Tobit estimates of Eq.3.2. Columns (4) and (5) report IV estimates of the same equation for the *Com_5* and *Com_10* samples, respectively. The indicators for financial literacy are normalized. Definitions of variables are displayed in Appendix 3.1. Bootstrap estimates of standard errors are computed with 200 replications and reported in parentheses. $***$ $p < 0.01$, $**$ $p < 0.05$, $*$ $p < 0.1$

Table 3.9 Robustness test 1: Considering the cognitive ability

	Probit				Tobit			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	FK_Score	FA_Indx	PRIDIT_Indx	Subj_FK	FK_Score	FA_Indx	PRIDIT_Indx	Subj_FK
Panel A: Baseline (N = 2,183)								
Basic FK	0.0152	0.0144	0.0132		0.0025	0.0021	-0.0024	
	(0.0248)	(0.0248)	(0.0230)		(0.0292)	(0.0293)	(0.0275)	
Advanced FK	0.1447***	0.1449***	0.1339***		0.1269***	0.1267***	0.1063***	
	(0.0299)	(0.0298)	(0.0331)		(0.0319)	(0.0318)	(0.0350)	
Subj_FK				0.0647**				0.0551
				(0.0314)				(0.0356)
WordTest	0.0019	0.0019	0.0020	0.0029**	0.0031*	0.0031*	0.0033**	0.0041**
	(0.0014)	(0.0014)	(0.0014)	(0.0015)	(0.0017)	(0.0017)	(0.0017)	(0.0017)
MathTest	-0.0001	-0.0002	0.0001	0.0008	0.0005	0.0004	0.0007	0.0010
	(0.0024)	(0.0024)	(0.0024)	(0.0025)	(0.0027)	(0.0027)	(0.0027)	(0.0027)
Panel B: IV (N = 1,771)								
Basic FK	-0.0610	-0.0605	-0.0551		-3.1592	-3.1921	-2.8545	
	(0.0561)	(0.0570)	(0.0526)		(2.3576)	(2.4104)	(2.2084)	
Advanced FK	0.2461***	0.2411***	0.2818***		7.0295**	6.8390**	7.8227**	
	(0.0852)	(0.0828)	(0.0994)		(3.4723)	(3.3852)	(3.9809)	
Subj_FK				0.1725*				7.0287*
				(0.0983)				(4.2550)
WordTest	0.0018	0.0018	0.0015	0.0027*	0.1642**	0.1649**	0.1657**	0.1928***
	(0.0015)	(0.0015)	(0.0016)	(0.0015)	(0.0699)	(0.0699)	(0.0711)	(0.0687)
MathTest	0.0008	0.0007	0.0005	0.0016	0.1022	0.1020	0.0939	0.0935
	(0.0031)	(0.0031)	(0.0032)	(0.0031)	(0.1327)	(0.1326)	(0.1336)	(0.1312)

Notes: This table reports a compressed version of the estimated marginal effects that incorporate cognitive ability in the baseline models. The other control variables are the same as in the baseline models. Panel A reports the baseline estimates, and Panel B reports the IV estimates. Bootstrap estimates of standard errors are computed with 200 replications and reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3.10 Robustness test 2: Alternative estimation methods (LMP and OLS)

	LPM				OLS			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	FK_Score	FA_Indx	PRIDIT_Indx	Subj_FK	FK_Score	FA_Indx	PRIDIT_Indx	Subj_FK
Panel A: Baseline (N = 2,183)								
Basic FK	-0.0152	-0.0177	-0.0026		-0.1478	-0.1618	-0.0207	
	(0.0232)	(0.023)	(0.0224)		(0.1694)	(0.1682)	(0.1613)	
Advanced FK	0.2079***	0.2157***	0.1315***		1.6854***	1.7427***	1.0661***	
	(0.0385)	(0.0393)	(0.0344)		(0.2806)	(0.2861)	(0.2479)	
Subj_FK				0.0562				0.3463
				(0.0344)				(0.2506)
Panel B: IV (N = 1,771)								
Basic FK	-0.0421	-0.0173	0.0013		-0.4112	-0.4316	-0.1804	
	(0.0450)	(0.0175)	(0.0164)		(0.3573)	(0.3608)	(0.3457)	
Advanced FK	0.3624***	0.1406***	0.1029***		3.1328***	3.1945***	2.4208***	
	(0.0812)	(0.0309)	(0.0309)		(0.6354)	(0.6329)	(0.6505)	
Subj_FK				0.0567*				1.3398**
				(0.0294)				(0.6519)

Notes: This table reports a compressed version of the LPM and CMP estimates for the participation equation, and the OLS and CMP estimates for the expenditure equation. The marginal effects from the Tobit models are estimated on the expected value of the truncated outcome, i.e., the expenditure on online shopping that is not less than zero. The other control variables are the same as in the full baseline models. Panel A reports the baseline estimates, and Panel B reports the IV estimates. Marginal effects are reported in the table. Bootstrap estimates of standard errors are computed with 200 replications and reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 3.11 Robustness test 3: Controlling for peer effects

	Probit					Tobit			
	(1)	(2)	(3)	(4)		(5)	(6)	(7)	(8)
VARIABLES	FK_Score	FA_Indx	PRIDIT_Indx	Subj_FK		FK_Score	FA_Indx	PRIDIT_Indx	Subj_FK
Panel A: Baseline (N = 995)									
Basic FK	0.0243	0.0232	0.0221		Basic FK	-0.0028	-0.0012	0.0024	
	(0.0276)	(0.0278)	(0.0253)			(0.0369)	(0.037)	(0.0354)	
Advanced FK	0.1602***	0.1607***	0.1511***		Advanced FK	0.1418***	0.1359***	0.0984*	
	(0.0320)	(0.0318)	(0.036)			(0.0458)	(0.0449)	(0.0541)	
Subj_FK				0.0802**	Subj_FK				-0.0045
				(0.0344)					(0.051)
D_peer	0.0713	0.0704	0.0737	0.0776	Lg_exp_peer	0.0001***	0.0001***	0.0001***	0.0001***
	(0.0477)	(0.0478)	(0.0481)	(0.0503)		(-0.0003)	(-0.0003)	(-0.0003)	(-0.0003)
Panel B: IV (N = 995)									
Basic FK	-0.0289	-0.0284	-0.0226		Basic FK	-3.7425	-4.1719	-3.7021	
	(0.0536)	(0.0545)	(0.051)			(0.4066)	(0.4123)	(0.3884)	
Advanced FK	0.2515***	0.2467***	0.2861***		Advanced FK	3.0163***	3.7484***	3.3158***	
	(0.0877)	(0.0854)	(0.1003)			(0.7758)	(0.7663)	(0.7885)	
Subj_FK				0.2554**	Subj_FK				-0.0272
				(0.1081)					(0.0521)
D_peer	0.1426*	0.1420*	0.1375*	0.1746**	Lg_exp_peer	0.0016***	0.0015***	0.0015***	0.0016***
	(0.0761)	(0.076)	(0.0771)	(0.0779)		(0.0003)	(0.0003)	(0.0003)	(0.0003)

Notes: This table reports a compressed version of the marginal effects of financial literacy after controlling for peer effects (*D_peer* and *Lg_exp_peer*). The other control variables are the same as in the full baseline models. Panel A reports the baseline estimates, and Panel B reports the IV estimates. Bootstrap estimates of standard errors are computed with 200 replications and reported in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1

Table 3.12 Online shopping participation: by age groups

	FK_Score		FA_Indx		PRIDIT_Indx		Subj_FK
Panel A	(1) Basic	(2) Advanced	(3) Basic	(4) Advanced	(5) Basic	(6) Advanced	(7)
FK×agegrp1	-0.0677	0.4325***	-0.0803	0.4332**	-0.0603	0.4017**	0.0725
	(0.1298)	(0.1636)	(0.1389)	(0.1692)	(0.1430)	(0.1664)	(0.1606)
FK×agegrp2	0.0499	0.2253***	0.0516	0.2238***	0.0502	0.2315***	0.1249
	(0.0551)	(0.0739)	(0.0586)	(0.0689)	(0.0561)	(0.0796)	(0.0867)
FK×agegrp3	0.0194	0.1130**	0.0199	0.1132**	0.0076	0.0983**	0.0374
	(0.0427)	(0.0469)	(0.0403)	(0.0461)	(0.0384)	(0.0462)	(0.0437)
FK×agegrp4	0.0096	0.0793	0.0077	0.0814*	0.0305	0.0715	0.0551
	(0.1244)	(0.0876)	(0.0474)	(0.0426)	(0.0526)	(0.0483)	(0.0391)
Observations	2,183	2,183	2,183	2,183	2,183	2,183	2,162

Notes: Marginal effects obtained using a Probit model are reported in this Table. The linear control variables are the same as in the baseline models. Bootstrap estimates of standard errors are computed with 200 replications and reported in parentheses. $***$ $p < 0.01$, $**$ $p < 0.05$, $*$ $p < 0.1$

Table 3.13 Online shopping expenditure: by age groups

	FK_Score		FA_Indx		PRIDIT_Indx		Subj_FK
Panel A	(1) Basic	(2) Advanced	(3) Basic	(4) Advanced	(5) Basic	(6) Advanced	(7)
FK×agegrp1	-0.0697	0.2061***	-0.0727	0.2003***	-0.0848	0.1712*	0.0476
	(0.0821)	(0.0773)	(0.0821)	(0.0772)	(0.0790)	(0.0891)	(0.1021)
FK×agegrp2	0.0499	0.2188***	0.0526	0.2159***	0.0448	0.2277***	0.0981
	(0.0650)	(0.0774)	(0.0653)	(0.0772)	(0.0606)	(0.0818)	(0.0767)
FK×agegrp3	0.0155	0.1444***	0.0161	0.1450***	0.0018	0.1240**	0.0515
	(0.0465)	(0.0541)	(0.0470)	(0.0538)	(0.0412)	(0.0585)	(0.0501)
FK×agegrp4	0.0221	0.1115**	0.0200	0.1138**	0.0510	0.1042*	0.0817
	(0.0356)	(0.0540)	(0.0348)	(0.0520)	(0.0418)	(0.0543)	(0.0535)
Observations	2,183	2,183	2,183	2,183	2,183	2,183	2,162

Notes: This Table reports Tobit estimates. The linear control variables are the same as in the baseline models. Bootstrap estimates of standard errors are computed with 200 replications and reported in parentheses. $***$ $p < 0.01$, $**$ $p < 0.05$, $*$ $p < 0.1$

Table 3.14 Online shopping participation: by income groups

	FK_Score		FA_Indx		PRIDIT_Indx		Subj_FK
Panel A	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Basic	Advanced	Basic	Advanced	Basic	Advanced	
FK×income_low	0.1688*	0.0650	0.1766*	0.0499	0.1208	0.1085	0.1168
	(0.0884)	(0.1046)	(0.0971)	(0.1169)	(0.0931)	(0.1198)	(0.1028)
FK×income_med	0.0358	0.1589***	0.0288	0.1659***	0.0304	0.1737***	0.0030
	(0.0450)	(0.0614)	(0.0457)	(0.0616)	(0.0407)	(0.0639)	(0.0569)
FK×income_high	-0.0294	0.1715***	-0.0287	0.1711***	-0.0205	0.1439***	0.0843*
	(0.0350)	(0.0425)	(0.0413)	(0.0427)	(0.0311)	(0.0479)	(0.0485)
Observations	2,183	2,183	2,183	2,183	2,183	2,183	2,162

Notes: Marginal effects obtained using a Probit model are reported in this Table. The linear control variables are the same as in the baseline models. Bootstrap estimates of standard errors are computed with 200 replications and reported in parentheses. $*** p < 0.01$, $** p < 0.05$, $* p < 0.1$

Table 3.15 Online shopping participation: by financial wealth

	FK_Score		FA_Indx		PRIDIT_Indx		Subj_FK
Panel A	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Basic	Advanced	Basic	Advanced	Basic	Advanced	
FK× fin_low	0.0832*	0.1378*	0.0798	0.1426**	0.0805*	0.1368**	0.1055*
	(0.0503)	(0.0714)	(0.0540)	(0.0652)	(0.0465)	(0.0674)	(0.0576)
FK× fin_med	0.0979	0.0469	0.0991	0.0365	0.0944	0.0693	-0.0053
	(0.0715)	(0.0978)	(0.0858)	(0.0969)	(0.0697)	(0.0955)	(0.0864)
FK× fin_high	-0.0498	0.1970***	-0.0496	0.1964***	-0.0491	0.1776***	0.0585
	(0.0401)	(0.0428)	(0.0356)	(0.0441)	(0.0411)	(0.0440)	(0.0424)
Observations	2,183	2,183	2,183	2,183	2,183	2,183	2,162

Notes: This Table reports Probit estimates. The linear control variables are the same as in the baseline models. Bootstrap estimates of standard errors are computed with 200 replications and reported in parentheses. $*** p < 0.01$, $** p < 0.05$, $* p < 0.1$

Table 3.16 Online shopping expenditure: by income groups

	FK_Score		FA_Indx		PRIDIT_Indx		Subj_FK
Panel A	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Basic	Advanced	Basic	Advanced	Basic	Advanced	
FK×inc_low	0.0441	-0.0901	0.0467	-0.0995	0.0202	-0.0983	-0.0196
	(0.0763)	(0.0823)	(0.0774)	(0.0825)	(0.0678)	(0.0927)	(0.0865)
FK×inc_med	-0.0161	0.1439***	-0.0237	0.1502***	-0.0142	0.1320***	0.0309
	(0.0386)	(0.0451)	(0.0382)	(0.0451)	(0.0368)	(0.0494)	(0.0528)
FK×inc_high	-0.0215	0.1922***	-0.0178	0.1889***	-0.0163	0.1667***	0.066
	(0.0398)	(0.0453)	(0.0403)	(0.0447)	(0.0375)	(0.051)	(0.0475)
Observations	2,183	2,183	2,183	2,183	2,183	2,183	2,162

Notes: Panel A reports the marginal effects of financial literacy on online shopping expenditure for each income group estimated using the Tobit model. The linear control variables are the same as in the baseline models. Panel B reports the statistics for the t-test on the difference between income groups, with the low-income group as reference. Bootstrap estimates of standard errors are computed with 200 replications and reported in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1

Table 3.17 Online shopping expenditure: by financial wealth

	FK_Score		FA_Indx		PRIDIT_Indx		Subj_FK
Panel A	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Basic	Advanced	Basic	Advanced	Basic	Advanced	
FK× fin_low	0.0881*	0.1173**	0.0854*	0.1242**	0.0864**	0.1019	0.1493***
	(0.0461)	(0.0592)	(0.0466)	(0.0596)	(0.0425)	(0.0649)	(0.0561)
FK× fin_med	0.0131	0.1255**	0.0136	0.1176*	-0.0136	0.1370**	-0.0356
	(0.0542)	(0.0631)	(0.0539)	(0.0632)	(0.0490)	(0.0641)	(0.0630)
FK× fin_high	-0.0626	0.1869***	-0.0604	0.1849***	-0.0551	0.1642***	0.0548
	(0.0456)	(0.0453)	(0.0452)	(0.0442)	(0.0447)	(0.0536)	(0.0481)
Observations	2,183	2,183	2,183	2,183	2,183	2,183	2,162

Notes: Panel A reports the marginal effects of financial literacy on online shopping expenditure for each financial wealth group estimated using the Tobit model. The linear control variables are the same as in the baseline models. Panel B reports the statistics for the t-test on the difference between financial wealth groups, with the first low-wealth group as reference. Bootstrap estimates of standard errors are computed with 200 replications and reported in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1

Appendix

Appendix 3.1: Definitions of the variables

Table A3.1 Definitions of the variables

Dependent variables	*Definitions*
D_Online	Dummy variable equals 1 if the respondent ever shopped online in the year before taking the interview, 0 otherwise.
Exp_Online	Continuous variable: annual expenditure of online shopping of the respondent.
Independent variables	
Individual-level	
Male	Dummy variable equals 1 if the respondent is male, 0 otherwise.
Edu_y	Continuous variable: number of years of education obtained by the respondent.
Married	Dummy variable equals 1 if the respondent is married or cohabitated.
Age dummies	Four dummies are generated regarding the age of the respondent. The cutoff points are 35, 45 and 60, with the age group 20-35 as the reference group.
Not-working	Dummy variable equals 1 if the respondent was not working at the time of the interview (including unemployment and student), and equals 0 if the respondent has agricultural work, or is employed or self-employed.
Risk_averse	Dummy variable equals 1 if the respondent is risk-averse and equals 0 if the respondent is risk-seeking or risk-neutral.
Health_Shock	Dummy variable equals 1 if the respondent reports worse health compared with last year and equals 0 if the respondent reports better or similar health status compared with last year.
Online_comm	Discrete variable: frequency of using the internet to communicate. The values represent 1. Every day, 2. Three to four times a week, 3. One or two times a week, 4. Two to three times per month, 5. Once a month, 6. Once in several months, and 7. Never.
Financial attitude	Discrete variable: Summary score of three Likert type questions (1 totally agree, 5 totally disagree) on the respondents' financial attitude towards "Spending money makes me more satisfied than saving money" "I intend to live in the present more and do not consider the future" "In my opinion, money is to be spent and we should spend it whenever we have money".
Location	Discrete variable: variable equals 1 if the family live in the western area of China, 2 if in the central area of China, 3 if in the coastal area of China.

Household level	
N_depchildren	Discrete variable: number of economically dependent children in the household.
N_depelderly	Discrete variable: number of economically dependent elderly in the household.
Net_HH_Income	Continuous variable: annual family income, including labour income and investment income.
Own_computer	Dummy variable equals 1 if the household own a computer, 0 otherwise.
Fin_asset	Continuous variable: household's total financial wealth, including cash at home, stocks, funds, state debt, trust products, foreign exchange products, futures, stock equity and other financial products.
Non-fin_asset	Continuous variable: household's total non-financial wealth, including fixed assets, durable assets, the value of the house and value of the land.
Community level	
Facility1	Continuous variable: a factor analysis index we computed based on the existence of Convenience/corner/department store, Kindergarten, Primary school, Middle School, Hospital / medical facility, Drug store, Buddhist / Taoist Temple, Ancestral hall, Church/ Mosque, Activity facility/community service center for the elderly, Nursing home/ Almshouse, Physical exercise facility, Playground, Bulletin board for village/community affairs, Informants' letterbox, Community Website/ Weibo in the community. The higher the index, the better the infrastructure in the village/community.
Facility2	Continuous variable: factor analysis index we computed based on whether the village/residential community has access to electricity, cable radio, satellite TV, postal service, telephone, cellphone signal/service, public road, access to the railway, tap water, pipeline gas, bus, subway, the transition from village community to the residential community, as well as whether the community establish enterprises run by local people, implement direct election in the community. The higher the index, the better the facility in the village/community.

Appendix 3.2: Wording of financial knowledge questions

Appendix 3.2.1 Wording of the Big Three questions (Lusardi and Mitchell, 2011a)

Suppose you had $100 in a savings account and the interest rate was 2% per year. After 5 years, how much do you think you would have in the account if you left the money to grow: more than $102, exactly $102, less than $102?

Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After 1 year, would you be able to buy more than, exactly the same as, or less than today with the money in this account?

Do you think that the following statement is true or false? "Buying a single company stock usually provides a safer return than a stock mutual fund."

Appendix 3.2.2: Wording of Basic and Advanced literacy questions in the CFPS

Source: Institute of Social Science Survey, 2015

- **Basic financial literacy questions**

(1) According to your estimation, what is the annual interest rate of one-year fixed deposit in the bank?

1. Lower than 1% 2. 1-5% 3. 5-10% 4. 10% and higher 9. Don't know

(2) Suppose you have a one-year fixed deposit of 10,000 yuan and the annual interest rate is 3%. If you don't withdraw within this period, how much money will you get on the due date?

1. 10300 yuan 2. More than 10300 yuan 3. Less than 10300 yuan 9. Don't know

(3) After the due date of deposit in the above question, if you continue to save the money as one-year fixed deposit and the annual interest rate is 3%, how much money will you have in this account after one year, including the principal and interests?

1. 10600 yuan 2. More than 10600 yuan 3. Less than 10600 yuan 9. Don't know

(4) If the annual interest rate of your savings account is 3%, and the inflation rate is 5%, how many goods you can buy using your savings in this account after one year?

1. More than today 2. The same as today 3. Less than today 9. Don't know

(5) Suppose Zhang San inherits 100,000 yuan today, and Li Si will inherit 100,000 yuan three years later. Who has a higher value of inheritance?

1. Zhang San has a higher value of inheritance

2. Li Si has a higher value of inheritance

3. They have the same value of inheritance

9. Don't know

- **Advanced financial literacy questions**

(6) In general, investments with high returns have high risks.

1. Right 5.Wrong 9. Don't know

(7) In general, the risk of investing in one-share stock is smaller than that of investing in stock fund.

1. Right 5. Wrong 9. Don't know

(8) Which of the following bank is responsible for making and carrying out monetary policies?

1. Bank of China 2. Industrial and Commercial Bank of China

3. People's Bank of China 4. China Construction Bank 9. Don't know

(9) Generally speaking, which of the following investment has the highest risk?

1. Bank savings 2. National debts 3. Stocks 4. Capital funds 9. Don't know

(10) If you purchase stocks of some company, that means:

1. No matter you hold these stocks for a short term or long term, you lend the money to the company anyway.

2. No matter you hold these stocks for a short term or long term, you are the stockholder of the company anyway.

3. You are the stockholder of the company when you hold these stocks for long term; and you lend

the money to the company when you hold these stocks for short term.

4. None of the above is correct

9. Don’t know

(11) Which one of the following is correct in terms of describing capital funds:

1. Low-price (low unit value) capital fund has a better future performance.

2. In general, the same capital fund could be invested in several assets; for example, one capital fund can be invested into both stocks and bonds.

3. In general, capital fund can provide a guaranteed rate of return based on past performance.

4. None of the above is correct

9. Don’t know

(12) Which of the following statement is correct about a bank’s financial products?

1. Same as with other risky investments, a bank’s financial products are also possible to suffer from loss.

2. As safe as savings, bank financial products at least will not lose.

3. The expected profits of bank financial products are actual profits.

4. None of the above is correct.

9. Don’t know

(13) Which of the following statements correctly describes the core function of stock market?

1. Stock market is helpful in predicting the profits of stocks

2. Stock market increases the prices of stocks

3. Stock market helps to make matches between stock buyers and sellers

4. None of the above is correct

9. Don’t know

Appendix 3.3: Financial Knowledge Indices

Appendix 3.3.1 Factor Analysis

1) Corresponding dummy variables are constructed for 13 financial knowledge question, which equals 1 if the answer is correct and 0 if the answer is incorrect or "Do not know".
2) Factor analysis is then applied to the basic and advanced sets of questions separately. For each of the indices, we first obtain the initial factor loadings and extract the factors with eigenvalues larger than one (following the Kaiser-Guttman Criterion). The factors kept according to the eigenvalues passes the parallel analysis and the Kaiser-Meyer-Olkin (KMO) statistics for both indices are larger than 0.8, indicating adequate sampling and the validity of factor analysis.[70]
3) Factor rotation is then implemented to produce orthogonal factors, so that to create indices without inter-correlated components.
4) We finally obtain two factor scores (indices), which are the weighted sums of the factors extracted from each set of questions.

[70] The KMO statistic is a measure of the proportion of variance among variables that might have common variance. KMO takes values between 0 and 1, with small values meaning that overall the variables have too little in common to warrant a factor analysis. A rule of thumb indicates that KMO values between 0.8 and 1 indicate the sampling is adequate (Kaiser, 1958).

Appendix 3.3.2 PRIDIT Indices

The PRIDIT approach includes a first step to calculate the RIDIT scores and a second step to conduct a principal component analysis on the RIDIT scores (Behrman *et al.*, 2012; Brockett *et al.*, 2002). Specifically, for each possible response *i* to question *t*[71], the RIDIT score (R_{ti}) is calculated as follows:

$$R_{ti} = \sum_{j<i} \widehat{p_{tj}} - \sum_{j>i} \widehat{p_{tj}}$$

where the first (second) component of the difference denotes the proportion of answers below (above) *i*. The score is then centred at 0. Taking the first financial knowledge question as an example, we have:

FK1

Value	Code	Number	Proportion	Proportion Below	Proportion Above	RIDIT
Correct	1	1,220	0.5588	0.4412	0	0.4412
Incorrect	0	963	0.4412	0	0.4412	-0.5588

In the same spirit, we calculate the RIDIT scores for all the financial knowledge questions, as outlined in the Table below. Finally, we calculate the PRIDIT indices by implementing principal component analysis on the RIDIT scores.

	Proportion of correct responses	RIDIT for Correct Answers	RIDIT for Incorrect/DK Answers
FK1	55.88	0.4412	-0.5588
FK2	44.58	0.5542	-0.4458
FK3	50.36	0.4964	-0.5036
FK4	59.82	0.4018	-0.5982
FK5	68.94	0.3106	-0.6894
FK6	86.40	0.136	-0.864
FK7	33.95	0.6605	-0.3395
FK8	32.57	0.6743	-0.3257
FK9	68.73	0.3127	-0.6873
FK10	15.86	0.8414	-0.1586
FK11	11.30	0.887	-0.113
FK12	29.13	0.7087	-0.2913
FK13	27.42	0.7258	-0.2742

[71] *i* denotes in turn a correct and an incorrect answer. In the former case, it takes value 1, and in the latter, 0.

Appendix 3.4: Distribution of indicators for financial knowledge across demographic groups

Table A3.4.1 Distribution of indicators for financial knowledge across demographic groups

		Basic Financial Knowledge			Advanced Financial Knowledge		
		(1)	(2)	(3)	(4)	(5)	(6)
Panel A: by Gender	Var	Summary score	FA index	PRIDIT	Summary score	FA index	PRIDIT
Female (N = 1,244)	Mean	0.5126	0.5173	0.4993	0.3392	0.3292	0.4248
	S.D.	0.3219	0.3232	0.3309	0.2165	0.2138	0.2242
Male (N = 939)	Mean	0.5839	0.5907	0.5726	0.4103	0.3965	0.5023
	S.D.	0.3098	0.3105	0.3206	0.2363	0.2344	0.2317
Panel B: by Age	Var	Summary score	FA index	PRIDIT	Summary score	FA index	PRIDIT
20-35 (N = 218)	Mean	0.6495	0.6555	0.6398	0.4691	0.4582	0.5418
	S.D.	0.2766	0.2761	0.2894	0.2237	0.2250	0.2054
36-45 (N = 480)	Mean	0.6253	0.6311	0.6080	0.4387	0.4266	0.5208
	S.D.	0.2903	0.2908	0.3026	0.2200	0.2186	0.2090
46-60 (N = 851)	Mean	0.5305	0.5361	0.5134	0.3663	0.3544	0.4579
	S.D.	0.3125	0.3137	0.3236	0.2264	0.2236	0.2286
60+ (N = 633)	Mean	0.4687	0.4741	0.4660	0.2952	0.2838	0.3883
	S.D.	0.3368	0.3388	0.3456	0.2117	0.2071	0.2358
Panel C: by Education	Var	Summary score	FA index	PRIDIT	Summary score	FA index	PRIDIT
Illiterate (N = 237)	Mean	0.3081	0.3101	0.2994	0.1869	0.1792	0.2623
	S.D.	0.3017	0.3021	0.3121	0.1731	0.1677	0.2151
Primary (N = 266)	Mean	0.3953	0.3979	0.3888	0.2737	0.2622	0.3690
	S.D.	0.3096	0.3115	0.3192	0.1899	0.1840	0.2140

Junior (N = 666)	Mean	0.5356	0.5416	0.5220	0.3480	0.3358	0.4446
	S.D.	0.2962	0.2973	0.3072	0.1919	0.1886	0.1996
Senior (N = 573)	Mean	0.6047	0.6118	0.5893	0.4079	0.3965	0.4986
	S.D.	0.2910	0.2911	0.3048	0.2096	0.2090	0.2003
College_above (N = 441)	Mean	0.7457	0.7535	0.7326	0.5554	0.5412	0.6299
	S.D.	0.2458	0.2438	0.2590	0.2242	0.2260	0.2035

Notes: This table reports the distribution of objective indicators for basic and advanced financial knowledge across gender, age groups and educational attainment. The indicators for basic and advanced financial knowledge are normalized to range from 0 to 1.

Appendix 3.5: Correlation matrices for key variables

	D_online	Online Exp	Basic FK	Advanced FK	Male	Edu_y	Married	Age	Not-working	Risk_ averse	Health_ shock
D_online	1										
Online Exp	0.674*	1									
Basic FK	0.244*	0.183*	1								
Advanced FK	0.339*	0.279*	0.53*	1							
Male	-0.078*	-0.051*	0.108*	0.154*	1						
Edu_y	0.367*	0.303*	0.401*	0.46*	0.101*	1					
Married	-0.009	0.003	0.105*	0.056*	0.112*	0.133*	1				
Age	-0.443*	-0.309*	-0.211*	-0.301*	0.104*	-0.388*	-0.092	1			
Not-working	-0.281*	-0.202*	-0.204*	-0.297*	-0.15*	-0.336*	-0.08	0.604*	1		
Risk_averse	-0.259*	-0.206*	-0.289*	-0.371*	-0.107*	-0.307*	-0.055	0.276*	0.235*	1	
Health_shock	-0.09*	-0.073*	-0.155*	-0.144*	-0.06*	-0.147*	-0.017	0.165*	0.14*	0.11*	1
Online_comm	0.224*	0.103*	0.229*	0.264*	0.057*	0.313*	0.088*	-0.231*	-0.195*	-0.219*	-0.074*
FA_score	0.079*	0.034	0.115*	0.104*	-0.025	0.071*	0.023	-0.192*	-0.16*	-0.065*	0.023
N_depchildren	0.129*	0.085*	0.012	0.071*	-0.05*	0.046*	0.041	-0.428*	-0.293*	-0.132*	-0.065*
N_depelderly	-0.197*	-0.132*	-0.076*	-0.143*	0.109*	-0.226*	0.055*	0.62*	0.471*	0.136*	0.097*
Own_computer	-0.033	-0.029	-0.016	-0.008	-0.014	-0.009	-0.03	0.022	0.027	0.021	-0.004
Net HH Income	0.184*	0.194*	0.287*	0.256*	0.034	0.281*	0.183	-0.032	-0.107*	-0.183*	-0.076*
Fin_asset	0.163*	0.142*	0.356*	0.292*	0.082*	0.229*	0.111	-0.014	-0.065*	-0.182*	-0.103*
Non-fin_asset	0.177*	0.173*	0.229*	0.206*	0.057*	0.2*	0.151	-0.049*	-0.097*	-0.204*	-0.088*
Facility1	0.008	0.004	0.018	0.023	0.048	0.048*	0.054	-0.071*	-0.082*	-0.037	-0.019
Facility2	-0.007	-0.048*	-0.04	-0.011	0.029	-0.031	0.062*	-0.028	-0.015	-0.006	0.021

	Online_comm	FA_score	N_depchildren	N_depelderly	Own_computer	Net HH Income	Fin_asset	Non-fin_asset	Facility1	Facility2
Online_comm	1									
FA_score	0.06*	1								
N_depchildren	0.057*	0.112*	1							
N_depelderly	-0.153*	-0.097*	-0.284*	1						
Own_computer	0.027	-0.023	0.006	-0.02	1					
Net HH Income	0.149*	-0.02	0.087*	0.062*	-0.058*	1				
Fin_asset	0.161*	0.043*	-0.056*	0.071*	-0.009	0.316*	1			
Non-fin_asset	0.143*	-0.021	0.113*	0.046*	-0.035	0.384*	0.265*	1		
Facility1	0.056*	-0.012	0.095*	-0.027	-0.016	0.005	-0.023	-0.01	1	
Facility2	-0.036	0.057*	0.034	0.015	-0.042	-0.01	-0.068*	0.038	0.028	1

Appendix 3.6: Validity tests for the instrumental variables

In this section, we discuss the tests for the validity of our instruments, namely the mother's duration of education and the community-level financial literacy evaluated within the *Com_5* sample.[72] The test results, based on a two stage least squares (2SLS) approach, are reported in Table A3.6.2 for online shopping expenditure[73].

We start with the participation equations. We first perform a set of endogeneity tests. If the IV specification passes the endogeneity test, i.e., rejects the null hypothesis of endogeneity test, then we cannot take the suspected endogenous variables as exogenous and have to tackle the endogeneity issue. Otherwise, we can rely on the simple Probit or Tobit estimates for interpretation. The last two rows in Table A3.6.1 indicate that we have to reject the null hypothesis of the endogeneity test, and have to adopt the IV approach for the models with objective financial literacy and online shopping participation. However, the null hypothesis of the endogeneity test is not rejected in the model with subjective financial literacy and online shopping participation, indicating the IV estimates may not as good as the Probit estimates in interpreting the relationship between subjective financial literacy and online shopping participation.

Next, for the three specifications using objective financial literacy (Models 1 to 3 in Table A3.6.1) that pass the endogeneity test, we further conduct the under-identification test (Kleibergen-Paap rk LM test) , weak identification test (Kleibergen-Paap rk Wald F test in case of heterogeneity in the error term) and over-identification test (Hansen J test).

For all the three models, we reject the null hypothesis of the under-identification test (the Kleibergen-Paap Chi-sq statistics equal 4.978, 8.140, 8.308, respectively). This indicates that all the three equations are identified, and the instruments are relevant.

Also, for all the three models, we cannot reject the null hypothesis of the overidentification test, i.e., the instruments are uncorrelated with the error, which indicates that the instruments are valid as they are uncorrelated with the error (the null hypothesis is that valid instruments are uncorrelated with the error).

[72] We also considered the father's duration of education as an instrument for the respondent's financial literacy, but this particular instrument failed most of the instrument validity tests. We therefore excluded this particular variable from our instrument set.

[73] These tests are based on estimates obtained using the 2SLS estimator because the CMP framework does not provide tests for the validity of the instruments.

However, all three models fail the weak-identification test: The Kleibergen-Paap Wald rk F statistics equal 1.664, 2737, 2.807, respectively in each of the three models, which is, in all cases, much lower than the rule of thumb critical value of 10. This implies that although the instruments for objective financial literacy are valid, their correlation with the endogenous regressors may be weak. This weak IV problem highlights the necessity of employing a CMP framework, which performs better than the 2SLS framework in case of weak instruments (Roodman, 2011).

The test results for the expenditure equation are reported in Table A3.6.2. As for the participation equation, these tests suggest that an IV approach has to be implemented for the models with objective financial literacy, whilst Tobit estimates are better for the models with subjective financial literacy. Furthermore, as in the participation equation, there appears to be a weak instrument problem.

Table A3.6.1 Participation equation

	(1)	(2)	(3)	(4)
VARIABLES	Summary score	FA indices	PRIDIT indices	Subjective score
Kleibergen-Paap Chi-sq	4.978	8.14	8.308	3.045
p-val Kleibergen-Paap Chi-sq	0.0830	0.0171	0.0157	0.385
Kleibergen-Paap F	1.664	2.737	2.807	1.003
Hansen J	0.0929	0.05	0.00317	13.47
p-val Hansen J	0.760	0.8228	0.955	0.00119
Endog	11.51	10.046	14.19	0.140
p-val Endog	0.00316	0.0066	0.000829	0.708

Notes: This table reports tests for instrument validity for the participation equation based on the *Com_5* sample, estimated using 2SLS. The control variables are the same as in the full baseline model. Robust standard errors are reported in parentheses.

Table A3.6.2 Expenditure equation

	(1)	(2)	(3)	(4)
VARIABLES	Summary score	FA indices	PRIDIT indices	Subjective score
Kleibergen-Paap Chi-sq	4.978	8.14	8.308	30.84
p-val Kleibergen-Paap Chi-sq	0.0830	0.0171	0.0157	2.01e-07
Kleibergen-Paap F	1.664	2.737	2.807	15.60
Hansen J	0.199	0.003	0.0218	7.153
p-val Hansen J	0.656	0.958	0.883	0.00748
Endog	9.362	12.006	11.81	1.357
p-val Endog	0.00927	0.0025	0.00273	0.244

Notes: This table reports tests for instrument validity for the expenditure equation based on the *Com_5* sample, estimated using 2SLS. The control variables are the same as in the full baseline models. Robust standard errors are reported in parentheses.

Appendix 3.7: Controlling for both objective and subjective financial literacy simultaneously

Table A3.7.1 Online shopping participation

	Probit			IV_Probit		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Sum_Score	FA	PRIDIT	Sum_Score	FA	PRIDIT
Subj_FK	0.0318	0.0315	0.0404	0.1128	0.1072	0.1293
	(0.0361)	(0.0360)	(0.0365)	(0.1012)	(0.1014)	(0.0995)
Basic_FK	0.0018	0.0001	-0.0022	-0.0530	-0.0517	-0.0511
	(0.0278)	(0.0278)	(0.0257)	(0.0557)	(0.0565)	(0.0527)
Advanced_FK	0.1476***	0.1498***	0.1220***	0.2479***	0.2435***	0.2741***
	(0.0312)	(0.0310)	(0.0357)	(0.0854)	(0.0833)	(0.0960)
Observations	2,162	2,162	2,162	2,162	2,162	2,162

Notes: Columns (1) to (3) report marginal effects of Probit estimates of Eq.3.1. Columns (4) to (6) report IV estimates of the same equation for the *Com_5* sample. All the indicators for financial literacy are normalized to range from 0 to 1. Definitions of variables are displayed in Appendix 1. Bootstrap estimates of standard errors are computed with 200 replications and reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A3.7.2 Online shopping expenditure

	Tobit			IV_Tobit		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Sum_Score	FA	PRIDIT	Sum_Score	FA	PRIDIT
Subj_FK	0.0267	0.0260	0.0344	0.0267	0.0260	0.0344
	(0.0324)	(0.0322)	(0.0326)	(0.0324)	(0.0322)	(0.0326)
Basic_FK	0.0118	0.0122	0.0067	0.0118	0.0122	0.0067
	(0.0283)	(0.0284)	(0.0264)	(0.0283)	(0.0284)	(0.0264)
Advanced_FK	0.1421***	0.1412***	0.1279***	0.1421***	0.1412***	0.1279***
	(0.0312)	(0.0310)	(0.0347)	(0.0312)	(0.0310)	(0.0347)
Observations	2,162	2,162	2,162	2,162	2,162	2,162

Notes: Columns (1) to (3) report Tobit estimates of Eq.3.2. Columns (4) and (5) report IV estimates of the same equation for the *Com_5* sample. The indicators for financial literacy are normalized. Definitions of variables are displayed in Appendix 1. Bootstrap estimates of standard errors are computed with 200 replications and reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix 3.8 Online shopping participation: full table

	Summary score		FA indices		PRIDIT indices		Subjective score	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Probit	IV_Com_5	Probit	IV_Com_5	Probit	IV_Com_5	Probit	IV_Com_5
Basic FK	0.0184	-0.0491	0.0177	-0.0479	0.0163	-0.0466		
	(0.0250)	(0.0545)	(0.0251)	(0.0552)	(0.0231)	(0.0518)		
Advanced FK	0.1507***	0.2539***	0.1509***	0.2489***	0.1411***	0.2859***		
	(0.0299)	(0.0831)	(0.0298)	(0.0808)	(0.0332)	(0.0962)		
Subj_FK							0.0639**	
							(0.0315)	
Male	-0.0701***	-0.5086***	-0.0698***	-0.0738***	-0.0684***	-0.0770***	-0.0628***	-0.0702***
	(0.0153)	(0.1128)	(0.0153)	(0.0161)	(0.0154)	(0.0167)	(0.0158)	(0.0158)
Edu_y	0.0179***	0.1298***	0.0179***	0.0192***	0.0188***	0.0188***	0.0210***	0.0209***
	(0.0024)	(0.0218)	(0.0024)	(0.0032)	(0.0024)	(0.0032)	(0.0024)	(0.0028)
Married	-0.0271	-0.1712	-0.0270	-0.0252	-0.0285	-0.0266	-0.0278	-0.0259
	(0.0220)	(0.1625)	(0.0221)	(0.0238)	(0.0218)	(0.0238)	(0.0218)	(0.0242)
Age(36-45)	-0.1234***	-0.8463***	-0.1233***	-0.1241***	-0.1257***	-0.1271***	-0.1262***	-0.1260***
	(0.0187)	(0.1460)	(0.0187)	(0.0206)	(0.0188)	(0.0207)	(0.0194)	(0.0211)
Age(45-60)	-0.2329***	-1.6867***	-0.2327***	-0.2475***	-0.2361***	-0.2505***	-0.2415***	-0.2575***
	(0.0198)	(0.1597)	(0.0198)	(0.0209)	(0.0198)	(0.0207)	(0.0204)	(0.0205)
Age(60+)	-0.3361***	-2.4443***	-0.3357***	-0.3588***	-0.3441***	-0.3648***	-0.3522***	-0.3735***
	(0.0374)	(0.2830)	(0.0376)	(0.0396)	(0.0379)	(0.0390)	(0.0389)	(0.0388)
Not-working	-0.0161	-0.1267	-0.0165	-0.0193	-0.0181	-0.0194	-0.0214	-0.0229
	(0.0199)	(0.1377)	(0.0199)	(0.0202)	(0.0199)	(0.0202)	(0.0203)	(0.0204)
Risk_Averse	-0.0259*	-0.2144*	-0.0256*	-0.0317*	-0.0297**	-0.0311*	-0.0330**	-0.0331**
	(0.0147)	(0.1120)	(0.0147)	(0.0164)	(0.0149)	(0.0163)	(0.0151)	(0.0160)
Health_Shock	0.0184	0.1399	0.0184	0.0204	0.0182	0.0214	0.0179	0.0243
	(0.0161)	(0.1095)	(0.0161)	(0.0160)	(0.0161)	(0.0161)	(0.0160)	(0.0165)
Online_Comm	0.0088***	0.0582***	0.0088***	0.0085***	0.0090***	0.0084***	0.0095***	0.0090***
	(0.0026)	(0.0192)	(0.0026)	(0.0028)	(0.0026)	(0.0028)	(0.0026)	(0.0028)
Financial attd	-0.0060	-0.0666	-0.0059	-0.0097	-0.0053	-0.0091	-0.0034	-0.0064

	(0.0061)	(0.0497)	(0.0061)	(0.0073)	(0.0061)	(0.0073)	(0.0061)	(0.0074)
N_depchildren	0.0001	-0.0821	0.0001	-0.0118	0.0019	-0.0119	0.0038	-0.0098
	(0.0179)	(0.1349)	(0.0179)	(0.0198)	(0.0181)	(0.0200)	(0.0179)	(0.0203)
N_depelderly	0.0141	0.1423	0.0142	0.0211	0.0154	0.0224	0.0154	0.0194
	(0.0131)	(0.0939)	(0.0132)	(0.0138)	(0.0131)	(0.0138)	(0.0130)	(0.0142)
Own_computer	0.0074	0.1714	0.0076	0.0256	0.0057	0.0196	0.0104	0.0330
	(0.0850)	(0.5456)	(0.0849)	(0.0801)	(0.0847)	(0.0809)	(0.0850)	(0.0815)
lg(Net HH Income)	0.0268***	0.2070***	0.0268***	0.0305***	0.0277***	0.0299***	0.0307***	0.0334***
	(0.0088)	(0.0764)	(0.0088)	(0.0112)	(0.0088)	(0.0112)	(0.0089)	(0.0113)
lg(Fin_asset)	0.0037***	0.0255**	0.0037***	0.0038**	0.0040***	0.0036*	0.0045***	0.0041***
	(0.0014)	(0.0121)	(0.0014)	(0.0018)	(0.0014)	(0.0018)	(0.0014)	(0.0016)
lg(Non-fin_asset)	0.0121**	0.0911**	0.0122***	0.0136**	0.0123***	0.0131**	0.0129***	0.0124**
	(0.0047)	(0.0361)	(0.0047)	(0.0053)	(0.0048)	(0.0053)	(0.0049)	(0.0054)
Facility1	-0.0084	-0.0974	-0.0073	-0.0133	-0.0145	-0.0195	-0.0203	-0.0260
	(0.0354)	(0.2559)	(0.0353)	(0.0376)	(0.0358)	(0.0376)	(0.0363)	(0.0375)
Facility2	0.0060	0.0343	0.0061	0.0054	0.0069	0.0057	0.0060	0.0061
	(0.0180)	(0.1161)	(0.0180)	(0.0170)	(0.0179)	(0.0171)	(0.0184)	(0.0172)
Western area	0.0417*	0.3111**	0.0420*	0.0495**	0.0392	0.0439*	0.0409	0.0454*
	(0.0245)	(0.1486)	(0.0244)	(0.0244)	(0.0249)	(0.0244)	(0.0259)	(0.0247)
Central area	-0.0079	-0.0596	-0.0074	-0.0079	-0.0090	-0.0115	-0.0057	-0.0038
Observations	2,183	1,771	2,183	1,771	2,183	1,771	2,183	1,771

Notes: This table reports the Probit and IV estimates for all the explanatory variables included in Eq.3.1. The dependent variable is whether the respondent shopped online in the past two year. The IV estimates are derived for the *Com_5* sample. Definitions of variables are displayed in Appendix A. Bootstrap estimates of standard errors are computed with 200 replications and reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Appendix 3.9 Online shopping expenditure: full table

	Summary score		FA indices		PRIDIT indices		Subjective score	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Tobit	IV_Com_5	Tobit	IV_Com_5	Tobit	IV_Com_5	Tobit	IV_Com_5
Basic FK	0.0085	-2.5255	0.0083	-2.5179	0.0035	-2.3453		
	(0.0291)	(2.3365)	(0.0293)	(2.3828)	(0.0273)	(2.2019)		
Advanced FK	0.1377***	7.7279**	0.1376***	7.5236**	0.1186***	8.6315**		
	(0.0317)	(3.4311)	(0.0315)	(3.3479)	(0.0350)	(3.9217)		
Subj_FK							0.0535	7.6313*
							(0.0359)	(4.2264)
Male	-0.0716***	-3.1844***	-0.0713***	-3.1554***	-0.0701***	-3.2324***	-0.0644***	-2.9940***
	(0.0147)	(0.6813)	(0.0147)	(0.6772)	(0.0148)	(0.6962)	(0.0150)	(0.6609)
Edu_y	0.0193***	0.8934***	0.0193***	0.8973***	0.0202***	0.8992***	0.0223***	0.9560***
	(0.0024)	(0.1364)	(0.0024)	(0.1360)	(0.0024)	(0.1390)	(0.0023)	(0.1228)
Married	-0.0271	-0.8833	-0.0270	-0.8901	-0.0288	-0.9543	-0.0303	-1.0142
	(0.0198)	(0.9228)	(0.0199)	(0.9227)	(0.0195)	(0.9228)	(0.0195)	(0.9263)
Age(36-45)	-0.1011***	-4.4246***	-0.1010***	-4.4147***	-0.1027***	-4.5347***	-0.1040***	-4.4834***
	(0.0158)	(0.8421)	(0.0158)	(0.8423)	(0.0158)	(0.8460)	(0.0163)	(0.8540)
Age(45-60)	-0.2084***	-9.4584***	-0.2082***	-9.4471***	-0.2110***	-9.5864***	-0.2176***	-9.8052***
	(0.0186)	(0.9516)	(0.0186)	(0.9515)	(0.0186)	(0.9554)	(0.0192)	(0.9520)
Age(60+)	-0.3288***	-14.7180***	-0.3285***	-14.7082***	-0.3345***	-14.9312***	-0.3429***	-15.1500***
	(0.0378)	(1.7442)	(0.0379)	(1.7450)	(0.0382)	(1.7488)	(0.0382)	(1.7208)
Not-working	-0.0210	-1.2297	-0.0214	-1.2594	-0.0225	-1.2376	-0.0245	-1.2537
	(0.0197)	(0.8581)	(0.0197)	(0.8577)	(0.0198)	(0.8599)	(0.0199)	(0.8622)
Risk_Averse	-0.0267*	-1.2747*	-0.0265*	-1.2828*	-0.0306**	-1.3018*	-0.0356**	-1.3249**
	(0.0145)	(0.6873)	(0.0145)	(0.6852)	(0.0146)	(0.6858)	(0.0147)	(0.6746)
Health_Shock	0.0082	0.2739	0.0080	0.2599	0.0081	0.3175	0.0061	0.3877
	(0.0155)	(0.6745)	(0.0155)	(0.6747)	(0.0155)	(0.6773)	(0.0155)	(0.6842)
Online_Comm	0.0106***	0.4631***	0.0106***	0.4637***	0.0107***	0.4549***	0.0111***	0.4705***
	(0.0026)	(0.1226)	(0.0026)	(0.1226)	(0.0026)	(0.1228)	(0.0027)	(0.1218)
Financial attd	-0.0052	-0.2997	-0.0052	-0.2944	-0.0047	-0.2822	-0.0029	-0.1567

	(0.0060)	(0.3029)	(0.0060)	(0.3030)	(0.0060)	(0.3044)	(0.0060)	(0.3076)
N_depchildren	-0.0082	-0.5280	-0.0082	-0.5314	-0.0083	-0.5285	-0.0115	-0.6297
	(0.0140)	(0.5747)	(0.0139)	(0.5752)	(0.0141)	(0.5775)	(0.0141)	(0.5739)
N_depelderly	0.0123	0.6673	0.0124	0.6722	0.0128	0.7058	0.0108	0.4942
	(0.0115)	(0.5220)	(0.0116)	(0.5220)	(0.0115)	(0.5249)	(0.0115)	(0.5291)
Own_computer	-0.0009	0.8313	-0.0007	0.8376	-0.0027	0.6659	0.0313	2.4177
	(0.0934)	(3.6186)	(0.0934)	(3.6184)	(0.0931)	(3.6472)	(0.0708)	(3.1046)
lg(Net HH Income)	0.0286***	1.4375***	0.0287***	1.4435***	0.0291***	1.4097***	0.0324***	1.5415***
	(0.0087)	(0.4653)	(0.0087)	(0.4650)	(0.0086)	(0.4672)	(0.0087)	(0.4613)
lg(Fin_asset)	0.0038***	0.1881**	0.0038***	0.1889**	0.0041***	0.1870**	0.0048***	0.2011***
	(0.0014)	(0.0744)	(0.0014)	(0.0742)	(0.0014)	(0.0763)	(0.0014)	(0.0664)
lg(Non-fin_asset)	0.0110**	0.5365**	0.0111**	0.5432**	0.0111**	0.5275**	0.0115**	0.4542**
	(0.0047)	(0.2238)	(0.0047)	(0.2240)	(0.0048)	(0.2238)	(0.0049)	(0.2268)
Facility1	-0.0209	-1.0952	-0.0207	-1.0840	-0.0226	-1.1562	-0.0263	-1.4088
	(0.0226)	(0.9632)	(0.0227)	(0.9632)	(0.0228)	(0.9686)	(0.0233)	(0.9753)
Facility2	0.0125	0.5664	0.0125	0.5734	0.0133	0.5914	0.0116	0.5844
	(0.0173)	(0.7224)	(0.0173)	(0.7223)	(0.0173)	(0.7251)	(0.0179)	(0.7249)
Western area	0.0534**	2.5402***	0.0537**	2.5602***	0.0506**	2.3132**	0.0517**	2.3092**
	(0.0251)	(0.9062)	(0.0251)	(0.9071)	(0.0254)	(0.9104)	(0.0264)	(0.9172)
Central area	-0.0025	-0.0866	-0.0022	-0.0616	-0.0036	-0.2127	-0.0011	0.1172
	(0.0159)	(0.6778)	(0.0159)	(0.6774)	(0.0160)	(0.6871)	(0.0163)	(0.6868)
Constant		-32.4074***		-32.4414***		-33.1077***		-34.3626***
		(5.1098)		(5.1114)		(5.1368)		(5.2247)
Observations	2,183	1,771	2,183	1,771	2,183	1,771	2,183	1,771

Notes: This table reports the OLS and IV estimates for all the explanatory variables included in Eq3.2. The dependent variable is whether the annual expenditure on online shopping in the past year. The IV estimates are derived for the *Com_5* sample. Definitions of variables are displayed in Appendix 3.1. Bootstrap estimates of standard errors are computed with 200 replications and reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Chapter Four: Financial Literacy and Indebtedness: Evidence from China

4.1 Introduction

Since the early 2000s, household debt has increased remarkably globally. Global household debt has reached $46.1 trillion as of the third quarter of 2018, accounting for around 19% of total debt and around 60% of global GDP (Bloomberg, 2019). The boom in global household debt is primarily fuelled by strong growth in the emerging markets, notably China. China's nominal household debt grew from $0.8 trillion to $5.4 trillion from 2008 to 2017, which renders it the world's second-highest household debt market (McKinsey, 2018).[74] Other developing countries, such as India, Mexico, Malaysia, and Chile have also experienced a growth in household debt by over 20% from 2016 to 2018 (Tiftik and Mahmood, 2019).

Unprecedented credit expansion, which substantially stimulates consumer spending and household debt, is widely witnessed in emerging economies as an efficient tool to fuel economic growth. However, high debt levels can create substantial vulnerabilities. From a micro perspective, high debt levels can impair the ability of households to smooth consumption (Alter, Feng and Valckx, 2018). Financial distress can also harm households' mental health by causing stress and anxiety (Kempson, 2002). From a macro perspective, high debt levels predict lower subsequent GDP growth, higher unemployment and even recessions in the medium and long run (Mian, Sufi and Verner, 2017). Therefore, understanding the determinants of high levels of debt and identifying the most vulnerable cohorts is of high social and political significance.

[74] During the same period, as the country with the highest household debt in the world, the US saw an increase in household debt from $14.1 trillion in 2008 to $14.9 trillion in 2017.

A large number of studies try to identify the determinants of high household debt levels, from either the supply or the demand side. On the demand-side, a factor which has been largely neglected is financial literacy. In an era with unprecedented innovation in financial services and financial instruments, people struggle to understand and manage complex new financial products; low levels of financial knowledge are widespread in the general population and around the world (e.g., Lusardi and Mitchell, 2011b; Klapper, Lusardi and Van Oudheusden, 2015). What makes it more problematic is that most people are unaware of their lack of financial knowledge, and overconfidence in financial literacy is witnessed in many countries (Lusardi and Mitchell, 2014).

On the other hand, financial literacy is vital in making sophisticated economic and financial decisions, and in obtaining sound economic and financial outcomes (e.g., Lusardi and Mitchelli, 2007a; van Rooij, Lusardi and Alessie, 2011; Anderson, Baker and Robinson, 2017; Lusardi, Michaud and Mitchell, 2017). However, little is known about the relationship between financial literacy and household borrowing behaviour, though the high household debt has become a global concern (Lusardi, Mitchell and Oggero, 2019). Among the limited literature on financial literacy and borrowing behaviours, most studies focus on the discussion regarding the correlation between financial literacy and the use of high-cost credit, such as credit card debt and alternative financial service (AFS), in the context of developed economies. Studies on developing countries, where financial markets are underdeveloped and where a large proportion of the population is financially excluded, are quite limited.

Therefore, this study aims to fill the gap by focusing on emerging markets to explore the extent to which financial literacy is related to households' indebtedness and their choice between formal and informal credit.

Among all the emerging economies, China poses a compelling case for this research question. First, China has experienced the fastest growth in household debt over the past decade

among the largest economies. Figure 4.1 displays the household debt-to-GDP ratio of five countries which have the world's largest household debt market. Despite still facing extremely high (US and UK) or moderately high (Germany and Japan) household debt-to-GDP ratios, the majority of developed countries have already witnessed slight declines in household debt since the onset of the global financial crisis. At the same time, the household debt-to-GDP ratio of China has surged from 18% in 2008 to 53% as of 2018 (PBC, N.D.), primarily fueled by the increased mortgage lending and the subsequent housing market boom. Consequently, much of the household savings are tied up with the real-estate sector leading to a decline in household consumption, which may have a detrimental effect on the economy in the long run (Figure 4.2).[75]

Second, China provides a perfect case for studying borrowing behaviours in a context where universal financial inclusion is not achieved. According to the 2014 Global Findex report, bank account ownership for Chinese adults was approximately 79 percent, compared to more than 90 percent in more advanced economies such as Japan, Korea and Germany (Demirguc-Kunt *et al.*, 2015). This resulted in a sizable household demand for services from non-bank financial service providers, which ask for higher cost (Fan, Wu and Yang, 2017; Zhang and Chen, 2019).

In this research, we make use of the 2014 wave of the China Family Panel Studies (CFPS) to study the association between financial literacy and overall indebtedness, and whether financial literacy influences households' choices between formal financing and informal financing. Specifically, we focus on both objective and subjective measures of household indebtedness. We investigate households' overall indebtedness and non-housing loans, from both extensive and intensive margins. In addition to the use of formal and informal credit, we

[75] As demonstrated in Figure 4.2, during the post-crisis period, the growth rate of household debt peaked at 43% in 2009, and became relatively steady since 2011. It is not surprising to find that the household debt grew in step with the fluctuations in housing prices. In tandem, the growth rate of household saving has been steadily decreasing; the growth rate of consumption per capita has kept going down, dropping from 23% in 2008 to 11% in 2009, and slowly decreasing to 9% in 2018.

also explore how financial literacy shapes households' preference over formal and informal credit if they intend to borrow, which has never been investigated in the literature. Overall, we find that households with higher levels of financial literacy are less likely to be indebted, show lower levels of debt and are less likely to report financial distress. We also observe significant evidence that financially literate households are less likely to use informal credit. However, financial literacy seems to be unrelated or even negatively related to the use of formal credit. When it comes to the intentions to borrow, financially literate households do demonstrate strong preferences for formal credit and against informal credit. These results are robust to using an Instrumental Variable (IV) approach and to several robustness tests. Finally, we find that the negative relationships between financial literacy and overall indebtedness, and between financial literacy and informal borrowing are primarily driven by the non-migrant and literate households. This research is of political significance in the background of excessively high debt levels of Chinese households. It is also of strategic significance to achieving financial inclusion in the wake of emerging debates on the risk of borrowing from informal financial companies.

The remainder of this chapter proceeds as follows. Section 4.2 summarizes the literature in the field of financial literacy and indebtedness. Section 4.3 develops hypotheses. Section 4.4 describes the dataset. Section 4.5 presents our main specifications, and discusses the estimation methodology. Section 4.6 reports and discusses the summary statistics and regression results. Section 4.7 performs robustness checks. Section 4.8 performs several extensions and Section 4.9 concludes this study.

4.2 Literature

Our study complements three strands of literature: the measurement of financial literacy, the relationship between financial literacy and indebtedness, and the relationship between financial literacy and households' holding of formal and informal credits.

4.2.1 Definition and measurement of financial literacy[76]

Financial literacy is a multi-dimensional concept encompassing individuals' understanding of financial knowledge, their attitudes toward and abilities to make sound financial decisions (Atkinson and Messy, 2016). Most empirical studies on the economic consequences of financial literacy focus on its cognitive dimension, using individuals' scores for financial knowledge tests as a proxy for financial literacy (see, e.g., Lusardi, 2009; Behrman *et al.*, 2012; Alessie, Van Rooij and Lusardi, 2011; Lusardi and Mitchell, 2011a). A widely used way of measuring financial literacy is based on the "Big Three" questions designed by Lusardi and Mitchell in the 2004 survey of the Health and Retirement Study (HRS) in the US.[77] The Big Three questions test three core financial concepts, including compound interest, the real interest rate and financial diversification. The inclusion of the Big Three questions in the longitudinal household-level questionnaire makes it possible to link individuals' financial literacy and financial behaviour with their personal and household characteristics.

More concepts related to competitive saving and investment decisions are considered in later studies, which aim to consider advanced financial. For instance, in the notable series of comparative analysis on worldwide financial literacy conducted by the OECD/INFE (e.g. OECD, 2016), seven questions covering the time-value of money, interest paid on loans,

[76] More details on the literature mentioned in this section is dicussed in Section 3.2.2 of Chapter Three.

[77] The Big Three questions are widely used in surveys and studies conducted in many countries, such as Australia, Germany, Italy, Sweden, Switzerland the Netherlands and Japan and so on (Lusardi and Mitchell, 2014).

interest compounding, risk-return nexus, inflation and risk diversification are asked. Lusardi, Mitchell and Curto (2014) analyse ten financial knowledge questions on capital markets, risk diversification, knowledge of fees as well as numeracy in the 2008 wave of the HRS. In their research on debt literacy and over-indebtedness, Lusardi and Tufano (2015) design a debt knowledge test including the concept of interest compounding, credit card debt payments and the time value of money. To this end, they take advantage of more complicated financial knowledge questions which are closely related to individuals' specific financial decisions.

Another challenge lies in the construction of a composite score that combines all sorts of information from the financial knowledge tests. The summary score, which counts the number of correct answers to the financial knowledge questions, is widely used in previous studies since it is convenient to calculate and unsophisticated to interpret (Lusardi, 2008; Anderson, Baker and Robinson, 2017; Hsiao and Tsai, 2018).

Some studies perform factor analysis to extract meaningful common factors lying behind the answers to financial knowledge questions (see, van Rooij, Lusardi and Alessie, 2011; Deuflhard, Georgarakos and Inderst, 2018). Via factor analysis, the authors are able to alleviate the potential multicollinearity problems and reduce the dimensions of a long list of financial questions.

However, both the summary score and the factor analysis index (FA index) weight each financial knowledge question equally, despite the fact that some questions are more informative in measuring financial sophistication regarding specific financial behaviours. In light of this specific concern, some scholars propose a two-step weighted scoring mechanism named PRIDIT (Behrman *et al.*, 2012; Lusardi, Mitchell and Curto, 2014).[78] In brief, this method

[78] The PRIDIT approach was developed by Brockett *et al.* (2002) to handle binary or categorical indicators that proxy for a difficult-to-observe dependent variable. Details on the rationale and construction of the PRIDIT index are discussed in Section 3.2.2 and Appendix 3.3.2 in Chapter Three.

weights each question according to how difficult it is for the respondents to answer it correctly, and then performs principle component analysis (PCA) on the weighted answers.

Though various questionnaires attempt to cover a wide range of financial concepts, they can hardly exhaust the list of topics that predict various financial behaviours. Therefore, some questionnaires (e.g., HRS and CFPS) augment the financial knowledge questions with a question asking about self-assessment of overall financial knowledge. Prior studies have documented an independent role of subjective financial literacy from objective financial literacy in predicting economic or financial behaviours, and the former has been found to be more predictive of retirement saving, precautionary saving and risk attitudes (Grable and Lytton, 1999; Lusardi and Mitchell, 2007a; Babiarz and Robb, 2014; Anderson, Baker and Robinson, 2017; Korkmaz *et al.*, 2019).

4.2.2 Financial literacy and borrowing behaviours

A growing literature analyses the links between financial literacy and financial decisions. Financially literate individuals are found to be more likely to plan for retirement, accrue higher retirement wealth and have a higher saving rate (e.g., Bernheim, Garrett and Maki, 2001; Lusardi and Mitchell, 2011a; Lusardi, Schneider and Tufano, 2011). Another strand of literature finds that higher financial literacy is associated with a higher probability to invest in the stock market, more diversified portfolios, and a higher portfolio return (e.g., Abreu and Mendes, 2010; van Rooij, Lusardi and Alessie, 2011; Bianchi, 2018). Moreover, financially literate households are also less likely to make mistakes such as under-diversification of investments, holding costly mortgages, paying high shares of fees and charges for credit cards, or default (Moore, 2003; Calvet, Campbell and Sodini, 2009; Gerardi, Goette and Meier, 2010; Disney and Gathergood, 2013; Lusardi and Tufano, 2015). As a result, by improving their

financial knowledge, individuals can reap substantial gains and make fewer mistakes in both the short- and long-term.

However, compared with the bourgeoning literature on financial literacy that focuses on the asset side of the household balance sheet, the discussion of financial literacy and households' borrowing behaviour is quite limited (Zinman, 2015).

4.2.2.1 Financial literacy and indebtedness

In the studies that explore financial literacy and borrowing behaviours, the primary concern is to examine whether less financially literate individuals are more inclined to carry debt, or more likely to carry high-cost credit. However, their results are often inadequate in identifying the causal inferences between financial literacy and indebtedness.

For instance, Brown and Graf (2013) examine whether Swiss households accrue adequate financial knowledge to make informative borrowing decisions. The Big Three questions are asked to a sample of 1,500 German-speaking individuals aged 20-74 in April 2011. For every single question, more than 70% of respondents give correct answers. The authors find that respondents who correctly answer all the Big Three questions are 10 pp more likely to carry a mortgage loan, but no significant difference is found in the incidence of consumer loan. This may be attributed to the authors' under-sampling of the low-income population with an immigration background, since this cohort may be screened out since they are not German-speaking.

Disney and Gathergood (2013) examine the association between financial literacy and households' credit portfolio based on a sample of 3,041 UK households. Financial literacy is measured by three financial knowledge questions on simple interest, compound interest and payment to credit card debt. The authors assert that those who show a lower level of financial

literacy are characterized by higher debt-to-income ratios. They also tend to hold a larger share of high-cost credit and end up with higher weighted average annual percentage interest rates (APR). The nexus between low financial literacy and high-cost borrowing is also documented in the US context by Lusardi and Scheresberg (2013).[79]

Lusardi and Tufano (2015) examine the extent to which over-indebtedness is related to financial literacy. Their analysis is performed on a sample of 1,000 US residents collected by a commercial market research company in 2007. Financial literacy is measured by three financial knowledge questions testing the respondents' understanding of interest compounding, payments to credit card debt and the time value of money, as well as a self-assessment of overall financial literacy. Over-indebtedness is measured by a self-assessed level of the debt burden. The authors find that those with higher subjective levels of financial literacy, or who understand interest compounding and the time value of money, are less likely to perceive being over-indebted.

In their research on the determinants of consumer debt and household net worth, French and McKillop (2016) investigate the two main parts of financial literacy, i.e., numerical skills and money management skills, separately. Their analysis focuses on UK credit union members, who are generally low-income and financially vulnerable, with either loans in arrears for more than nine weeks or loans that had been rescheduled. Compared to Agarwal and Mazumder (2013) who find that mathematical skills help people avoid high-cost borrowing, French and McKillop (2016) assert that, while numeracy has no impact on consumers' borrowing behaviours, money management skills are important to reduce the debt-to-income ratio, the use of high-cost credit and the number of different lenders. Money management skills also raise the households' net worth. These findings are robust to using an IV approach, as well as a

[79] Lusardi and Scheresberg (2013) use the Big Three questions to measure financial literacy.

constrained sample with only the household members who are in charge of financial decision-making.

Lusardi, Mitchell and Oggero (2019) evaluate the factors associated with older Americans' debt patterns and how these patterns evolve over time. Their analysis on the nexus between financial literacy and indebtedness is based on a sample of Americans aged 56-61 and 62-66 collected from the 2012 and 2015 waves of the National Financial Capability Study (NFCS). Financial literacy is measured by the number of correct answers to five financial knowledge questions that test the respondents' understanding of numeracy, knowledge of inflation, risk diversification, mortgages and basic asset pricing. The authors find that, for both cohorts, an additional correct answer to the financial knowledge questions is associated with a lower probability of self-reported over-indebtedness and financial fragility.[80]

Similar results are also witnessed in emerging markets, though the literature is very scant. For example, Grohmann (2018) investigates whether financial literacy leads to informative use of credit cards among the Asian middle class. The analysis is performed on a sample of 491 Bangkok adults who are younger than 60 and meet the criteria of the middle class in terms of income.[81] Several variables relevant to childhood experiences are employed as IVs for financial literacy, including the educational attainment and financial literacy of the parents, whether the parents taught budgeting and encouraged savings during the respondent's childhood, whether the respondent received economic education in school and the respondent's ownership of a bank account before 18. As revealed by the authors, financially literate individuals use credit

[80] In the NFCS, respondents are asked if they think they have too much debt (responses range from 1 to 7), and whether they could probably or certainly not come up with $2,000 within a month. In Lusardi, Mitchell and Oggero (2019), the former question is used as a measure of self-reported over-indebtedness, and the latter is used to measure whether the household is financially fragile.

[81] The cohort below 60 is considered to be financially active, since the mandatory retirement age is 60 in Thailand.

cards more informatively: they are more likely to know the interest rate on the credit card debt, and are less likely to have difficulty in meeting their repayment obligations.

Feng *et al.* (2019) investigate the association between financial literacy and households' holding of secured and unsecured debt in the Chinese context, based on a sample of 27,205 households from the 2015 wave of CHFS. Using the Big Three questions to measure financial literacy, the authors find that financially literate households are more likely to hold secured debt and less likely to hold unsecured debt.[82] Conditional on the positive holding of debt, the amount of either type of debt held by the household is higher for more financially literate households. The authors explain such a positive relationship reflects the financial constraints faced by the financially illiterate households. This argument may not stand since, though financial constraint may hinder households from getting secured debt as it requires collateral, it may not affect the approval of unsecured debt. Moreover, since the authors do not control for endogeneity issues, reverse causality may exist. In this case, the positive relationship between financial literacy may actually go the other way around; i.e., borrowing experiences improve households' financial literacy related to borrowing.

4.2.2.2 Financial literacy and formal/informal credit

Despite the extensive literature on financial literacy and overall indebtedness or high-cost borrowing behaviours, little is known about whether financial literacy affects individuals' choices between formal loans and informal loans. This topic is of interest in the context of developing countries, where financial inclusion is still at an early stage, and a large number of

[82] Feng *et al.* (2019) define secured debt as mortgage loans, car loans and any guaranteed loans related to agricultural and business activities. Unsecured debt includes credit card charges, credit loans for agricultural and business activities, student loans, and other loans. Unsecured debt is in general more costly than secured debt.

people face limited access to formal credit. Indeed, most of the research on formal and informal financing focuses on the emerging markets, especially China.

Cull *et al.* (2015) explore the determinants of the choices between formal and informal loans of Chinese households, based on a sample of 28,100 households drawn from the 2013 wave of the China Household Finance Survey (CHFS). Four proxies, including whether the respondent prefers risk-free returns compared with higher returns with high risk, the experience of financial education, whether the respondent understands compound interest and the real interest rate, are used to measure financial literacy.[83] The authors find that higher financial literacy tends to shift individuals' credit portfolios towards formal loans. More specifically, urban respondents who prefer risk-free returns compared with higher returns with high risk are more likely to borrow from banks and less likely to carry informal loans. Having attended financial classes raises the probability of holding bank loans and the level of bank loans of both rural and urban households. Being able to calculate compound interest increases rural respondents' probability of holding bank loans and the level of bank loans. Finally, understanding the concept of the real interest rate reduces the probability that rural households carry informal loans.

Using the same database, Lyons, Grable and Zeng (2019) also examine the links between financial literacy and formal and informal borrowing of Chinese households. Financial literacy is measured by whether the respondent has attended financial courses, whether the respondent pays attention to financial information, as well as their responses to the Big Three questions. Similar to Cull *et al.* (2015), the authors also control for supply-side variables, including social networks, the use of mobile technology and the infrastructure in the residence area. Overall,

[83] Formal loans in their research refer to bank loans, and informal loans refer to loans from relatives and friends and other informal lenders. A long-term loan refers to debt with maturity of more than one year. Among the four indicators for financial literacy, the "Risk for Return" measures the personal traits beyond financial literacy, such as risk attitude, which are found to be associated with borrowing behaviour (e.g., Lusardi and Scheresberg, 2013; Feng *et al.*, 2019).

the authors find that financially literate households are more likely to hold bank loans and less likely to hold non-bank loans. However, these findings are more pronounced for financially included respondents. Those who are illiterate, migrants, and those living in rural areas are less likely to be benefit from financial literacy in making borrowing decisions. Moreover, the relationship between financial literacy and formal/informal borrowing are consistent regardless of the purpose of debt, i.e., loans for business, agriculture, education or housing.

While Grohmann, Klühs and Menkhoff (2018) study the relationship between formal/informal borrowing and financial literacy on the country level, based on the World Bank's Findex data which documents the level of financial literacy for 143 countries. The authors find no significant relationship between financial literacy and the proportion of the population that borrow from formal financial institutions, but they do find a significantly negative relationship between financial literacy and the proportion of people who borrow from informal financial companies. These results are robust to an IV approach using the level of numeracy of primary school children as an instrument for country-level financial literacy.

4.2.3 Other determinants of indebtedness

An extensive literature has examined the determinants of indebtedness and borrowing behaviours.

Compared with the married or cohabiting, single people or lone parents are more likely to encounter over-indebtedness and report financial distress. This can be explained considering that the exploitation of economies of scale after marriage and cohabitation lower the debt demand of a household (Kempson, 2002; Del Rio and Young, 2006; Keese, 2009).

Education plays an independent role from financial literacy in explaining households' indebtedness. Li, Rozelle, and Zhang (2004) suggest that a minimum level of education is

required for borrowers to understand the terms of micro-loans and the passbooks for the micro-credit accounts. Lusardi, Mitchell and Oggero (2019) find that US respondents with at least high-school education are considerably less likely to be over-indebted or financially fragile.

Financial attitudes may shape borrowing behaviours, but the findings are mixed depending on the outcome variables. For instance, Meier and Sprenger (2010) reveal that individuals who are present-biased are more likely to hold consumer loans. A potential reason is that such individuals tend to have high discount rates and value current consumption relatively high. They are more likely to borrow to fulfil the demand for consumption in the current period. Brown and Graf (2013) find that myopic individuals are less likely to hold mortgages, presumably because they are less likely to make a long-term financial plan, hence they are less motivated to buy property as a means of investment.[84]

Risk attitude is found to closely correlate with borrowing behaviour. Risk seeking individuals are more likely to use high-cost debt instruments, such as AFS (Lusardi and Scheresberg, 2013). Such people are also more likely to bear the debt, and have higher levels of debt, regardless of the type of debt (Feng *et al*., 2019). Conversely, people who are risk-averse are less likely to have debt larger than their annual income.

Social networks may shape households' borrowing behaviour in two ways. On the one hand, social networks can affect households' borrowing behaviour through peer effects. Georgarakos, Haliassos and Pasini (2014) find that the higher the perception of the income of the social circle, the more likely it is that the household has loans. One potential explanation is that households try to catch up with their peers in terms of living and spending standards. Meanwhile, they expect higher income, such as their peers in the future, which would enable them to service the debt. On the other hand, large social networks comprising parents, siblings, relatives and

[84] Brown and Graf (2013) define a household who fully or partially agrees to the statement "I live for the present and don't think about my financial future" as myopic.

friends, imply an informal insurance mechanism and adequate access to informal finance. Prior studies have found that respondents with more siblings are more likely to take on informal finance (Cull et al., 2015). Credit provided by the social network is particularly important in economies where formal finance markets are not well established, and in the context where people face severe financial constraints (Gan, Xu and Yao, 2012; Cull *et al*., 2015, 2016).

Households with more children are more likely to have home loans or report excessive debt, partly because they are more likely to become homeowners and, consequently, have larger burdens from mortgages (Keese, 2009; Lusardi, Mitchell and Oggero, 2019).

While most of the studies find a significant determinant role of income in borrowing behaviour, the direction of its effect can differ depending on the contexts and the outcome variables being investigated. For instance, Betti *et al*. (2007) assert that in European countries where people have broader access to consumer credit, high-income households are more likely to be overindebted compared with low-income households. The authors attribute this relationship to the existence of market imperfection, where the high-income households enjoy much better access to credit. Similarly, Keese (2009) finds that German households with higher income are more likely to hold consumer credit and housing loans. The author explains that high-income households are more capable of servicing debt, and their debt level is more tolerable compared with the low-income household. By contrast, Lusardi, Mitchell and Oggero (2019) assert that among US individuals at the verge of retirement, those with higher income are less likely to be indebted.

Income shocks are another reason for both arrears and financial difficulties (Keese, 2009; Lusardi, Mitchell and Oggero, 2019). A job loss increases the probability of holding consumer credit and home loans, while being unemployed does not influence the incidence of loans. This implies that households who experience job loss are more likely to rely on credit to smooth out consumption, but only in the short term (Keese, 2009). The income shocks to households are

also reflected in the increasing debt burden: Those who experienced a large and unexpected income drop in the previous year are more likely to define themselves as over-indebted and financially fragile (Lusardi, Mitchell and Oggero, 2019). In the same vein, people who have experienced health shocks are found more likely to be indebted (Mohanan, 2012), since the shocks bring large out-of-pocket healthcare expenses on top of a potential loss in income.

4.2.4 Research gaps in the literature

The literature on financial literacy and households' borrowing behaviour is still limited (Zinman, 2015). Most of the relevant studies focus on the extent of indebtedness and the use of high-cost credit in developed countries. Little research has been conducted on emerging markets, which have been experiencing a boom in household debt. Consequently, existent studies have paid little attention to households' borrowing behaviour in terms of formal and informal credit, which is of specific interest to emerging economies where financial markets are underdeveloped, and a large population does not have access to formal credit. Moreover, only a few studies on financial literacy and indebtedness are able to draw the causal inference.

This research tries to fill this gap by exploring how financial literacy is related to Chinese households' indebtedness, and how it affects households' choice between formal and informal credit.

4.3 Development of Hypotheses

The life-cycle model posits that individuals smooth consumption across stages of life. More specifically, the model predicts that individuals borrow before entering the labour market, accumulate wealth during the working life, and dissave after they have retired (Modigliani and Brumberg, 1954; Browning and Crossley, 2001). Yet, it is found that similar households approach retirement with very different wealth levels (e.g., Bernheim and Garrett, 2003; Lusardi and Mitchell, 2011a; Lusardi, Michaud and Mitchell, 2017). The literature on retirement wealth suggests that financially literate individuals save for retirement and accumulate higher wealth (e.g., Lusardi and Mitchell, 2011a). Consequently, financially literate individuals are less likely to hold debt and/or hold lower levels of debt, presumably because they will have accumulated adequate resources and will find borrowing unnecessary. Moreover, financially literate individuals are found to be more sophisticated in financial behaviours. For instance, they set up long-term financial planning, assess the affordability of expenditures, and monitor household budgets, and so on (Agarwalla *et al.*, 2015). Therefore, they are more aware of the costs and consequences of indebtedness, and will avoid borrowing if possible. In light of these considerations, we put forward the following hypotheses:

H1: Financially literate households are less likely to be indebted, and have a lower level of debt.

Although a higher level of financial literacy may reduce the probability and level of overall indebtedness, it may affect formal credit and informal credit in different ways. Previous studies have shown that financial literacy enhances voluntarily financial inclusion, such as taking up bank accounts, online banking and digital financial services (Servon and Kaestner, 2008;

Königsheim, Lukas and Nöth, 2017; Grohmann, Klühs and Menkhoff, 2018). Therefore, financially literate individuals are likely to have wider access to formal credit compared with the less privileged.

While informal credit from relatives and friends can be relatively cheap in terms of the interest rate, its hidden cost can be high. In the spirit of mutual aid, the lenders are willing to accept lower monetary interest rates because they expect broader potential reciprocity, such as emergency medical aid, valuable information sharing, or emotional support, in the long term (Nahapiet and Ghoshal, 1998; Franzen and Hangartner, 2006; Lin, 2017). Informal credit from other private lenders, such as informal financial companies, can also turn out to be very costly, in terms of high interest rates. Since financially literate households are less likely to use high-cost credit, they may also be less likely to take on informal credit.

Therefore, we propose the following two hypotheses regarding the holding and levels of formal loans and informal loans, respectively:

H.2A: Households with higher levels of financial literacy are more likely to borrow from formal sources and less likely to borrow from informal sources.

H.2B: Households with higher levels of financial literacy hold higher levels of formal loans and lower levels of informal loans.

4.4 Dataset and Sample Construction

Data for this study are collected from the 2014 wave of the CFPS. The CFPS is a nationally representative longitudinal survey of Chinese communities, households, and individuals, administered by Peking University. The 2014 wave comprises data from 12,000 families and 32,307 respondents in 621 communities across 25 provinces. Detailed information on socio-demographics, household composition, household consumption and wealth, as well as individuals' financial literacy are collected.

In the CFPS, only the family member who is most familiar with the family's financial situation is asked the financial literacy questions. Therefore, this research is limited to the 3,380 respondents who answered the 13 financial knowledge questions. We further restrict our sample to respondents aged 18 and over, and who report being most familiar with the economic condition in the household. We also drop observations that have missing values on the variables included in our models. After making these adjustments, our final sample contains 2,237 observations.[85] Since some of the variables we use in this study, such as income, wealth and consumption, are sensitive to outliers, we winsorize the top and bottom 1% of these variables.

4.5 Methodology and Specifications

4.5.1 Dependent variables

The CFPS asks several questions on household's borrowing behaviour in the Family Assets module. This enables us to obtain a thorough and full insight into the households' indebtedness. We use two sets of dependent variables. The first measures whether the households are

[85] All the variables used in this research are listed in Appendix 4.1 and will be discussed in detail in Section 4.5.1.3.

indebted, and the levels of household debt. The second set measures the households' preferred sources of credit.

4.5.1.1 Use of debt

We construct six dummy variables reflecting the households' use of debt from either formal sources or informal sources.

The first dummy variable we create (*D_debt*) aims to measure the household's overall indebtedness, considering both housing loans and non-housing loans. The variable equals 1 if the household has any types of outstanding debt owed to the bank, relatives, friends or other individuals and institutions other than banks (such as private loan institution), and 0 otherwise.

We then focus on the non-housing debt to further verify our hypothesis about the relationship between financial literacy and indebtedness. To this end, we construct the second dummy variable (*D_NH_loan*), which equals 1 if the household has non-housing loans, and 0 otherwise.

In our sample, 25.6% of the respondents report holding either mortgages or non-housing loans, and 12.7% of the respondents report holding loans other than mortgages (See Table 4.1).

Next, for overall indebtedness, we further construct two variables to distinguish between formal loans and informal loans. We construct a dummy variable *D_formal*, which equals one if the household has mortgages or non-housing loans owed to the banks, and 0 otherwise. We construct another dummy variable, *D_informal*, which equals one if the household holds mortgages or non-housing loans borrowed from its friends, relatives, other individuals or institutions other than banks (informal institutions hereafter), and 0 otherwise. In our sample, around 12.2% of the respondents report that their families ever borrowed from banks, and 17.7%

report that they ever borrowed from friends/relatives, other individuals or informal institutions (See Table 4.1).

Similarly, we also create two dummy variables, namely *D_NH_bank* and *D_NH_inform,* to identify whether the household owes formal non-housing loans or informal non-housing loans, respectively. In our sample, only 3.7% of the respondents report that they ever owed non-housing loans to banks, and 10.2% report that they ever owed non-housing loans to their friends/relatives, other individuals or informal institutions (See Table 4.1).

In the CFPS questionnaire, if the household admits holding one type of debt, they are further asked about the amount of this type of debt it holds. Based on these questions, we further create six continuous variables on the amount of debt held by the household, each corresponding to one of the six dummy variables mentioned above. More specifically, we use *Total_debt* and *Total_NH_loan* to measure the total amount of household debt and the amount of non-housing loans, respectively. The respondents in our sample have an average debt level at 29,300 RMB (see Table 4.1).

We then divide total debt into formal and informal components, and construct two variables: *Debt_formal* and *Debt_informal*. Similarly, we construct *Debt_NH_bank* to measure non-housing loans from banks and *Debt_NH_inform* to measure non-housing loans from friends, relatives and other informal sources. Compared with informal loans, a larger proportion of total debt comes from banks (16,945 vs 11,106 RMB). In contrary, a larger proportion of the non-housing loans are informal loans (2,163 vs 4,256 RMB, see Table 4.1).

We discuss debt holdings in more detail in Section 4.6.1.

4.5.1.2 Attitudes towards formal and informal loans

Our second set of variables measures individuals' attitudes towards the source of the funds. While the use of debt represents the alignment of the supply of and demand for funds, the preferences reflect what households' real attitudes towards formal and informal credits are in the absence of external constraints such as access to credit. Such preferences are more closely related to the households' real demand. As such, they are likely to be shaped by their background characteristics, such as financial literacy.

The relevant question asked in the CFPS is the following:

*If your family needs to borrow a large amount of money (e.g. purchasing a house, operational turnover), the first choice of borrowing money is from*_____

1. Relatives 2. Friends 3. Bank 4. Formal financial institution other than bank 5. Individual or private loan institution 6. Will not borrow money under any condition.

The households who choose the first, second or fifth options are considered to prefer borrowing from informal sources. Those who choose the third or fourth option are considered to prefer borrowing from formal sources. Accordingly, we create two dummy variables, namely *Pref_bank* and *Pref_inform*, to indicate the household's preferred source of finance.

4.5.2 Financial literacy

4.5.2.1 Objective financial knowledge

The financial literacy module in the CFPS includes five questions testing basic financial concepts such as the working of simple interest and compound interest (Questions 1, 2 and 3),

the real interest rate and inflation (Question 4) and the time value of money (Question 5).[86] These concepts are essential to everyday economic and financial decision-making, financial planning, and borrowing behaviours (Stango and Zinman, 2009a; Disney and Gathergood, 2013; van Rooij, Lusardi and Alessie, 2011; Lusardi, Mitchell and Oggero, 2019). Hence, we construct a score for the individual's financial literacy, which counts the number of correct answers to the five questions.[87]

4.5.2.2 Subjective financial knowledge

In addition to asking questions about specific concepts of financial knowledge, the questionnaire of CFPS 2014 also asks respondents to assess their financial knowledge. The wording of the relevant question is as follows:

"In your opinion, your financial knowledge is, compared to the people of your age: 1. Way above the average level 2. Above the average level 3. About average level 4. Below the average level 5. Way below the average level".

Based on this question, we construct a subjective measure of financial knowledge which ranges from 1, indicating way below the average, to 5, indicating way above the average.

[86] The exact wording of the five relevant questions is displayed in Appendix 3.2.2 in Chapter Three. The 2nd and 3rd questions are identical to the corresponding questions used in the OECD report (OECD, 2016) and the DNB Household Survey (DHS) (e.g., van Rooij, Lusardi and Alessie, 2011). The 4th and 5th questions are almost identical to those in the DHS survey. The statistics relative to the financial knowledge questions in the OECD and DHS questionnaires appear in Section 3.4.2.2 of Chapter Three. They are therefore not discussed in detail here.

[87] Each of the questions gives four options including one option saying "Don't know", which we consider as an as incorrect answer, i.e., the respondents score 0 if they reply "Don't know".

4.5.3 Baseline specifications

4.5.3.1 Baseline specifications and estimation method

We start with a set of simple models for the binary dependent variables indicating whether the household is indebted and whether it prefers formal or informal credit. Our baseline model takes the following form:

$$Y_{i,j} = 1\big(\beta_{1,j} * FK_B_i + \beta_{2,j} * Subj_FK_i + \gamma_j * X_i + Province\ dummies + u_{i,j} > 0\big), (i = 1, \dots, N; j = 1, \dots, 6) \quad \textit{Eq. 4.1}$$

where $Y_{i,j}$ represents the jth binary dependent variable for household i. FK_B_i and $Subj_FK_i$ denote the objective and subjective level of financial knowledge of the head of household i, respectively.

X_i denotes control variables. Following the literature on indebtedness and borrowing behaviour (Disney and Gathergood, 2013; Lusardi and Scheresberg, 2013; Lusardi and Tufano, 2015; Lyons, Grable and Zeng, 2019), we consider:

(i) Individual-level variables such as age, gender, marital status, education, cognitive ability, migrant status, working status, risk attitude, financial attitude, whether the respondent has experienced a health shock in the past two years, whether the respondent uses a mobile phone.

(ii) Household-level variables such as whether the household resides in a rural area, has close connections with relatives, the number of dependent children and dependent elderly, house ownership, household income and total assets (including financial assets and non-financial assets).

(iii) Community-level variables such as the facilities and infrastructure in the community where the household resides.[88]

Finally, $u_{i,j}$ denotes the error term.

For each dependent variable, we estimate three models which differ in the explanatory variables included. In the first model, for each dependent variable, we include only the objective and subjective measures of financial literacy as explanatory variables. In the second model, we additionally control for individual-level variables. In the third model, we include a full set of control variables, including individual-, household- and community-level variables. Provincial dummies are controlled for in all the specifications.

For Hypothesis 1 to hold, we expect to observe negative estimates of $\beta_{1,j}$ and $\beta_{2,j}$. For Hypothesis 2A and 2B to hold, we expect $\beta_{1,j}$ and $\beta_{2,j}$ to be negative in the models for informal loans, and positive in the models for formal loans.

As for the control variables, following the literature (e.g., Lusardi and Scheresberg, 2013; Cull *et al.*, 2015), we expect households with heads who are young, male, unmarried and risk-seeking to be more likely to take on debt, since they are in general more risk-taking than their counterparts. We expect that those who are illiterate and with lower cognitive ability are more likely to borrow, since they are less likely to understand the risk and cost of carrying debt. We also expect that those who are migrants, unemployed, have experienced health shock, use mobile phones, hold less favourable financial attitudes (e.g., money is to be spent), have more dependent children, and a lower level of wealth and income are more likely to borrow to make ends meet.

[88] The two community variables are constructed in the same way as in Chapter Three (see Page 110). Definitions of all the control variables are displayed in Appendix 4.1. We also highlight some of the control variables in Section 4.5.3.2.

Among the households who use credit, we expect those with heads who are female, illiterate, with lower cognitive ability, migrant, risk-seeking to be more likely to borrow from informal sources compared with borrowing from banks, since they may be unaware of the higher risk and cost of informal loans. We expect those who are unemployed, have experienced health shocks or have a lower level of wealth and income more likely to borrow from informal sources since they are less likely to provide adequate collaterals for bank loans. We also expect those who have close relatives more likely to have informal loans from these relatives. We also expect those who live in communities with worse infrastructure and facilities more likely to choose informal loans since they may not have access to bank's credit services.

Similarly, the baseline model for the intensive margins, i.e., the amount of debt, will take the following form:

$$Y_{i,k} = \text{Max}(0, \beta_{0,k} + \beta_{1,k} * FK_B_i + \beta_{2,k} * Subj_{FK_i} + \gamma_k * X_i + Province\ dummies + u_{i,k})$$

$$(i = 1, \dots, N;\ k = 1, \dots, 6), \qquad \textit{Eq. 4.2}$$

where $Y_{i,k}$ denotes the amount of a specific type of debt k held by household i, expressed in logarithm. All control variables are the same as in Eq. 4.1. For Hypothesis 1 to hold, we expect $\beta_{1,k}$ and $\beta_{2,k}$ to be negative. For Hypothesis 2A and 2B to hold, we expect $\beta_{1,j}$ and $\beta_{2,j}$ to be negative in the models for informal loans, and positive in the models for formal loans. The signs of the control variables are expected to be similar to those discussed with reference to Eq.4.1

4.5.3.2 Discussion of selected control variables

In this section, we highlight some control variables worth considering in the context of emerging markets, but are not included in most of the literature on financial literacy and debt.

First, we include a dummy variable indicating whether the household is a rural-to-urban migrant. Rural-to-urban migrants are commonly identified as those who are registered in a rural area, i.e. hold an agricultural hukou, but currently reside in an urban area which does not match their hukou registration (e.g., Zhao, 1999; Keung Wong, Li and Song, 2007; Gong *et al.*, 2012). In the Chinese context, migrants are considered to be financially excluded and are found to be less knowledgeable on the use of formal loans, while non-migrants are more likely to borrow from banks and less likely to borrow from informal sources such as friends, relatives and informal institutions (Lyons, Grable and Zeng, 2019).

As discussed in Section 4.2, most of the empirical studies use individuals' financial knowledge to proxy for financial literacy, despite the fact that financial literacy is a multi-facet concept that covers not only financial knowledge, but also financial attitudes and financial behaviours (Atkinson and Messy, 2016). Taking this into consideration, we include three dummy variables for financial attitudes towards spending and saving. Specifically, in the 2014 wave of the CFPS, three questions on the respondents' agreement with the following statements are asked: "*Spending money makes me more satisfied than saving money", "I intend to live in the present more and do not consider the future*" and "*In my opinion, money is to be spent and we should spend it whenever we have money*". Five options are given for each of the questions, ranging from 1 (Totally inapplicable) to 5 (Totally applicable). Those who answer "Totally inapplicable" are considered to hold the most sophisticated financial attitudes that lead to sound financial decisions. Accordingly, for each of the financial attitude questions, we include a discrete variable ranging from 1 to 5.

Infrastructure (either physical, technological or informational) may also shape households' borrowing behaviour, especially in developing countries where financial inclusion is still at its early stage (Lyons, Grable and Zeng, 2017; Lyons, Grable and Zeng, 2019). Therefore, we construct two indices to measure the adequacy of facilities and infrastructure in the community by performing factor analysis (see details in Table A4.1).

Technological innovations bring new customers, potentially including millions of unbanked mobile phone owners, into the financial market (Cohen and Nelson, 2011). Given the dramatic change in the landscape of financial services brought by digital finance, we also include a variable indicating whether the respondent uses a mobile phone. This may be important, especially in explaining the trade-off between formal and informal finance.

The pairwise correlations among all the explanatory variables are reported in Table A.4.2.1. The correlations are generally low, and we do not expect multicollinearity to be problematic in our research.

4.5.4 Estimation methodology

We estimate the models with binary dependent variables using Probit models, and the models for debt levels using Tobit models. However, the Probit and Tobit estimates of Eq.4.1 and Eq.4.2 may be biased as the relationship between financial knowledge and indebtedness may be subject to endogeneity issues. First, financial literacy may be measured with substantial error. The responses to the financial knowledge questions may result from guessing (van Rooij, Lusardi and Alessie, 2011) and the self-perceived level of financial literacy is subject to justification bias (Hastings, Madrian and Skimmyhorn, 2013). Second, unobserved time-invariant individual characteristics, such as innate ability and motivation, may influence

households' debt behaviour and, at the same time, households' financial knowledge. In any of these cases, the observed financial literacy may be correlated with the error term. As a consequence, the coefficient/marginal effect associated with financial literacy may suffer from downward bias (Hyslop and Imbens, 2001). Also, reverse causality may occur if people who intend to borrow enhance their financial literacy in the process. In that case, it would be that borrowing decisions affect financial literacy, rather than the other way around. This may overestimate the effect of financial literacy on debt behaviour (Hilgert, Hogarth and Beverly, 2003).

To tackle this problem, we adopt the IV method. Following Behrman *et al.* (2012) and Jappelli and Padula (2017), the first IV we select for the respondent's financial knowledge is the mother's duration of education. It has been found in the literature that higher levels of parental education are associated with higher scores on financial knowledge questions (Lusardi, 2009, P.263; Jappelli and Padula, 2017). Additionally, it is likely that parental education is related to one's financial behaviours only via its effect on one's financial literacy, after considering the full set of control variables.

Following Bucher-Koenen and Lusardi (2011) and Calcagno and Monticone (2015), the second IV we select is the average level of financial literacy in the community excluding the respondent.[89] On the one hand, the community level of financial literacy is likely to increase individual financial knowledge through social interactions. On the other hand, financial literacy of the others is beyond the control of the respondents. Similar to Chapter Three, we only consider the communities which have at least five respondents (the Com_5 sample, with 1,771

[89] A very scant literature investigates the relationship between financial literacy and indebtedness, and very few papers in this literature make use of an IV approach. Therefore, we refer to the literature on retirement planning, saving behaviours, financial advice and consumption to select our IVs.

observations) when performing the IV method, to address the concerns on the representativeness of the community average.

The IV estimates are first derived using the two-stage least square (2SLS) estimation method. However, since both the binary and continuous dependent variables in this study are censored, we also exploit a conditional mixed-process (CMP) framework. In a CMP framework, the second stage (where we regress the dependent variable on the IVs) allows us to capture the censored nature of the dependent variables and produces more efficient estimates (Roodman, 2011).

4.6 Main Results and Discussion

4.6.1 Summary statistics

4.6.1.1 The use of debt

Columns 1 to 4 of Table 4.1 report descriptive statistics of the key variables for the full sample with 2,237 observations. Columns 5 to 10 further report the statistics by indebtedness, preference for formal loans and preference for informal loans, respectively.

The first panel of Table 4.1 presents the statistics of the dependent variables. We can see that 25.6% of the sample are in debt, and 12.7% of the sample have a non-housing loan. The average amount of total debt is around 29,300 CNY, and the average amount of non-housing debt is around 7,874 CNY. Compared with the average annual household income of 61,743 CNY, the debt burden does not seem to be problematic. However, for those who are indebted (Column 6), the average amounts of total debt and non-housing debt are around 114,306 CNY and 30,719 CNY, respectively. These figures are rather high compared with the average annual household income of those who are indebted (60,219 CNY).

A point worth noting is that, for the full sample of the 2014 wave of the CFPS, the total debt per capita is approximately 9,459 CNY, whilst the bank debt per capita is approximately 4,519 CNY.[90] These figures are much lower than the corresponding figure reported by the PBC (PBC, 2014, 2015). Specifically, the PBC documents that the bank loans per capita was approximately 14,595 CNY by the end of 2013 and CNY 16,926 by the end of 2014. A potential reason is that Chinese households may be reluctant to report indebtedness, as borrowing for consumption purposes is frowned upon in Chinese society. A similar underreporting of consumer loans is

[90] These figures are calculated for all the respondents who participated in the 2014 wave of CFPS and are not included in the summary statistics.

also found by Karlan and Zinman (2008) in the US and by Brown and Graf (2013) in Switzerland.

Similar to previous studies in the Chinese context (e.g., Cull *et al.*, 2015; Fan, Wu and Yang, 2017), Chinese households rely heavily on informal loans. Putting mortgages and non-housing loans together, 12.2% of the full sample report holding formal loans and 17.7% report holding informal loans. As for solely non-housing loans, only 3.7% of the full sample are in debt to banks, and 10.2% owe loans to relatives, friends or other individuals or informal institutions. The average total formal loan of the full sample is around 16,945 CNY, higher than the average total informal loan of 11,106 CNY. This is as expected since most of the formal loans are borrowed in the form of mortgages. As for the non-housing loans, on average, people hold less in formal loan (2,163 CNY) than in the informal loans (4,226 CNY).

4.6.1.2 Attitudes towards formal and informal loans

As for the attitudes towards formal and informal finance, around 25.8% of the respondents in the full sample prefer to borrow from banks, and 57% prefer to borrow from informal sources. It seems that households' attitudes towards formal and informal loans are related to the actual use of debt, but only on a limited scale. The correlations between the preference for formal loans and the probability of holding any type of formal loan is only 0.1624.[91] The correlations between the preference for informal loans and the probability of holding any type of informal loan is only 0.1873. Both correlations are significant at the 1% level. The mismatch between attitude and actual use of debt reflects the fact that people's borrowing behaviour may be constrained by external factors, such as access to formal and informal credits.

[91] The correlation matrix for all the dependent variables is reported in Table A4.2.2.

4.6.1.3 Financial literacy

The second panel of Table 4.1 reports the average score of financial literacy. On average, the respondents in our sample can answer less than three out the five financial knowledge questions correctly, and perceive their financial knowledge level as below average. Households who are in debt, on average, score lower in the financial knowledge test and have a lower perceived level of financial knowledge (Columns 5 and 6). For households who prefer formal finance, the average scores for either the objective or subjective financial knowledge are higher than those who prefer informal finance (Columns 8 and 10).

Table 4.2 reports the distribution of the answers to each financial knowledge question. We can see that only 55.30% of respondents know the annual interest rate of a one-year fixed deposit (Question 1). This percentage is strikingly low since China is well-known for its high saving rate. About 43.9% of our sample can calculate simple interest (Question 2), but, surprisingly, around 50% of our sample can calculate compound interest (Question 3). While only 25.7% of the sample answer both the 2nd and 3rd questions correctly. About 59.28% of respondents understand the concept of real interest (Question 4), and 68.84% are aware of the time value of money (Question 5). Overall, 14.09% of respondents cannot answer any of the questions correctly, and less than 15% can answer all the questions correctly. These statistics reflect a generally low level of financial knowledge in our sample, compared with that reported by Alessie, Van Rooij and Lusardi (2011) for the Netherlands, where the correct rate for each single financial knowledge question is over 70%, and 40.2% of the sample correctly answered all the five questions.

The distribution of subjective financial knowledge is reported in Table 4.3. We observe that 39.92% of respondents report their financial knowledge as about average and more than half of the respondents report their financial knowledge as below average (54.45%). However,

potential misperception may exist. As reported in Table A.4.2.1, the pairwise correlation between the objective and subjective level of financial knowledge is only 0.271, significant at the 1% level. This implies that subjective financial knowledge may measure something beyond the test score, which confirms the necessity to include subjective and objective measures of financial literacy simultaneously in our models.

Next, we plot the dependent variables by the objective financial knowledge score (see Figure 4.3). Graphs (1) to (3) illustrate declining trends in the probability of holding debt and holding informal loans, as well as an upward trend in the probability of holding formal loans, as the test score increases. For the non-housing loans, we observe declining trends in the probability of holding non-housing loans (Graph 4) and informal non-housing loans (Graph 6) as the objective financial knowledge score increases. But we do not observe a clear trend in the incidence of formal non-housing loans (Graph 5). We also observe an increasing preference for formal loans as the objective level of financial knowledge rises (Graph 7). But we do not find a significant trend in households' preference for informal loans along the objective financial knowledge scores (Graph 8).

Graphs (9) to (14) show the average debt level for each objective financial knowledge score. In general, the total/non-housing amounts of debt (Graphs 9 and 12) and the amount of formal total/non-housing loans (Graphs 10 and 13) increase with the objective level of financial knowledge. However, we do not observe significant trends in the levels of informal total/non-housing loans as objective financial knowledge score increases (Graphs 11 and 14).

4.6.1.4 Control variables

The third to fifth panels of Table 4.1 report the summary statistics for the individual-, household- and community-level control variables, respectively.

By comparing the control variables for the households who are indebted (Columns 5) with those who are not (Column 6), we observe that the indebted households are more likely to have younger heads, be married, migrant, be employed or self-employed, be risk-seeking, use mobile phones, have close relatives, be a large household, have more dependent children and less dependent elderly, as well as lower total assets.[92]

Compared with those who prefer informal credit (Column 10), the households who prefer formal credit (Column 8) are more likely to be headed by males who are married, literate, have higher cognitive ability, live in an urban area, are employed, are risk-seeking, use mobile phones and and have higher total assets and higher income.

4.6.2 Financial literacy and indebtedness

4.6.2.1 Are the financially literate less likely to borrow?

4.6.2.1.1 Probability of indebtedness

The marginal effects of financial literacy in models for whether the household is indebted and whether it holds non-housing loans are reported in the first and second panels of Table 4.4, respectively. Standard errors are adjusted for heteroskedasticity, allowing for clustering at the community level.

As discussed in Section 4.5.3, we start with three Probit models for the likelihood of having loans, which differ in the control variables included. The first three columns of Table 4.4 report Probit estimates for the specifications which include (i) financial knowledge only, (ii) financial knowledge and individual-level covariates, and (iii) financial knowledge and the full set of

[92] Though not reported for brevity, the differences in the other control variables between the two groups are insignificant. The same applies to the next paragraph.

control variables, respectively, as explanatory variables. This same structure also applies to Table 4.5 and Table 4.6. Province dummies are controlled for in all the specifications hereafter.

In Table 4.4, we observe that those households with a more financially literate head are less likely to hold any types of loans, and also less likely to hold non-housing loans (Column 3, Panels 1 and 2). These effects of objective financial knowledge persist regardless of which control variables are included. More specifically, households who correctly answer one more financial knowledge question are, on average, 1.56 pp less likely to hold any types of debt (Column 3, Panel 1), and 1.19 pp less likely to hold non-housing loans (Column 3, Panel 2). By contrast, we only observe weak evidence that one point higher in the score of subjective financial literacy is associated with a 1.92 pp higher likelihood of being indebted (Column 3, Panel 1). The findings from the baseline specifications support our Hypothesis 1, according to which respondents with higher financial literacy are less likely to hold debt. Our results are consistent with Feng *et al.* (2019) in the Chinese context, who find that financially literate households are less likely to hold unsecured debts (which cover a variety of non-housing loans).[93]

We then exploit the IV approach to tackle the potential endogeneity issue. The estimated marginal effects of the second stage regression are also reported in Table 4.4. Specifically, Column (4) reports the IV estimates of the second stage of the 2SLS estimation, and Column (5) reports the CMP estimates, using OLS in the first stage and Probit in the second stage. The respondent's mother's duration of education and community-level financial knowledge are used as IVs. Several validity tests for the IVs are performed, and the results for the tests are reported in Appendix 4.3. The 2SLS and CMP estimates remain statistically negative for both

[93] See more details in Footnote 82.

outcome variables, and are even larger than the Probit estimates. Hence, our Hypothesis 1 is further verified by the IV approach.

The results also reveal that the objective and subjective levels of financial literacy affect the probability of indebtedness differently; the former is more predictive. Yet, while subjective financial literacy shows a weak effect on indebtedness when controlling for the full set of control variables and using the IV approach, the marginal effects of objective financial literacy are always significant at the 5% or 1% level. One explanation for the weak predictive power of subjective financial literacy compared with the objective measures is that the actual use of debt is subject to the scrutiny from the lender regarding the borrower's capacity to repay, which makes how good they actually are matters more than how they think they are.

The Probit estimates for the other explanatory variables are reported in Table A4.4, which are as expected and are in line with the literature (Keese, 2009; Meier and Sprenger, 2010; Feng *et al.*, 2019; Lusardi, Mitchell and Oggero, 2019). More specifically, we find that households with heads who are younger, risk-seeking, have experienced health shocks, use mobile phones, and households with more dependent children, and live in communities with worse facilities are more likely to borrow (Column 1). Similar results are found for the holding of non-housing loans, except that we find no significant marginal effect for risk-seeking and facility quality. In addition, we find that households who have lower income or have heads who are self-employed are more likely to hold non-housing loans (Column 2). One point worth noting is that, in stark contrast with financial literacy, we hardly find any significant role of cognitive ability and general literacy when we controlling for household income and wealth. This implies that, if we hold various SES constant, it is financial literacy, rather than education or cognitive ability, that determines whether one hold debt and the type of debt, i.e., formal loans and informal loans.

4.6.2.1.2 Amount of outstanding debt

We further examine the relationship between financial knowledge and the outstanding amount of debt of the households. Tobit estimates for the marginal effects of financial knowledge are reported in the first three columns of Table 4.5. All the marginal effects are estimated on the expected value of the truncated outcome, i.e., the amount of debt that is not less than zero. Focusing on total debt and non-housing debt, the estimates suggest that, on average, a one-point higher financial knowledge score is associated with a 2.10 pp lower total amount of debt (Column 3, Panel 1), and a 1.21 pp lower amount non-housing debt (Column 3, Panel 2). Therefore, the second part of Hypothesis 1 is verified.

Column (4) reports the 2SLS estimates. Column (5) reports the estimates for the CMP framework, in which the first stage is estimated by OLS, and the second stage is estimated by a Tobit model. The results of the validity tests for the IVs are reported in Appendix 4.3. As shown in Panels 1 and 2 of Table 4.5, these estimates further confirm the baseline result that financial knowledge is negatively related to the amount of total and non-housing debt.

These findings are in line with Disney and Gathergood (2013) and French and McKillop (2016), who find that financially literate households tend to have a lower debt-to-income ratio in the US and UK contexts. However, our findings are inconsistent with Feng *et al.* (2019), who find that, in the Chinese context, the level of financial literacy is associated with a higher amount of secured or unsecured household debt. The authors attribute the positive relationship between financial literacy and debt level to the financial constraints faced by financially illiterate households. However, while financial constraints may hinder the financially illiterate households from using secured debt, it may not impede the use of unsecured debt which does not require collateral. Moreover, the positive association between financial literacy and debt levels detected by Feng *et al*. (2019) may result from reverse causality, an issue which they do

not address in their research. Apart from that financial literacy leads to lower likelihood of indebtedness and debt levels, there is also such a situation where various borrowing experiences reversely improve households' financial literacy on borrowing. Therefore, if the negative association in the former direction is lower than the positive association in the latter direction in magnitude, we will only observe a positive association between financial literacy and debt levels, as is observed by Feng *et al*. (2019).

The Tobit estimates for the other explanatory variables are reported in Table A4.5. Similar to the case of the extensive margins, we find that households with heads who are younger, risk-seeking, having experienced health shocks, use mobile phones, and households with more dependent children, and live in communities with worse facilities on average show higher debt levels. However, we hardly observe any evidence that cognitive ability and general literacy are associated with debt levels. Similar patterns are found for the non-housing loans, except that we do not find risk-seeking households hold more non-housing loans. But additionally, we observe that those who are self-employed and have lower income show higher non-housing debt levels.

Overall, the baseline and IV estimates in Table 4.4 and Table 4.5 suggest that financial knowledge is an important predictor of indebtedness: Respondents who have low financial knowledge are more likely to be overall indebted and hold higher levels of total debt. This result still holds when we focus on only non-housing loans. Therefore, Hypothesis 1 is satisfied.

4.6.2.2 Formal loans and informal loans

4.6.2.2.1 Financial literacy and the use of formal/informal loans

Next, we investigate the extent to which financial knowledge influences households' use of formal loan and informal loans.

Panels 3 and 4 of Table 4.4 report the Probit and IV estimates for the marginal effects of financial knowledge. We observe no significant evidence that subjective financial literacy is related to *D_formal,* but we do observe that households with higher levels of objective financial knowledge are less likely to hold bank loans (Column 3 and 4, Panel 3).[94] Specifically, one more correct answer to basic financial knowledge question is associated with a 1.43 pp lower likelihood of holding formal loans (Column 3, Panel 3). This finding is further confirmed by looking at only the non-housing debt (Column 3 and 4, Panel 5). However, these findings contradict the formal credit part of our Hypothesis 2A, presumably because that financially literate households refrain from borrowing in the first place, so that the positive relationship between financial literacy and formal borrowing is offset by the overall negative relationship between financial literacy and borrowing.

This inference can be further clarified by the debt position across levels of objective financial knowledge. Figure 4.4 plots the fraction of respondents who do not hold any debt (non-indebtedness), hold formal loans only, hold informal loans only and hold both types of loans, respectively, by their score on basic financial knowledge questions. The graph shows that, except for those who score the lowest in basic financial knowledge questions, the respondents with higher levels of basic financial knowledge are less likely to hold debt (the blue area). We can also find from the graph that, though the formal loan holdings of those with

[94] Since the specification for *D_formal* does not reject the endogeneity test, we rely on the Probit estimates to explain the relationship between subjective financial literacy and *D_formal*.

higher levels of objective financial literacy tend to be higher, the trend may not be as significant as the trend in non-indebtedness (the blue area). Therefore, the positive association between financial literacy and the incidence of formal borrowing may be exceeded by the negative association between financial literacy and indebtedness, which results in a negative net effect of financial literacy on formal borrowing.

In terms of informal loan holding, we observe one additional correct answer to the financial knowledge test is associated with 1.82 pp lower likelihood of holding informal loans (Column 3, Panel 4). Moreover, the IV estimates for objective financial literacy in the specifications for *D_informal* are significantly negative at the 1% level (Columns 4 and 5, Panel 4). These findings support the informal credit part of our Hypothesis 2A, according to which financial literacy reduces the probability of informal borrowing.

We also observe that households with a higher objective level of financial knowledge are less likely to hold informal non-housing loans (Column 3 to 5, Panel 6). More specifically, one additional correct answer to the financial knowledge test is associated with a 1.3 pp lower likelihood of holding informal non-housing loans (Column 3, Panel 6). This conclusion is confirmed by the 2SLS and CMP estimates (Columns 4 and 5, Panel 6). Therefore, we provide further evidence to support the second half of Hypothesis 2A, according to which households with higher levels of objective financial literacy are less likely to hold informal (non-housing) debt.

Overall, in line with Grohmann, Klühs and Menkhoff (2018) and Lyons, Grable and Zeng (2019), our results suggest a negative relationship between financial literacy and the use of informal loans, but we do not find significant evidence that objective or subjective levels of financial literacy increase the probability of holding formal non-housing loans. Instead, we find negative relationships between financial literacy and the holding of formal loans in several cases. This may be explained bearing in mind that financially literate households refrain from

borrowing in the first place, though they may prefer formal credit to informal credit when they have to borrow in the future.

The Probit estimates for the other explanatory variables are reported in Columns (3) to (6) in Table A4.4. The results are as expected and in line with the literature (Cull *et al.*, 2015; Lyons, Grable and Zeng, 2019). In general, we find that younger households with more dependent children are more likely to borrow from either formal or informal sources. Specifically, we find those who have higher cognitive abilities, are self-employed, do not possess a house, have higher wealth are more likely to use formal credit. While those who are risk-seeking, experienced health shocks, use mobile phones, own houses and possess less assets are more likely to use informal credit. Surprisingly, we find that having close relatives raises the possibility of holding non-housing bank loans, but has no significant association with the use of informal loans.

4.6.2.2.2 Financial literacy and the amount of formal/informal loans

We further investigate the relationship between financial knowledge and the amount of formal and informal loans. The relevant estimates are reported in Panels 3 to 6 in Table 4.5. Focusing on the full baseline regressions, we find that one additional correct answer to the financial knowledge test is associated with a 1.70 pp lower amount of total informal debt (Column 3, Panel 4), and with a 1.23 pp lower amount of informal non-housing debt (Column 4, Panel 5). The negative relationship between financial literacy and informal loans is also supported by the IV estimates (Columns 4 and 5, Panels 4 and 6).

In contrast to Cull *et al.* (2015) who find higher levels of bank loans for financially literate households, we observe negative relationship between financial literacy and total formal debt. This may be because the positive association between financial literacy and the levels of formal

loan is offset by the negative relationship between financial literacy and the overall indebtedness, as we find the estimate of Basic FK in the model for the total amount of debt is nearly twice that in the model for the amount of formal debt. Therefore, in informal credit part of Hypothesis 2B is verified, while the formal credit part is not.

The signs for the other explanatory variables are similar to the findings in the case of extensive margins. In general, households who are younger and have more dependent children tend to hold higher levels of formal loans and informal loans. Specifically, those who have a higher cognitive ability, are self-employed, and possess higher levels of assets, tend to hold higher levels of formal loans. Those who are risk-seeking, experienced health shocks, use mobile phones and possess less wealth show higher levels of informal loans.

4.6.2.2.3 Does financial literacy shape the attitude towards formal/informal loans?

Next, we further analyse the determinants of two variables indicating households' attitude towards formal and informal credit. The estimates are reported in Table 4.6.

The baseline results suggest that, compared with the objective level of financial knowledge, the subjective level of financial knowledge matters more in the context of households' attitude towards credit sources (Column 3, Panels 1 and 2). Specifically, on average, a one-point higher score of perceived financial knowledge is associated with a 3.47 pp higher likelihood that the households prefer formal loans (Column 3, Panel 1), and with a 4.83 pp lower likelihood that the household will prefer informal loans (Column 3, Panel 2). Therefore, the baseline results provide further support to Hypothesis 2A. It also implies that subjective financial literacy may prevent households from borrowing informally by restraining their preference for informal loans in the first place.

The 2SLS and CMP estimates are reported in Columns (5) and (6), and the results for the validity tests of the IV approach are reported in Appendix 4.3. We cannot reject the null hypothesis of the endogeneity test for *Pref_formal,* indicating we can rely on the probit estimates for interpretation. Yet, we reject the null hypothesis of the endogeneity test for *Pref_informal*, indicating that financial knowledge should be taken as endogenous and the IV approach should be applied. However, the IV estimates provide no significant evidence that higher subjective financial knowledge is associated with a lower households' preference for informal finance (Column 4, Panel 2).

In summary, we find significant evidence in support of Hypothesis 2A, that a higher level of subjective financial knowledge is associated with a preference for bank loans and a preference against informal finance. Compared to the actual use of formal and informal loans, preferences are less likely to be subject to conditions such as access to credit and the lender's judgement about the borrower's capability to repay. Hence, it is not surprising that subjective levels of financial literacy are more predictive than objective measures in predicting people's preferences.

The signs of the other control variables are as expected. Households who are married, do not own house but possess higher wealth show a strong preference for formal credit. Those who are younger, with lower verbal skills, have close relatives, own a house but possess lower wealth prefer to borrow from friends, relatives or other informal sources if they have to borrow in the future.

4.7 Robustness Tests

4.7.1 Alternative estimation methods

To check if the findings are sensitive to the selection of estimation techniques, we alternatively estimate the extensive margins of debt using a linear probability model (LPM), and the intensive margins using OLS. IV estimates for the extensive and intensive margins are also derived from a CMP framework, and both the first stage and the second stage are estimated by OLS. The estimated marginal effects of financial knowledge are reported in Table 4.7. For brevity, we do not report the marginal effects of other covariates.

We first focus on the overall indebtedness of the households. Panel 1 and Panel 2 report the marginal effects of financial knowledge in models aimed at assessing the determinants of the probability that the households owe specific types of loans and the amount of outstanding specific types of loans, respectively. The estimates using alternative estimation methods show similar results to the baseline estimates. More specifically, the LPM and CMP_LPM estimates suggest a negative association between financial knowledge and the probability of overall indebtedness, as well as the probability of holding non-housing debt (Columns 1 and 2, Panel 1). Additionally, the OLS and CMP_OLS estimates suggest that a higher level of objective financial knowledge is associated with a lower amount of total debt and of non-housing debt (Columns 1 and 2, Panel 2). These findings provide additional support for Hypothesis 1.

Next, we find a significant negative association between objective financial knowledge and the probability of holding any type of informal loans and, specifically, non-housing informal loans (Columns 5 and 6, Panel 1). Households with higher objective financial knowledge are also found to hold less total informal debt and non-housing informal debt (Columns 5 and 6, Panel 2). While subjective levels of financial knowledge seem to be more predictive in the context of formal loans: Those with a higher perceived level of financial knowledge are less

likely to borrow from banks and hold fewer bank loans. This finding contradicts the formal credit part of our Hypotheses 2A and 2B. Therefore, Hypotheses 2A and 2B are only partly satisfied, suggesting a negative relationship between financial literacy and informal credit.

As for households' preferences over formal and informal loans, similar to the baseline results, we find the subjective measure to be more predictive than the objective measures of financial knowledge. More specifically, a one point higher of subjective financial knowledge is associated with a preference for formal loans and a preference against informal loans (Columns 1 to 4, Panel 3). This provides additional evidence in favour of Hypotheses 2A and 2B.

4.7.2 An alternative measure of objective financial literacy

Following Behrman *et al.* (2012), we construct an alternative measure of objective financial literacy, using the two-step PRIDIT method. The advantage of the PRIDIT method is that it gives a higher reward to the correct answers to difficult questions, and a higher penalty to the incorrect answers to simple questions. The way to calculate the PRIDIT score is described in Appendix 3.3.2 in Chapter Three. The estimation results are reported in Table 4.8 to Table 4.10. For all regressions in this section, subjective financial literacy remains the same as in the main analysis.

Table 4.8 reports the estimated marginal effects of financial knowledge for the binary dependent variables. The five columns display the Probit, OLS, 2SLS, CMP_Probit and CMP_LPM estimates, respectively.[95] Again, we find significant evidence that the PRIDIT score for objective financial knowledge is associated with a lower probability of owing any

[95] CMP_Probit and CMP_LPM indicate that the main equation is estimated by a Probit model or by OLS, respectively. In the following paragraphs, CMP_Tobit and CMP_OLS indicates that the main equation is estimated by a Tobit model or by OLS, respectively.

type of debt and owing non-housing debt, regardless of the estimation methods (Panels 1 and 2). Therefore, Hypothesis 1 is partly satisfied.

Compared with the baseline estimates reported in Table 4.4, we find stronger evidence that financial literacy is related to households' use of formal credit, but the sign of the relationship is negative, which contradicts the formal credit part of Hypothesis 2A (Panels 3 and 5). However, our results again support the informal credit part of Hypothesis 2A. More specifically, the probabilities of holding informal (non-housing) debt are found to be significantly lower as objective financial knowledge goes up, regardless of the estimation methods (Panels 4 and 6).

The marginal effects of financial knowledge along the intensive margins are reported in Table 4.9. Tobit, OLS, 2SLS, CMP_Tobit and CMP_OLS estimates are reported in order in Columns (1) to (5). Similar to the baseline results, we find significant evidence that the PRIDIT score for objective financial knowledge is negatively associated with total debt and non-housing debt (Panels 1 and 2), which is in line with Hypothesis 1.

The results for formal and informal credit are similar to the baseline results. More specifically, we observe a significantly negative relationship between the PRIDIT score for objective financial knowledge and total outstanding formal loans (Panel 3), which contradicts the formal credit part of Hypothesis 2B. However, we find that a higher PRIDIT score for objective financial knowledge is significantly associated with a lower amount of overall informal loans and non-housing informal loans (Panels 4 and 6). Once again, no significant predictive power is found for subjective financial knowledge in any of the cases, except that, in the Tobit model, it is negatively correlated with non-housing bank loans at the 10% significance level (Column 1, Panel 5). Therefore, the informal credit part of Hypothesis 2B is supported.

Finally, we estimate households' borrowing preferences in relation to the level of financial knowledge. Columns (1) to (5) in Table 4.10 report the Probit, OLS, 2SLS, CMP_Probit and CMP_LPM estimates, respectively. Again, we find significant predictive power for subjective financial knowledge, but not for objective financial knowledge. Similar to the baseline result, we find that a higher level of subjective financial knowledge is associated with a preference for formal loans and a preference against informal loans (Columns 1 and 2). However, the IV estimates for subjective financial literacy are insignificant in all the specifications for attitudes towards formal credit (Columns 3 to 5, Panel 1). The IV estimates for subjective financial literacy are significantly negative in two of the specifications for the attitudes towards informal credit (Columns 3 and 5, Panel 2). Overall, these findings provide further evidence to support Hypothesis 2A and 2B.

In conclusion, our main results are robust to using PRIDIT scores for objective financial knowledge; Hypothesis 1 is further supported, and the informal loan parts of Hypotheses 2A and 2B are verified. It is worth mentioning that the estimates for subjective financial literacy we derive from this section are of similar significance and magnitude to the baseline models where we use the summary score for objective financial literacy. This may be attributed to the relatively independent roles of objective and subjective financial literacy in the context of borrowing preferences.

4.7.3 An alternative measure of debt level

Next, we consider an alternative indicator for financial pressure, which measures the capacity of bearing credit. Specifically, following Stango and Zinman (2009b) and Lusardi, Mitchell and Oggero (2019), we scale the amount of specific types of debt by household income.

A high debt-to-income ratio implies that the household is vulnerable to financial and other shocks (e.g. unemployment shocks and health shocks) to their disposable income.

The Tobit, OLS, 2SLS, CMP_Tobit and CMP_OLS estimates are reported in Table 4.11. Objective financial knowledge is measured by counting the number of correct answers to the five financial knowledge questions.

Focusing on the first two panels, we find that objective financial literacy is still a significant predictor of a lower debt level (Panels 1 and 2), which supports our Hypothesis 1. As for the choice between formal and informal loans, we observe that households with higher subjective financial literacy, on average, have a lower total/non-housing debt-to-income ratio, but the evidence is only shown in the Tobit and OLS estimates (Column 1, Panel 3; Columns 1 and 2, Panel 5). Again, this finding contradicts the formal credit part of Hypothesis 2B. For informal loans, we find that households with higher objective financial knowledge, on average, hold a lower level of informal debt (and of non-housing informal debt), relative to their annual income (Panels 4 and 6). These findings are in line with the informal credit part of Hypothesis 2B. We also observe a significantly positive CMP_OLS estimate of subjective financial knowledge in the model for informal loans-to-income ratio (Column 5, Panel 4), which is difficult to interpret economically. One potential explanation is that people with higher subjective financial literacy believe in their capability of earning higher (investment) income and paying back, and hence, would like to take on more debt to make ends meet. However, people may overestimate their actual levels of financial literacy. Since informal loans are subject to less scrutiny compared to formal credit from banks, and therefore easier for them to obtain, they are more likely to take on large amounts of informal loans relative to current income. Nevertheless, the sign of this estimate contradicts our findings from all the other models, where we hardly find any significant role of subjective financial literacy in predicting the use of loans. Therefore, this inference requires more evidence to make sense.

4.7.4 An alternative measure of indebtedness: Financial distress

Individuals' perceptions of their indebtedness (which measure the extent to which an individual perceives debt to be manageable or not) can differ from the objective measures of debt status. While the objective debt status may not be indicative of how households understand their debt level, the subjective measure of indebtedness is better at capturing people's concern about debt conditional on their income and capability to pay (Tay *et al.*, 2017). Inspired by Lusardi and Tufano (2015) and Disney and Gathergood (2013), we seek to understand how financial literacy is related to whether people experience difficulties paying off their debt. To this end, we construct a dummy variable for financial distress, which takes a value of 1 when the respondent reports facing financial difficulties, and 0 otherwise.[96]

The Probit and IV estimates for the marginal effects of financial literacy in models looking at the determinants of the probability of facing financial difficulties are reported in Table 4.12. The Probit estimates for the marginal effects of both objective and subjective measures of financial literacy are always significant, even when including a full set of control variables (Column 3). Specifically, one additional correct answer to the financial knowledge test is associated with a 3.14 pp lower probability of feeling financial distress (Column 3). And one higher point for the subjective level of financial knowledge is associated with a 3.20 pp lower probability of feeling financial distress (Column 3). The negative associations are confirmed by the IV estimates.[97] These findings suggest that a higher levels of financial literacy can mitigate financial distress, which verifies our Hypothesis 1 from a subjective perspective of debt level.

[96] This dummy is based on the answer provided to the following question: What is the main difficulty faced by your family currently? [Multiple choice] 1. Financial difficulty 2. Housing difficulty 3. Children's education 4. Caring for the elderly 5. Medical treatment and care of patients 6. Strained family relationship 7. Unemployment/waiting for jobs of family members 77. Other 78. No difficulty

[97] The results for the instrument validity tests of the IV approach are reported in Appendix 4.3.

4.8 Conclusion

As financial markets have opened up over the past two decades, people gained unprecedented access to credit. The global high level of household indebtedness in the wake of the financial crisis and sub-prime crisis calls for a better understanding of households' borrowing behaviours. Among all the factors that relate to high levels of indebtedness, low financial literacy has attracted the attention of both the academics and the policymakers recently.

Using data from the 2014 wave of the CFPS, this paper investigates the effect of financial literacy on the indebtedness of Chinese households. We use both objective and subjective measures of financial literacy. The empirical results suggest that households with heads who have a higher level of objective financial literacy are less likely to borrow and hold a lower debt level. This finding is in line with a variety of existing studies in the context of developed countries (e.g., Disney and Gathergood, 2013; Lusardi and Tufano, 2015; Lusardi, Mitchell and Oggero, 2019), and is confirmed by our analysis on solely non-housing loans.

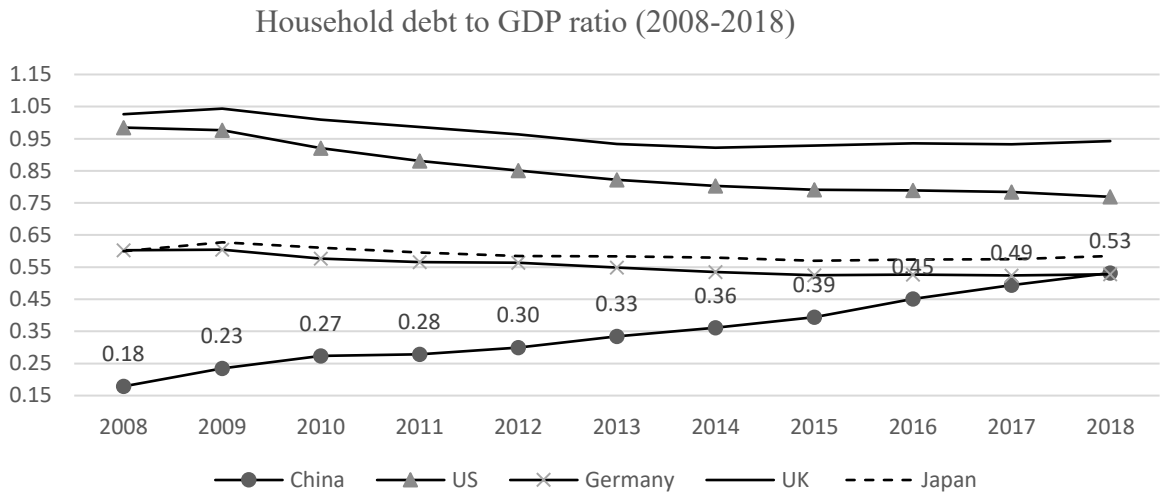
We further distinguish between formal credit and informal credit. Similar to prior studies in the Chinese context (e.g., Cull *et al.*, 2015; Lyons, Grable and Zeng, 2019), we find significant evidence that financially literate households are less likely to have informal loans and hold lower levels of informal loans. However, different from these studies which find positive relationships between financial literacy and formal credit holdings, we observe negative relationships in several cases. A potential explanation is that the positive associations between financial literacy and the incidence/levels of formal loans are offset by the negative relationship between financial literacy and overall indebtedness. By contrast, the associations between financial literacy and households' attitudes towards formal and informal credit are both significantly identified: Households with higher financial literacy prefer to borrow from banks

and are reluctant to borrow from informal sources in case they need to borrow in the future. These results are robust to using an IV approach and various robustness checks.

Our empirical results imply that financial literacy is effective in improving debt management and enhancing financial inclusion in terms of borrowing decisions. This implication may also apply to other emerging countries where formal financial markets are underdeveloped.

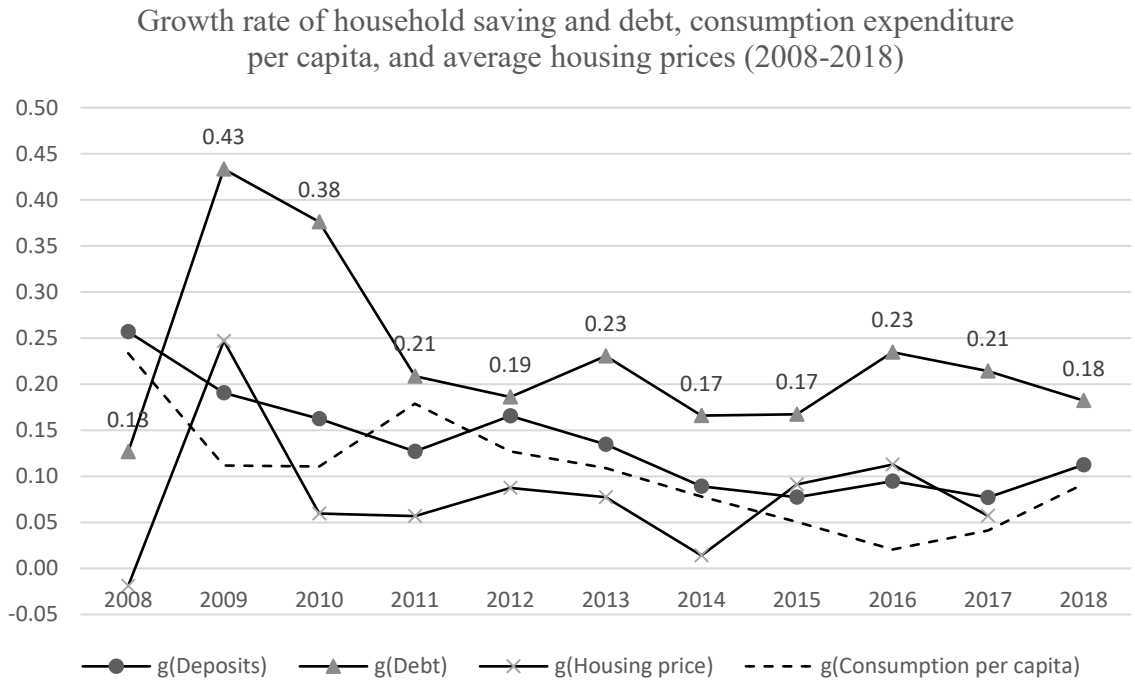
This research has several limitations. First, due to the limitation of the data, we are not able to adequately address the factors that are related to the supply of formal financial services, such as the regional development of the financial market and the access to bank services. Moreover, in the CFPS, we cannot isolate the loans from friends and relatives from the loans owed to other individuals or informal institutions. However, these two sources of loans can differ substantially in terms of risk and cost, and people may have distinct preferences over them. Finally, similar to Chapter Three, the sample we use in this analysis may not be representative for the whole Chinese population.

Figures and Tables



Sources: PBC and CEIC Data.

Figure 4.1 Household debt to GDP ratio: Five countries (2008-2018)

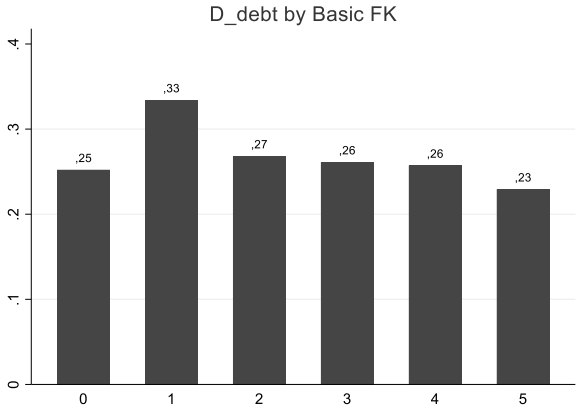


Sources: PBC and NBSC.

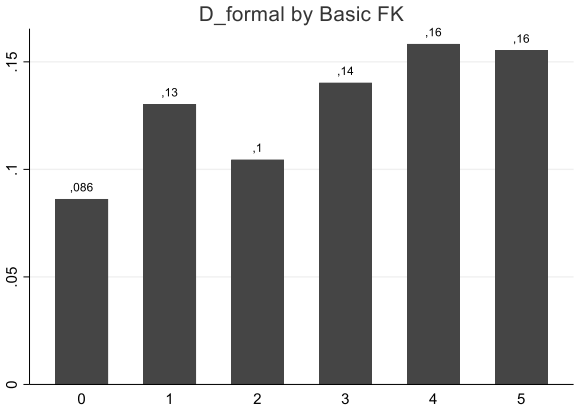
Figure 4.2 Household saving, debt, consumption, and housing prices: China (2008-2018)

Figure 4.3 Dependent variables by financial knowledge test score (0-5)

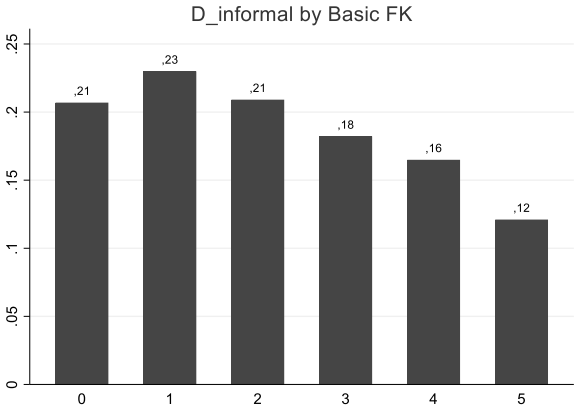
(1)



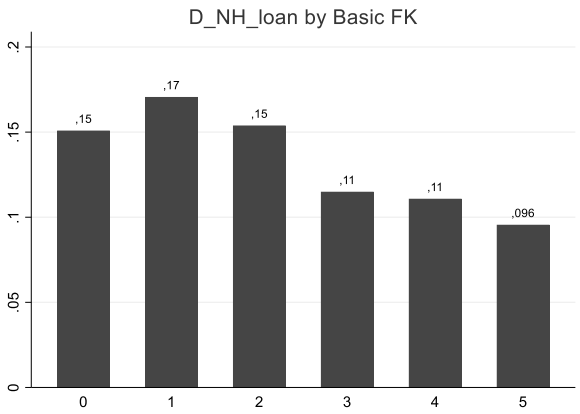
(2)



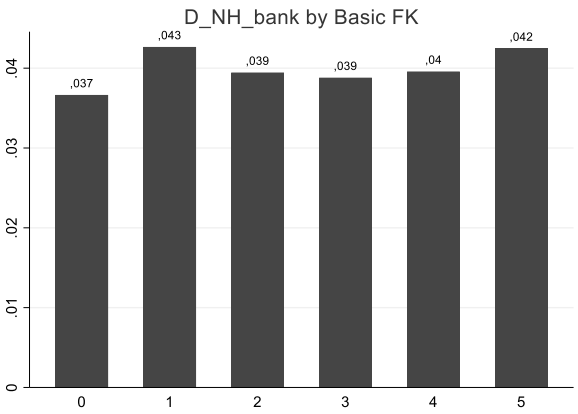
(3)



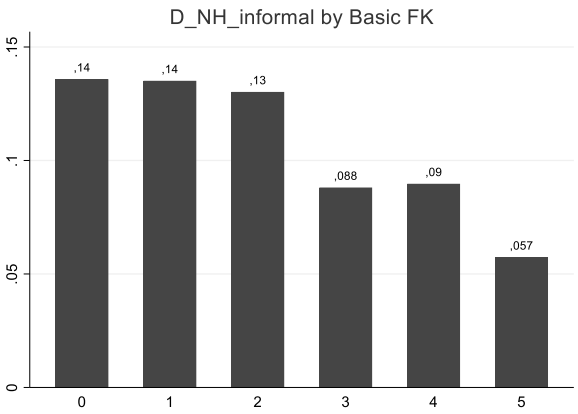
(4)



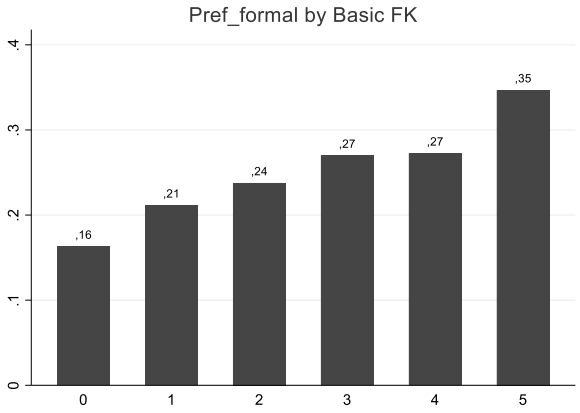
(5)



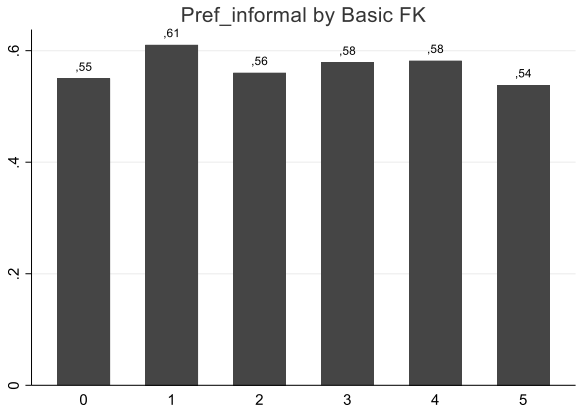
(6)



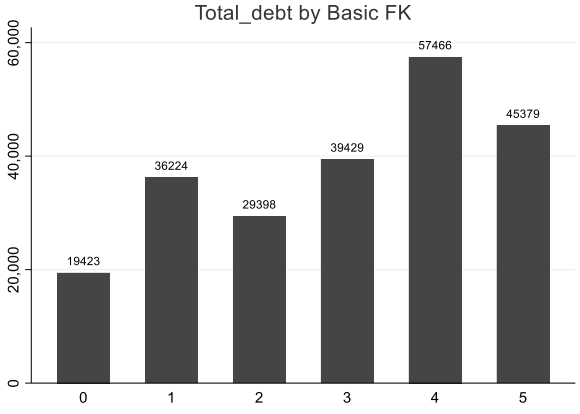
(7)


Pref_formal by Basic FK
0
,1
,2
,3
,4
0
1
2
3
4
5
,16
,21
,24
,27
,27
,35

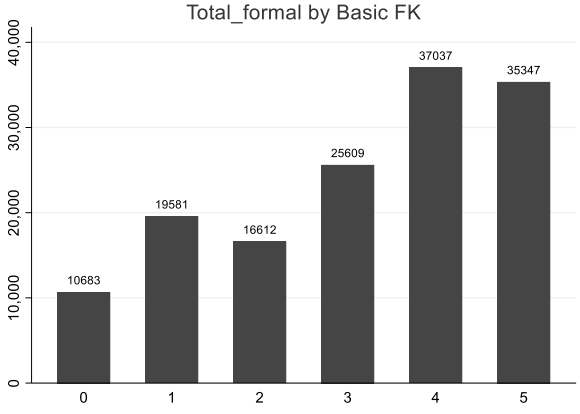
(8)


Pref_informal by Basic FK
0
,2
,4
,6
0
1
2
3
4
5
,55
,61
,56
,58
,58
,54

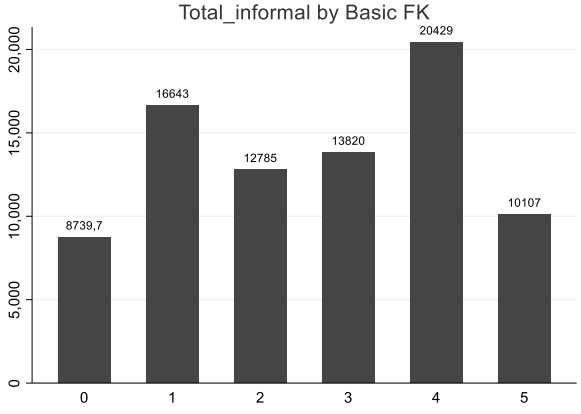
(9)


Total_debt by Basic FK
0
20,000
40,000
60,000
0
1
2
3
4
5
19423
36224
29398
39429
57466
45379

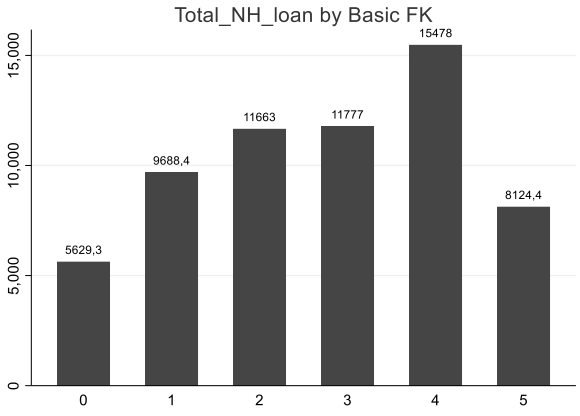
(10)


Total_formal by Basic FK
0
10,000
20,000
30,000
40,000
0
1
2
3
4
5
10683
19581
16612
25609
37037
35347

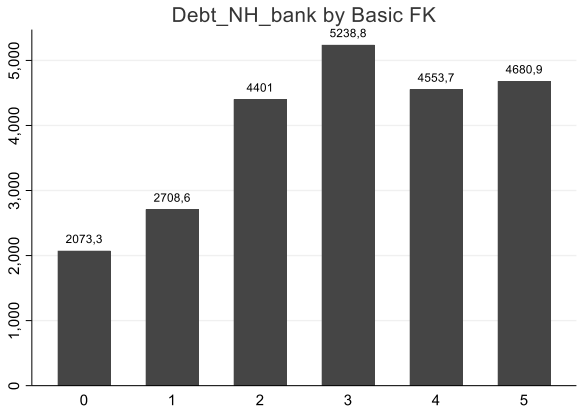
(11)


Total_informal by Basic FK
0
5,000
10,000
15,000
20,000
0
1
2
3
4
5
8739,7
16643
12785
13820
20429
10107

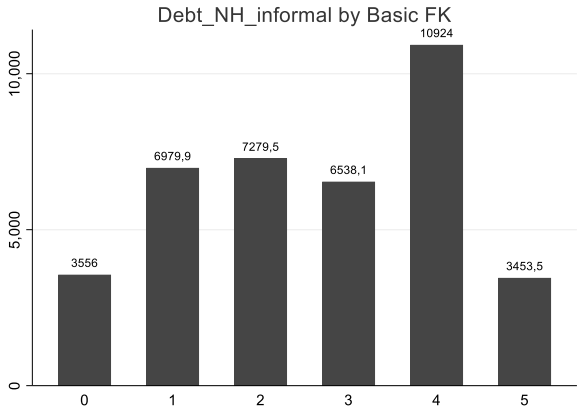
(12)

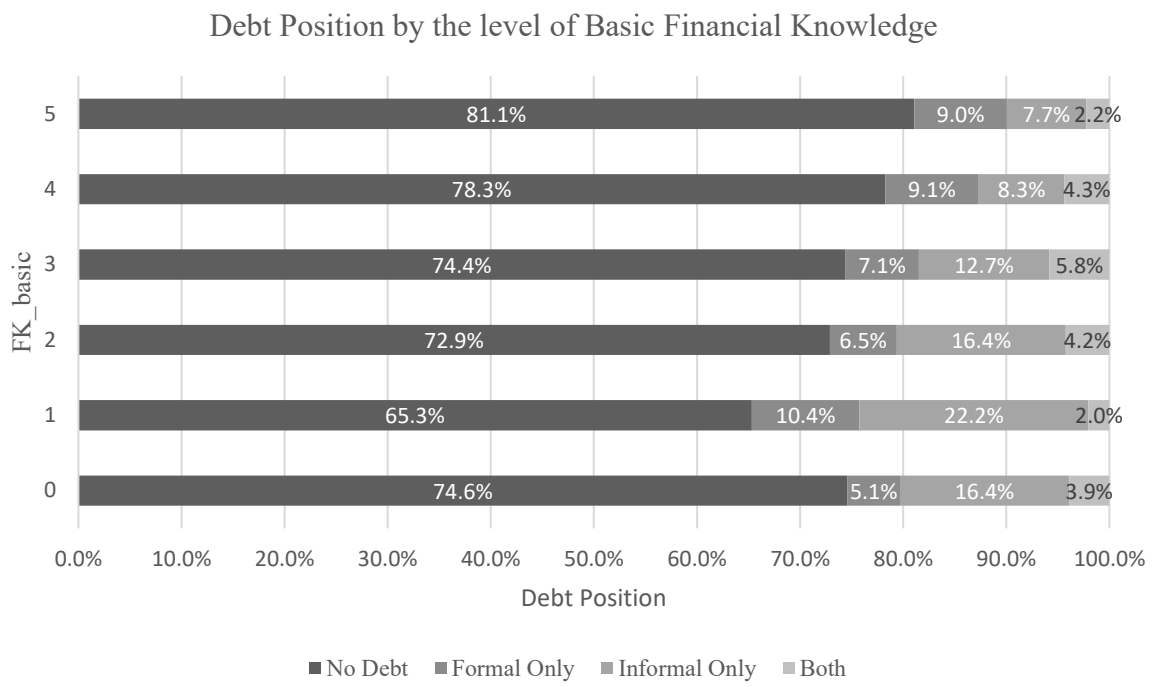

Total_NH_loan by Basic FK
0
5,000
10,000
15,000
5629,3
9688,4
11663
11777
15478
8124,4
0
1
2
3
4
5

(13)


Debt_NH_bank by Basic FK
0
1,000
2,000
3,000
4,000
5,000
2073,3
2708,6
4401
5238,8
4553,7
4680,9
0
1
2
3
4
5

(14)


Debt_NH_informal by Basic FK
0
5,000
10,000
3556
6979,9
7279,5
6538,1
10924
3453,5
0
1
2
3
4
5



Source: CFPS (2014)

Figure 4.4 Debt position by the level of basic financial knowledge

Table 4.1 Summary statistics of key variables (N = 2,237)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Full sample				D_debt=0	D_debt=1	Pref _bank=0	Pref _bank=1	Pref _inform=0	Pref _inform=1
Variable	Mean	Std. Dev.	Min	Max	N=1,664	N=573	N=1,660	N=577	N=958	N=1,279
Panel 1: Dependent Variable										
D_debt	0.256	0.437	0	1	0	1	0.252	0.271	0.202	0.298
D_formal	0.122	0.328	0	1	0	0.477	0.097	0.197	0.140	0.109
D_informal	0.177	0.382	0	1	0	0.692	0.197	0.121	0.094	0.241
D_NH_loan	0.127	0.333	0	1	0	0.494	0.132	0.112	0.087	0.157
D_NH_bank	0.037	0.190	0	1	0	0.144	0.029	0.060	0.043	0.033
D_NH_inform	0.102	0.302	0	1	0	0.397	0.115	0.066	0.054	0.138
Pref_bank	0.258	0.438	0	1	0.253	0.272	0	1	0.603	0
Pref _inform	0.572	0.495	0	1	0.540	0.664	0.771	0	0	1
Fin_diff	0.291	0.455	0	1	0.212	0.522	0.306	0.248	0.223	0.342
Total debt	29,300.14	78,355.64	0	400,000	0	114,306	25,663	40,000	28,601	29,938
Debt_formal	16,945.44	59,635.76	0	335,000	0	66,108	12,142	30,888	21,761	13,403
Debt_informal	11,106.37	34,824.47	0	200,000	0	43,328	12,248	7,922	5,764	15,152
Total_NH_loan	7,874.23	30,891.86	0	200,000	0	30,719	7,499	9,019	6,334	9,059
Debt_NH_bank	2,163.19	13,116.54	0	100,000	0	8,435	1,511	4,055	2,657	1,801
Debt_NH_inform	4,225.99	17,029.93	0	110,000	0	16,487	4,861	2,438	2,060	5,865
Panel 2: Financial literacy										
Basic FK	2.769	1.563	0	5	2.839	2.568	2.679	3.022	2.817	2.731
Subj_fk	2.400	0.784	1	5	2.419	2.345	2.353	2.538	2.470	2.349
Panel 3: Individual characteristics										
Age	52.242	13.116	18	89	53.864	47.536	52.957	50.153	54.851	50.272
Male	0.438	0.496	0	1	0.434	0.449	0.429	0.467	0.453	0.428
Married	0.863	0.344	0	1	0.853	0.891	0.844	0.916	0.859	0.865
Illiterate	0.116	0.321	0	1	0.118	0.113	0.128	0.081	0.116	0.116
Mathtest14	13.523	5.629	0	24	13.473	13.669	13.115	14.683	13.625	13.441
Wordtest14	22.351	8.697	0	34	22.249	22.648	21.746	24.112	22.614	22.164
Migrant	0.116	0.321	0	1	0.097	0.172	0.114	0.122	0.106	0.124
Unemployed	0.414	0.493	0	1	0.455	0.293	0.444	0.328	0.464	0.376

Self-employed	0.125	0.331	0	1	0.100	0.196	0.118	0.143	0.110	0.136
Risk seeking	0.022	0.147	0	1	0.016	0.040	0.019	0.033	0.024	0.021
Risk averse	0.663	0.473	0	1	0.674	0.629	0.690	0.583	0.676	0.653
FA_feeling	3.579	1.279	1	5	3.568	3.612	3.574	3.598	3.554	3.600
FA_present	3.153	1.331	1	5	3.116	3.262	3.143	3.188	3.086	3.206
FA_spend	3.433	1.297	1	5	3.420	3.471	3.425	3.455	3.360	3.488
Health shock	0.333	0.471	0	1	0.315	0.383	0.339	0.314	0.338	0.329
Use Mobile	0.913	0.281	0	1	0.892	0.974	0.900	0.950	0.891	0.930
E-shopper	0.165	0.371	0	1	0.154	0.199	0.143	0.229	0.162	0.168
Panel 4: Household characteristics										
Rural	0.048	0.214	0	1	0.047	0.050	0.052	0.036	0.033	0.059
Close_relative	0.567	0.496	0	1	0.559	0.589	0.560	0.588	0.543	0.586
N_depchildren	0.932	0.725	0	5	0.858	1.144	0.920	0.966	0.838	1.002
N_depelderly	0.625	0.824	0	3	0.708	0.383	0.656	0.536	0.755	0.528
Own house	0.872	0.334	0	1	0.872	0.872	0.867	0.888	0.866	0.878
Asset_net	585,361	792,698	0	3,300,000	602,138.20	523,232.30	523,581.90	749,737.10	713,616.40	483,248.70
Inc_net	61,743	47,369	500	219,300	62,268.75	60,219.39	58,015.10	72,416.60	67,099.48	57,714.23
Panel 5: Community level										
Facility_indx1	0.061	0.214	-1	1	0.050	0.092	0.056	0.075	0.057	0.064
Facility_indx2	0.044	0.421	-1	1	0.050	0.029	0.036	0.068	0.044	0.044

Notes: Definitions of all the variables can be found in Appendix 4.1.

Table 4.2 Basic financial knowledge

Panel A: Percentages of answers (N = 2,237)

	Correct	Incorrect	DK
1. Annual interest rate of one-year fixed deposit	55.30	9.38	35.32
2. One-year interest and principal	43.90	34.2	21.90
3. Interest compounding	49.71	26.51	23.78
Both 2 & 3	25.70		-
4. Purchasing power (inflation)	59.28	19.75	20.97
5. Time value of money	68.84	15.02	16.14

Panel B: Distribution of correct answers (N = 2,237)

Correct	None	1	2	3	4	5
Percentage	14.09	12.82	15.40	20.35	23.03	14.31

Notes: Panel A reports the percentages of respondents who answered correctly, incorrectly and "do not know". Panel B reports the distribution of the number of correct answers.

Table 4.3 Subjective assessment of financial knowledge

Subjective Score (N = 2,237)

Subjective Score	1 (way below)	2	3	4	5 (way above)	DK
Percentage	11.94	42.51	39.92	5.01	0.63	1.00

Notes: This table reports the distribution of the self-perceived level of financial knowledge.

Table 4.4 Probability of indebtedness and financial literacy

VARIABLES	(1) Probit	(2) Probit	(3) Probit	(4) IVreg	(5) CMP
Panel 1: D_debt					
Basic FK	-0.0042	-0.0217***	-0.0256***	-0.1109*	-0.3052***
	(0.0050)	(0.0056)	(0.0064)	(0.0611)	(0.0914)
Subj_FK	-0.0163	-0.0118	-0.0154	-0.0812	-0.1838
	(0.0110)	(0.0112)	(0.0131)	(0.1171)	(0.1663)
Panel 2: D_NH_loan					
Basic FK	-0.0096**	-0.0147***	-0.0150***	-0.1255**	-0.3232***
	(0.0039)	(0.0044)	(0.0049)	(0.0491)	(0.1102)
Subj_FK	-0.0192**	-0.0147*	-0.0112	0.0704	-0.0025
	(0.0079)	(0.0081)	(0.0086)	(0.0881)	(0.2167)
Panel 3: D_formal					
Basic FK	0.0149***	-0.0076*	-0.0143***	0.0367	-0.0517
	(0.0039)	(0.0046)	(0.0050)	(0.0464)	(0.1089)
Subj_FK	0.0013	-0.0003	-0.0049	-0.1554*	-0.3624**
	(0.0089)	(0.0092)	(0.0102)	(0.0934)	(0.1829)
Panel 4: D_informal					
Basic FK	-0.0132***	-0.0170***	-0.0182***	-0.1596***	-0.3863***
	(0.0045)	(0.0051)	(0.0058)	(0.0595)	(0.0866)
Subj_FK	-0.0189**	-0.0112	-0.0104	-0.0156	0.0284
	(0.0092)	(0.0095)	(0.0111)	(0.1094)	(0.1775)
Panel 5: D_NH_bank					
Basic FK	0.0016	-0.0042	-0.0051*	-0.0503**	-0.2452
	(0.0023)	(0.0028)	(0.0030)	(0.0251)	(0.1527)
Subj_FK	-0.0050	-0.0051	-0.0089*	0.0284	-0.0749
	(0.0044)	(0.0043)	(0.0046)	(0.0473)	(0.3417)
Panel 6: D_NH_informal					
Basic FK	-0.0120***	-0.0125***	-0.0131***	-0.1214***	-0.3394***
	(0.0036)	(0.0039)	(0.0043)	(0.0466)	(0.1037)
Subj_FK	-0.0186***	-0.0141*	-0.0067	0.0218	-0.0026
	(0.0070)	(0.0074)	(0.0083)	(0.0850)	(0.2073)
Individual control	No	Yes	Yes	Yes	Yes
Household control	No	No	Yes	Yes	Yes
Province Fixed Effect	Yes	Yes	Yes	Yes	Yes
Observations	3,233	3,012	2,237	2,020	2,020

Notes: This table reports the estimated marginal effects associated with financial literacy in regressions explaining the probability of having various types of debt. Columns (1) – (3) report estimates obtained from Probit models. Column (4) reports 2SLS estimates and Column (5) reports the CMP-Probit estimates using mother's education and community-level financial literacy as instruments for financial literacy. The marginal effects of the control variables are reported in Appendix 4.4. Definitions of all variables are displayed in Appendix 4.1. Standard errors are clustered at the community level and reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4.5 Debt level and financial literacy

VARIABLES	(1) Tobit	(2) Tobit	(3) Tobit	(4) IVreg	(5) CMP
Panel 1: Total_debt					
Basic FK	-0.0022	-0.0205***	-0.0250***	-1.1790**	-0.6140**
	(0.0049)	(0.0054)	(0.0062)	(0.5891)	(0.2544)
Subj_FK	-0.0137	-0.0092	-0.0137	-0.7883	-0.3508
	(0.0108)	(0.0109)	(0.0128)	(0.9426)	(0.4572)
Panel 2: Total_NH_loan					
Basic FK	-0.0090**	-0.0143***	-0.0145***	0.2000	-0.4545**
	(0.0039)	(0.0043)	(0.0049)	(0.4572)	(0.2055)
Subj_FK	-0.0186**	-0.0140*	-0.0105	-1.6647**	0.0006
	(0.0079)	(0.0081)	(0.0087)	(0.7472)	(0.3688)
Panel 3: Debt_formal					
Basic FK	0.0153***	-0.0075	-0.0143***	-1.3724***	0.0006
	(0.0039)	(0.0046)	(0.0050)	(0.5301)	(0.1792)
Subj_FK	0.0022	0.0007	-0.0042	0.1952	-0.4245
	(0.0089)	(0.0092)	(0.0100)	(0.8073)	(0.3229)
Panel 4: Debt_informal					
Basic FK	-0.0122***	-0.0163***	-0.0170***	-1.1832***	-0.6818***
	(0.0044)	(0.0050)	(0.0055)	(0.4413)	(0.2188)
Subj_FK	-0.0178*	-0.0107	-0.0096	0.5873	0.0556
	(0.0091)	(0.0094)	(0.0106)	(0.6698)	(0.3979)
Panel 5: Debt_NH_bank					
Basic FK	0.0016	-0.0043	-0.0048*	-0.4397*	-0.0422
	(0.0023)	(0.0027)	(0.0029)	(0.2431)	(0.1087)
Subj_FK	-0.0046	-0.0045	-0.0086*	0.2331	-0.0086
	(0.0045)	(0.0043)	(0.0045)	(0.3569)	(0.1933)
Panel 6: Debt_NH_informal					
Basic FK	-0.0116***	-0.0122***	-0.0123***	-1.0599***	-0.4759***
	(0.0036)	(0.0038)	(0.0042)	(0.4041)	(0.1801)
Subj_FK	-0.0180***	-0.0137*	-0.0059	0.1657	-0.0400
	(0.0069)	(0.0073)	(0.0080)	(0.6263)	(0.3253)
Individual control	No	Yes	Yes	Yes	Yes
Household control	No	No	Yes	Yes	Yes
Province Fixed Effect	Yes	Yes	Yes	Yes	Yes
Observations	3,235	3,013	2,237	2,020	2,020

Notes: This table reports the estimated marginal effects associated with financial literacy in regressions explaining the outstanding debt held by the households. The dependent variables are the logarithm of the amount of debt. Columns (1) – (3) report estimates obtained from Tobit models. Column (4) reports 2SLS estimates and Column (5) reports the CMP- Tobit estimates using mother's education and community-level financial literacy as instruments for financial literacy. The marginal effects of the control variables are reported in Appendix 4.5. Definitions of all variables are displayed in Appendix 4.1. Standard errors are clustered at the community level and reported in parentheses. $***$ $p < 0.01$, $**$ $p < 0.05$, $*$ $p < 0.1$.

Table 4.6 Borrowing preferences and financial literacy

VARIABLES	(1) Probit	(2) Probit	(3) Probit	(4) IVreg	(5) CMP
Panel 1: Pref_formal					
Basic FK	0.0247***	0.0023	-0.0006	0.0295	0.0802
	(0.0055)	(0.0063)	(0.0072)	(0.0599)	(0.0875)
Subj_FK	0.0296***	0.0289***	0.0347***	0.0609	0.0111
	(0.0091)	(0.0098)	(0.0118)	(0.1111)	(0.1560)
Panel 2: Pref_informal					
Basic FK	0.0079	0.0037	0.0020	-0.0946	0.0035
	(0.0062)	(0.0071)	(0.0081)	(0.0712)	(0.0771)
Subj_FK	-0.0328***	-0.0329***	-0.0383***	-0.2063	-0.0837
	(0.0113)	(0.0117)	(0.0142)	(0.1336)	(0.1397)
Individual control	No	Yes	Yes	Yes	Yes
Household control	No	No	Yes	Yes	Yes
Province Fixed Effect	Yes	Yes	Yes	Yes	Yes
Observations	3,233	3,012	2,237	2,020	2,020

Notes: This table reports the estimated marginal effects associated with financial literacy in regressions explaining households' preference over formal and informal credit. Columns (1) – (3) report estimates obtained from Probit models. Column (4) reports 2SLS estimates and Column (5) reports the CMP-Probit estimates using mother's education and community-level financial literacy as instruments for financial literacy. The marginal effects of the control variables are reported in Appendix 4.4. Definitions of all variables are displayed in Appendix 4.1. Standard errors are clustered at the community level and reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4.7 Robustness test 1: Alternative estimation methods (LPM and OLS)

	(1)	(2)	(3)	(4)	(5)	(6)
Panel 1						
	D_debt		D_formal		D_informal	
	LPM	CMP_LMP	LPM	CMP_LMP	LPM	CMP_LMP
Basic FK	-0.0145***	-0.0890***	-0.0046	-0.0204**	-0.0120**	-0.0675***
	(0.0055)	(0.0201)	(0.0034)	(0.0098)	(0.0054)	(0.0198)
Subj_FK	-0.0205*	-0.0225	-0.0422**	-0.1213*	-0.0502	-0.0635
	(0.0104)	(0.0377)	(0.019)	(0.0718)	(0.0412)	(0.1479)
	D_NH_loan		D_NH_bank		D_NH_informal	
	LPM	CMP_LMP	LPM	CMP_LMP	LPM	CMP_LMP
Basic FK	-0.0107**	-0.0444**	-0.0033	-0.0088	-0.0100**	-0.0487***
	(0.0050)	(0.0183)	(0.0035)	(0.0100)	(0.0043)	(0.0164)
Subj_FK	-0.0093	0.0178	-0.0066	0.0096	-0.0052	0.0064
	(0.0082)	(0.0339)	(0.0049)	(0.0184)	(0.0076)	(0.0307)
Panel 2						
	Total_debt		Debt_formal		Debt_informal	
	OLS	CMP_OLS	OLS	CMP_OLS	OLS	CMP_OLS
Basic FK	-0.1454**	-0.8948***	-0.0497	-0.1626	-0.1109*	-0.6482***
	(0.0605)	(0.2133)	(0.0365)	(0.1063)	(0.0589)	(0.2083)
Subj_FK	-0.2160*	-0.2015	-0.1160**	-0.2855	-0.1122	-0.1318
	(0.1119)	(0.3988)	(0.0536)	(0.1931)	(0.1092)	(0.3863)
	Total_NH_loan		Debt_NH_bank		Debt_NH_informal	
	OLS	CMP_OLS	OLS	CMP_OLS	OLS	CMP_OLS
Basic FK	-0.1031*	-0.4113**	-0.0266	-0.0130	-0.0959**	-0.4506***
	(0.0534)	(0.1897)	(0.0377)	(0.1031)	(0.0451)	(0.1668)
Subj_FK	-0.0842	0.1697	-0.0600	0.1037	-0.0431	0.0249
	(0.0865)	(0.3512)	(0.0519)	(0.1888)	(0.0782)	(0.3094)
Panel 3						
	Pref_formal		Pref_informal			
	LPM	CMP_LMP	LPM	CMP_LMP		
Basic FK	0.0023	0.0239	0.0004	-0.0077		
	(0.007)	(0.022)	(0.0082)	(0.0269)		
Subj_FK	0.1452***	0.0988	-0.1746***	-0.298		
	(0.0474)	(0.1618)	(0.0581)	(0.1975)		
# of Obs	2,237	2,020	2,237	2,020		

Notes: This table reports the estimated marginal effects of financial literacy on the dependent variables using alternative estimation methods. Amounts of debt are log-transformed. Estimates for the other covariates are not reported for brevity but available upon request. Definitions of all variables are displayed in Appendix 4.1. Standard errors are clustered at the community level and reported in parentheses. $***$ $p < 0.01$, $**$ $p < 0.05$, $*$ $p < 0.1$.

Table 4.8 Robustness test 2: Probability of holding debt and PRIDIT score for basic financial knowledge

VARIABLES	(1) Probit	(2) LPM	(3) IVreg	(4) CMP_Probit	(5) CMP_LPM
Panel 1: D_debt					
Basic FK	-0.0120***	-0.0108***	-0.0528*	-0.1183***	-0.0217*
	(0.0029)	(0.0030)	(0.0286)	(0.0436)	(0.0112)
Subj_FK	-0.0158	-0.0171	-0.0409	-0.2198	-0.0510
	(0.0133)	(0.0129)	(0.1075)	(0.1618)	(0.0420)
Panel 2: D_NH_loan					
Basic FK	-0.0066***	-0.0060**	-0.0594**	-0.1599***	-0.0226**
	(0.0023)	(0.0024)	(0.0237)	(0.0532)	(0.0093)
Subj_FK	-0.0100	-0.0099	0.0755	-0.0770	-0.0083
	(0.0085)	(0.0082)	(0.0812)	(0.2180)	(0.0355)
Panel 3: D_formal					
Basic FK	-0.0062**	-0.0056**	0.0157	0.0251	0.0023
	(0.0025)	(0.0025)	(0.0216)	(0.0470)	(0.0076)
Subj_FK	-0.0063	-0.0083	-0.1015	-0.3909**	-0.0358
	(0.0102)	(0.0104)	(0.0840)	(0.1699)	(0.0286)
Panel 4: D_informal					
Basic FK	-0.0090***	-0.0079***	-0.0728***	-0.1465***	-0.0261***
	(0.0026)	(0.0026)	(0.0279)	(0.0444)	(0.0101)
Subj_FK	-0.0106	-0.0093	-0.0228	-0.0882	-0.0395
	(0.0111)	(0.0104)	(0.1007)	(0.1769)	(0.0379)
Panel 5: D_NH_bank					
Basic FK	-0.0021	-0.0018	-0.0269**	-0.1602***	-0.0083*
	(0.0015)	(0.0016)	(0.0128)	(0.0599)	(0.0050)
Subj_FK	-0.0089*	-0.0075	0.0391	0.0855	0.0048
	(0.0047)	(0.0049)	(0.0460)	(0.2899)	(0.0189)
Panel 6: D_NH_informal					
Basic FK	-0.0063***	-0.0058***	-0.0536**	-0.1651***	-0.0221***
	(0.0020)	(0.0020)	(0.0220)	(0.0486)	(0.0084)
Subj_FK	-0.0050	-0.0048	0.0194	0.0257	-0.0147
	(0.0081)	(0.0076)	(0.0766)	(0.1994)	(0.0318)
Individual control	Yes	Yes	Yes	Yes	Yes
Household control	Yes	Yes	Yes	Yes	Yes
Province Fixed Effect	Yes	Yes	Yes	Yes	Yes
Observations	2,237	2,237	2,020	2,020	2,020

Notes: This table reports the estimated marginal effects associated with financial literacy in regressions explaining the probability of having various types of debt. The PRIDIT measure of actual financial literacy is used. Columns (1) and (2) report the Probit and LPM estimates. Columns (3) to (5) report the 2SLS estimates, CMP_Probit and CMP_LPM estimates, respectively, using mother's education and community-level financial literacy as instruments for financial literacy. Estimates for other covariates are not reported but available upon request. Definitions of all variables are displayed in Appendix 4.1. Standard errors are clustered at the community level and reported in parentheses. $***$ $p < 0.01$, $**$ $p < 0.05$, $*$ $p < 0.1$.

Table 4.9 Robustness test 2: Debt level and PRIDIT score for basic financial knowledge

VARIABLES	(1) Tobit	(2) OLS	(3) IVreg	(4) CMP_Tobit	(5) CMP_OLS
Panel 1: Total_debt					
Basic FK	-0.0117***	-0.1152***	-0.5213	-1.3925**	-0.2077*
	(0.0029)	(0.0338)	(0.3182)	(0.5653)	(0.1210)
Subj_FK	-0.0142	-0.1791	-0.5328	-2.8657	-0.5030
	(0.0130)	(0.1457)	(1.1971)	(2.0453)	(0.4545)
Panel 2: Total_NH_loan					
Basic FK	-0.0065***	-0.0589**	-0.6487**	-1.9461***	-0.2209**
	(0.0023)	(0.0255)	(0.2539)	(0.6662)	(0.0962)
Subj_FK	-0.0093	-0.0920	0.7624	-1.4127	-0.1005
	(0.0086)	(0.0864)	(0.8688)	(2.4554)	(0.3640)
Panel 3: Debt_formal					
Basic FK	-0.0062**	-0.0656**	0.1934	0.5416	0.0445
	(0.0026)	(0.0283)	(0.2474)	(0.8209)	(0.0849)
Subj_FK	-0.0056	-0.0873	-1.0119	-6.7424**	-0.3586
	(0.0100)	(0.1219)	(0.9586)	(3.1716)	(0.3207)
Panel 4: Debt_informal					
Basic FK	-0.0083***	-0.0731***	-0.7424**	-2.5444***	-0.2527**
	(0.0026)	(0.0281)	(0.2986)	(0.8487)	(0.1049)
Subj_FK	-0.0099	-0.0832	-0.4526	0.7725	-0.3942
	(0.0106)	(0.1109)	(1.0749)	(3.0963)	(0.3945)
Panel 5: Debt_NH_bank					
Basic FK	-0.0019	-0.0144	-0.2938**	-2.7435*	-0.0443
	(0.0015)	(0.0179)	(0.1431)	(1.4791)	(0.0516)
Subj_FK	-0.0087*	-0.0701	0.5299	1.7133	0.0451
	(0.0045)	(0.0517)	(0.4986)	(6.2490)	(0.1935)
Panel 6: Debt_NH_informal					
Basic FK	-0.0060***	-0.0564***	-0.5511**	-2.4500***	-0.2080**
	(0.0020)	(0.0213)	(0.2287)	(0.8737)	(0.0843)
Subj_FK	-0.0042	-0.0399	0.0683	0.1252	-0.1816
	(0.0079)	(0.0780)	(0.7990)	(3.1336)	(0.3185)
Individual control	Yes	Yes	Yes	Yes	Yes
Household control	Yes	Yes	Yes	Yes	Yes
Province Fixed Effect	Yes	Yes	Yes	Yes	Yes
Observations	2,237	2,237	2,020	2,020	2,020

Notes: This table reports the estimated marginal effects associated with financial literacy in regressions explaining the outstanding debt held by the households. The dependent variables are the logarithm of the amount of debt. PRIDIT measure of actual financial literacy is used. Columns (1) and (2) report the Tobit and OLS estimates. Columns (3) to (5) reports the 2SLS estimates, CMP_Tobit and CMP_OLS estimates, respectively, using mother's education and community-level financial literacy as instruments for financial literacy. Estimates for other covariates are not reported but available upon request. Definitions of all variables are displayed in Appendix 4.1. Standard errors are clustered at the community level and reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4.10 Robustness test 2: Borrowing preference and PRIDIT score for basic financial knowledge

VARIABLES	(1) Probit	(2) LPM	(3) IVreg	(4) CMP_Probit	(5) CMP_LPM
Panel 1: Pref_formal					
Basic FK	0.0007	0.0010	0.0068	0.0481	0.0115
	(0.0034)	(0.0034)	(0.0283)	(0.0419)	(0.0109)
Subj_FK	0.0361***	0.0343***	0.0874	0.1354	0.0279
	(0.0116)	(0.0118)	(0.1028)	(0.1569)	(0.0410)
Panel 2: Pref_informal					
Basic FK	0.0003	0.0003	-0.0384	-0.0042	-0.0014
	(0.0038)	(0.0039)	(0.0338)	(0.0379)	(0.0134)
Subj_FK	-0.0402***	-0.0410***	-0.2301*	-0.2335	-0.0834*
	(0.0143)	(0.0146)	(0.1230)	(0.1426)	(0.0502)
Individual control	Yes	Yes	Yes	Yes	Yes
Household control	Yes	Yes	Yes	Yes	Yes
Province Fixed Effect	Yes	Yes	Yes	Yes	Yes
Observations	2,237	2,237	2,020	2,020	2,020

Notes: This table reports the estimated marginal effects associated with financial literacy in regressions explaining households' preference over formal and informal credit. PRIDIT measure of actual financial literacy is used. Columns (1) and (2) report the Probit and LPM estimates. Columns (3) to (5) report the 2SLS estimates, CMP_Probit and CMP_LPM estimates, respectively, using mother's education and community-level financial literacy as instruments for financial literacy. Estimates for other covariates are not reported but available upon request. Definitions of all variables are displayed in Appendix 4.1. Standard errors are clustered at the community level and reported in parentheses. $***$ $p < 0.01$, $**$ $p < 0.05$, $*$ $p < 0.1$.

Table 4.11 Robustness test 3: Using the debt-to-income ratio as a dependent variable

VARIABLES	(1) Tobit	(2) OLS	(3) IVreg	(4) CMP_Tobit	(5) CMP_OLS
Panel 1: Total_debt					
Basic FK	-0.0139***	-0.3300	0.6084	-12.8014***	-1.2792
	(0.0049)	(0.3150)	(2.0330)	(3.0970)	(0.8655)
Subj_FK	-0.0136	-0.2091	4.6210	2.4999	2.1049
	(0.0086)	(0.4457)	(3.6709)	(5.8758)	(1.6415)
Panel 2: Total_NH_loan					
Basic FK	-0.0120***	-0.0809	-0.8054	-12.5279***	-0.5561
	(0.0042)	(0.2340)	(0.3422)	(4.2374)	(0.6201)
Subj_FK	-0.0059	-0.2377	3.4258	4.0630	0.7964
	(0.0082)	(0.3378)	(2.4116)	(7.8974)	(1.1721)
Panel 3: Debt_formal					
Basic FK	-0.0036	-0.2681	0.2353	-4.8206	-0.6002
	(0.0036)	(0.2857)	(1.6468)	(5.1334)	(0.6954)
Subj_FK	-0.0096**	-0.4625	3.2237	-0.6957	0.6863
	(0.0042)	(0.3306)	(3.1114)	(10.7152)	(1.3154)
Panel 4: Debt_informal					
Basic FK	-0.0096**	-0.0790	0.5238	-8.2503***	-1.2994**
	(0.0045)	(0.0862)	(0.5866)	(1.8881)	(0.5361)
Subj_FK	-0.0090	0.2003	1.6771	2.6587	2.8184***
	(0.0068)	(0.2316)	(1.3989)	(3.8138)	(1.0626)
Panel 5: Debt_NH_bank					
Basic FK	-0.0022	-0.0120	-0.3564	-8.4846	-0.1609
	(0.0030)	(0.1157)	(0.4600)	(5.6277)	(0.2803)
Subj_FK	-0.0079**	-0.2373*	1.1383	8.8263	0.2083
	(0.0039)	(0.1406)	(1.0720)	(11.6693)	(0.5302)
Panel 6: Debt_NH_informal					
Basic FK	-0.0094***	-0.0346	-0.0349	-5.5944***	-0.2387
	(0.0033)	(0.0525)	(0.2902)	(1.7437)	(0.2270)
Subj_FK	-0.0013	0.1264	1.0043	0.6495	0.2939
	(0.0062)	(0.1454)	(0.9785)	(3.2356)	(0.4300)
Individual control	Yes	Yes	Yes	Yes	Yes
Household control	Yes	Yes	Yes	Yes	Yes
Province Fixed Effect	Yes	Yes	Yes	Yes	Yes
Observations	2,237	2,237	2,020	2,020	2,020

Notes: This table reports the estimated marginal effects associated with financial literacy in regressions explaining the outstanding debt-to-income ratios. Columns (1) and (2) report the Tobit and OLS estimates. Columns (3) to (5) report the 2SLS estimates, CMP_Tobit and CMP_OLS estimates, respectively, using mother's education and community-level financial literacy as instruments for financial literacy. Estimates for other covariates are not reported but available upon request. Definitions of all variables are displayed in Appendix 4.1. Standard errors are clustered at the community level and reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4.12 Robustness test 4: Using financial distress as a depenent variable

Fin_diff					
	(1)	(2)	(3)	(4)	(5)
VARIABLES	Probit	Probit	Probit	IVreg	CMP
Basic FK	-0.0439***	-0.0374***	-0.0299***	-0.0756	-0.1798**
	(0.0052)	(0.0058)	(0.0062)	(0.0621)	(0.0902)
Subj_FK	-0.0424***	-0.0302***	-0.0302**	-0.2349**	-0.1545
	(0.0106)	(0.0116)	(0.0129)	(0.1143)	(0.1566)
Individual control	No	Yes	Yes	Yes	Yes
Household control	No	No	Yes	Yes	Yes
Province Fixed Effect	Yes	Yes	Yes	Yes	Yes
Observations	3,233	3,012	2,237	2,020	2,020

Notes: This table reports the estimated marginal effects associated with financial literacy in regressions explaining whether households encounter financial distress. Columns (1) – (3) report estimates obtained from Probit models. Column (4) reports 2SLS estimates and Column (5) reports the CMP-Probit estimates using mother's education and community-level financial literacy as instruments for financial literacy. Definitions of all variables are displayed in Appendix 4.1. Standard errors are clustered at the community level and reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix

Appendix 4.1 Definitions of the variables

Table A4.1 Definitions of the variables

Dependent variables	***Definitions***
D_debt	Dummy: equals 1 if household has outstanding debts to repay, 0 otherwise.
D_formal	Dummy: equals 1 if household is in debt to banks, 0 otherwise.
D_informal	Dummy: equals 1 if household is in debt to friends, relatives or other informal sources, 0 otherwise.
D_NH_loan	Dummy: equals 1 if household has outstanding non-housing debts to repay, 0 otherwise.
D_NH_bank	Dummy: equals 1 if household has outstanding non-housing bank loans, 0 otherwise.
D_NH_inform	Dummy: equals 1 if household has outstanding non-housing loans owed to friends, relatives or other informal sources, 0 otherwise.
Pref _bank	Dummy: equals 1 if household prefers to borrow from bank or other formal financial institutions, in case of house purchasing or operational turnover, 0 otherwise.
Pref _inform	Dummy: equals 1 if household prefers to borrow from friends, relatives or other informal sources, in case of house purchasing or operational turnover, 0 otherwise.
Fin_diff	Dummy: equals 1 if household currently faces financial difficulties, 0 otherwise.
Total debt	Continuous: Net housing and non-housing debt owed by the household.
Debt_formal	Continuous: Net housing and non-housing debt owed to banks by the household.
Debt_informal	Continuous: Net housing and non-housing debt owed to friends, relatives or other informal sources by the household.
Total_NH_loan	Continuous: Total outstanding non-housing debts (i.e. formal non-housing debts)
Debt_NH_bank	Continuous: Outstanding non-housing debt owed to bank (i.e. formal non-housing debts)
Debt_NH_inform	Continuous: Outstanding non-housing debt owed to friends, relatives and other informal sources (i.e. informal non-housing debts)
Independent variables	
Individual-level	
FK_b	Continuous: score for basic financial knowledge
Subj_fk	Discrete: Self-perceived financial knowledge compared with people of similar age. The score ranges from 1 (Way below average) to 5 (Way above average).
Age	Continuous: Age of the respondent
Male	Dummy: equals 1 if the respondent is male, 0 otherwise.
Married	Dummy: equals 1 if the respondent is married or cohabitate, 0 otherwise.
Illiterate	Dummy: equals 1 if the respondent is illiterate, 0 otherwise.
Mathtest14	Discrete: Number of correct answers in math test.
Wordtest14	Discrete: Number of correct answers in verbal test.
Rural	Dummy: equals 1 if the respondent lives in rural area, 0 otherwise.
Migrant	Dummy: equals 1 if the respondent is registered in a rural area but resident in an urban area, 0 otherwise.

Unemployed	Dummy: equals 1 if the respondent is currently not working (includes students and retired people).
Self-employed	Dummy: equals 1 if the respondent is self-employed (includes agricultural work), 0 otherwise.
Risk seeking	Dummy: equals 1 if the respondent is risk seeking, 0 otherwise.
Risk averse	Dummy: equals 1 if the respondent is risk averse, 0 otherwise.
FA_feeling	Discrete: respondent's attitude towards "Spending money makes me more satisfied than saving money", ranging from 1 (totally agree) to 5 (totally disagree).
FA_present	Discrete: the respondent's attitude towards "I intend to live in the present more and do not consider the future", ranging from 1 (totally agree) to 5 (totally disagree).
FA_spend	Discrete: the respondent's attitude towards "In my opinion, money is to be spent and we should spend it whenever we have money", ranging from 1 (totally agree) to 5 (totally disagree).
Health shock	Dummy: equals 1 if the respondent reports worse health during last year, 0 otherwise.
Use Mobile	Dummy: equals 1 if the respondent uses a mobile phone, 0 otherwise.
Household level	
Close_relative	Dummy: equals 1 if the respondent frequently contacts non-resident relatives (at least once a month), 0 otherwise.
N_depchildren	Discrete: number of economically dependent children in the household.
N_depelderly	Discrete: number of economically dependent elderly in the household.
Own house	Dummy: equals 1 if the household owns residential house or other houses, 0 otherwise.
Asset_net	Continuous: household's net wealth, including financial assets (cash, stocks, funds, state debt, trust products, foreign exchange products, futures, stock equity and other financial products) and non-financial assets (fixed assets, durable assets, house and land).
Inc_net	Continuous: annual family income including labour income and investment income.
Community level	
Facility1	Continuous variable: a factor analysis index we computed based on the existence of Convenience/corner/department store, Kindergarten, Primary school, Middle School, Hospital / medical facility, Drug store, Buddhist / Taoist Temple, Ancestral hall, Church/ Mosque, Activity facility/community service center for the elderly, Nursing home/ Almshouse, Physical exercise facility, Playground, Bulletin board for village/community affairs, Informants' letterbox, Community Website/ Weibo in the community. The higher the index, the better the infrastructure in the village/community.
Facility2	Continuous variable: factor analysis index we computed based on whether the village/residential community has access to electricity, cable radio, satellite TV, postal service, telephone, cellphone signal/service, public road, access to the railway, tap water, pipeline gas, bus, subway, the transition from village community to the residential community, as well as whether the community establish enterprises run by local people, implement direct election in the community. The higher the index, the better the facility in the village/community.

Appendix 4.2 Pairwise correlation matrix for the explanatory variables

Table A.4.2.1 Pairwise correlation matrix for the control variables

	FK_b	*Subj_FK*	*Age*	*Male*	*Married*	*Illiterate*	*Mathtest14*	*Wordtest14*	*Rural*	*Migrant*
FK_b	1									
Subj_FK	0.271***	1								
Age	-0.194***	-0.0445*	1							
Male	0.0939***	0.0990***	0.0771***	1						
Married	0.121***	0.00211	-0.144***	0.103***	1					
Illiterate	-0.262***	-0.140***	0.289***	-0.0552*	-0.135***	1				
Mathtest14	0.400***	0.174***	-0.339***	0.101***	0.101***	-0.581***	1			
Wordtest14	0.395***	0.142***	-0.349***	0.0893***	0.0970***	-0.528***	0.710***	1		
Rural	-0.0789***	-0.0384	0.0208	0.0349	0.0155	0.130***	-0.137***	-0.145***	1	
Migrant	-0.0917***	-0.0159	-0.140***	0.0132	0.0477*	0.0664**	-0.151***	-0.0923***	-0.0666**	1
Unemployed	-0.176***	-0.0879***	0.579***	-0.174***	-0.117***	0.202***	-0.286***	-0.258***	0.00923	-0.112***
Self-employed	-0.0683**	0.00894	-0.164***	0.0858***	0.0812***	0.0153	-0.0503*	-0.0134	0.0842***	0.232***
Risk seeking	0.004	0.107***	-0.0369	0.0997***	0.00537	-0.0231	0.0317	0.0510*	-0.0278	0.00657
Risk averse	-0.277***	-0.218***	0.277***	-0.0979***	-0.0651**	0.162***	-0.253***	-0.255***	0.0127	-0.00691
FA_feeling	0.0491*	-0.0371	-0.0680**	0.0164	0.0125	-0.00246	0.0126	0.0341	0.0520*	0.0111
FA_present	0.0826***	-0.00616	-0.211***	-0.0379	0.0243	-0.0485*	0.0623**	0.0647**	0.00378	0.0910***
FA_spend	0.0894***	-0.028	-0.127***	-0.0185	0.0275	-0.0545*	0.0621**	0.0759***	0.00165	0.0168
Health shock	-0.132***	-0.0962***	0.140***	-0.0597**	-0.0173	0.0927***	-0.113***	-0.120***	-0.0222	-0.0181
Use Mobile	0.0926***	0.0681**	-0.374***	0.0718***	0.107***	-0.243***	0.262***	0.266***	-0.014	0.0355
E-shopper	0.228***	0.116***	-0.414***	-0.0615**	0.00991	-0.158***	0.313***	0.307***	-0.0686**	0.00252
Close_relative	0.119***	0.0759***	-0.0818***	-0.0434*	0.0613**	-0.0445*	0.102***	0.100***	-0.0212	-0.00343
N_depchildren	0.0262	0.0530*	-0.427***	-0.0102	0.0691**	-0.0386	0.0440*	0.0417	0.0509*	0.188***
N_depelderly	-0.0643**	0.0151	0.603***	0.0949***	0.0131	0.182***	-0.231***	-0.186***	0.0193	-0.100***
Own house	0.108***	0.0972***	-0.106***	0.0435*	0.122***	-0.0516*	0.0897***	0.0778***	0.0166	0.0352
Lg_asset_net	0.158***	0.147***	-0.0306	0.0438*	0.122***	-0.0248	0.0641**	0.0749***	0.00286	-0.00659
Lg_inc_net	0.284***	0.136***	-0.00169	0.0166	0.144***	-0.181***	0.252***	0.230***	-0.00901	-0.212***
Facility1	-0.011	0.0139	-0.111***	0.04	0.0500*	-0.0112	-0.041	-0.0243	-0.154***	0.223***
Facility2	-0.0122	0.0226	-0.0582**	0.032	0.0504*	-0.00657	-0.0286	-0.0239	-0.0625**	0.0756***

	Unemployed	*Self-employed*	*Risk seeking*	*Risk averse*	*FA_feeling*	*FA_present*	*FA_spend*	*Health shock*	*Use Mobile*	*E-shopper*
Unemployed	1									
Self-employed	-0.311***	1								
Risk seeking	-0.0627**	0.0242	1							
Risk averse	0.226***	-0.0462*	-0.209***	1						
FA_feeling	-0.0538*	-0.0207	-0.0445*	0.0229	1					
FA_present	-0.174***	0.0287	0.00343	-0.104***	0.290***	1				
FA_spend	-0.104***	-0.00994	-0.0204	-0.0452*	0.392***	0.459***	1			
Health shock	0.121***	-0.0698**	-0.0115	0.0962***	0.0367	-0.01	0.0229	1		
Use Mobile	-0.280***	0.0794***	0.0246	-0.148***	-0.00315	0.0819***	0.0483*	-0.0741***	1	
E-shopper	-0.241***	0.0239	0.0279	-0.259***	-0.0208	0.0776***	0.0263	-0.0759***	0.126***	1
Close_relative	-0.0579**	0.00656	0.018	-0.0812***	-0.00558	0.028	0.0442*	-0.0234	0.0836***	0.111***
N_depchildren	-0.302***	0.217***	0.022	-0.154***	0.0342	0.154***	0.0854***	-0.0606**	0.141***	0.101***
N_depelderly	0.447***	-0.126***	-0.0495*	0.129***	-0.0263	-0.103***	-0.0884***	0.0808***	-0.264***	-0.180***
Own house	-0.0926***	0.0295	0.0133	-0.128***	0.00427	0.0216	-0.00618	-0.025	0.0573**	0.135***
Lg_asset_net	-0.0583**	0.00748	0.0154	-0.133***	-0.0171	-0.0026	-0.0321	-0.0444*	0.0267	0.106***
Lg_inc_net	-0.0417	-0.250***	0.0487*	-0.173***	-0.0477*	0.0202	0.00126	-0.0476*	0.0506*	0.169***
Facility1	-0.150***	0.157***	-0.00571	-0.00917	-0.0221	0.0504*	0.0137	-0.0183	0.0152	0.00801
Facility2	-0.038	0.0415	-0.0227	-0.0287	0.0468*	0.0284	0.0315	0.0142	0.0308	0.00151

	Close_relative	N_depchildren	N_depelderly	Own house	Lg_asset_net	Lg_inc_net	Facility1	Facility2
Close_relative	1							
N_depchildren	0.0435*	1						
N_depelderly	-0.0705**	-0.274***	1					
Own house	0.0575**	0.109***	-0.0479*	1				
Lg_asset_net	0.0645**	0.103***	0.0302	0.779***	1			
Lg_inc_net	0.127***	0.0592**	0.0794***	0.168***	0.232***	1		
Facility1	0.00464	0.133***	-0.0669**	0.00619	-0.00235	-0.0714***	1	
Facility2	0.0313	0.0417	-0.0148	0.0297	0.0834***	-0.00216	0.191***	1

Table A.4.2.2 Pairwise correlation matrix for the bivariate dependent variables

	D_debt	*D_formal*	*D_informal*	*D_NH_loan*	*D_NH_bank*	*D_NH_inform*	*Prep_bank*	*Prep_inform*	*Fin_diff*
D_debt	1								
D_formal	0.6500*	1							
D_informal	0.7889*	0.2001*	1						
D_NH_loan	0.6423*	0.2718*	0.6504*	1					
D_NH_bank	0.3321*	0.5104*	0.1406*	0.5188*	1				
D_NH_inform	0.5654*	0.0798*	0.7167*	0.8803*	0.1505*	1			
Pref_bank	0.0372	0.1624*	-0.0781*	-0.015	0.0849*	-0.0675*	1		
Pref_inform	0.1020*	-0.0589*	0.1873*	0.1004*	-0.0277	0.1372*	-0.6731*	1	

Appendix 4.3 Validity tests for instrumental variables

In this section, we discuss the results for the validity tests for the IVs, i.e., mother's duration of education and community-level financial literacy (See Table A.4.3).

We first perform a set of endogeneity tests. The rejection of the null hypothesis of the endogeneity test indicates that the suspected endogenous regressors cannot be treated as exogenous and further work should be done to tackle the endogenous issue. Otherwise, we can rely on the Probit or Tobit estimates. The results for the endogeneity test (in Rows 6 and 7) suggest that we cannot reject the null hypothesis for the specifications with *D_formal*, *D_NH_bank, Debt_NH_bank* and *Pref_formal,* and we rely on the Probit or Tobit estimates for these specifications.

Next, for all the specifications that reject the null hypothesis of the endogeneity test, we perform the under-identification test (Kleibergen-Paap rk LM test), weak identification test (Kleibergen-Paap rk Wald F test in case of heterogeneity) and overidentification test (Hansen J test).

For all the specifications, we reject the null hypothesis for the under-identification test, implying that all the equations are identified, and the instruments are relevant (See Rows 1 and 2 in each panel). For all the specifications, we cannot reject the null hypothesis for the overidentification test, implying that the instruments are uncorrelated with the error terms in the main equations (see Rows 4 and 5 in each panel). All the specifications with binary dependent variables (Panels 1 and 3) fail the weak-identification test, since the Kleibergen-Paap Wald rk F statistics (equal 8.583) are lower than the rule-of-thumb critical value of 10, implying the existence of a weak IV problem. This highlights the necessity of employing a CMP framework for the IV approach, which performs better than the 2SLS framework in case of weak IV (Roodman, 2011). While all the specifications with debt levels (Panel 2) pass the weak-identification test, with KP-F statistics of 15.05.

Table A.4.3 Validity test for instrument variables

	Panel 1	(1)	(2)	(3)	(4)	(5)	(6)
		D_debt	D_NH_loan	D_formal	D_informal	D_NH_bank	D_NH_informal
(1)	KP Chi-sq	24.63	24.63	24.63	24.63	24.63	24.63
(2)	p-val KP Chi-sq	4.48E-06	4.48E-06	4.48E-06	4.48E-06	4.48E-06	4.48E-06
(3)	KP F	8.583	8.583	8.583	8.583	8.583	8.583
(4)	Hansen J	0.445	0.49	0.0808	0.0289	0.381	0.331
(5)	p-val Hansen J	0.505	0.484	0.776	0.865	0.537	0.565
(6)	Endog	11.22	6.925	3.849	9.987	4.282	6.343
(7)	p-val Endog	0.00366	0.0314	0.146	0.00678	0.118	0.0419
	Panel 2	(7)	(8)	(9)	(10)	(11)	(12)
		Total_debt	Debt_formal	Debt_informal	Total_NH_loan	Debt_NH_bank	Debt_NH_informal
(1)	KP Chi-sq	43.76	43.76	43.76	43.76	43.74	43.76
(2)	p-val KP Chi-sq	3.15E-10	3.15E-10	3.15E-10	3.15E-10	3.18E-10	3.15E-10
(3)	KP F	15.05	15.05	15.05	15.05	15.05	15.05
(4)	Hansen J	0.538	0.623	0.0805	0.179	0.0103	0.0856
(5)	p-val Hansen J	0.463	0.43	0.777	0.672	0.919	0.77
(6)	Endog	17.73	12.35	9.783	8.549	4.868	6.137
(7)	p-val Endog	0.000141	0.00208	0.00751	0.0139	0.0877	0.0465
	Panel 3	(13)	(14)				
		Pref_formal	Pref_informal				
(1)	KP Chi-sq	24.44	24.44				
(2)	p-val KP Chi-sq	4.94E-06	4.94E-06				
(3)	KP F	8.517	8.517				
(4)	Hansen J	0.9	1.837				
(5)	p-val Hansen J	0.343	0.175				
(6)	Endog	1.138	7.631				
(7)	p-val Endog	0.566	0.022				

Appendix 4.4 Probit models: Full table

Table A4.4 Determinants of the probability of indebtedness and borrowing preference: Full table

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	D_debt	D_NH_loan	D_formal	D_informal	D_NH_bank	D_NH_informal	Pref_formal	Pref_informal
Basic FK	-0.0256***	-0.0150***	-0.0143***	-0.0182***	-0.0051*	-0.0131***	-0.0006	0.0020
	(0.0064)	(0.0049)	(0.0050)	(0.0058)	(0.0030)	(0.0043)	(0.0072)	(0.0081)
Subj_FK	-0.0154	-0.0112	-0.0049	-0.0104	-0.0089*	-0.0067	0.0347***	-0.0383***
	(0.0131)	(0.0086)	(0.0102)	(0.0111)	(0.0046)	(0.0083)	(0.0118)	(0.0142)
Age	-0.0048***	-0.0022***	-0.0023***	-0.0033***	-0.0004	-0.0020***	-0.0018*	-0.0057***
	(0.0011)	(0.0007)	(0.0008)	(0.0010)	(0.0004)	(0.0007)	(0.0011)	(0.0012)
Male	0.0251	0.0149	0.0101	0.0016	0.0027	0.0091	0.0117	-0.0028
	(0.0187)	(0.0140)	(0.0139)	(0.0180)	(0.0075)	(0.0141)	(0.0200)	(0.0226)
Married	0.0081	-0.0017	0.0409	-0.0085	-0.0029	0.0067	0.0760**	-0.0239
	(0.0315)	(0.0210)	(0.0277)	(0.0253)	(0.0134)	(0.0190)	(0.0296)	(0.0330)
Illiterate	0.0214	0.0375	-0.0216	0.0176	-0.0003	0.0362*	0.0467	0.0116
	(0.0356)	(0.0255)	(0.0336)	(0.0278)	(0.0185)	(0.0217)	(0.0404)	(0.0408)
Mathtest	0.0012	-0.0014	0.0042**	-0.0019	0.0005	-0.0015	0.0044	-0.0011
	(0.0026)	(0.0018)	(0.0020)	(0.0022)	(0.0013)	(0.0017)	(0.0027)	(0.0028)
Wordtest	-0.0003	0.0020*	0.0009	0.0003	0.0014*	0.0012	0.0020	-0.0032*
	(0.0016)	(0.0012)	(0.0013)	(0.0014)	(0.0008)	(0.0011)	(0.0017)	(0.0019)
Migrant	0.0402	0.0182	0.0115	0.0479*	-0.0146	0.0280	0.0014	-0.0039
	(0.0315)	(0.0227)	(0.0212)	(0.0254)	(0.0124)	(0.0196)	(0.0286)	(0.0353)
Unemployed	0.0212	0.0107	0.0109	0.0094	-0.0023	0.0094	-0.0114	0.0097
	(0.0232)	(0.0166)	(0.0185)	(0.0217)	(0.0106)	(0.0156)	(0.0254)	(0.0297)
Self-employed	0.0422	0.0341*	0.0548***	0.0056	0.0274***	0.0068	0.0149	0.0058
	(0.0297)	(0.0194)	(0.0200)	(0.0250)	(0.0095)	(0.0193)	(0.0307)	(0.0344)
Risk Seeking	0.1207**	0.0563	0.0044	0.1499***	0.0091	0.0565	0.0565	-0.0419
	(0.0526)	(0.0395)	(0.0394)	(0.0448)	(0.0217)	(0.0372)	(0.0641)	(0.0742)
Risk Averse	-0.0082	-0.0067	-0.0223	0.0024	0.0022	-0.0118	-0.0021	-0.0354
	(0.0217)	(0.0168)	(0.0152)	(0.0193)	(0.0087)	(0.0160)	(0.0213)	(0.0266)
FA_Feeling	0.0013	-0.0014	0.0019	0.0009	-0.0038	0.0010	0.0000	0.0007
	(0.0080)	(0.0058)	(0.0058)	(0.0072)	(0.0030)	(0.0056)	(0.0081)	(0.0090)

FA_Time	0.0019	-0.0059	-0.0020	0.0036	-0.0039	-0.0017	0.0004	-0.0068
	(0.0077)	(0.0061)	(0.0058)	(0.0070)	(0.0033)	(0.0056)	(0.0074)	(0.0087)
FA_Spend	-0.0046	-0.0007	0.0004	-0.0003	0.0041	-0.0031	0.0001	0.0144
	(0.0081)	(0.0058)	(0.0059)	(0.0072)	(0.0033)	(0.0055)	(0.0078)	(0.0088)
Health Shock	0.0640***	0.0560***	0.0127	0.0662***	0.0014	0.0612***	0.0226	-0.0122
	(0.0176)	(0.0124)	(0.0143)	(0.0154)	(0.0089)	(0.0116)	(0.0190)	(0.0213)
Use Mobile	0.1186**	0.0859***	0.0378	0.1182***	0.0135	0.0704**	0.0256	0.0219
	(0.0490)	(0.0330)	(0.0386)	(0.0401)	(0.0282)	(0.0293)	(0.0401)	(0.0437)
Rural	-0.0404	-0.0279	-0.0550	-0.0018	0.0076	-0.0404	0.0132	0.0727
	(0.0424)	(0.0326)	(0.0463)	(0.0348)	(0.0198)	(0.0271)	(0.0444)	(0.0703)
Closerelative	0.0270	0.0136	0.0195	0.0030	0.0236***	-0.0050	-0.0024	0.0471**
	(0.0165)	(0.0140)	(0.0131)	(0.0154)	(0.0087)	(0.0127)	(0.0184)	(0.0212)
N_depchild	0.0487***	0.0296***	0.0234**	0.0456***	0.0093	0.0319***	-0.0160	0.0129
	(0.0141)	(0.0100)	(0.0113)	(0.0125)	(0.0063)	(0.0094)	(0.0156)	(0.0163)
N_depold	-0.0240	-0.0043	-0.0168	0.0022	-0.0031	-0.0014	-0.0100	-0.0084
	(0.0148)	(0.0103)	(0.0123)	(0.0134)	(0.0077)	(0.0094)	(0.0147)	(0.0157)
Own house	-0.0679	-0.0983	-0.2472	0.2694**	-0.2734***	0.0608	-0.4874***	0.5449***
	(0.1614)	(0.0889)	(0.1839)	(0.1139)	(0.0665)	(0.0712)	(0.1338)	(0.1673)
Lg_asset_net	0.0062	-0.0005	0.0257*	-0.0247***	0.0203***	-0.0129**	0.0386***	-0.0417***
	(0.0126)	(0.0069)	(0.0142)	(0.0086)	(0.0050)	(0.0054)	(0.0102)	(0.0131)
Lg_inc_net	-0.0090	-0.0241***	0.0126	-0.0152*	-0.0099**	-0.0189***	0.0124	0.0036
	(0.0111)	(0.0074)	(0.0093)	(0.0091)	(0.0043)	(0.0066)	(0.0117)	(0.0124)
Facility1	0.0164	0.0036	0.0229	-0.0031	0.0321	-0.0017	0.0110	-0.0034
	(0.0621)	(0.0431)	(0.0481)	(0.0528)	(0.0218)	(0.0397)	(0.0472)	(0.0608)
Facility2	-0.0582**	-0.0143	-0.0405*	-0.0190	-0.0096	-0.0064	-0.0035	0.0436
	(0.0267)	(0.0166)	(0.0207)	(0.0228)	(0.0089)	(0.0167)	(0.0227)	(0.0298)
Observations	2,237	2,237	2,237	2,237	2,237	2,237	2,237	2,237

Notes: This table reports the baseline Probit estimates of the marginal effects of the explanatory variables. Dependent variables are listed in the header of the table. Standard errors are clustered at community level and reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Appendix 4.5 Tobit models: Full table

Table A4.5 Determinants of debt levels: Full table

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Total_debt	Total_NH_loan	Debt_formal	Debt_informal	Debt_NH_bank	Debt_NH_informal
Basic FK	-0.0250***	-0.0145***	-0.0143***	-0.0170***	-0.0048*	-0.0123***
	(0.0062)	(0.0049)	(0.0050)	(0.0055)	(0.0029)	(0.0042)
Subj_FK	-0.0137	-0.0105	-0.0042	-0.0096	-0.0086*	-0.0059
	(0.0128)	(0.0087)	(0.0100)	(0.0106)	(0.0045)	(0.0080)
Age	-0.0048***	-0.0022***	-0.0023***	-0.0031***	-0.0004	-0.0019***
	(0.0011)	(0.0007)	(0.0008)	(0.0009)	(0.0004)	(0.0007)
Male	0.0252	0.0158	0.0096	0.0019	0.0030	0.0091
	(0.0182)	(0.0137)	(0.0139)	(0.0172)	(0.0073)	(0.0135)
Married	0.0068	-0.0030	0.0413	-0.0087	-0.0031	0.0061
	(0.0309)	(0.0208)	(0.0278)	(0.0244)	(0.0130)	(0.0182)
Illiterate	0.0210	0.0374	-0.0191	0.0139	0.0026	0.0335*
	(0.0342)	(0.0245)	(0.0337)	(0.0261)	(0.0173)	(0.0204)
Mathtest	0.0017	-0.0013	0.0044**	-0.0018	0.0006	-0.0013
	(0.0025)	(0.0018)	(0.0020)	(0.0021)	(0.0012)	(0.0016)
Wordtest	-0.0003	0.0020*	0.0008	0.0003	0.0014*	0.0012
	(0.0015)	(0.0012)	(0.0013)	(0.0013)	(0.0008)	(0.0010)
Migrant	0.0378	0.0173	0.0097	0.0442*	-0.0133	0.0260
	(0.0290)	(0.0218)	(0.0202)	(0.0235)	(0.0115)	(0.0185)
Unemployed	0.0243	0.0124	0.0118	0.0088	-0.0025	0.0103
	(0.0228)	(0.0164)	(0.0186)	(0.0205)	(0.0107)	(0.0149)
Self-employed	0.0440	0.0357*	0.0543***	0.0097	0.0281***	0.0088
	(0.0276)	(0.0190)	(0.0190)	(0.0233)	(0.0091)	(0.0185)
Risk Seeking	0.0995**	0.0522	-0.0039	0.1383***	0.0054	0.0585*
	(0.0459)	(0.0382)	(0.0371)	(0.0388)	(0.0203)	(0.0353)
Risk Averse	-0.0121	-0.0067	-0.0240	0.0021	0.0015	-0.0117
	(0.0211)	(0.0166)	(0.0149)	(0.0184)	(0.0082)	(0.0153)
FA_Feeling	0.0012	-0.0020	0.0020	0.0002	-0.0041	0.0007

	(0.0076)	(0.0057)	(0.0058)	(0.0068)	(0.0029)	(0.0054)
FA_Time	0.0024	-0.0051	-0.0021	0.0043	-0.0030	-0.0011
	(0.0074)	(0.0060)	(0.0058)	(0.0065)	(0.0032)	(0.0053)
FA_Spend	-0.0038	-0.0003	0.0005	0.0006	0.0037	-0.0028
	(0.0078)	(0.0058)	(0.0058)	(0.0067)	(0.0032)	(0.0053)
Health Shock	0.0610***	0.0545***	0.0120	0.0625***	0.0017	0.0583***
	(0.0169)	(0.0123)	(0.0141)	(0.0144)	(0.0087)	(0.0111)
Use Mobile	0.1202**	0.0850**	0.0391	0.1151***	0.0120	0.0666**
	(0.0508)	(0.0337)	(0.0400)	(0.0391)	(0.0276)	(0.0287)
Rural	-0.0415	-0.0289	-0.0551	-0.0042	0.0081	-0.0402
	(0.0422)	(0.0321)	(0.0473)	(0.0330)	(0.0189)	(0.0258)
Closerelative	0.0247	0.0140	0.0184	0.0039	0.0229***	-0.0046
	(0.0162)	(0.0139)	(0.0130)	(0.0147)	(0.0085)	(0.0122)
N_depchild	0.0494***	0.0301***	0.0250**	0.0443***	0.0086	0.0316***
	(0.0131)	(0.0095)	(0.0112)	(0.0113)	(0.0061)	(0.0086)
N_depold	-0.0247*	-0.0053	-0.0167	0.0016	-0.0023	-0.0021
	(0.0148)	(0.0105)	(0.0123)	(0.0130)	(0.0075)	(0.0092)
Own house	-0.0702	-0.1015	-0.2436	0.2438**	-0.2653***	0.0622
	(0.1615)	(0.0888)	(0.1837)	(0.0997)	(0.0643)	(0.0680)
Lg_asset_net	0.0069	0.0003	0.0255*	-0.0222***	0.0195***	-0.0123**
	(0.0126)	(0.0069)	(0.0141)	(0.0075)	(0.0048)	(0.0052)
Lg_inc_net	-0.0039	-0.0218***	0.0140	-0.0130	-0.0079*	-0.0165***
	(0.0102)	(0.0070)	(0.0089)	(0.0084)	(0.0042)	(0.0061)
Facility1	0.0173	0.0042	0.0256	-0.0025	0.0297	-0.0004
	(0.0608)	(0.0426)	(0.0471)	(0.0502)	(0.0210)	(0.0376)
Facility2	-0.0578**	-0.0169	-0.0392**	-0.0207	-0.0080	-0.0077
	(0.0260)	(0.0164)	(0.0200)	(0.0220)	(0.0086)	(0.0160)
Observations	2,237	2,237	2,237	2,237	2,237	2,237

Notes: This table reports the baseline Tobit estimates of the marginal effects of the explanatory variables. Dependent variables are listed in the header of the table. Standard errors are clustered at community level and reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Chapter Five: Conclusion

5.1 Summary of the Findings

In Chapter Two, we explore the relationship between health shocks and cash holding of middle-aged and aged rural residents in China, based on a three-period panel from CHARLS, with 8,123 observations in each wave. Using a set of dynamic response models and fractional response models, we find that, for middle-aged and elderly people living in rural China, health shocks are associated with a higher probability of holding only cash in their financial portfolios, along with a higher proportion of their financial portfolios held in cash. These relationships either disappear or largely decline in magnitude after controlling for financial wealth. We also augment our baseline specification with interactions of health shocks with the type of health insurance reimbursement, access to banks and indicators for a large family. We find that the positive relationships between health shocks and cash holding are more pronounced for those who cannot get immediate reimbursement, do not have access to bank accounts, and live alone or with their spouse.

In Chapter Three, we study the association between financial literacy and online shopping decisions based on a cross-sectional sample with 2,183 observations from the 2014 wave of CFPS. We estimate a set of Probit models for the participation in online shopping, and OLS models for the amount of online shopping expenditure. We also apply the IV approach in a CMP framework (Roodman, 2011), using parental education and community-level financial literacy as instruments for individuals' financial literacy. We find that respondents with higher financial literacy show a higher probability of shopping online and spend more online. We then interact financial literacy with age groups dummies and income and wealth gradients. We find that the detected effects are stronger for younger respondents, and for those with higher levels

of income and wealth. These results are robust to using alternative measures of financial literacy, alternative estimation methods, and alternative models that consider peer effects and cognitive ability.

In Chapter Four, we analyse how financial literacy can improve households' debt status and shape households' preferences over formal and informal credit. Our analysis is based on a cross-sectional sample consisting of 2,237 observations from the 2014 wave of the CFPS. A set of Probit and Tobit models are estimated. An IV approach is further applied in both 2SLS and CMP frameworks, using parental education and community-level financial literacy as instruments for individuals' financial literacy. Overall, we find that financially literate households are less likely to be indebted, show lower levels of debt, and are less likely to report financial distress. Such households are less likely to use informal credit. We also observe that financially literate households show a strong preference for formal credit in case they have to borrow in the future, but we find no evidence that they are more likely use formal credit in practice. This may be explained bearing in mind that the financially literate are less likely to be indebted in the first place. These results are robust to using alternative measures of financial literacy and indebtedness, and alternative estimation methods.

5.2 **Policy Implications**

This thesis puts forward several policy implications.

Chapter Two finds that the onset of a health shock is associated with a higher probability of holding only cash as financial assets and a higher share of cash in households' financial portfolios. Such a relationship operates through financial wealth, and is more pronounced for those who cannot get immediate reimbursement for healthcare expenses, do not have access to banks, or live alone or with their spouse. These findings imply that policymakers can protect

households from health shocks by reducing out-of-pocket costs of healthcare, providing immediate reimbursement of healthcare costs and improving access to financial services. Moreover, it is worth noting that the parents of the 'one-child generation' have either already retired or close to retirement, and, as such, are faced with increasing health risks. This calls for a supportive policy that can protect the 'one-child generation' who lack a family-based safety net against parental health shocks.

In Chapter Three, we document that consumers with higher financial literacy are more likely to shop online and spend more on online shopping. Such patterns are more pronounced for young consumers and for those with higher income and wealth. These findings suggest that policymakers should not only be aware of the benefits of financial literacy in the context of taking up online shopping, but should also be aware that such benefits may be more pronounced for the young and the better-off. Therefore, financial education may be more effective at early stages of life, and should consider the potentially detrimental consequences of online shopping, such as indebtedness.

In Chapter Four, we observe that financially literate households are less likely to be indebted, have lower debt level, and are less likely to report financial distress. These households also show a preference for formal credit and against informal credit. This implies that financial literacy is effective in improving debt management and enhancing financial inclusion in terms of formal credit for the general population.

5.3 **Limitations and Future Research**

In Chapter Two, though we have applied rigorous econometric methods to control for the initial condition issue and state dependence, we cannot draw a causal inference. Additionally, due to the limitation in the data, we cannot identify the exact timing of the health shock and

the way that the patient paid for the medicine and treatment, which limits interpretation. Future research can search for sophisticated instruments for health shocks to identify potential causality between health shocks and cash holdings. Moreover, we could extend our research by using more recent waves of the CHARLS database (when they become available) to investigate the dynamics of the relationship in a longer time span. Future research could also look at the transitions in other households' economic behaviours following a health shock, such as education and property investments and borrowing decisions, to obtain a full picture of household decisions in the presence of health shocks.

For Chapter Three, due to the limitation of the database, we cannot verify the potential channels that link financial literacy and online shopping, i.e., the adoption of bank accounts and digital payment tools. Furthermore, our analysis is restricted to only the respondents who answered the financial knowledge questions in the 2014 wave of the CFPS, leading to a loss of observations. Therefore, our final sample may not be representative of the whole population. We could extend our research by looking at whether online shopping leads to overspending and indebtedness in the future.

In Chapter Four, due to the limitation of the data, we are not able to adequately address the factors that are related to the supply of formal financial services. Moreover, in the CFPS, we cannot isolate the loans from friends and relatives from the loans owed to other individuals or informal institutions. However, these two sources of loans can differ substantially in terms of risk and cost, and people may have distinct preferences over them. Finally, similar to Chapter Three, the sample we use in this analysis may not be representative for the whole Chinese population. Future research can also look into households' holding of mortgages, and explore alternative databases on financial literacy and household finance in China, or in other emerging countries.

Reference

Abreu, M. & Mendes, V. 2010. Financial literacy and portfolio diversification. *Quantitative Finance,* 10**,** 515-528.

Agarwal, S. & Mazumder, B. 2013. Cognitive abilities and household financial decision making. *American Economic Journal: Applied Economics,* 5**,** 193-207.

Agarwalla, S. K., Barua, S. K., Jacob, J. & Varma, J. R. 2015. Financial literacy among working young in urban India. *World Development,* 67**,** 101-109.

Alba, J., Lynch, J., Weitz, B., Janiszewski, C., Lutz, R., Sawyer, A. & Wood, S. 1997. Interactive home shopping: consumer, retailer, and manufacturer incentives to participate in electronic marketplaces. *Journal of Marketing,* 61**,** 38-53.

Alessie, R., Van Rooij, M. & Lusardi, A. 2011. Financial literacy and retirement preparation in the Netherlands. *Journal of Pension Economics & Finance,* 10**,** 527-545.

Allgood, S. & Walstad, W. B. 2016. The effects of perceived and actual financial literacy on financial behaviors. *Economic Inquiry,* 54**,** 675-697.

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

Alter, A., Feng, A. X. & Valckx, N. 2018. *Understanding the Macro-Financial Effects of Household Debt: A Global Perspective*, International Monetary Fund.

Ameriks, J. & Zeldes, S. P. 2004. How Do Household Portfolio Shares Vary with Age. Working paper, Columbia University.

Anderson, A., Baker, F. & Robinson, D. T. 2017. Precautionary savings, retirement planning and misperceptions of financial literacy. *Journal of Financial Economics,* 126**,** 383-398.

Atella, V., Brunetti, M. & Maestas, N. 2012. Household Portfolio Choices, Health Status and Health Care Systems: A Cross-Country Analysis Based on Share. *Journal of Banking & Finance,* 36**,** 1320-1335.

Atkinson, A. & Messy, F.-A. 2011. Assessing financial literacy in 12 countries: an OECD/INFE international pilot exercise. *Journal of Pension Economics & Finance,* 10**,** 657-665.

Atkinson, A. & 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, OECD Publishing.

Atkinson, A., Monticone, C. & Mess, F. 2016. *OECD/INFE international survey of adult financial literacy competencies* [Online]. Available: www.oecd.org/finance/OECD-INFE-International-Survey-of-Adult-Financial-Literacy-Competencies.pdf.

Au, N. & Johnston, D. W. 2014. Self-Assessed Health: What Does It Mean and What Does It Hide? *Social Science & Medicine,* 121**,** 21-28.

Babiarz, P. & Robb, C. A. 2014. Financial literacy and emergency saving. *Journal of Family Economic Issues,* 35**,** 40-50.

Baker, M., Stabile, M. & Deri, C. 2004. What Do Self-Reported, Objective, Measures of Health Measure? *Journal of Human Resources,,* 39(4), pp.1067-1093.

Baltagi, B. H. 1995. *Econometric Analysis of Panel Data*, Wiley New York.

Banks, J., Blundell, R., Bozio, A. & Emmerson, C. 2011. Disability, Health and Retirement in the United Kingdom. *National Bureau of Economic Research*, No.W17049.

Barnay, T. 2016. Health, Work and Working Conditions: A Review of the European Economic Literature. *The European Journal of Health Economics,* 17**,** 693-709.

Bech, P., Olsen, L. R., Kjoller, M. & Rasmussen, N. K. 2003. Measuring Well-Being Rather Than the Absence of Distress Symptoms: A Comparison of the Sf-36 Mental Health Subscale and the Who-Five Well-Being Scale. *International Journal of Methods in Psychiatric Research,* 12**,** 85-91.

Behrman, J. R., Mitchell, O. S., Soo, C. & Bravo, D. 2012. How financial literacy affects household wealth accumulation. *American Economic Review: Papers & Proceedings*, 102 (3), 300-304.

Benitez-Silva, H. & Ni, H. 2008. Health Status and Health Dynamics in an Empirical Model of Expected Longevity. *Journal of Health Economics,* 27**,** 564-584.

Benzoni, L., Collin-Dufresne, P. & Goldstein, R. S. 2007. Portfolio Choice over the Life-Cycle When the Stock and Labor Markets Are Cointegrated. *The Journal of Finance,* 62**,** 2123-2167.

Berkowitz, M. K. & Qiu, J. 2006. A Further Look at Household Portfolio Choice and Health Status. *Journal of Banking & Finance,* 30**,** 1201-1217.

Bernheim, B. D. & Garrett, D. M. 2003. The effects of financial education in the workplace: Evidence from a survey of households. *Journal of Public Economics,* 87**,** 1487-1519.

Bernheim, B. D., Garrett, D. M. & Maki, D. M. 2001. Education and saving: The long-term effects of high school financial curriculum mandates. *Journal of Public Economics,* 80**,** 435-465.

Bertocchi, G., Brunetti, M. & Torricelli, C. 2011. Marriage and Other Risky Assets: A Portfolio Approach. *Journal of Banking & Finance,* 35**,** 2902-2915.

Betti, G., Dourmashkin, N., Rossi, M. & Ping Yin, Y. 2007. Consumer over-indebtedness in the EU: measurement and characteristics. *Journal of Economic Studies*, 34, 136-156.

Bianchi, M. 2018. Financial literacy and portfolio dynamics. *The Journal of Finance,* 73**,** 831-859.

Bloomberg. 2019. *Global Debt of $244 Trillion Nears Record Despite Faster Growth* [Online]. Available: https://www.bloomberg.com/news/articles/2019-01-15/global-debt-of-244-trillion-nears-record-despite-faster-growth [Accessed 12th, Aug 2019].

Bosnjak, M., Galesic, M. & Tuten, T. 2007. Personality determinants of online shopping: Explaining online purchase intentions using a hierarchical approach. *Journal of Business Research,* 60**,** 597-605.

Bowling, A. & Windsor, J. 2008. The Effects of Question Order and Response-Choice on Self-Rated Health Status in the English Longitudinal Study of Ageing (Elsa). *Journal of Epidemiology & Community Health,* 62**,** 81-85.

Brashear, T. G., Kashyap, V., Musante, M. D. & Donthu, N. 2009. A profile of the internet shopper: evidence from six countries. *Journal of Marketing Theory and Practice,* 17**,** 267-282.

Bressan, S., Pace, N. & Pelizzon, L. 2014. Health Status and Portfolio Choice: Is Their Relationship Economically Relevant? *International Review of Financial Analysis,* 32**,** 109-122.

Brockett, P. L., Derrig, R. A., Golden, L. L., Levine, A. & Alpert, M. 2002. Fraud classification using principal component analysis of RIDITs. *Journal of Risk and Insurance,* 69**,** 341-371.

Bross, I. D. 1958. How to use ridit analysis. *Biometrics***,** 18-38.

Brown, M. & Graf, R. 2013. Financial literacy and retirement planning in Switzerland. *Numeracy,* 6**,** 6.

Browning, M. & Crossley, T. F. 2001. The life-cycle model of consumption and saving. *Journal of Economic Perspectives,* 15**,** 3-22.

Bucher-Koenen, T. & Lusardi, A. 2011. Financial literacy and retirement planning in Germany. *Journal of Pension Economics & Finance,* 10**,** 565-584.

Bucher-Koenen, T., Lusardi, A., Alessie, R. & Van Rooij, M. 2017. How financially literate are women? An overview and new insights. *Journal of Consumer Affairs,* 51**,** 255-283.

Bünnings, C. 2017. Does New Health Information Affect Health Behaviour? The Effect of Health Events on Smoking Cessation. *Applied Economics,* 49**,** 987-1000.

Bursztyn, L., Ederer, F., Ferman, B. & Yuchtman, N. 2012. Understanding peer effects in financial decisions: Evidence from a field experiment. *National Bureau of Economic Research,* No. W18241.

Calcagno, R. & Monticone, C. 2015. Financial literacy and the demand for financial advice. *Journal of Banking & Finance,* 50**,** 363-380.

Calvet, L. E. & Sodini, P. 2014. Twin Picks: Disentangling the Determinants of Risk‐Taking in Household Portfolios. *The Journal of Finance,* 69**,** 867-906.

Calvet, L. E., Campbell, J. Y. & Sodini, P. 2009. Measuring the Financial Sophistication of Households. *American Economic Review,* 99**,** 393-98.

Calvet, L. E., Campbell, J. Y. & Sodini, P. 2009. Measuring the financial sophistication of households. *American Economic Review,* 99**,** 393-98.

Campbell, J. Y. 2006. Household finance. *Journal of Finance,* 61**,** 1553-1604.

Cardak, B. A. & Wilkins, R. 2009. The Determinants of Household Risky Asset Holdings: Australian Evidence on Background Risk and Other Factors. *Journal of Banking & Finance,* 33**,** 850-860.

Castellacci, F. & Tveito, V. 2018. Internet use and well-being: A survey and a theoretical framework. *Research policy*, 47, 308-325.

Charness, G. & Gneezy, U. 2012. Strong Evidence for Gender Differences in Risk Taking. *Journal of Economic Behavior & Organization,* 83**,** 50-58.

Chase, S., Gjertson, L. & Collins, J. M. 2011. Coming up with cash in a pinch: Emergency savings and its alternatives. *CFS Issue Brief*, 6.

Chetty, R., Sándor, L. & Szeidl, A. 2017. The Effect of Housing on Portfolio Choice. *The Journal of Finance,* 72**,** 1171-1212.

China Internet Network Information Center (CNNIC). 2018. Available: http://www.cnnic.cn/hlwfzyj/hlwxzbg/hlwtjbg/201803/P020180305409870339136.pdf [Accessed 15 Dec 2018].

Christelis, D., Jappelli, T. & Padula, M. 2010. Cognitive Abilities and Portfolio Choice. *European Economic Review,* 54**,** 18-38.

Christiansen, C., Joensen, J. S. & Rangvid, J. 2015. Understanding the Effects of Marriage and Divorce on Financial Investments: The Role of Background Risk Sharing. *Economic Inquiry,* 53**,** 431-447.

Clark, R. L., Morrill, M. S. & Allen, S. G. 2012. The role of financial literacy in determining retirement plans. *Economic Inquiry,* 50**,** 851-866.

CNNIC 2016. Report on E-commerce Market in China (2015). [Online]. Available: http://www.cnnic.cn/hlwfzyj/hlwxzbg/dzswbg/201606/P020160721526975632273.pdf [Accessed 10th May 2019].

CNNIC 2015. The 35th China Statistical Report on Internet Development. [Online]. Available: http://www.cnnic.cn/hlwfzyj/hlwxzbg/201502/P020150203551802054676.pdf [Accessed 10th Feb 2020].

CNNIC 2019. The 43rd China Statistical Report on Internet Development. [Online]. Available: http://www.cnnic.cn/hlwfzyj/hlwxzbg/hlwtjbg/201902/P020190318523029756345.pdf [Accessed 10th Feb 2020].

Cohen, M. & Nelson, C. 2011. Financial literacy: a step for clients towards financial inclusion. *Global Microcredit Summit***,** 14-17.

Coile, C. & Milligan, K. 2009. How Household Portfolios Evolve after Retirement: The Effect of Aging and Health Shocks. *Review of Income and Wealth,* 55**,** 226-248.

Cole, S. A., Paulson, A. L. & Shastry, G. K. 2012. Smart money: The effect of education on financial behavior. *Harvard Business School Finance Working Paper*.

Cull, R., Gan, L., Gao, N. & Xu, L. C. 2015. *Dual credit markets and household access to finance: evidence from a representative Chinese household survey*, The World Bank.

Cull, R., Gan, L., Gao, N. & Xu, L. C. 2016. *Social Capital, Finance, and Consumption: Evidence from a Representative Sample of Chinese Households*, The World Bank.

Czaja, S. J., Hammond, K., Blascovich, J. J. & Swede, H. 1989. Age related differences in learning to use a text-editing system. *Behaviour & Information Technology,* 8**,** 309-319.

Dai, W., Arnulf, J. K., Iao, L., Wan, P. & Dai, H. 2019. Like or want? Gender differences in attitudes toward online shopping in China. *Psychology & Marketing*, 36, 354-362.

Dano, A. M. 2005. Road Injuries and Long-Run Effects on Income and Employment. *Health Economics,* 14**,** 955-970.

Decker, S. & Schmitz, H. 2016. Health Shocks and Risk Aversion. *Journal of Health Economics,* 50**,** 156-170.

Demirguc-Kunt, A., Klapper, L., Singer, D. & Van Oudheusden, P. 2015. *The Global Findex Database 2014: Measuring Financial Inclusion around the World.* The World Bank.

Demirguc-Kunt, A., Klapper, L., Singer, D., Ansar, S. & Hess, J. 2018. *The Global Findex Database 2017: Measuring Financial Inclusion and the Fintech Revolution*, The World Bank.

Deuflhard, F., Georgarakos, D. & Inderst, R. 2018. Financial literacy and savings account returns. *Journal of the European Economic Association,* 17**,** 131-164.

Dinkova, M., Kalwij, A. & Alessie, R. 2018. Know more, spend more? The impact of financial literacy on household consumption. *Working Paper*.

Disney, R. & Gathergood, J. 2013. Financial literacy and consumer credit portfolios. *Journal of Banking & Finance,* 37**,** 2246-2254.

Dittmar, H., Long, K. & Meek, R. 2004. Buying on the Internet: Gender differences in on-line and conventional buying motivations. *Sex Roles,* 50**,** 423-444.

Dohmen, T. J., Falk, A., Huffman, D., Sunde, U., Schupp, J. & Wagner, G. G. 2005. Individual Risk Attitudes: New Evidence from a Large, Representative, Experimentally-Validated Survey.

Doolin, B., Dillon, S., Thompson, F. & Corner, J. L. 2005. Perceived risk, the Internet shopping experience and online purchasing behavior: A New Zealand perspective. *Journal of Global Information Management,* 13**,** 66-88.

Dowd, J. B. & Todd, M. 2011. Does Self-Reported Health Bias the Measurement of Health Inequalities in U.S. Adults? Evidence Using Anchoring Vignettes from the Health and Retirement Study. *The Journals of Gerontology: Series B,* 66B**,** 478-489.

Drakos, K. & Konstantinou, P. T. 2013. Investment Decisions in Manufacturing: Assessing the Effects of Real Oil Prices and Their Uncertainty. *Journal of Applied Econometrics,* 28**,** 151-165.

Edwards, R. D. 2008. Health Risk and Portfolio Choice. *Journal of Business & Economic Statistics,* 26**,** 472-485.

Elmendorf, D. W. & Kimball, M. S. 2000. Taxation of Labor Income and the Demand for Risky Assets. *International Economic Review,* 41**,** 801-832.

eMarketer. 2018. *Worldwide Retail and Ecommerce Sales: eMarketer's Estimates for 2016–2021* [Online]. Available: https://www.emarketer.com/Report/Worldwide-Retail-Ecommerce-Sales-eMarketers-Estimates-20162021/2002090 [Accessed 10th May 2019].

eMarketer. 2019. *What's Driving the Top Five Retail Ecommerce Markets Worldwide?* [Online]. Available: https://www.emarketer.com/content/whats-driving-the-top-five-retail-ecommerce-markets-worldwide [Accessed 10th, May 2019].

Fan, E. & Zhao, R. 2009. Health Status and Portfolio Choice: Causality or Heterogeneity? *Journal of Banking & Finance,* 33**,** 1079-1088.

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

Feng, X., Lu, B., Song, X. & Ma, S. 2019. Financial literacy and household finances: A Bayesian two-part latent variable modeling approach. *Journal of Empirical Finance,* 51**,** 119-137.

Franzen, A. & Hangartner, D. 2006. Social networks and labour market outcomes: The non-monetary benefits of social capital. *European Sociological Review,* 22**,** 353-368.

French, D. & McKillop, D. 2016. Financial literacy and over-indebtedness in low-income households. *International Review of Financial Analysis,* 48**,** 1-11.

García-Gómez, P. 2011. Institutions, Health Shocks and Labour Market Outcomes across Europe. *Journal of Health Economics,* 30**,** 200-213.

García-Gómez, P., Van Kippersluis, H., O'Donnell, O. & Van Doorslaer, E. 2013. Long-Term and Spillover Effects of Health Shocks on Employment and Income. *Journal of Human Resources,* 48**,** 873-909.

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

Gerardi, K., Goette, L. & Meier, S. 2010. *Financial literacy and subprime mortgage delinquency: Evidence from a survey matched to administrative data*, DIANE Publishing.

Gertler, P. & Gruber, J. 2002. Insuring Consumption against Illness. *American Economic Review,* 92**,** 51-70.

Goldman, D. & Maestas, N. 2013. Medical Expenditure Risk and Household Portfolio Choice. *Journal of Applied Econometrics,* 28**,** 527-550.

Gong, P., Liang, S., Carlton, E. J., Jiang, Q., Wu, J., Wang, L. & Remais, J. V. 2012. Urbanisation and health in China. *The Lancet,* 379**,** 843-852.

Grable, J. & Lytton, R. H. 1999. Financial risk tolerance revisited: the development of a risk assessment instrument☆. *Financial Services Review,* 8**,** 163-181.

Grewal, D., Iyer, G. R. & Levy, M. 2004. Internet retailing: enablers, limiters and market consequences. *Journal of Business Research,* 57**,** 703-713.

Grohmann, A. 2018. Financial literacy and financial behavior: Evidence from the emerging Asian middle class. *Pacific-Basin Finance Journal,* 48**,** 129-143.

Grohmann, A., Klühs, T. & Menkhoff, L. 2018. Does financial literacy improve financial inclusion? Cross country evidence. *World Development*, 111, 84-96.

Gunasekara, F. I., Carter, K. & Blakely, T. 2012. Comparing Self-Rated Health and Self-Assessed Change in Health in a Longitudinal Survey: Which Is More Valid? *Social Science & Medicine,* 74**,** 1117-1124.

Gustman, A. L., Steinmeier, T. L. & Tabatabai, N. 2012. Financial knowledge and financial literacy at the household level. *American Economic Review,* 102**,** 309-13.

Halko, M.-L., Kaustia, M. & Alanko, E. 2012. The Gender Effect in Risky Asset Holdings. *Journal of Economic Behavior & Organization,* 83**,** 66-81.

Hanson, T. A. & Olson, P. M. 2018. Financial literacy and family communication patterns. *Journal of Behavioral and Experimental Finance*.

Hasan, B. 2010. Exploring gender differences in online shopping attitude. *Computers in Human Behavior,* 26**,** 597-601.

Hastings, J. S., Madrian, B. C. & Skimmyhorn, W. L. 2013. Financial literacy, financial education, and economic outcomes. *Annual Review of Economics,* 5**,** 347-373.

Häubl, G. & Trifts, V. 2000. Consumer decision making in online shopping environments: The effects of interactive decision aids. *Marketing Science,* 19**,** 4-21.

Heaton, J. & Lucas, D. 2000. Portfolio Choice in the Presence of Background Risk. *The Economic Journal,* 110**,** 1-26.

Heckman, J. J. 1982. Heterogeneity and State Dependence. *National Bureau of Economic Research Reprint Series,* No. 298.

Hernández, B., Jiménez, J. & José Martín, M. 2011. Age, gender and income: do they really moderate online shopping behaviour? *Online Information Review,* 35**,** 113-133.

Hernandez-Quevedo, C., Jones, A. M. & Rice, N. 2008. Persistence in Health Limitations: A European Comparative Analysis. *Journal of Health Economics,* 27**,** 1472-1488.

Hilgert, M. A., Hogarth, J. M. & Beverly, S. G. 2003. Household financial management: The connection between knowledge and behavior. *Federal Reserve Bulletin 89 Fed. Res. Bull. (2003)* 89**,** 309.

Hsiao, Y.-J. & Tsai, W.-C. 2018. Financial literacy and participation in the derivatives markets. *Journal of Banking & Finance,* 88**,** 15-29.

Huang, W., Lei, X., Ridder, G., Strauss, J. & Zhao, Y. 2013. Health, Height, Height Shrinkage, and Ses at Older Ages: Evidence from China. *American Economic Journal: Applied Economics,* 5**,** 86-121.

Hui, T. K. & Wan, D. 2007. Factors affecting Internet shopping behaviour in Singapore: gender and educational issues. *International Journal of Consumer Studies,* 31**,** 310-316.

Hung, A., Parker, A. M. & Yoong, J. 2009. Defining and measuring financial literacy. *RAND Labor and Population Working Paper Series,* WR-708.

Huston, S. J. 2010. Measuring financial literacy. *Journal of Consumer Affairs,* 44**,** 296-316.

Hyslop, D. R. & Imbens, G. W. 2001. Bias from classical and other forms of measurement error. *Journal of Business & Economic Statistics,* 19**,** 475-481.

Institute of Social Science Survey, Peking University, 2015, "*China Family Panel Studies (CFPS)*", https://doi.org/10.18170/DVN/45LCSO, Peking University Open Research Data Platform, V37

Islam, A. & Maitra, P. 2012. Health Shocks and Consumption Smoothing in Rural Households: Does Microcredit Have a Role to Play? *Journal of Development Economics,* 97**,** 232-243.

Ivković, Z., Sialm, C. & Weisbenner, S. 2008. Portfolio Concentration and the Performance of Individual Investors. *Journal of Financial and Quantitative Analysis,* 43**,** 613-655.

Jappelli, T. & Padula, M. 2017. Consumption growth, the interest rate, and financial sophistication. *Journal of Pension Economics and Finance,* 16**,** 348-370.

Jones, A. M., Rice, N. & Zantomio, F. 2016. Acute Health Shocks and Labour Market Outcomes. *University Ca'Foscari of Venice, Dept. of Economics Research Paper Series,* (9).

Kaiser, H. F. 1958. The varimax criterion for analytic rotation in factor analysis. *Psychometrika,* 23**,** 187-200.

Karlan, D. & Zinman, J. 2008. Lying about borrowing. *Journal of the European Economic Association,* 6**,** 510-521.

Karlan, D., Ratan, A. L. & Zinman, J. 2014. Savings by and for the Poor: A Research Review and Agenda. *Review of Income and Wealth,* 60**,** 36-78.

Keese, M. 2009. Triggers and determinants of severe household indebtedness in Germany. *Ruhr Economic Paper,* 150.

Kempson, E. 2002. *Over-indebtedness in Britain*, Department of Trade and Industry London.

Keung Wong, D. F., Li, C. Y. & Song, H. X. 2007. Rural migrant workers in urban China: living a marginalised life. *International Journal of Social Welfare,* 16**,** 32-40.

Klapper, L., Lusardi, A. & Van Oudheusden, P. 2015. Financial literacy around the world. Standard & Poor's Ratings Services Global Financial Literacy Survey., Access mode: http://media. mhfi. com/documents/2015-Finlit_paper_17_F3_SINGLES. pdf.

Ko, H., Jung, J., Kim, J. & Shim, S. W. 2004. Cross-cultural differences in perceived risk of online shopping. *Journal of Interactive Advertising,* 4**,** 20-29.

Königsheim, C., Lukas, M. & Nöth, M. 2017. Financial knowledge, risk preferences, and the demand for digital financial services. *Schmalenbach Business Review*, 18, 343-375.

Korkmaz, A. G., Yin, Z., Yue, P. & Zhou, H. 2019. Does Financial Knowledge Alleviate Risk Attitude and Risk Behavior Inconsistency? *ASSA Annual Meeting.* Atlanta, GA, USA.

Kuoppamäki, S.-M., Taipale, S. & Wilska, T.-A. 2017. The use of mobile technology for online shopping and entertainment among older adults in Finland. *Telematics and Informatics,* 34**,** 110-117.

Lai, S., Shen, C., Xu, Y., Yang, X., Si, Y., Gao, J., Zhou, Z. & Chen, G. 2018. The distribution of benefits under China's new rural cooperative medical system: evidence from western rural China. *International Journal for Equity in Health*, 17, 137.

Lei, X. & Lin, W. 2009. The New Cooperative Medical Scheme in Rural China: Does More Coverage Mean More Service and Better Health? *Health Economics,* 18**,** S25-S46.

Li, N. & Zhang, P. 2002. Consumer online shopping attitudes and behavior: An assessment of research. *AMCIS 2002 Proceedings***,** 74.

Li, H., Rozelle, S. & Zhang, L. 2004. Micro‐credit programs and off‐farm migration in China. *Pacific Economic Review*, 9, 209-223.

Lian, J.-W. & Yen, D. C. 2014. Online shopping drivers and barriers for older adults: Age and gender differences. *Computers in Human Behavior,* 37**,** 133-143.

Liang, L., Langenbrunner, J. C. & World, B. 2013. Long March to Universal Coverage : Lessons from China.

Liao, Z. & Cheung, M. T. 2001. Internet-based e-shopping and consumer attitudes: an empirical study. *Information & Management,* 38**,** 299-306.

Limayem, M., Khalifa, M. & Frini, A. 2000. What makes consumers buy from Internet? A longitudinal study of online shopping. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans,* 30**,** 421-432.

Lin, N. 2017. Building a network theory of social capital. *Social Capital.* Routledge.

Lindeboom, M., Llena-Nozal, A. & van der Klaauw, B. 2016. Health Shocks, Disability and Work. *Labour Economics,* 43**,** 186-200.

Lindelow, M. & Wagstaff, A. 2005. *Health Shocks in China: Are the Poor and Uninsured Less Protected?*, The World Bank.

Love, D. A. & Smith, P. A. 2010. Does Health Affect Portfolio Choice? *Health Economics,* 19**,** 1441-1460.

Lusardi, A. & Mitchell, O. 2007 a. Financial literacy and retirement planning: New evidence from the Rand American Life Panel. Michigan Retirement Research Center Research Paper No. WP157.

Lusardi, A. & Mitchell, O. S. 2007b. Baby boomer retirement security: The roles of planning, financial literacy, and housing wealth. *Journal of Monetary Economics*, 54, 205-224.

Lusardi, A. & Mitchell, O. S. 2011a. Financial literacy and planning: Implications for retirement wellbeing. *National Bureau of Economic Research,* No. W17078.

Lusardi, A. & Mitchell, O. S. 2011b. Financial literacy around the world: an overview. *Journal of Pension Economics & Finance,* 10**,** 497-508.

Lusardi, A. & Mitchell, O. S. 2014. The economic importance of financial literacy: Theory and evidence. *Journal of Economic Literature,* 52**,** 5-44.

Lusardi, A. & Scheresberg, C. d. B. 2013. Financial literacy and high-cost borrowing in the United States. *National Bureau of Economic Research*. No. W18969

Lusardi, A. & Tufano, P. 2015. Debt literacy, financial experiences, and overindebtedness. *Journal of Pension Economics & Finance,* 14**,** 332-368.

Lusardi, A. 2008. Financial literacy: an essential tool for informed consumer choice? *National Bureau of Economic Research,* No. W14084.

Lusardi, A. 2009. Overcoming the saving slump: How to increase the effectiveness of financial education and saving programs, University of Chicago Press.

Lusardi, A. 2011. Americans' financial capability. *National Bureau of Economic Research,* No. W17103.

Lusardi, A., Michaud, P.-C. & Mitchell, O. S. 2017. Optimal financial knowledge and wealth inequality. *Journal of Political Economy,* 125**,** 431-477.

Lusardi, A., Mitchell, O. S. & Curto, V. 2010. Financial literacy among the young. *Journal of Consumer Affairs,* 44**,** 358-380.

Lusardi, A., Mitchell, O. S. & Curto, V. 2014. Financial literacy and financial sophistication in the older population. *Journal of Pension Economics & Finance,* 13**,** 347-366.

Lusardi, A., Mitchell, O. S. & Oggero, N. 2019. Debt and financial vulnerability on the verge of retirement. *Journal of Money, Credit and Banking (Forthcoming)*.

Lusardi, A., Schneider, D. J. & Tufano, P. 2011. Financially fragile households: Evidence and implications. *National Bureau of Economic Research,* No. W17072.

Lyons, A. C., Grable, J. E., & Zeng, T. (2017). Infrastructure, urbanization, and the financial inclusion of Chinese households. *ADBI Working Paper Series*, No. 767. Tokyo, Japan: Asian Development Bank Institute.

Lyons, A. C., Grable, J. E. & Zeng, T. 2019. Impacts of Financial Literacy on the Loan Decisions of Financially Excluded Households in the People's Republic of China. ADBI Working Paper *Series,* No. 923.

Lyons, A. C., Palmer, L., Jayaratne, K. S. & Scherpf, E. 2006. Are we making the grade? A national overview of financial education and program evaluation. *Journal of Consumer Affairs,* 40**,** 208-235.

Mahmood, M. A., Bagchi, K. & Ford, T. C. 2004. On-line shopping behavior: Cross-country empirical research. *International Journal of Electronic Commerce,* 9**,** 9-30.

McKinsey, G. i. a. 2018. *The dynamics of debt — a global exploration* [Online]. Available: https://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/visualizing-global-debt [Accessed 1st, Sep 2019].

Meier, S. & Sprenger, C. 2010. Present-biased preferences and credit card borrowing. *American Economic Journal: Applied Economics,* 2**,** 193-210.

Mian, A., Sufi, A. & Verner, E. 2017. Household debt and business cycles worldwide. *The Quarterly Journal of Economics,* 132**,** 1755-1817.

Mitra, S., Palmer, M., Mont, D. & Groce, N. 2016. Can Households Cope with Health Shocks in Vietnam? *Health Economics,* 25**,** 888-907.

Modigliani, F. & Brumberg, R. 1954. Utility analysis and the consumption function: An interpretation of cross-section data. *Post-Keynesian Economics,* 1**,** 338-436.

Mohanan, M. 2013. Causal Effects of Health Shocks on Consumption and Debt: Quasi-Experimental Evidence from Bus Accident Injuries. *Review of Economics and Statistics,* 95**,** 673-681.

Moore, D. L. 2003. *Survey of financial literacy in Washington State: Knowledge, behavior, attitudes, and experiences*, Washington State Department of Financial Institutions.

Moretti, E. 2011. Social learning and peer effects in consumption: Evidence from movie sales. *The Review of Economic Studies,* 78**,** 356-393.

Nahapiet, J. & Ghoshal, S. 1998. Social capital, intellectual capital, and the organizational advantage. *Academy of Management Review,* 23**,** 242-266.

NBSC. 2019. Available: http://data.stats.gov.cn/easyquery.htm?cn=A01&zb=A0706&sj=201903 [Accessed].

OECD 2016. OECD/INFE International Survey of Adult Financial Literacy Competencies. *OECD, Paris*.

OECD. 2017. *G20/OECD INFE report on adult financial literacy in G20 countries* [Online]. Available: file:///C:/Users/lxt488/Downloads/2017%20OECD_INFE%20Report%20on%20Adult%20Financial%20Literacy%20in%20G20%20Countries%20(1).pdf [Accessed 29 Jan 2019].

PBC. 2009. *China Rural Finance Service Report (2009) (in Chinese)* [Online]. Available: http://www.pbc.gov.cn/zhengcehuobisi/125207/125227/125960/126049/126080/2881691/index.html [Accessed 12, May 2019].

PBC. 2014. *Sources & Uses of Credit Funds of Financial Institutions (by Sectors) (in RMB and Foreign Currency)* [Online]. Available: http://www.pbc.gov.cn/eportal/fileDir/defaultCurSite/resource/cms/2015/07/2013s01a.htm [Accessed 25th, June 2019].

PBC. 2015. *China Rural Finance Service Report (2014) (in Chinese)* [Online]. Available: http://www.cff.org.cn/zgjrlt/cffyj98/91983/91995/98543/20151103093651837 56.pdf [Accessed 18, June 2018].

PBC. 2015. *Sources & Uses of Credit Funds of Financial Institutions (by Sectors) (in RMB and Foreign Currency)* [Online]. Available: http://www.pbc.gov.cn/eportal/fileDir/defaultCurSite/resource/cms/2015/07/2014s01a.htm [Accessed 25th, June 2019].

PBC. 2018. *China Rural Finance Service Report (2017) (in Chinese)* [Online]. Available: http://www.pbc.gov.cn/zhengcehuobisi/125207/125227/125960/126049/3564964/3564228/2018062221381937583.pdf [Accessed 18, May 2019].

PBC. N.D. *Statistics and analysis* [Online]. Available: http://www.pbc.gov.cn/diaochatongjisi/116219/116319/index.html [Accessed 20th July 2019].

Perea y Monsuwé, T., Dellaert, B. G. & De Ruyter, K. 2004. What drives consumers to shop online? A literature review. *International Journal of Service Industry Management,* 15**,** 102-121.

Pérez-Hernández, J. & Sánchez-Mangas, R. 2011. To have or not to have Internet at home: Implications for online shopping. *Information Economics and Policy,* 23**,** 213-226.

Prina, S. 2015. Banking the Poor Via Savings Accounts: Evidence from a Field Experiment. *Journal of Development Economics,* 115**,** 16-31.

Rabe-Hesketh, S. & Skrondal, A. 2013. Avoiding Biased Versions of Wooldridge's Simple Solution to the Initial Conditions Problem. *Economics Letters,* 120**,** 346-349.

Rantanen, T., Volpato, S., Luigi Ferrucci, M., Eino Heikkinen, M., Fried, L. P. & Guralnik, J. M. 2003. Handgrip Strength and Cause-Specific and Total Mortality in Older Disabled Women: Exploring the Mechanism. *Journal of the American Geriatrics Society,* 51**,** 636-641.

Räsänen, P. & Koiranen, I. Changing patterns of ICT use in Finland–The senior citizens' perspective. International Conference on Human Aspects of IT for the Aged Population, 2016. Springer, 226-237.

Remund, D. L. 2010. Financial literacy explicated: The case for a clearer definition in an increasingly complex economy. *Journal of Consumer Affairs,* 44**,** 276-295.

Rodgers, S. & Harris, M. A. 2003. Gender and e-commerce: An exploratory study. *Journal of Advertising Research,* 43**,** 322-329.

Roodman, D. 2011. Fitting fully observed recursive mixed-process models with cmp. *Stata Journal,* 11**,** 159-206.

Rosen, H. S. & Wu, S. 2004. Portfolio Choice and Health Status. *Journal of Financial Economics,* 72**,** 457-484.

Scheerder, A., van Deursen, A. & van Dijk, J. 2017. Determinants of Internet skills, uses and outcomes. A systematic review of the second-and third-level digital divide. *Telematics and Informatics,* 34**,** 1607-1624.

Schultz, T. P. & Tansel, A. 1997. Wage and Labor Supply Effects of Illness in Côte D'ivoire and Ghana: Instrumental Variable Estimates for Days Disabled. *Journal of Development Economics,* 53**,** 251-286.

Sekita, S. 2011. Financial literacy and retirement planning in Japan. *Journal of Pension Economics & Finance,* 10**,** 637-656.

Servon, L. J. & Kaestner, R. 2008. Consumer financial literacy and the impact of online banking on the financial behavior of lower‐income bank customers. *Journal of Consumer Affairs,* 42**,** 271-305.

Smith, D., Menon, S. & Sivakumar, K. 2005. Online peer and editorial recommendations, trust, and choice in virtual markets. *Journal of Interactive Marketing,* 19**,** 15-37.

Smith, J. P. 1999. Healthy Bodies and Thick Wallets: The Dual Relation between Health and Economic Status. *Journal of Economic Perspectives,* 13**,** 145-166.

Smith, J. P. 2004. Unraveling the Ses: Health Connection. *Population and Development Review,* 30**,** 108-132.

Smith, J. P. 2005. Consequences and Predictors of New Health Events. *Analyses in the Economics of Aging.* University of Chicago Press.

Sohn, S. 2017. A contextual perspective on consumers' perceived usefulness: The case of mobile online shopping. *Journal of Retailing and Consumer Services,* 38**,** 22-33.

Stango, V. & Zinman, J. 2009a. Exponential growth bias and household finance. *The Journal of Finance,* 64**,** 2807-2849.

Stango, V. & Zinman, J. 2009b. What do consumers really pay on their checking and credit card accounts? Explicit, implicit, and avoidable costs. *American Economic Review,* 99**,** 424-29.

Statista. 2015. *Gross merchandise volume of China's e-commerce market from 2010 to 2019 (in trillion yuan)* [Online]. Available: https://www.statista.com/statistics/278552/gross-merchandise-volume-of-the-e-commerce-industry-in-china/ [Accessed 22 Feb 2019].

Strauss, J. & Thomas, D. 1998. Health, Nutrition, and Economic Development. *Journal of Economic Literature,* 36**,** 766-817.

Sui, Y. & Niu, G. 2018. The Urban–Rural Gap of Chinese Household Finance. *Emerging Markets Finance and Trade*, 54, 377-392.

Susskind, A. M. 2004. Electronic commerce and world wide web apprehensiveness: An examination of consumers' perceptions of the world wide web. *Journal of Computer-Mediated Communication,* 9**,** JCMC931.

Tay, L., Batz, C., Parrigon, S. & Kuykendall, L. 2017. Debt and subjective well-being: The other side of the income-happiness coin. *Journal of Happiness Studies,* 18**,** 903-937.

Thaichon, P. 2017. Consumer socialization process: The role of age in children's online shopping behavior. *Journal of Retailing and Consumer Services,* 34**,** 38-47.

The World Bank. 2017. *The Global Findex Database (2017)* [Online]. Available: https://globalfindex.worldbank.org/#GF-ReportChapters [Accessed 18, Jun 2018].

Tiftik, E. & Mahmood, K. 2019. Global Debt Monitor: Devil in the details *Institute of International Finance*.

Tobin, J. 1958. Liquidity preference as behavior towards risk. *The Review of Economic Studies*, 25, 65-86.

Trocchia, P. J. & Janda, S. 2000. A phenomenological investigation of Internet usage among older individuals. *Journal of Consumer Marketing,* 17**,** 605-616.

van Rooij, M. C., Lusardi, A. & Alessie, R. J. 2012. Financial literacy, retirement planning and household wealth. *Economic Journal,* 122**,** 449-478.

van Rooij, M., Lusardi, A. & Alessie, R. 2011. Financial literacy and stock market participation. *Journal of Financial Economics*, 101, 449-472.

Wagstaff, A. 2007. The Economic Consequences of Health Shocks: Evidence from Vietnam. *Journal of Health Economics,* 26**,** 82-100.

Wang, R. J.-H., Malthouse, E. C. & Krishnamurthi, L. 2015. On the go: How mobile shopping affects customer purchase behavior. *Journal of Retailing*, 91, 217-234.

Wang, Z., Li, X., Chen, M. & Si, L. 2018. Social Health Insurance, Healthcare Utilization, and Costs in Middle-Aged and Elderly Community-Dwelling Adults in China. *International Journal for Equity in Health,* 17**,** 17.

Ware Jr, J. E. & Gandek, B. 1998. Overview of the Sf-36 Health Survey and the International Quality of Life Assessment (Iqola) Project. *Journal of Clinical Epidemiology,* 51**,** 903-912.

Wooldridge, J. M. 2005. Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity. *Journal of Applied Econometrics,* 20**,** 39-54.

World Bank. 2017a. *Population Ages 65 and above, Total* [Online]. Available: https://data.worldbank.org/indicator/SP.POP.65UP.TO?locations=CN [Accessed 02/08 2019].

World Bank. 2017b. Population Ages 65 and above (% of Total Population).

Worldmeters. n.d. *China Population (Live)* [Online]. Available: https://www.worldometers.info/world-population/china-population/ [Accessed 14, May 2019].

Wu, W., Rong, P. & Xu, Q. 2011. Health Status and Household Portfolio Choice. *Economic Research Journal***,** S1.

Xiao, J. J., Chen, C. & Chen, F. 2014. Consumer financial capability and financial satisfaction. *Social Indicators Research,* 118**,** 415-432.

Xie, E. & Sun, G. 2012. The Effect of Health Shocks on Portfolio Choice of Elderly Households [J]. *Journal of Population and Development,* 4**,** 47-55.

Yaohui Zhao; John Strauss; Gonghuan Yang, 2015, "*China Health and Retirement Longitudinal Study（2011 baseline）*", https://doi.org/10.18170/DVN/IZU8OF, Peking University Open Research Data Platform, V1

Yaohui Zhao; John Strauss; Gonghuan Yang, 2015, "*China Health and Retirement Longitudinal Study（2013 wave2）*", https://doi.org/10.18170/DVN/AQIU6C, Peking University Open Research Data Platform, V1

Yaohui Zhao; John Strauss; Gonghuan Yang, 2017, "*China Health and Retirement Longitudinal Study（2015 wave3）*", http://charls.pku.edu.cn/, Peking University Open Research Data Platform, V1

Yu, H. 2015. Universal Health Insurance Coverage for 1.3 Billion People: What Accounts for China's Success? *Health Policy,* 119**,** 1145-1152.

Zhang, J., Gan, L., Xu, L. C. & Yao, Y. 2014. Health Shocks, Village Elections, and Household Income: Evidence from Rural China. *China Economic Review,* 30**,** 155-168.

Zhang, M. L. & Chen, M. S. 2019. *China's Digital Economy: Opportunities and Risks*, International Monetary Fund.

Zhao, Y. 1999. Leaving the countryside: rural-to-urban migration decisions in China. *American Economic Review,* 89**,** 281-286.

Zhong, H. 2011. Effect of Patient Reimbursement Method on Health‐Care Utilization: Evidence from China. *Health Economics,* 20**,** 1312-1329.

Zhou, L., Dai, L. & Zhang, D. 2007. Online shopping acceptance model-A critical survey of consumer factors in online shopping. *Journal of Electronic Commerce Research,* 8.

Zinman, J. 2015. Household debt: Facts, puzzles, theories, and policies. *Economics,* 7**,** 251-276.