The lion or the ostrich: how do facing credit constraints affect Chinese household stock assets?

by

Che Chen

Submitted to

Department of Economics and Business

Central European University

In partial fulfilment of the requirements for the degree of

Master of Arts in Economics

Supervisor: Professor András Danis

Vienna, Austria

2022

Abstract

This paper examines the impact of credit constraints on stock holdings per household with the introduction of financial micro-level data on Chinese households. The results of the empirical study show that those households which are facing the credit constraints dilemma have a significantly lower proportion of stock ownership compared to other households, while the main channels between credit constraints and household stock holdings are examined and it is found that credit constraints reduce household stock assets by reducing household total wealth. Further, the paper also classifies credit constraints into two different types of constraints, supply-oriented and demand-oriented, and the regression results show that demand-oriented credit constraints have a greater impact on household stock holdings than supply-oriented credit constraints. In the conclusion section of this paper, recommendations are made to mitigate the sharp rise of income inequality in China against the above findings.

Acknowledgements

As another summer has approached, it is different from the same season last year, since I will be farewell to CEU soon. The acknowledgements section is the final part I wrote in my thesis, and by the time I reach this point, I am already filled with emotions, as if the joy I felt when I received the acceptance to CEU two years ago is still vivid in my mind.

It rained heavily in Vienna yesterday evening, while I was lying in bed listening to the sound of the rain and reflecting on my journey of studying in this city. I would like to express my gratitude to CEU for not only providing me with financial support for my studies, but also for offering such a flawless platform to learn solid knowledge, connect with professors in close relationship, and make friends from all over the world. During my second year at CEU, I took a PhD course related with finance from my thesis advisor Dr. András Danis. It shaped the direction I wanted to explore in the future and was an opportunity for me to choose the field of finance as my thesis topic. Within the months of writing this paper, I thank my advisor for his patience and guidance in refining the errors and omissions in my work, and this thesis is the fruit of our discussions through multiple meetings.

During these two years at CEU, my interest in research has been thoroughly sparked, and I would like to thank Dr. Florian Exler, Dr. Geoffrey Castillo, Dr. Sergey Lychagin, and Dr. Zhenghua Liu for their great support in helping me apply to the target institutions and making my dream of pursuing further studies after graduation come true. All along, I have been a

lucky enough person to receive the full backing from my parents, relatives and many teachers and friends on my way of education. From Qingdao to Singapore, then from Singapore to Vienna, and after this summer vacation, I will leave for the next city from Vienna. With the convenience of modern technology, I, a boy from a small county in Fujian, have already completed the "travel ten thousand miles" as the ancient Chinese used to say. I would like to end with a Latin verse that caught my fancy, thanking myself for not giving up even when I encountered difficulties along the way.

Macte nova virtute, puer, sic itur ad astra.

-- *Aeneid*

By Virgil

Table of contents

1. Introduction

Have you ever fantasized about striking gold overnight by speculating in stocks like Jordan Belfort? Many of you who have seen the movie "The Wolf of Wall Street" have probably had a similar dream, which was created by the magic of stocks that are high risk yet accompanied by high returns. Since its inception in Amsterdam in the 17th century, the stock market has been evolving through lengthy development (Stringham and Curott, 2015) and now become an important part of the financial market. Households are the most dominant participants in Chinese stock market, with more than 80% of the total (Zhang et al., 2021). As the field of household finance has received considerable scholarly attention, a growing number of papers have explored the factors that affect household financial market participation rate (Zou and Deng, 2019) and attempted to explain the phenomenon that the vast majority of families do not invest in high-risk financial assets, a violation of standard portfolio prediction results (Guiso and Sodini, 2013).

The importance of stocks as a representative of risky financial assets cannot be overstated. What are the main factors that affect the decision of the households on whether to participate in the stock market? How much of the assets a household allocates to stocks is also influenced and limited by what? As the share of financial assets in total household assets has increased, the topic of asset allocation has become a hotspot for research in finance (Shum and Faig, 2006). The data employed for this thesis comes from a representative micro-level financial survey for Chinese families, which was conducted in 2017 and is the so-called China

Household Finance Survey (Gan et al., 2016). At present, China has moved into one of the most inequality countries in the world due to the dramatic acceleration of income inequality despite the gradual eradication of poverty after rapid economic growth (Jain-Chandra et al., 2018). How this inequality can be ameliorated was my initial goal for writing this paper, and the reason why I choose financial micro-level data of Chinese households as the subject of my study. The high-risk and high-return characteristics of stocks is well-known, and a reasonable allocation of household stock assets can help boost household wealth and thus reduce the gap between rich and poor Chinese households. Figures 1a and 1b[1] show the overall market size and the average market value of investor account assets in the Chinese stock market over recent years. The large size of the stock market and the high average stock values of investors are symbols of the booming Chinese stock market, and given that households are the most important players in this market as mentioned in the previous paragraph, the topic of investigating the main factors influencing household stock asset holdings is certainly of great importance.

The historical studies on influencing drivers of household stock market participation or stock investment shares are mainly focused on individual characteristics (Campbell, 2006), family characteristics (Betermier et al., 2017), financial literacy (Arrondel et al., 2015), housing (Chetty et al., 2017), risk preferences (Ameriks et al., 2020), and other factors. Additionally, credit constraints may also be one of the crucial factors. By comparing portfolio behavior across countries, Guiso et al. (2000) found that as affected by credit constraints, older

[1] All figures and tables mentioned in this paper are available in the appendix section.

households prefer to invest in financial assets, while younger households prefer the real assets. Campbell and cocco (2003) also found that credit constraints can affect the choice of portfolio made by investors. Since then, Kozak and Sosyura (2015) used micro-level data to detect that acquisition of credit makes a favorable effect on increasing the stock market involvement with respect to households. Summarizing the arguments of researchers who have studied credit constraints, the major causes of credit constraints are identified as high transaction costs (Hoff & Stiglitz, 1990) and the so-called "discouraged borrowers" (Jappelli, 1990). The high transaction costs include the interest on the loan, the procedure and the waiting time for the loan application, while the "discouraged borrowers" are the applicants who do not submit the application for fear of being rejected and the emergence for such type of borrowers is mainly due to information asymmetry.

Nowadays, credit constraints have become an integral part of the world described by economists (Banerjee and Duflo, 2014). The specific definition and measurement for credit constraints would be described in the following data section. The early theory that correlated with credit constraints was introduced by Flavin (1981), and after several years of development, the research systems for credit constraints and household financial asset choice are now established more systematically overseas, whereas research on the relationship between them is still lacking adequate analysis since household finance in China is still in the early stage of development (Wang, 2016). Also, the majority of the literature on the impact of credit constraints over household stocks has focused on how credit constraints affect household stock market participation, with more limited research investigating the effect of credit constraints on household stock assets ownership. In the title of this paper, I choose to use the terms lion and ostrich to refer to two classes of households with different attitudes toward stock. The lion refers to those "brave" households that own more risky assets such as stocks, while the ostrich refers to those households that tend to be conservative and hold only a small amount of stocks, or even do not hold a stock account. Therefore, to fill the gap in the literature of relevant field, this paper would adopt the empirical research methods by applying OLS and instrumental variables approach for the selected financial micro-level data of Chinese households to reveal the influence of credit constraints on the holdings of stocks assets of these different households and the underlying impact mechanisms. Moreover, I referred to the classification categories of Boucher et al. (2008), which classify credit constraints into demand-oriented credit constraints and supply- oriented credit constraints, and compare the impact magnitude of both credit constraints for the stock assets ownership of Chinese households by examining the empirical results. Meanwhile, this paper employs methods such as winsorizing to guarantee the robustness of the outcomes.

The remainder sections of my thesis is structured as follows: Section 2 is the data section, which introduces the data I choose, the specific definitions of the variables, and the descriptive statistics of the data; Hypothesis development is scheduled in the third section, in which I propose several hypotheses related to the research topic, and those would be verified one by one in the later sections through empirical methods; Section 4 is the section on model that I construct several OLS models as baseline models, but given the potential endogeneity issues, I also adopt an instrumental variable and build several 2SLS models to address the issues; Section 5 presents the analysis of the results, for which I regress each of the constructed models and systematically analyze the results; the last section draws conclusions, which mainly discusses and summarizes my findings obtain through the analysis of the results.

2. Data

2.1 Source of the data

The data used for the empirical study in this paper are from the 2017 China Household Finance Survey (CHFS). It is a sampling project conducted within mainland China by the Survey and Research Center for China Household Finance, and the initial survey was launched 11 years ago and has been followed up every two years afterwards. It is a three-stage stratified random sampling among county-level, community-level, and household-level, and aims to collect information related to the micro-level of household finance, including the demographic characteristics, household assets and liabilities, family consumption and income, insurance and social security, etc. This one-year period data that I introduced covers 29 provinces/autonomous regions/municipalities, more than 300 counties, and the sample contains data associated with approximately 40,000 households. Since my research question is focus on the household level rather than the individual level, my sample only includes information related to the head of the household and removes the data of other household members. In addition, although there is no age limit for the household head, I set the minimum age of the household head to 16 years old because the topic I want to discuss requires the respondent to be a person with fully civil capacity. The final dataset contains a total of 33,990 related profiles of household heads.

2.2 Definition of variables

My thesis examines the effect of credit constraints over stock holdings of household. However, there are many variables (including credit constraints) that are not directly available from the 2017 CHFS raw dataset, hence I need to process the data to construct the required dependent, explanatory, and control variables. The following is a detailed description of the variables that I have constructed.

2.2.1. Dependent variable

The dependent variable within the study is the log-transformation for household stock ownership. Respondents in the CHFS were asked about the number of stocks their families held at the time of the interview and the current market value of those stocks. Due to the large number of missing values in the sample, I defined those respondents whose families did not possess a stock account as having zero household stock assets to fill in the missing values, since the absence of a stock account implied that the household did not hold any stocks. In addition, considering the large data variation between the samples of household stock assets, I take the logarithm of the household stock assets as the dependent variable and the calculation formula is shown below[2]:

$$\text{log\_stock\_assets} = \log(\text{market values of household stock assets} + 1)$$

2.2.2. Explanatory variable

The core explanatory variable in my thesis is credit constraints. This is a dummy variable that takes two specific values of 0 and 1. I adopt the approach of Jappelli (1990) to measure the credit constraints by defining two categories of household as suffering from credit constraints dilemma: those who were rejected for a loan application and who fail to apply the loan out of

[2] The market value of stocks owned by households in the right-hand side of the equation is first added by 1 and then taken logarithmically mainly because there are many values of 0 in the variable.

phobia about rejection. In 2017 CHFS, respondents were asked about the reason why their family did not attempt to apply for the loan from a bank/credit union to obtain the needed funds in the following areas: industry and commerce, housing, health care, education, car purchase, marriage and funeral, etc. If the respondent selected "Applied but was not approved" or "Estimated that the loan application would not be approved", meaning that the family of the respondent was facing credit constraints, hence the credit constraints variable was assigned a value of 1. Other households that did not face credit constraints were assigned a value of 0.

2.2.3. Control variables

Apart from the key explanatory variable selected for the empirical research, there are many other factors that can affect the household stock assets holdings. With reference to the previous literature (mentioned in the introduction section), I choose two categories of variables, personal characteristics of the household head and family characteristics, as control variables. Among them, the variables at the level of personal characteristics for household heads include age of the household head, age squared, gender, education level, married or not, health status, unemployed information, household registration, happiness status, the level of concern they usually paid to economic and financial information, and their investment risk preference.

Same as the credit constraints variable, the dummy variables associated with gender, unemployment and household registration all take two specific values of 0 or 1 only. The health of the household head is measured by the rating of respondent for his or her health status in the survey, which ranges from 0 to 4, with higher values indicating that the respondent is (self-perceived to be) healthier. Regarding the attention that household head paid for economic and financial information, the survey included a relevant question that asked the respondent, and the higher the value of this variable, the higher the concern of respondent for the economic and financial information. In addition, the investment risk preference of respondent is measured by the question " Which investment project you will be most willing to choose if a lump sum of cash if given for investment?" The higher the value, the more risk-lover the respondent is, and the lower the value, the more risk-averse the respondent tends to be. 2017 CHFS only asked new respondents about their concerns for the economic and financial information, also the investment risk preferences only for the new respondents, so I call the past CHFS data to fill in the missing values of these two variables[3].

Meanwhile, at the household characteristics level, the control variables include the family size, housing status, the log-transformation for total household income and total household assets. Among these variables, the variable relevant to the housing status is a binary variable that signed the value of 1 if the survey participant owns one or more than one house, and 0 otherwise. The specific definitions of each variable and the ranges of values are shown in Table 1.

[3] I call the CHFS data for 2011, 2013, and 2015 to fill in the missing values for some of the variables.

2.3 Descriptive statistics

Table 2 reveals descriptive statistics for the sample of 33,990 household heads, reporting the mean, standard deviation, minimum and maximum values of the variables. As shown in Table 2, the percentage of female household heads in the sample is 20.18%, the unemployed household heads account for 14.29%, and there are slightly more rural than urban household heads. The mean for logarithm of household stock assets is only 0.60 RMB[4], which is a quite small number, implying that the average stock assets owned by households in the sample is small[5], and this is consistent with the average attention that household heads in the sample paid for economic and financial information and their average investment risk preference of 1.02 and 0.92, respectively. Most of the household heads did not choose to purchase risky financial assets such as stocks because they were less concerned about economic and financial information and were tend to be risk-averse. Considering the representative sample of CHFS, this suggests that most Chinese households possessed extremely limited stock assets at 2017. Also, the mean value of the credit constraints variable is about 0.07, indicating that 7.13% of the households in the sample lacked access to credit support from formal financial institutions such as banks. The average value of the variables in the descriptive statistics table also presents that the mean age among the household heads in the sample is over 50 years old and most of them were married, satisfied with their lives and considered themselves to be in better health than their peers. In addition, most of the household heads have a family size consisting of three or four individuals, their average education level is roughly between middle school

[4] As I mentioned in the notes to Table 2, all money-related variables are measured in RMB, such as stock assets variable. However, it is also important to note that these variables are not meaningful in units once they are logarithmized, since the purpose of logarithmization is mainly for comparing growth rates.

[5] The reason for this is that most households own zero stocks, which is why the mean value of the log stock assets variable is small. In Table 2 I also add the specific numbers and proportion of households with stock assets for reference.

level and high school level, and more than 90% of the households in the sample own a house or even more than one[6].

[6] In China, the government provides economical affordable housing for the household with low-income. Also, the excessive disparity in housing prices between different cities and different areas may also contribute to the large mean value of this variable, since people in areas with high housing prices are offered the option of buying a house at areas where they can afford. However, note that an extremely high ownership rate of housing in the sample does not mean that households no longer need to buy additional houses, there are still households that want to buy more of them.

3. Hypotheses

Previous studies have shown that exposure to credit constraints reduces the rate of household stock market participation, but research on the effect of credit constraints over household stock holdings is more limited. Although I expect those households that subject to credit constraints would have lower stock assets, given the cultural differences, households with credit constraints in China may respond differently to their stock holdings than households in other countries. To verify my guesses, I need to develop hypotheses first and test them on a one-by-one basis. In this section, I would propose hypotheses on the direction of the influence of credit constraints over household stock assets ownership, the main channels through which credit constraints affect household stock holdings, and the magnitude of the impact of the two different credit constraints on household stock assets ownership, respectively. The verification of the hypotheses would be performed in the subsequent sections.

Consider the following scenario. There are two households with similar headship and family characteristics, A and B. They each have \$700,000 in bank savings and \$300,000 in stock assets, and they both have the same goal of owning a family house. The heads of households A and B simultaneously applied to the bank for a loan, but only household A received the necessary loan that was requested, while loan application of household B was rejected by the bank. Thus, in the definition of credit constraints that I employ, household A is not constrained in credit, while household B is facing a credit constraints dilemma. At this point, with sufficient amount available, family A takes out all of its savings and loan for the purchase of a desirable house, while family B does not have adequate amount for the purchase of a house because its loan application was rejected, so family B decides to take out a portion of its savings for the purchase of more stocks after deliberation. Under the circumstances, whether or not the household faces credit constraints directly affected the plans of the different households. The household facing credit constraints has to choose the suboptimal option of buying stocks as a last resort since it could not obtain a bank loan to purchase the desired target asset. This leads me to the first hypothesis of this paper:

H1a: Credit constraints have a direct effect on the increase in household stock assets, which means that households facing credit constraints significantly increase their holdings of stock assets.

The case envisioned above may yield the opposite result. I still assume that there are two households with similar heads of household and family characteristics, A and B. Their goals, savings, and stock assets remain the same, and the result of applying for a loan from the bank also follows that only the application of household A is approved. In this case, family A takes out all of its savings and the loan it applied for to buy a house, while family B does not receive a loan from the bank, but after deliberation, family B decides to sell all of its stock assets to raise the amount needed to buy a house. In this case, the credit constraint has a direct effect on the reduction of the stock assets holdings by the household, so I propose a second hypothesis contrary to H1a:

H1b: Credit constraints have a direct effect on the decrease in household stock assets, which means that households facing credit constraints significantly reduce their holdings of stock assets.

For the completely opposing hypotheses H1a and H1b mentioned above, I prefer to construct OLS and 2SLS models to test the hypotheses by examining the significance and sign of the estimated coefficients of credit constraints after regression. Further, if I verify the existence of a positive or negative effect of credit constraints on household stock ownership, I could also classify credit constraints into supply-side credit constraints and demand-side credit constraints as I mentioned in the Introduction section. Those who are rejected for the loan application from formal financial institutions can be classified as households who are facing supply-oriented credit constraints, while those who are not confident of successfully qualifying for the loan application are classified as households subject to demand-oriented credit constraints. The former is caused by the decision of the formal financial institution not to provide credit support after considering factors such as the credit level and repayment ability of the household, while the latter is caused by other potential factors such as the lack of confidence of the household or cognitive bias in the loan application. The two-way choice between the household and the financial institution determines whether the demand for credit by the household is satisfied, and which credit constraints affect the stock assets of households more considerably. H2a to H2c are three possible hypotheses:

H2a: The effect of demand-side credit constraints on household stock assets are greater than that of supply-side credit constraints.

H2b: The effect of demand-side credit constraints on household stock assets are less than that of supply-side credit constraints.

H2c: The effect of demand-side credit constraints on household stock assets are equal to the effect caused by supply-side credit constraints.

In addition, besides the potential direct effect of credit constraints, credit constraints may also affect household stock assets through mediating variables. Households facing credit constraints held less income and assets in comparison to households without credit constraints (Jappelli and Cox, 1990), and credit constraints may reduce the tolerance of households for risky assets by reducing household total assets and thus their holdings of risky assets such as stocks. This is a reasonable pathway since stock holdings are part of total household assets, and thus variations in total household assets might cause shifts in stock assets. However, with respect to total household income, I suspect that it is not a mediating variable in the way that credit constraints affect household stock ownership, for the reason that credit constraints do not, in theory, affect total household income. Although these two variables are correlated, that is because household income is one of the criteria considered for repayment ability, so that household income affects the credit constraints, but the inverse relationship does not theoretically hold. Considering the above scenarios which I have assumed, I also develop the following hypotheses H3 and H4:

H3: Total household income is not one of the important channels through which credit constraints affect household holdings of stock assets.

H4: One of the important channels through which credit constraints affect household stock assets is by lowering household total assets and thus reducing household holdings of stock assets.

4. Models

4.1 Ordinary least squares (OLS) method

With the purpose of exploring the influence of credit constraints on household stock assets, for the first step, I construct three OLS models as baseline models with the following equation:

$$\log(\text{stock assets}_i) = a_0 + a_1 \text{credit constraints}_i + a_2 X_i + u_i \quad (1)$$

In equation (1), $\log(\text{stock assets}_i)$ is the explained variable, which is the logarithm of stock assets held by the i-th household. $\text{credit constraints}_i$ is a dummy variable and also the core explanatory variable of this paper, which refers to whether the i-th household faces credit constraints. X_i are control variables, of which the first baseline model includes 10 control variables: age, age squared, gender, education level, married or not, health status, unemployed information, family size, housing status, and household registration. The second baseline model adds three more variables than the first model. These three variables related to the individual characteristics of the household head, that is, concerns for the economic and financial information, investment risk preference, and happiness status. The third baseline model adds two more variables based on the second model and these two variables are related to family characteristics, that is, logarithm of total household income and logarithm of total household assets. Additionally, to test the magnitude of the effect that supply credit constraints and demand credit constraints have on the household stock assets, I add two additional OLS models by replacing the credit constraints variable with two dummy variables which are so-called supply-oriented credit constraints and demand- oriented credit constraints, respectively.

Running regression on the baseline model, if the coefficient on the credit constraint variable in the regression is significant at the 1%, 5%, or 10% level, does it mean that credit constraints has a direct effect on household stock assets? Does it mean that hypothesis H1a or hypothesis H1b is successfully verified? If the absolute value of the estimated coefficient of the demand credit constraints variable is greater than that of the supply credit constraint, does that means the hypothesis H2a have also been successfully tested? Under desirable conditions, dependent variable is significantly correlated with explanatory variables in the analysis of regression, but if there is an endogeneity problem, the regression results favouring the insinuated causation in the structural equation cannot be trusted (Chenhall and Moers, 2007). Omitted variable, simultaneity, measurement error, and selection are the four causes of endogeneity problems summarized by Wooldridge (2010).

The handling of endogeneity has been specifically embedded in quantitative analysis of economics since a while now (Ashenfelter and Card, 2001). Unobserved characteristics would potentially affect credit constraint variables. Nevertheless, due to the limitations of the data[7], I cannot construct the panel dataset and perform the method of fixed effect to eliminate the effect of unobserved characteristics, for example, the unobserved personality of the household head. Moreover, if the endogeneity problem caused by omitted variables is not addressed, the

[7] The limitations of the data are mainly due to the fact that the CHFS survey is conducted every two years, and the questions of the survey changed each time, making some years of the survey data not contain the variables which I need.

error term will include variables related to credit constraints, which will lead me to overestimate or underestimate the impact of credit constraints over household stock assets, and that means the estimated coefficients of this variables in the regression results of baseline models 1-3 deviate from the true values. Similarly, this issue leads to inaccurate coefficient estimates for the demand-oriented and supply-oriented credit constraints variables in baseline models 4-5 and unconvincing test results for hypotheses H2a to H2c.

4.2 Identification strategy

To avoid the impact of potential endogeneity issues on the results of the empirical analysis, I need to pick an identification strategy to solve the issue. Commonly used identification strategies include Instrumental Variable (IV) approach, Regression Discontinuity Design (RDD), Propensity Score Matching (PSM), Difference in Differences (DID) method, and so on. In this paper, I have adopted IV method to solve the potential issues. However, not any variable can be taken as an IV. The selection of instrumental variable requires the satisfaction of both instrument relevance and instrument exogeneity, that is:

$$\text{cov}(\text{IV*credit\_constraints}) \neq 0 \text{ and } \text{cov}(\text{IV*error term}) = 0$$

According to the arguments of Stock and Watson (2003), the IV chosen for the credit constraints should be satisfied these two conditions: 1. There should be a significant degree of correlation between IV and credit constraints variable to make a strong instrument; 2. To avoid correlation of IV with the error term, IV cannot be the explanatory variable of the model for the dependent variable. I select bank density[8] as the IV for the credit constraints and it is calculated by the following equation:

$$\text{Bank density}_i = \frac{\text{number of banks located in the province where respondent i lives}}{\text{area of the province where respondent i lives}}$$

Higher local bank density can facilitate access to credit and reduce the chance of credit constraints (Rajan and Ramcharan, 2011; Rossi and Trucchi, 2016), thus satisfying the instrument relevance property mentioned above. Furthermore, the IV need to satisfy another assumption, namely the exclusion restriction (Jones, 2015). This assumption requires that the selected IV is uncorrelated with the disturbance term, but the disturbance term is unobservable and thus the exogeneity of the IV needs to be examined in terms of the correlation between z and the dependent variable, which means that IV cannot be an explanatory variable for the dependent variable of the model as I mentioned above (instrument exogeneity condition). The dependent variable of household stock assets in this paper is not affected by bank density due to the fact that the distribution of bank branches is exogenous for both households and individuals (Gao et al., 2020). Meanwhile, stock accounts in mainland China are mainly enabled via two ways, offline to the business offices of securities companies or online Application (APP), which are not associated with bank density, so the exclusion restriction assumption of the IV is also satisfied[9]. After selecting the

[8] The CHFS data does not include data of bank density, which is obtained by my personal calculation. Data for the number of banks in each province are obtained from the China Banking and Insurance Regulatory Commission at https://xkz.cbirc.gov.cn/jr/

[9] The exclusion restriction assumption for the instrumental variable would break down in one possible scenario. Areas with higher bank density are likely to be wealthier and therefore local households hold more stocks, but the inclusion of variables related to household wealth in my control variables could solve this potential problem.

appropriate IV, I use the Two-Stage least squares (2SLS) approach, which is performed in two stages, and apply the OLS method twice for fitting.

4.2.1 First-stage of 2SLS method

In the previous I construct several OLS models as baseline models in which the dependent variable is the log-transformation of stock assets owned by the household and the core explanatory variable is a dummy variable for whether the household is facing the credit constraints. However, in the first stage of 2SLS, the independent variable is the bank density that I select as an IV, and the credit constraints turns out to be the dependent variable, as shown in equation (2). Additionally, the random disturbance term ε_i in equation (2) is the unobservable factor that constitutes the source of endogeneity.

$$\text{credit constraints}_i = \beta_0 + \beta_1 X_i + \beta_2 \text{Bank density}_i + \varepsilon_i \quad (2)$$

By regressing equation (2) with the OLS method, I can obtain the estimated values of the variables in equation (3). The predicted value for the dependent variable in equation (3), $\widehat{\text{credit constraints}_i}$, is not correlated with the random disturbance term ε_i in equation (2). Note that the disturbance term ε_i here is different from the disturbance term u_i in equation (1).

$$\widehat{\text{credit constraints}_i} = \widehat{\beta_0} + \widehat{\beta_1} X_i + \widehat{\beta_2} \text{ Bank density}_i \quad (3)$$

4.2.2 Second-stage of 2SLS method

The credit constraint variable is equal to the predicted value in equation (3) combined with the random disturbance term ε_i in equation (2), that is:

$$\text{credit constraints}_i = \widehat{\text{credit constraints}_i} + \varepsilon_i \quad (4)$$

In the second stage of the 2SLS approach, I regress the explained variable that associated with household stock assets on the control variables and the estimates of the credit constraints variable obtained from the first stage, by bringing equation (4) into equation (1) and estimating it using the OLS method, and sorting it out to derive the following[10]:

$$\log\left(\text{stock assets}_i\right) = a_0 + a_1 \widehat{\text{credit constraints}_i} + a_2 X_i + u_i'$$

[10] For the whole derivation process, please refer to the appendix.

5. Results

5.1 Baseline models regression results

Columns 1 to 3 of Table 3a present the regression results for the first three OLS models I build. In model 1, the p-value of the credit constraints variable is significant at the 5% level, but the R-squared of the model is only 0.085. In model 2, I add some variables related to the personal characteristics of the household head as control variables, including the concerns of household head to economics and financial information, investment risk preference, and their happiness status. The p-value of the credit constraints variable is significant at the 1% level now and the R-squared value of model 2 has increased, from 0.085 in model 1 to 0.132. The R-squared value ranges from 0 to 1, indicating how much of the variation in the dependent variable of the model is explained by the independent variables. 0.132 means that 13.2% of the variation in log(stock assets) iwthin regression model 2 is explained by the explanatory variable of credit constraints and also the control variables related to the personal characteristics of the household head.

In model 3, I add some additional variables related to family characteristics to the control variables based on model 2, including the log-transformation for total household income and total household assets. The R-squared of model 3 shows that the credit constraints variable, the individual characteristics of the household head variables, and the variables of household characteristics collectively explain 14.9% of the amount of variation in log(stock assets). The p-value of the credit constraints variable in model 3 is significant at the 1% level, suggesting that whether or not a household is exposed to credit constraints directly affects their holdings of stock assets, when other variables are controlled to be constant. The magnitude of the coefficients of the core explanatory variables in the model captures the extent to the impact of whether a household faces credit constraints on the stock assets hold by the whole family. Examining Table 3a, I find that after adding the logarithm of total household income and the logarithm of total household assets as control variables, the coefficient on the credit constraints variable in Model 3 is -0.120, revealing a 12.00% change in the proportion of household holdings of stock assets for those households facing credit constraints compare to those without credit constraints when all other variables are fixed. A negative coefficient on the credit constraints variable demonstrates that the effect of the credit constraints variable on household stock assets is negative, in the sense that lack of credit support from formal financial institutions reduces household holdings of stock assets.

Moreover, majority of the control variables in model 3 are significant, except for two variables, the married status of the household head and their unemployed information. The most significant control variable that has the greatest effect on household stock assets is whether the household owns one or more than one house[11], followed by the degree of concerns for economics and financial information by the household head, but these two control variables have different directions of influence on household stock assets. A positive

[11] Given that both hypotheses H1a and H1b are based on housing purchases (but not limited to houses), while the majority of households in the data own at least one house, I also construct a subsample of household heads who do not own housing property. In this subsample, I construct OLS models and include all control variables (except for the housing dummy variable), and the core explanatory variables are credit constraint, demand credit constraint, and supply credit constraint, respectively. By looking at the p-values of the estimated coefficients for the explanatory variables in the first three columns of Table 3b, I find that none of them are significant and that the results in columns 3 to 5 of Table 3a cannot be replicated in the subsample. Considering the possible endogeneity problem that makes the estimation results of these models unreliable, the fourth column of Table 3b presents the regression results of the 2SLS model for the subsample, where the estimated coefficients of the credit constraints variable remains significant at the 1% level.

coefficient of the concerns over economics and financial information implies that the more the household head pays attention to such type of information, the more stock assets the household will hold. In contrast, the coefficient on household housing status is negative, and one possible explanation is that most households possess relatively limited money, and investing in property means that they have less money to spend on stock assets or the others. Other important control variables that affect household equity assets are the household registration of the household head, their investment risk preference, and the logarithm of total household assets.

In comparison to the coefficients of the key explanatory variable in models 1-3, the absolute value of the coefficient in model 2 which adds three additional household head characteristics variables is larger than the absolute value of the coefficient in model 1, while the absolute value of the coefficient in model 3 after adding two more family characteristics variables as control variables based on model 2 is smaller than the absolute value of the coefficient in model 2. These changes in coefficients suggest that there may be mediating variables among these added control variables, in that credit constraints variable has an effect on household stock assets through mediating variables. I will analyze the specific mechanism of influence in the later sections. While comparing the coefficients of the key explanatory variable in columns 4 and 5, the absolute value of the estimated coefficient of the demand-side credit constraints variable is slightly larger than that of the supply-side credit constraint variable, also both of them are significant at 1% level. These representing that both credit constraints have a significant negative impact on household stock assets, but the effect of the demand credit constraint on household stock assets is larger than that derived from the supply credit constraint variable.

Besides, as I discuss in the model section that potential endogeneity issues can cause the estimated coefficients of the key explanatory variables to deviate from the true values, leading me to overestimate or underestimate the impact of facing credit constraints on household stock assets, and the estimates of the magnitude of the impact from the two different types of credit constraints are not enough convincing. Thus, by analyzing the regression results of the baseline model, hypothesis H1b and H2a are initially verified, yet I need to perform further verification of H1b after solving the endogeneity issues through the IV approach/2SLS method, while the further verification of H2a I will do in the robustness test section.

5.2 2SLS regression results

In the first-stage regression results of the 2SLS (the first column of Table 4a), the coefficient of the bank density that I select as an IV is negative and significant at the 1% level, suggesting that when the bank density of the province where the respondent is located is higher, the less likely the household of the respondent would face credit constraints. To determine whether my selected bank density is a weak IV or not, I also measure the IV using an F-test method. The criterion used in this paper is the rule of thumb given by Stock and Yogo (2005), which determines the weakness of the IV according to whether the F-statistic in the first stage of 2SLS is less than 10. The first column of Table 4a reports an F-statistic in the first stage of 2SLS, far exceeding the benchmark of 10, and thus the instrumental variable I choose, bank density, is not a weak instrument.

The second column of Table 4a illustrates the regression results for the second stage of 2SLS, where the absolute value of coefficient on the credit constraints variable is significantly raised and also significant at the 1% level by employing the IV approach and adding the household head personal characteristics variable and the family characteristics variable as control variables. Comparing the regression results in the baseline model 3, the estimated coefficient of credit constraint obtained using the 2SLS method declines to -21.39, indicating that holding the other factors fixed, the proportion of stock assets of households suffering from credit constraints decreases by 2139% compare to households without credit constraints. One possible reason for the large proportional decline is the smaller average stock asset base of households exposed to credit constraints.

The above ratio may be difficult to comprehend and when I replace the dependent variable with the value of household stock assets (without logarithmization), the coefficient on the credit constraints variable becomes --328437 and remains significant at the 1% level (as shown in the Table 4b), implying that, holding all other variables constant, households suffering from credit constraints have on average RMB 328,437 less in household stock assets compared to households without credit constraint. The reduction in stock assets may seem considerable, but when I compare this figure to the average amount of assets in the securities accounts of Chinese investors over RMB 400,000 in 2017(refer to Figure 1b), all appears to make sense. Households suffering from credit constraints may sell off most of their risky financial assets[12] in exchange for cash to buy target assets (such as house, car, etc.) since they cannot apply for loans from formal financial institutions like banks, while households without credit constraints can still keep or even increase their investment in risky assets because they have a higher risk tolerance due to the credit support from formal financial institutions.

Hypothesis H1b is strengthened by the analysis of the 2SLS regression results, which confirm that credit constraints have a significant negative effect on the proportion of household stock asset holdings. Noted that both in the baseline model 3 and in the regression results of the 2SLS second stage, the variables of the married status of household head and their unemployed profile are insignificant. Also, the control variables that are not significant in the regression results of the 2SLS second stage include the logarithm of total household income, and therefore hypothesis H3 is likely to hold, but further analysis is needed, and I would verify hypothesis H3 in the section for the mediating variables.

As an additional note, the table does not report R-squared values in either of the first or the second stages of regression results for the 2SLS method due to the fact that most software applies equation (5) to calculate the R-squared after the IV estimation. However, since the SSR of the IV is likely to be larger than the SST, there is a chance that the R-squared in the IV estimation would be negative (Wooldridge, 2015). Based on this, reporting the magnitude of the R-squared is not meaningful for the IV approach.

[12] The hoarded liquidity of financial assets can serve the demand for future ready money (Holmström and Tirole, 2002). Financial assets offer superior liquidity than fixed assets (such as housing), signifying they are much easier to realize.

$$R^2 = 1 - \frac{SSR}{SST} \quad (5)$$

5.3 Mediating variables

Exogenous increases in credit constraints affect the depth of household participation in the stock market, in other words, the amount of stock assets owned by the family. To explore the mechanism hiding behind the relationship, I follow the method of Persico et al.(2004) to explore the causality, by examining how the estimated coefficient of the credit constraints variable changes after the addition of various control variables. If the estimated coefficient of the credit constraint variable varies significant as a result of adding a specific control variable, it is possible that the credit constraint variable affects household stock assets by shaping that specific control variable and thus the stock assets of the family, which means that the effect of the credit constraints variable on household stock assets is likely to be a consequence of affecting the specific control variable.

Table 5 presents the results of the second stage regression of the 2SLS method with the addition of various kinds of control variables. Among the columns, column m1 is the baseline model and contains the same control variables as the baseline model 1. The following columns m2 to m6 add a single specific control variable separately, and column m7 contains all the control variables (in line with the results in the second column of Table 4a). Hereafter I would like to analyze the mechanism of the credit constraints on household stock assets in terms of five dimensions: the concerns that household head focus on economics and financial information, investment risk preference, happiness status, log-transformation for total household income and total household assets. I determine the existence of the mechanism, the importance of the mechanism, and the direction of the impact provided by the mechanism through changes in the estimated coefficient of the credit constraints. By observing columns m2, m3 and m6, the coefficients of the added mediating variables are all significantly positive at the 1% level, showing that holding the other factors fixed, the more the household head pays attention to economics and financial information, the more total stock assets household holds. Also, the more the household head tends to be a risk lover, or the increase in total household assets can also somewhat increase the family stock assets. The negative sign of the estimated coefficient of the happiness variable in column m4 suggests that this variable has a negative direction of influence on household stock assets, while the coefficient on the logarithm of total household income in column m5 is insignificant, so I verify the hypothesis H3 and will not include the logarithm of total household income in the rest of the analysis.

There are four potential paths for the mechanism, with credit constraints positively/negatively affecting the mediating variable and the mediating variable positively/negatively impacting household stock assets. Reviewing the regression results in Table 5, in comparison with column m1, I find that the estimated coefficients of the credit constraints variables in the columns that added the concerns of household head to the economics and financial information, investment risk preference, and happiness status these three variables varied slightly, demonstrating that these are not significant mediating variables. In contrast, the estimated coefficient on the credit constraints becomes significantly larger in column m6 where the logarithm of total household assets is added as a mediating variable, indicating that this is an important channel between the two negatively correlated variables, credit constraint and total household assets. I have derived from the analysis above a positive relationship between the mediating variable and household stock assets, which seems reasonable: the more household assets would include a larger share of household stock holdings given that stocks, as risky financial assets, are also part of household assets. And regarding the relationship between credit constraints and the logarithm of total household assets, it could be verified through the correlation test (available in the Table 6). The test result for the correlation between logarithm of total household assets and the credit constraints variable is negative and significant at 1% level, thus confirming the negative relationship between these two variables.

All in all, I examine the mediating mechanism behinds the impact from credit constraints by adding various kinds of mediating variables and find that one potential channel for credit constraints to reduce household stock asset by reducing total household assets, then the hypothesis H4 is successfully verified. The findings also suggest that although the three variables of the attention that household head paid to economics and financial information, investment risk preference, and happiness status are also significant variables for credit constraints to affect stock assets hold by families, the degree of the effect is small and they are not the main mechanism through which credit constraints affect household stock assets holdings. Moreover, I also verify the hypothesis that the logarithm of total household income is not one of the significant mediating variables of credit constraints affecting total stock ownership of family.

6. Tests

6.1 Endogeneity test

Commonly used tests for endogeneity include the Hausman test and the Durbin-Wu-Hausman (DWH) test. In order to obtain heteroskedasticity-robust standard error estimates, all regression models in this thesis are adjusted for heteroskedasticity using the "robust" command in Stata. And since the Hausman test is not applicable in the existence of heteroskedasticity, I choose to use the heteroskedasticity robust method of DWH test to examine the endogeneity.

The DWH test is a widely used specification test for the endogeneity of IV regressions, and possessing strong instruments is one of the prerequisites for the DWH test (Guo, et al., 2018). The DWH test using 2SLS variance estimators emerges with distorted size under the null hypothesis when there is a weak instrumental variable issue (Staiger and Stock, 1997), which is the reason why I first perform an F-test for testing the weakness or strength of the instrumental variable in the previous. Table 7 demonstrates the null hypothesis and the corresponding p-values for the DWH test. The null hypothesis is that all variables are exogenous and I can reject it at the 1% level of significance due to the p-value of 0. The rejection of the null hypothesis that all explanatory variables are exogenous implies that the original explanatory variable is endogenous when the IV is selected appropriately, hence my decision to use the identification strategy to address the endogeneity issue is the correct and necessary procedure.

6.2 Robustness check

Robustness tests are currently quite common in empirical studies, where researchers modify the regression specification in some manner to detect changes in the coefficient estimates of core variables, and coefficient estimates that do not change considerably can be used as evidence that the coefficients are robust (Lu and White, 2014). Popular robustness tests include replacing or adding variables, split-sample regression, and changing the sample size to examine whether the results are still "robust". In fact, the scope of robustness testing contains lots of methods, such as the Hausman test, the over-identification test as I mentioned above and the DWH test which I already applied. Therefore, in this part I will perform a robustness test by varying the sample size, which is also a procedure to evaluate the H1c hypothesis.

In the introduction section I mention that despite rapid economic growth of China in recent years, income inequality has been gradually rising, and thus the CHFS data I use for the study include the outliers in the sample due to the representativeness of CHFS. Given that the presence of extreme values would affect the overall average of each variable, which in turn may have impact on the regression coefficients. Consequently, in order to reduce outlier disturbances and thus affect the regression results, this paper continues to estimate the impact of credit constraints on household stock assets using the 2SLS method after a 1% upper and lower winsorizing process for all variables. As compared to the regression results in Table 4b, the estimated coefficients of the credit constraints variables in the second-stage regressions of 2SLS after the winsorizing treatment (refer to the column 7 of Table 8) do not vary substantially and remain significant at the 1% level, denoting that holding the other factors fixed, the proportion of stock holdings is significantly lower for households who are facing credit constraints compare to the others. For the results of the above robustness tests confirm the robustness of the findings in this thesis. I also conduct robustness check on the main channels through which credit constraints affect household stock assets and found that the estimated coefficients in the columns of Table 8 are not significantly different from those in the columns of Table 5, confirming that total household assets are an important mediating variable between credit constraints and household stock assets holdings.

Moreover, to further verify whether hypothesis H2a holds, I conduct a two-stage regression using the instrumental variable I picked in the identification strategy section. Since replacing econometric methods is also a common robustness test, in order to validate that the outcomes are robust, I use the treatment effect model (TEM) developed by Maddala (1983), which addresses the issue of self-selection bias induced by omitted unobservables. The major difference between the TEM and 2SLS models is that the TEM replaces the OLS model in the first stage of the 2SLS with a Probit model and calculates the hazard ratio, thus adding the hazard ratio to the OLS model in the second stage for regression. The Probit model is shown in equation (8). Note that the dependent variable Y in the Probit model is no longer the credit constraints variable at this time, but the supply-side credit constraints variable or the demand-side credit constraints variable[13]. Φ is the cumulative distribution function of the standard normal distribution.

[13] I construct two treatment effect models where one dependent variable is the supply-side credit constraints and the other is the demand-side credit constraints. Both variables are dummy variables and they take the specific value of 0 or 1.

$$\Pr(Y = 1|\text{Bank density}) = \Phi(\beta_0 + \beta_1 \text{Bank density}) \quad (6)$$

If Y = 1, then the household faces supply-oriented or demand-oriented credit constraints; conversely, the household does not suffer from supply or demand side credit constraints. Furthermore, the hazard ratio can be calculated from equation (7), where φ is the probability density function of the standard normal distribution and $\hat{Y}$ is the predicted value of Y after the Probit regression of equation (6). The calculated ratio will be brought into the OLS model to adjust for the bias arising from endogeneity issues, which is added the hazard ratio to the equation (1) and ran the regression.

$$\text{Hazard ratio} = \begin{cases} \frac{\varphi(\hat{Y})}{\Phi(\hat{Y})}, & Y = 1 \\ \frac{-\varphi(\hat{Y})}{\Phi(-\hat{Y})}, & Y = 0 \end{cases} \quad (7)$$

Table 9 shows the regression results for the second stage of the TSM. Comparing the estimated coefficients of the key explanatory variables in the first and second columns, both the two types of credit constraints have a significant negative effect on household stock assets, but the larger absolute value of the estimated coefficient of the demand-side credit constraints variable indicates that this type of credit constraints have a greater effect on household stock ownership, and the hypothesis of H2a is successfully verified. The results of the above tests ensure the robustness of the outcomes in this paper.

7. Conclusion

This thesis examines the impact of credit constraints on the stock asset holdings of household by introducing the 2017 CHFS, a micro-level financial data of Chinese households, to comprehensively explore not only the direction in which credit constraints affect household stock assets, but also to identify the main channels between credit constraints and household stock ownership. Additionally, this paper refers to the previous works of scholars to further classify credit constraints into demand-oriented credit constraints and supply-oriented credit constraints, so as to compare the magnitude of the effects that the two types of credit constraints on household stock holdings.

As a representative risky asset, stocks are able to achieve an increase in household wealth if families can implement a reasonable allocation to them, which is an effective pathway to reduce income inequality among Chinese households. This paper captures the direction of credit constraints affecting household stock assets through OLS and 2SLS as empirical methods, and the results suggest that those households facing credit constraints dilemma have significantly lower proportion of stock ownership compared to those without credit constraints. To test the mechanism by which credit constraints affect household stock assets, I also build multiple 2SLS models by adding mediating variables one by one and find that credit constraints reduce the stock holdings of Chinese households by decreasing household wealth. For more, I compare the two types of credit constraints and discover that the demand-oriented credit constraints have a significantly higher impact on household stock assets than the supply-oriented credit constraints. All the above results are guaranteed to be robust by passing the relevant robustness tests.

Enhancing the supply of credit from financial institutions allows households to better access the credit support they needed, facilitating them to choose their optimal plan without credit constraints, and more appropriately allocating their assets, leading to the growth of household wealth. In the robustness check section, I also verify that the demand credit constraints have a greater negative impact on household stock holdings than the supply credit constraints, and that this situation would be improved by refining the credit market function to minimize the "discouraged borrowers" phenomenon caused by information asymmetry at the root. All in all, alleviating the credit constraints faced by households is an essential way to ameliorate the current sharp rise in inequality of China.

There also exists room for deeper improvement in this study. The effect of credit constraints on the share of stock assets over total household assets is not examined within this paper, mainly because there may be a correlation between the instrumental variable of bank density that I have chosen and total household assets[14], such that an increase in bank density makes households less likely to face credit constraints dilemma, and thus households are more probable to secure credit support for the purchase of housing, car, and other assets that are part of the household total assets. In the future, I look forward to discovering suitable

[14] To study the effect of credit constraints on the share of stock assets, the dependent variable for the OLS and 2SLS second-stage models I have constructed needs to become the share of stock assets over total household assets. Given the theoretical possibility of correlation between the variable I pick as instrument and the share of stock assets, bank density is not an appropriate IV to address for such topic.

instrumental variables to explore the causal relationship between credit constraints and stock assets shares in depth and to enrich the literature for related fields.

Bibliography

Ameriks, J., Kézdi, G., Lee, M., & Shapiro, M. D. (2020). Heterogeneity in Expectations, Risk Tolerance, and Household Stock Shares: The Attenuation Puzzle. *Journal of Business & Economic Statistics*, *38*(3), 633–646. https://doi.org/10.1080/07350015.2018.1549560

Arrondel, L., Debbich, M., & Savignac, F. (2015). Stockholding in France: The role of financial literacy and information. *Applied Economics Letters*, *22*(16), 1315–1319. https://doi.org/10.1080/13504851.2015.1026578

Ashenfelter, O., & Card, D. (2010). *Handbook of Labor Economics*. Elsevier.

Banerjee, A. V., & Duflo, E. (2014). Do Firms Want to Borrow More? Testing Credit Constraints Using a Directed Lending Program. *The Review of Economic Studies*, *81*(2), 572–607. https://doi.org/10.1093/restud/rdt046

Betermier, S., Calvet, L. E., & Sodini, P. (2017). Who Are the Value and Growth Investors? *The Journal of Finance*, *72*(1), 5–46. https://doi.org/10.1111/jofi.12473

Boucher, S. R., Carter, M. R., & Guirkinger, C. (2008). Risk Rationing and Wealth Effects in Credit Markets: Theory and Implications for Agricultural Development. *American Journal of Agricultural Economics*, *90*(2), 409–423. https://doi.org/10.1111/j.1467-8276.2007.01116.x

Campbell, J. Y. (2006). Household Finance. *The Journal of Finance*, *61*(4), 1553–1604. https://doi.org/10.1111/j.1540-6261.2006.00883.x

Campbell, J. Y., & Cocco, J. F. (2003). Household Risk Management and Optimal Mortgage Choice*. *The Quarterly Journal of Economics*, *118*(4), 1449–1494. https://doi.org/10.1162/003355303322552847

Chenhall, R. H., & Moers, F. (2007). The Issue of Endogeneity within Theory-Based, Quantitative Management Accounting Research. *European Accounting Review*, *16*(1), 173–196. https://doi.org/10.1080/09638180701265937

Chetty, R., Sándor, L., & Szeidl, A. (2017). The Effect of Housing on Portfolio Choice. *The Journal of Finance*, *72*(3), 1171–1212. https://doi.org/10.1111/jofi.12500

Cox, D., & Jappelli, T. (1990). Credit Rationing and Private Transfers: Evidence from Survey Data. *The Review of Economics and Statistics*, *72*(3), 445–454. https://doi.org/10.2307/2109352

Flavin, M. A. (1981). The Adjustment of Consumption to Changing Expectations About Future Income. *Journal of Political Economy*, *89*(5), 974–1009. https://doi.org/10.1086/261016

Gan, L., Yin, Z., & Tan, J. (2016). *Report on the development of household finance in rural China (2014)*. Singapore: Springer Science Business Media.

Gao, N., Ma, Y., & Xu, L. C. (2020). *Credit Constraints and Fraud Victimization: Evidence from a Representative Chinese Household Survey* [Working Paper]. World Bank. https://doi.org/10.1596/1813-9450-9460

Guiso, L., Jappelli, T., & Haliassos, M. (2000). *Household Portfolios: An International Comparison* (SSRN Scholarly Paper No. 245805). Social Science Research Network. https://doi.org/10.2139/ssrn.245805

Guiso, L., & Sodini, P. (2013). Chapter 21 - Household Finance: An Emerging Field. In G. M. Constantinides, M. Harris, & R. M. Stulz (Eds.), *Handbook of the Economics of Finance* (Vol. 2, pp. 1397–1532). Elsevier. https://doi.org/10.1016/B978-0-44-459406-8.00021-4

Guo, Z., Kang, H., Cai, T. T., & Small, D. S. (2018). Testing endogeneity with high dimensional covariates. *Journal of Econometrics*, *207*(1), 175–187. https://doi.org/10.1016/j.jeconom.2018.07.002

Hoff, K., & Stiglitz, J. E. (1990). Introduction: Imperfect Information and Rural Credit Markets—Puzzles and Policy Perspectives. *The World Bank Economic Review*, *4*(3), 235–250. https://doi.org/10.1093/wber/4.3.235

Holmström, B., & Tirole, J. (2001). LAPM: A Liquidity-Based Asset Pricing Model. *The Journal of Finance*, *56*(5), 1837–1867. https://doi.org/10.1111/0022-1082.00391

Jain-Chandra, S., Khor, N., Schauer, J., Mano, R. C., Wingender, P., & Zhuang, J. (2018). *Inequality in China—Trends, Drivers and Policy Remedies* (SSRN Scholarly Paper No. 3221211). Social Science Research Network. https://papers.ssrn.com/abstract=3221211

Jappelli, T. (1990). Who is Credit Constrained in the U. S. Economy?*. *The Quarterly Journal of Economics*, *105*(1), 219–234. https://doi.org/10.2307/2937826

Jones, D. (2015). *The Economics of Exclusion Restrictions in IV Models* (Working Paper No. 21391; Working Paper Series). National Bureau of Economic Research. https://doi.org/10.3386/w21391

Kozak, S., & Sosyura, D. (2015). *Access to Credit and Stock Market Participation* (SSRN Scholarly Paper No. 2594976). Social Science Research Network. https://doi.org/10.2139/ssrn.2594976

Lu, X., & White, H. (2014). Robustness checks and robustness tests in applied economics. *Journal of Econometrics*, *178*, 194–206. https://doi.org/10.1016/j.jeconom.2013.08.016

Maddala, G. S. (1983). *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press. https://doi.org/10.1017/CBO9780511810176

Persico, N., Postlewaite, A., & Silverman, D. (2004). The Effect of Adolescent Experience on Labor Market Outcomes: The Case of Height. *Journal of Political Economy*, *112*(5), 1019–1053. https://doi.org/10.1086/422566

Rajan, R. G., & Ramcharan, R. (2011). Land and Credit: A Study of the Political Economy of Banking in the United States in the Early 20th Century. *The Journal of Finance*, *66*(6), 1895–1931. https://doi.org/10.1111/j.1540-6261.2011.01694.x

Rossi, M., & Trucchi, S. (2016). Liquidity constraints and labor supply. *European Economic Review*, *87*, 176–193. https://doi.org/10.1016/j.euroecorev.2016.05.001

Shum, P., & Faig, M. (2006). What explains household stock holdings? *Journal of Banking & Finance*, *30*(9), 2579–2597. https://doi.org/10.1016/j.jbankfin.2005.11.006

Staiger, D., & Stock, J. H. (1994). *Instrumental Variables Regression with Weak Instruments* (Working Paper No. 151; Technical Working Paper Series). National Bureau of Economic Research. https://doi.org/10.3386/t0151

Stock, J., & Watson, M. (2003). *Introduction to Econometrics*. Prentice Hall.

Stock, J., & Yogo, M. (2005). Testing for Weak Instruments in Linear IV Regression. In D. W. K. Andrews (Ed.), Identification and Inference for Econometric Models (pp. 80–108). Cambridge University Press.

Stringham, E. P., & Curott, N. A. (2015). *On the Origins of Stock Markets*. The Oxford Handbook of Austrian Economics.

Wang, G. (2016). Credit Constraints and Household Selection of Financial Assets. *Open Journal of Social Sciences*, *4*(2), 47–56. https://doi.org/10.4236/jss.2016.42009

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data, second edition*. MIT Press.

Wooldridge, J. M. (2015). *Introductory Econometrics: A Modern Approach*. Cengage Learning.

Zou, J., & Deng, X. (2019). Financial literacy, housing value and household financial market participation: Evidence from urban China. *China Economic Review*, *55*, 52–66. https://doi.org/10.1016/j.chieco.2019.03.008

Appendix

Figure 1a Total market value of Shanghai and Shenzhen stock exchange (Unit: trillion RMB)

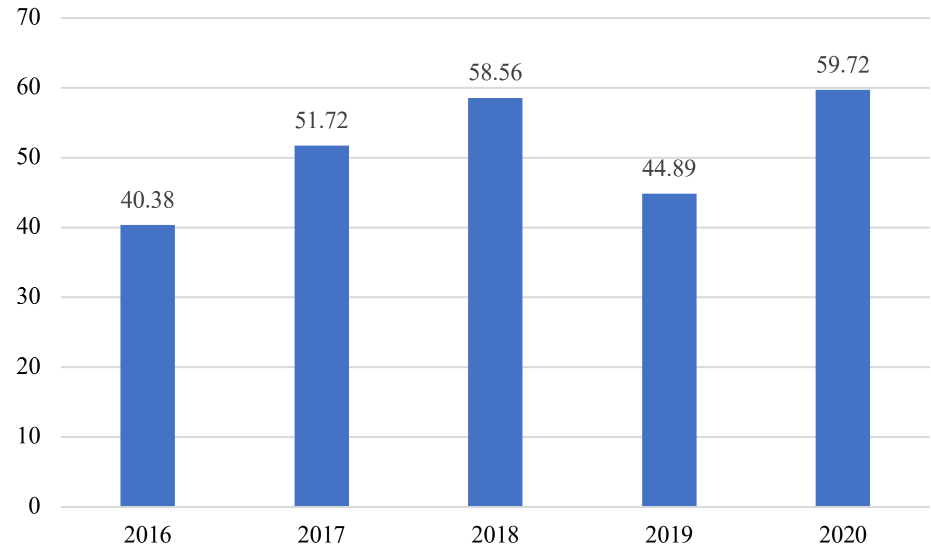
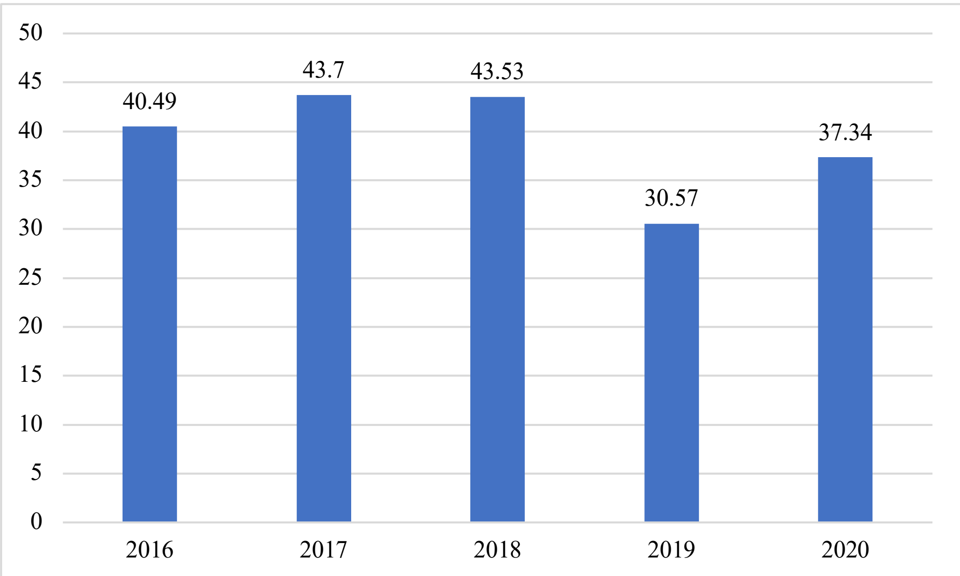


Figure 1b Average market value of the accounts for Chinese stock investors (Unit: 10,000 RMB)



Note: The data in Figures 1a and 1b are both from eastmoney choice data, and for each year I have chosen the data from January as representative of the whole year. The Beijing Stock Exchange was founded in 2021, and since both figures end in January 2020, the relevant data for it was not included in the above two graphs.

Table1 Variable definitions

Variable category	Variable name	Variable definition
Dependent variable	log_stock assets	The logarithm of household stock assets, calculated by the formula: log_assets = log(household stock assets + 1)
Key explanatory variable	credit_constraints	= 1, if the family is facing credit constraints = 0, if the family is not exposed to credit constraints
Control variable - Household head characteristic variable	age	= 2017 - year of birth that the household head was born
	age2	= age * age
	gender	= 1, if the gender of the household head is woman = 0, if the gender of the household head is man
	married	= 1, the head of the household is married = 0, otherwise
	health	The self-perceived health status of the household head compared to their peers. The value of this variable ranges from 0 to 4, with higher values indicating that the head of household perceives his or her health to be better than the peers.
	unemployment	= 1, the head of the household is unemployed = 0, otherwise
	education_level	=1, never received any education =2, Elementary school education =3, Junior high school education =4, High school education =5, Specialized secondary school/Technical high school =6, College/Higher vocational school =7, Bachelor education =8, Masters education =9, PhD education
	hh_registration	= 1, the head of the household owns the rural registration = 0, otherwise
	info_concern	The variable that measures the degree of concern that household head pay to economics and financial information usually, with values ranging from 0 to 4. Higher values mean that household head are more concerned with such information.

	risk_preference	Investment risk preference variable. Values range from 0 to 4, with higher values indicating that household heads tend to be more risk-lovers and lower values tend to be more risk-averse.
	happiness	= 3, Extremely Happiness = 2, Happiness = 1, Average = 0, Unhappiness
Control variable - Family characteristic variable	family_size	Size of the family
	own_house	= 1, the head of the household owns one or more houses = 0, otherwise
	log_income	The logarithm of household total income, calculated by the formula: log_income = log(total household income + 1)
	log_asset	The logarithm of household total assets, calculated by the formula: log_asset = log(total household assets + 1)

Table 2 Descriptive statistics

Name of the variables	Observations	Mean	Standard deviation	Min	Max
credit_constraints	33,990	0.07	0.26	0	1
log_stock_assets	33,990	0.60	2.52	0	14.51
stock_ownership	33,990	0.06	0.23	0	1
age	33,990	55.00	14.10	16	117
age2	33,990	3,224	1,567	256	13,689
gender	33,990	0.20	0.40	0	1
married	33,990	0.85	0.35	0	1
health	33,990	2.41	1.01	0	4
unemployment	33,990	0.14	0.35	0	1
education_level	33,990	3.44	1.66	1	9
family_size	33,990	3.67	2.25	1	30
own_house	33,990	0.91	0.29	0	1
hh_registration	33,990	0.53	0.50	0	1
info_concern	33,990	1.02	1.09	0	4
risk_preference	33,990	0.92	1.17	0	4
happiness	33,990	1.89	0.78	0	3
log_income	33,990	10.71	1.55	0	15.96
log_assets	33,990	12.65	1.85	0	17.97

Note: 1.All figures in Table 2 for descriptive statistics are rounded and retained to two decimal places, therefore the figures may slightly differ from what is described in the corresponding chapter. 2.stock_ownership is a dummy variable whose value of 0 means that the household has no stock ownership, while the opposite (= 1) means that the household holds stock. 3.The number of households in the sample with stock assets (meaning that stock holdings are not zero) is 1,864. 4.All money-related variables measured in RMB.

2SLS second stage equation organizing process

Plug equation (4) into equation (1)

$$\log(\text{stock assets}_i) = a_0 + a_1 \text{credit constraints}_i + a_2 X_i + u_i$$

$$\log(\text{stock assets}_i) = a_0 + a_1 (\text{credit } \widehat{\text{constraints}_i} + \varepsilon_i) + a_2 X_i + u_i$$

$$\log(\text{stock assets}_i) = a_0 + a_1 \text{credit } \widehat{\text{constraints}_i} + a_2 X_i + (a_1 \varepsilon_i + u_i)$$

$$\log(\text{stock assets}_i) = a_0 + a_1 \text{credit } \widehat{\text{constraints}_i} + a_2 X_i + u_i'$$

Note for Table 3a and 3b: 1.The dependent variable (DV) for all models is log_stock_assets. 2.For EV (explanatory variable): EV1 is credit constraints, EV2 is the supply-oriented credit constraints and EV3 is demand-oriented credit constraints. 3.*** p<0.01, ** p<0.05, * p<0.1.

Table 3a OLS models regression results

Variables	(1)Model1-DV	(2)Model2-DV	(3)Model3-DV	(4)Model4-DV	(5)Model5-DV
EV1	-0.0717**	-0.167***	-0.120***		
	(0.0342)	(0.0344)	(0.0343)		
EV2				-0.116***	
				(0.0406)	
EV3					-0.125***
					(0.0473)
age	0.0629***	0.0714***	0.0671***	0.0671***	0.0670***
	(0.00611)	(0.00606)	(0.00599)	(0.00599)	(0.00599)
age2	-0.000553***	-0.000577***	-0.000532***	-0.000532***	-0.000530***
	(5.32e-05)	(5.23e-05)	(5.18e-05)	(5.18e-05)	(5.17e-05)
gender	0.137***	0.182***	0.153***	0.154***	0.155***
	(0.0414)	(0.0401)	(0.0396)	(0.0396)	(0.0396)
married	0.133***	0.155***	0.0494	0.0498	0.0502
	(0.0402)	(0.0391)	(0.0390)	(0.0390)	(0.0390)
health	0.0362***	0.0212*	-0.0343***	-0.0337***	-0.0328***
	(0.0124)	(0.0124)	(0.0125)	(0.0125)	(0.0124)
unemployment	-0.0368	-0.0333	0.0440	0.0438	0.0449
	(0.0290)	(0.0286)	(0.0288)	(0.0288)	(0.0288)
education_level	0.337***	0.250***	0.188***	0.189***	0.189***
	(0.0134)	(0.0125)	(0.0123)	(0.0123)	(0.0123)
family_size	-0.0216***	-0.0297***	-0.0449***	-0.0452***	-0.0452***
	(0.00411)	(0.00419)	(0.00432)	(0.00432)	(0.00433)
own_house	0.0913**	0.0562	-0.431***	-0.431***	-0.432***
	(0.0458)	(0.0445)	(0.0496)	(0.0496)	(0.0496)
hh_registration	-0.532***	-0.474***	-0.272***	-0.273***	-0.276***
	(0.0272)	(0.0263)	(0.0254)	(0.0255)	(0.0254)
info_concern		0.376***	0.350***	0.350***	0.349***
		(0.0149)	(0.0144)	(0.0144)	(0.0144)
risk_preference		0.277***	0.260***	0.260***	0.260***
		(0.0137)	(0.0134)	(0.0134)	(0.0134)
happiness		-0.0640***	-0.0814***	-0.0809***	-0.0804***
		(0.0158)	(0.0157)	(0.0157)	(0.0157)
log_income			0.0547***	0.0549***	0.0550***
			(0.00690)	(0.00689)	(0.00690)
log_assets			0.203***	0.203***	0.203***
			(0.00812)	(0.00812)	(0.00812)
Constant	-2.175***	-2.734***	-4.906***	-4.916***	-4.918***
	(0.193)	(0.194)	(0.214)	(0.214)	(0.215)

Observations	33,990	33,990	33,990	33,990	33,990
R-squared	0.085	0.132	0.149	0.149	0.149

Table 3b Regression results for the sub-sample

Variables	(1)DV	(2)DV	(3)DV	(4)DV
EV1	0.0145			-25.13***
	(0.127)			(9.732)
EV2		-0.00823		
		(0.153)		
EV3			0.0640	
			(0.190)	
age	0.0566***	0.0566***	0.0565***	0.145**
	(0.0142)	(0.0142)	(0.0142)	(0.0643)
age2	-0.000374***	-0.000375***	-0.000374***	-0.00124**
	(0.000125)	(0.000125)	(0.000125)	(0.000619)
gender	0.178*	0.178*	0.179*	-0.492
	(0.102)	(0.101)	(0.101)	(0.345)
married	0.0321	0.0318	0.0321	-0.268
	(0.109)	(0.109)	(0.109)	(0.316)
health	-0.0259	-0.0263	-0.0259	-0.616**
	(0.0400)	(0.0400)	(0.0398)	(0.260)
unemployment	0.0674	0.0672	0.0672	-0.0608
	(0.0905)	(0.0905)	(0.0905)	(0.352)
education_level	0.133***	0.133***	0.133***	0.119
	(0.0372)	(0.0372)	(0.0372)	(0.0885)
family_size	-0.0612***	-0.0610***	-0.0613***	0.216
	(0.0210)	(0.0210)	(0.0210)	(0.139)
hh_registration	-0.318***	-0.317***	-0.319***	0.786
	(0.0719)	(0.0720)	(0.0718)	(0.494)
info_concern	0.399***	0.399***	0.399***	0.642***
	(0.0503)	(0.0503)	(0.0504)	(0.173)
risk_preference	0.239***	0.239***	0.239***	0.311**
	(0.0432)	(0.0432)	(0.0432)	(0.122)
happiness	-0.0847*	-0.0847*	-0.0844*	-0.179
	(0.0472)	(0.0472)	(0.0471)	(0.165)
log_income	0.0242*	0.0241*	0.0243*	-0.148
	(0.0145)	(0.0144)	(0.0145)	(0.106)
log_assets	0.238***	0.238***	0.238***	0.119
	(0.0240)	(0.0239)	(0.0240)	(0.0818)
Constant	-4.681***	-4.678***	-4.682***	-1.416
	(0.565)	(0.564)	(0.565)	(2.064)
Observations	3,184	3,184	3,184	3,184
R-squared	0.151	0.151	0.151	

Table 4a Regression result of the 2SLS model

Variables	(1)First-stage credit_constraints	(2)second-stage log_stock_assets
credit_constraints		-21.39***
		(2.340)
bank_density	-0.124***	
	(0.00954)	
age	0.00160**	0.102***
	(0.000637)	(0.0154)
age2	-2.19e-05***	-0.00103***
	(5.56e-06)	(0.000141)
gender	-0.0225***	-0.348***
	(0.00319)	(0.0943)
married	-0.00932**	-0.142
	(0.00434)	(0.102)
health	-0.0188***	-0.435***
	(0.00160)	(0.0563)
unemployment	-0.00634	-0.0792
	(0.00437)	(0.0986)
education_level	-0.00530***	0.0729***
	(0.000995)	(0.0273)
family_size	0.00584***	0.0922***
	(0.000797)	(0.0232)
own_house	-0.00254	-0.338***
	(0.00535)	(0.124)
hh_registration	0.0338***	0.477***
	(0.00335)	(0.110)
info_concern	0.00734***	0.513***
	(0.00145)	(0.0387)
risk_preference	0.00543***	0.381***
	(0.00138)	(0.0351)
happiness	-0.0127***	-0.342***
	(0.00199)	(0.0532)
log_income	-0.00330***	-0.0253
	(0.00117)	(0.0275)
log_assets	-0.00212**	0.106***
	(0.00104)	(0.0253)
Constant	0.179***	-0.747
	(0.0227)	(0.702)
Observations	33,990	33,990
F-statistics	170.308	

Note: *** p<0.01, ** p<0.05, * p<0.1

Table 4b Regression result of 2SLS with the dependent variable being stock assets

Variables	stock_assets
credit_constraints	-328,437***
	(56,065)
age	1,973***
	(276.2)
age2	-18.74***
	(2.604)
gender	-5,882***
	(1,769)
married	-2,537
	(1,627)
health	-6,551***
	(1,213)
unemployment	771.8
	(1,737)
education_level	1,167*
	(603.4)
family_size	1,328***
	(433.3)
own_house	-6,629***
	(2,099)
hh_registration	11,000***
	(2,324)
info_concern	7,446***
	(808.1)
risk_preference	6,588***
	(695.5)
happiness	-4,515***
	(994.2)
log_income	-282.2
	(459.3)
log_assets	2,283***
	(468.0)
Constant	-37,813**
	(14,692)
Observations	33,990

Note: *** p<0.01, ** p<0.05, * p<0.1

Table 5 Multiple 2SLS model regression results

Variables	m1 DV	m2 DV	m3 DV	m4 DV	m5 DV	m6 DV	m7 DV
EV1	-24.32***	-24.51***	-24.45***	-24.00***	-24.00***	-19.59***	-21.39***
	(2.189)	(2.153)	(2.166)	(2.149)	(2.313)	(2.246)	(2.340)
age	0.100***	0.0960***	0.123***	0.0942***	0.0997***	0.0904***	0.102***
	(0.0171)	(0.0171)	(0.0173)	(0.0168)	(0.0170)	(0.0145)	(0.0154)
age2	-0.00114***	-0.00111***	-0.00126***	-0.00105***	-0.00114***	-0.00100***	-0.00103***
	(0.000155)	(0.000155)	(0.000156)	(0.000151)	(0.000155)	(0.000135)	(0.000141)
gender	-0.486***	-0.447***	-0.452***	-0.465***	-0.479***	-0.383***	-0.348***
	(0.103)	(0.101)	(0.102)	(0.101)	(0.103)	(0.0919)	(0.0943)
married	-0.238**	-0.249**	-0.210*	-0.154	-0.244**	-0.230**	-0.142
	(0.116)	(0.116)	(0.116)	(0.114)	(0.115)	(0.0973)	(0.102)
health	-0.493***	-0.522***	-0.511***	-0.430***	-0.489***	-0.431***	-0.435***
	(0.0612)	(0.0612)	(0.0612)	(0.0576)	(0.0620)	(0.0562)	(0.0563)
unemployment	-0.103	-0.0769	-0.124	-0.110	-0.0916	-0.0502	-0.0792
	(0.109)	(0.110)	(0.110)	(0.108)	(0.110)	(0.0902)	(0.0986)
education_level	0.199***	0.0942***	0.153***	0.199***	0.196***	0.184***	0.0729***
	(0.0288)	(0.0302)	(0.0293)	(0.0285)	(0.0283)	(0.0247)	(0.0273)
family_size	0.124***	0.116***	0.119***	0.121***	0.120***	0.0887***	0.0922***
	(0.0236)	(0.0234)	(0.0235)	(0.0232)	(0.0250)	(0.0213)	(0.0232)
own_house	-0.0923	-0.157	-0.0975	-0.0569	-0.0964	-0.419***	-0.338***
	(0.125)	(0.126)	(0.126)	(0.124)	(0.124)	(0.116)	(0.124)
hh_registration	0.440***	0.501***	0.483***	0.438***	0.442***	0.381***	0.477***
	(0.119)	(0.119)	(0.119)	(0.117)	(0.117)	(0.104)	(0.110)
info_concern		0.635***					0.513***
		(0.0416)					(0.0387)
risk_preference			0.520***				0.381***
			(0.0389)				(0.0351)
happiness				-0.391***			-0.342***
				(0.0576)			(0.0532)
log_income					0.0231		-0.0253
					(0.0302)		(0.0275)
log_assets						0.154***	0.106***
						(0.0231)	(0.0253)
Constant	0.724	0.696	-0.437	1.225**	0.482	-1.169**	-0.747
	(0.580)	(0.578)	(0.571)	(0.594)	(0.686)	(0.591)	(0.702)
Observations	33,990	33,990	33,990	33,990	33,990	33,990	33,990

Note:1.The dependent variable (DV) for all models is log_stock_assets. 2.EV is the abbreviation for the explanatory variable. EV1 is credit constraints. 3.*** p<0.01, ** p<0.05, * p<0.1.

Table 6 Correlation test between credit constraints and log of household assets

	log_assets	credit_constraints
log_assets	1.0000	
credit_constraints	-0.0754***	1.0000

Note: *** p<0.01, ** p<0.05, * p<0.1

Table 7 Durbin-Wu-Hausman test result

Tests of endogeneity	
Ho: variables are exogenous	
Robust score chi2(1)	= 151.764 (p = 0.0000)
Robust regression F(1,33972)	= 159.027 (p = 0.0000)

Table 8 The results of robustness tests for previous 2SLS models

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	m1	m2	m3	m4	m5	m6	m7
Variables	DV	DV	DV	DV	DV	DV	DV
EV1	-24.10***	-24.27***	-24.22***	-23.78***	-23.82***	-19.37***	-21.30***
	(2.168)	(2.133)	(2.146)	(2.129)	(2.338)	(2.267)	(2.387)
age	0.114***	0.111***	0.136***	0.107***	0.114***	0.102***	0.113***
	(0.0174)	(0.0174)	(0.0176)	(0.0171)	(0.0174)	(0.0149)	(0.0159)
age2	-0.00128***	-0.00125***	-0.00139***	-0.00117***	-0.00127***	-0.00111***	-0.00114***
	(0.000158)	(0.000158)	(0.000159)	(0.000154)	(0.000159)	(0.000139)	(0.000146)
gender	-0.482***	-0.444***	-0.448***	-0.461***	-0.476***	-0.381***	-0.349***
	(0.102)	(0.100)	(0.101)	(0.0998)	(0.103)	(0.0914)	(0.0943)
married	-0.251**	-0.263**	-0.220*	-0.168	-0.255**	-0.235**	-0.145
	(0.115)	(0.115)	(0.115)	(0.113)	(0.114)	(0.0964)	(0.101)
health	-0.490***	-0.517***	-0.507***	-0.427***	-0.487***	-0.427***	-0.433***
	(0.0607)	(0.0607)	(0.0607)	(0.0571)	(0.0619)	(0.0561)	(0.0565)
unemployment	-0.0892	-0.0637	-0.110	-0.0969	-0.0802	-0.0436	-0.0761
	(0.108)	(0.109)	(0.108)	(0.107)	(0.109)	(0.0889)	(0.0979)
education_level	0.198***	0.0923***	0.151***	0.197***	0.195***	0.182***	0.0723***
	(0.0290)	(0.0304)	(0.0295)	(0.0287)	(0.0285)	(0.0247)	(0.0275)
family_size	0.131***	0.122***	0.125***	0.128***	0.126***	0.0927***	0.0993***
	(0.0246)	(0.0244)	(0.0244)	(0.0242)	(0.0267)	(0.0225)	(0.0250)
own_house	-0.104	-0.168	-0.108	-0.0688	-0.107	-0.418***	-0.340***
	(0.124)	(0.125)	(0.125)	(0.123)	(0.123)	(0.115)	(0.124)
hh_registration	0.426***	0.487***	0.469***	0.424***	0.429***	0.370***	0.461***
	(0.117)	(0.118)	(0.118)	(0.116)	(0.115)	(0.103)	(0.109)
info_concern		0.628***					0.509***
		(0.0411)					(0.0387)
risk_preference			0.510***				0.374***
			(0.0384)				(0.0351)
happiness				-0.383***			-0.336***
				(0.0570)			(0.0530)
log_income					0.0212		-0.0406
					(0.0359)		(0.0328)
log_assets						0.156***	0.108***
						(0.0248)	(0.0272)
Constant	0.366	0.330	-0.777	0.895	0.143	-1.500**	-0.909
	(0.577)	(0.575)	(0.570)	(0.591)	(0.722)	(0.598)	(0.739)
Observations	33,990	33,990	33,990	33,990	33,990	33,990	33,990

Note:1.The dependent variable (DV) for all models is log_stock_assets. 2.EV is the abbreviation for the explanatory variable. EV1 is credit constraints. 3.*** p<0.01, ** p<0.05, * p<0.1.

Table 9 Treatment effect model regression results

Variables	(1) log_stock_assets	(2) log_stock_assets
supply_side_credit_constraints	-6.749***	
	(0.783)	
demand_side_credit_constraints		-14.24***
		(2.128)
age	0.0675***	0.0673***
	(0.00598)	(0.00598)
age2	-0.000547***	-0.000543***
	(5.16e-05)	(5.16e-05)
gender	0.147***	0.146***
	(0.0395)	(0.0395)
married	0.0543	0.0500
	(0.0389)	(0.0389)
health	-0.0388***	-0.0371***
	(0.0125)	(0.0124)
unemployment	0.0438	0.0452
	(0.0287)	(0.0288)
education_level	0.186***	0.187***
	(0.0123)	(0.0123)
family_size	-0.0381***	-0.0398***
	(0.00431)	(0.00434)
own_house	-0.356***	-0.374***
	(0.0493)	(0.0495)
hh_registration	-0.271***	-0.276***
	(0.0254)	(0.0254)
info_concern	0.354***	0.353***
	(0.0144)	(0.0144)
risk_preference	0.264***	0.262***
	(0.0134)	(0.0134)
happiness	-0.0737***	-0.0764***
	(0.0156)	(0.0156)
log_income	0.0480***	0.0491***
	(0.00689)	(0.00689)
log_assets	0.177***	0.183***
	(0.00790)	(0.00805)
hazard_ratio	3.156***	5.786***
	(0.377)	(0.882)
Constant	-4.225***	-4.368***
	(0.225)	(0.225)
Observations	33,990	33,990
R-squared	0.154	0.152

Note: *** p<0.01, ** p<0.05, * p<0.1