How Does Financial Literacy Affect Mortgage Default?*

Sumit Agarwal
National University of Singapore
Souphala Chomsisengphet
Office of the Comptroller of the Currency
Yunqi Zhang
National University of Singapore

This version: April 2015

Abstract

This paper uses a dataset from one of the leading subprime lenders in America, which contains detailed information on borrower, loan, and property characteristics. We find that borrowers from the financial industry, who have higher financial literacy, perform better in loan choice. More importantly, the financial industry borrowers are less likely to default. This effect cannot be explained by borrower characteristics such as income and education, loan terms, property characteristics, or geographic effects. The results are robust to various measures of delinquency. We also find there are variations in this effect of financial literacy for different types of borrowers or different kinds of loans. Our results indicate that financial literacy plays an important part in repayment behavior and have helpful policy implications.

Key words: financial literacy, mortgages, default, financial crisis, household finance
JEL classification: D10, R20, G01, G21

* The views expressed in this paper are those of the authors and may not reflect those of the Office of the Comptroller of the Currency.

1. Introduction

Households have been increasingly involved in financial markets and have access to increasing numbers of new financial products and services. Before the outbreak of the subprime crisis, there had been an increasing number of people holding consumer credit and mortgages. Alternative financial services have also grown fast in recent years.[1] As a consequence, people need to make more financial decisions than before. Many financial products are complex and have high requirements on investors' financial sophistication. People's financial literacy, however, are normally poor.[2] Researchers cast doubt on the assumption that people can make financial decisions rationally, which stimulates studies on financial literacy. There are some comments that the spike of mortgage default rates in the subprime crisis are, to a large extent, attributed to individuals' poor financial literacy (Akerlof and Shiller, 2010; Boeri and Guiso, 2008). However, empirical evidence on the role of financial literacy in mortgage default is scarce.

In this paper we use unique proprietary data containing detailed information on borrower characteristics and mortgage origination and performance, and comprehensively analyze how financial literacy affects mortgage borrowers' repayment behavior. Financial literacy means "peoples' ability to process economic information and make informed decisions about financial planning, wealth accumulation, pensions, and debt" (Lusardi and Mitchell, 2013). We hypothesize that people with better financial literacy can make better financial planning such that their income and wealth is sufficient to pay their periodical debt services, so they are less likely to default.

To measure financial literacy, we construct a dummy variable based on whether the borrower works in the financial industry. We control for socioeconomic characteristics including education level and family income to address potential correlation between financial literacy and socioeconomic factors. We also consider mortgage contract terms and disentangle the effects of financial literacy on financial budgeting from the effects of financial literacy on loan choices. Additionally, we take

[1] See FINRA Investor Education Foundation (2009), Lusardi (2011), and Lusardi and Mitchell (2013).
[2] See the review by Agarwal et al. (2011), and Lusardi and Mitchell (2013).

into account the geographical effects and the property-type fixed effects. We find that the risk-free rate and interest rate spread for the high-literacy borrowers are about 1.5 basis points and 8.3 basis points respectively lower than those of the other borrowers, and the high-literacy borrowers are 21.1% less likely to have prepayment penalty.

As for the default, we find that borrowers from the financial industry are 14% less likely to default than borrowers from the non-financial industry. This lower likelihood to default cannot be explained by socioeconomic factors such as education and income, nor can it be explained by mortgage characteristics, location and origination year.

Another finding is that the effect of financial literacy on default changes when some loan characteristics or borrowers' demographic characteristics change. Financial literacy has a large impact on the repayment behavior of ARM borrowers, male borrowers, and borrowers between the ages of 35 and 50.

We compare the mark-to-market loan-to-value ratio (LTV) of high-literacy borrowers' delinquent mortgages and non-high-literacy borrowers' delinquent mortgages. We find no difference in the mark-to-market LTV and eliminate the possibility that the difference in the defaults rates of high and low financial literacy borrowers comes from their different incentives to strategically default.

We construct two other dependent variables for robustness tests. We compute the proportion of the time when a mortgage is in delinquency, and the ratio of the number of a borrower's missed payments to the number of the borrower's total payments. We get similar results using the different delinquency measures.

This paper contributes to the literature in the following aspects. First, many studies on mortgage default have emerged since the subprime crisis of 2007. These studies focus on the role of loan characteristics, trigger events like unemployment and divorce, equity position, debt-to-income ratio, credit history, local housing markets, and macroeconomic conditions in mortgage default.[3] However, there is little research on how financial literacy affects mortgage default. To our knowledge, there is only one paper on this issue. Gerardi, Goette, and Meier (2010) conducted a survey on the subprime mortgage borrowers in Massachusetts, Connecticut, and Rhode Island. They

[3] See the review by Jones and Sirmans (2014).

obtained information on about three hundred borrowers' financial literacy and matched this information to the mortgage data. They found that financial literacy has statistically significant correlation with the borrowers' default.

Their research, however, is constrained by the limitations of the survey data. First, the sample is too small – there are only 332 respondents. Properties of estimation based on this small sample are not as good as one based on a large sample. Also, the response rate is only 4.4%, which casts doubt on its representativeness. Meanwhile, in the sample, respondents and non-respondents in total only cover three states. If we only consider respondents, the regions covered by the data may be even fewer (this information was not revealed in the paper). This limited geographic coverage would further challenge its representativeness in that there may be heterogeneity across different regions. Additionally, their survey may lead to selection bias. For example, a person refuses to take an interview maybe because her hourly wage is high and the opportunity cost to take an interview is high. Therefore the survey may exclude some high-income people, which is implied by the larger loan size for the non-respondents according to the summary statistics of the paper. Moreover, the measurement error in the paper would affect the estimation results. For instance, high-income earners – who are likely to have higher levels of financial literacy – tend to underreport their income. As a consequence, the low default rates due to high income may be captured by the coefficient of the literacy variable in the regression, i.e., the estimated coefficient is upward biased.

We overcome these limitations with the proprietary data. We have hundreds of thousands of observations from 8,900 cities across fifty states in America; the proprietary data avoid the selection problem and measurement error, giving us access to richer and more accurate socioeconomic information, while being able to address potential endogeneity problems.

Second, we add to the existing literature on financial literacy and financial education programs. There is worldwide evidence that households' financial literacy is poor.[4] That challenges the rational individual assumption in conventional economic

[4] See Agarwal et al. (2010), Annamaria Lusardi and Mitchell (2009), Agnew et al. (2011), Lusardi and Mitchell (2011), Lusardi, Mitchell, and Curto (2010), Bateman et al. (2012),Banks and Oldfield (2007),

models and spurs innovation in both theoretical and empirical research. In theoretical research, some recent works incorporate financial literacy in the model of consumers' financial decision (Delavande, Rohwedder, and Willis, 2008; Jappelli and Padula, 2013; Hsu, 2011; Lusardi, Michaud, and Mitchell, 2013). On other hand, many studies provide empirical evidence on the association between financial literacy and household behavior. Several studies document the relation between financial literacy and individual's behavior on assets management.[5] At the same time, the effects of financial literacy on liabilities have also been discovered. Agarwal et al. (2009) find that people at different ages have different performance regarding interest rates of a wide range of financial products. Moore (2003), Lusardi and Tufano (2009), Lusardi and Scheresberg (2013), and Mottola (2013) argue that less financially literate borrowers are more likely to have high-cost debts; Agarwal and Mazumder (2013) find people with higher cognitive ability make less mistakes in using credit card and loan choice; Stango and Zinman (2009) conclude that less financially sophisticated borrowers have larger-size loans and less wealth. Bucks and Pence (2008) find that people are not aware of the specific contents of their loan contracts, particularly the interest rate. These studies only indirectly indicate that literacy may affect loan performance. Agarwal et al. (2011) review the literature on financial counseling and education. There is some evidence that financial education programs are correlated with improved financial decisions. However, no study gives strong evidence that the improved outcomes are due to better financial literacy obtained from the education programs. For example, Agarwal et al. (2014) study a mandatory mortgage counseling program and find that decline in default is because of the threat of oversight and the transaction cost of counseling instead of the information acquired during counseling. So this stream of literature still cannot confirm the relation between financial literacy and loan performance. Our paper directly documents the link between financial literacy and default, thus improving our understanding of this issue.

The rest of this article is organized as follows. Section 2 describes our data and

Behrman et al. (2012), Cole, Sampson, and Zia (2011), Hastings and Mitchell (2011), Lusardi and Tufano (2009), and McArdle, Smith, and Wills (2009).

[5] See Banks and Oldfield (2007), Christelis, Jappelli, and Padula (2010), Lusardi and Mitchell (2007), McArdle, Smith, and Wills (2009), and van Rooij, Lusardi, and Alessie (2011).

the methodology and Section 3 presents the results. Section 4 concludes the paper and discusses policy implications.

2. Data and Methodology

2.1. Data

The data used in this paper are obtained from one of the leading subprime lenders in the United States of America, which originated subprime loans all over the United States. The data consist of residential mortgage loans originated from 1996 to 2007 across 8,900 cities and 51 states in the United States. The data contain extensive information on borrower, loan, and property characteristics. First, it has detailed information on each borrower's demographic characteristics, financial status, and credit rating. We know the borrower's age, gender, marital status, years of residence, and employment status. We also have information on the co-borrower's education, gender, etc. We have the combined monthly income of the borrower and the co-borrower, which is more useful than only the borrower's income in predicting repayment behavior. Meanwhile, the data contain the debt-to-income ratio, which equals the total monthly housing expenses for the borrower and co-borrower combined divided by the total monthly income for the borrower and co-borrower combined, and times one hundred; this debt-to-income ratio measures the debt burden more accurately. We have FICO score, a proxy we normally use to measure credit quality. Most importantly, we know the borrower's occupation and we are going to use this information to proxy financial literacy. The details are explained in Section 2.2.

Second, we have detailed information on both loan origination and performance. For origination, we have information on the interest rate, the loan amount, the loan term, the combined LTV, the interest rate type (ARM or FRM), the lien status, the documentation type, the servicer's name, whether the loan has prepayment penalty, whether the loan is a subprime mortgage and the purpose of the loan (to purchase house or refinance existing mortgage). The data also keep a record of the performance history. We know, for each month, whether the borrower has payments overdue, as well as how many payments have been missed. Last, the data provide information on

property characteristics. The data contain the occupancy status, the property type, and the postal code of the property.

The proprietary data have several advantages. For one thing, the large sample of the data enables us to get consistent estimates, which is not feasible using survey data. Second, the data cover 8,900 cities and 51 states. Thus its representativeness is much better than the survey data and we can address geographical heterogeneity. In addition, this proprietary data can avoid the sample selection problem and measurement error in survey data, and generate unbiased estimator. Last, this dataset has rich and accurate information on different dimensions of borrowers' socioeconomic characteristics. This enables us to control for the variables correlated with financial literacy and make causality inference between financial literacy and loan performance. For example, this sample includes not only the borrower's income but also the co-borrower's income. The debt-to-income ratio is computed based on the total income and the expense of the borrower and the co-borrower. These variables are more accurate in measuring the borrowers' affordability to the loan and generate more accurate estimates.

2.2. Methodology

The main research question of the paper is how financial literacy affects default. We use the following logistic regression and least squares regression to test the relation between financial literacy and mortgage default:

$$Pr\,(D_i = 1) = 1 - 1/(1 + \exp(\alpha + \beta * L_i + \Sigma\lambda_{ij} * x_{ij} + \varepsilon_i)). \quad (1)$$

$$Y_i = \alpha + \beta * L_i + \Sigma\lambda_{ij} * x_{ij} + \varepsilon_i. \quad (2)$$

In the logistic regression (1), the dependent variable D is a dummy measuring mortgage delinquency. We define default as when a loan is in delinquency. For each month, we can identify whether a borrower defaults. In the baseline specification, we consider whether a borrower defaults during the course of the mortgage tenure, regardless of how many times the borrower defaults. Therefore, in our baseline model, the dependent variable D equals 1 if the borrower defaults at least once and 0 otherwise. In a robustness test we also construct continuous dependent variables based on the frequency of delinquency and the extent of delinquency, which are explicitly

introduced in Section 3.4. The corresponding least squares regression is equation (2).

The most important independent variable L is the measure of financial literacy. With the unique dataset we know whether or not a borrower works in the financial industry: the dataset includes a variable briefly describing the borrower's job; we can infer whether the borrower works in an industry related to finance. We define the financial industry as jobs related to the knowledge of interest rate computation, risk management, macroeconomics, financial planning or other aspects important to liabilities management. We take into account banking, insurance, fixed-income security, stock and derivatives investment, risk analysis and management, accounting, asset management, financial regulation, economists, professor of economics-related disciplines, and so on. People of these professions are more likely to have received good training on financial knowledge and skills, and hence, have higher financial literacy. We construct a dummy variable for financial literacy, which equals 1 if the borrower works in the financial industry and 0 otherwise. Note that we merely use this dummy to measure the borrower's financial literacy and we do not require the causality relation between the borrower's profession and financial literacy level. It is possible that some people are talented in finance and choose to join the financial industry. That possibility does not affect our results.

x_j comprises a set of controls from borrower characteristics to property characteristics. First, borrowers with better financial literacy are more likely to choose suitable mortgages. This is one way finance literacy decreases the probability of default (Moore, 2003; Lusardi and Tufano, 2009; Lusardi and Scheresberg, 2013; Mottola, 2013). Another mechanism is conditional on certain mortgage contract borrowers with better financial literacy being able to make better decisions on consumption and investment such that they are able to pay off the debts. We include mortgage term variables including the loan amount, the annual percentage interest rate, the adjustable rate mortgage (ARM) dummy, the first lien dummy, the interest-only loan dummy, the documentation status, the subprime dummy, the combined LTV, and the debt-to-income ratio in the regression to disentangle the second mechanism from the first one.

Second, we pick out the effects of education on repayment behavior. The

borrower's education level is absent in our dataset. Nevertheless, our dataset includes the co-borrower's years of schooling and we use it as the proxy for the borrower's education, as Mare (1991) points out that people with similar education level tend to marry each other. In our sample, 93.2% of the borrowers were married and the difference in the ages of the borrower and the co-borrower is less than ten years. We consider them couples. For the other 6.8%, the co-borrowers are most likely to be the borrowers' relatives or close friends who have close relationship with the borrowers. They tend to have similar education levels. In an unreported robustness test, we restrict our sample to those 93.2% for which we think the borrower and co-borrower are couples and get similar results.

Third, the borrower's profession is correlated with certain socioeconomic characteristics. For instance, borrowers from the financial industry may earn higher incomes which leads to lower default rates, and one may argue the effects of financial literacy on default are effectively the effects of income on repayment behavior. To address this issue, in the regression we include combined monthly income, including the income of both the borrower and co-borrower to measure the borrower's ability to pay off the debt more accurately.

Fourth, we also control for the geographic fixed effects and origination year. The financial industry may cluster in certain areas (such as New York City) and market conditions in these regions may be different from those in others. Additionally, credit supply may be different in different periods of time. In some years, lenders relax loan screening standard and mortgages originated in these years are more likely to enter delinquency (Mian and Sufi, 2009).

We also control for other variables associated with the borrower's repayment behavior: FICO score, the purpose of borrowing the mortgage, occupation status, property type, and servicer fixed effects. We further control for other demographic characteristics including gender, borrower age, marital status, borrower residence years, self-employed or not, and citizenship to further wipe out potential omitted variables problems.

In addition to default, we also study how financial literacy affects borrowers' loan choices:

$$C_i = \alpha + \beta * L_i + \Sigma\lambda_{ij} * x_{ij} + \varepsilon_i \quad (3)$$

or $$Pr\,(C_i = 1) = 1 - 1/(1 + \exp(\alpha + \beta * L_i + \Sigma\lambda_{ij} * x_{ij} + \varepsilon_i)). \quad (4).$$

The dependent variable C is the loan characteristics. It could be dichotomous or continuous. For example, it could be the interest rate.[6] Similar to equation (1) and (2), L is the financial literacy dummy and x_j comprises all the controls.[7] Equation (3) and (4) explore whether high financial literacy changes borrowers' loan choice behavior. In the case of interest rate, for instance, the coefficient of L indicates whether the high-financial literacy borrowers choose loans with lower interest rates.

2.3. Descriptive Statistics

Table 1 shows the summary statistics of the demographic characteristics, financial literacy dummy, loan origination, and performance information for the full sample.[8] 3.6% of the borrowers in our sample work in the financial industry. 98.9% of the loans are subprime loans, a source of the financial crisis in 2007. 71.3% of the borrowers are males, in part reflecting the gender disparity in socioeconomic status.

Then we divide the full sample into two subsamples based on the financial literacy dummy: one comprises borrowers working in the financial industry; the other includes borrowers working in the non-financial industry. Table 2 presents the summary statistics of these two subsamples. The default rate for the financial-industry borrowers is 9.7%, 15.5% lower than the default rate for the non-financial-industry borrowers. The difference in the default rates implies financial literacy may have an impact on default. The average FICO score is higher for the financial-industry borrowers. That is consistent with the correlation of financial literacy and FICO score in the literature: "households that can maintain a high FICO score show that they have the discipline and knowledge to plan their financial matters effectively" (Amromin et

[6] Other dependent variables of loan choice regression are introduced in details in section 3.1.

[7] x_j in equation (2) is slightly different from the one in equation (1): it does not include the dependent variable. For example, if the dependent variable in equation (2) is interest rate, then we have to exclude interest rate from x_j, while in equation (1) x_j always includes interest rate.

[8] We restrict the value of the variables within reasonable boundaries as follows: 0<borrower's age < 80; 0< borrower's residence years < 100, 1<debt-to-income ratio <100, 300< FICO score < 850, 0<co-borrower's years of schooling < 40, combined LTV >1, and combined monthly income < 30000.

al., 2011). The correlation indicates that borrowers' professions are correlated with financial literacy, supporting the effectiveness of our literacy dummy. The co-borrower's years of schooling, a proxy for the borrower's education, is similar across the two subsamples, suggesting that the difference in the default rates is not likely to come from difference in education. The two subsamples have different average combined income: borrowers from the financial industry earn 27.7% more than their counterparts from the non-financial industry. Almost all the loan characteristics are similar across these two subsamples, except that the financial-industry borrowers have bigger loan size, probably because of their higher income. We control for all these variables in our regression to eliminate the disturbances from these factors.

3. Empirical Results

3.1. Loan Choices

Before the analysis of mortgage borrowers' repayment behavior, we first explore how financial literacy affects their loan choices. Borrowers with higher financial literacy are expected to be smarter in loan choices, and they are more likely to have loans with lower cost. First we examine how financial literacy affects the interest rate spread of a loan. Interest rate spread is the difference between the total interest rate of a loan and the risk-free rate. It is the additional return required by the lender for bearing the borrower's risk, and determines the cost of the loan. We use the yield on U.S. Treasury securities at 30-year constant maturity as the risk-free rate, and subtract it from the note rate of a mortgage to get the spread. We also consider prepayment penalty. If a loan has prepayment penalty, the borrower would be charged a punitive fee if the loan is prepaid. Since the right to prepay is a put option and has value for the borrower, prepayment penalty decreases the value of the option and increases the loan cost. If borrowers from the financial industry perform better in loan choices, they are less likely to have prepayment penalty.

We use equation (2) to test our hypothesis. We use generalized least squares regression for spread and logistic regression for prepayment penalty. We control for all variables in Table 5, except interest rate for spread regression and prepayment

penalty, and prepayment penalty years for prepayment penalty regression. The results are in Table 3. Estimates for control variables are omitted. In Column (1), the spread for borrowers from the financial industry are 4.6 basis points lower than the spread for borrowers from non-financial industry. In Column (2), we report the odds ratio to display the magnitude of the effects of financial literacy.[9] Borrowers from the financial industry are 21.1% less likely to have prepayment penalty. The results indicate that borrowers with higher financial literacy tend to have low-cost loans and are consistent with our hypothesis.

Then we redefine high-literacy borrowers: instead of considering all people from the financial industry, we focus on the following people: senior staff (such as team leader, vice president or CEO), people with professional certificates (such as CFA or CPA), borrowers whose jobs require high financial knowledge level (such as analysts or financial planners), and people with loan-related jobs (such as mortgage brokers or loan officers). They tend to have higher financial literacy than the average level of the financial industry and should have much lower spread. We define these people as high-financial literacy borrowers and repeat the regression (1). The results are in Column (1), Table 4. We report the coefficients of FICO and combined LTV for comparison. The magnitude of the coefficient increases to 8.3 basis points, up from 4.6 basis points in Table 3. We then replace the dependent variable by risk-free rate and re-estimate the coefficient in Column (2), Table 4. The risk-free rate for high financial-literacy borrowers is 1.5 basis points lower than that of the other borrowers. The lower risk-free rate indicates that the high financial-literacy borrowers do a better job in timing the mortgage market: they are more likely to borrow mortgages when market interest rate is low. The magnitude of the coefficient (1.5 basis points) is not as large as that of the coefficient of spread, but much larger than that of the coefficient of FICO score and combined LTV, which are important factors in the mortgage pricing. This indicates that financial literacy has significant impact on mortgage borrowers' ability to timing the mortgage market. Besides, if we consider both the spread and the risk-free rate, the total interest rate for high financial-literacy borrowers is 9.8 basis

[9] The number of observations of Column (2) is lower than that of Column (1) because some MSAs do not have defaults.

points lower. Normally that increases the borrower's saving by ten thousand dollars.

One concern is there is an alternative explanation of the lower spread for the high-literacy borrowers: the loan officers know they are from the financial industry and are less risky, and hence offer them lower spreads. We cannot rule out this possibility and we do not make a strong claim that higher literacy make borrowers choose lower spreads. However, this alternative explanation does not affect our argument regarding the choice of prepayment penalty and market timing. In addition, the explanation that loan officers offer low spreads for lower risk of high-literacy borrowers supports the main idea of this paper that high financial literacy leads to low default risk.

3.2. The Baseline Result

To rigorously analyze how financial literacy affects borrowers' repayment behavior, we conduct logistic regression (1). We have rich information on borrower, loan, and property characteristics, and can disentangle the effects of financial literacy on repayment behavior from the effects of other factors. We can also test the extent to which this effect is realized by affecting borrowers' financial budgeting rather than loan choice.

First of all, we regress the default dummy on the financial literacy measure. We include FICO score, purpose of borrowing the loan, occupancy-status, and property type, servicer and origination year fixed effect. Column (1) of Table 5 presents the results. We report the odds ratio rather than the coefficient to measure the marginal effects of financial literacy. The effect of financial literacy dummy is negative and statistically significant at 1% level, and consistent with our hypothesis and what Table 2 suggests. Additionally, the magnitude of the effect is important: the odds ratio is about 0.87, meaning borrowers from the financial industry are about 13% less likely to default, holding everything else constant.

Next we take into account factors that are correlated with financial literacy and may affect borrowers' repayment behavior, to eliminate alternative explanations.

3.2.1. Mortgage Characteristics

Literature shows that individuals with better financial literacy choose loans with lower cost (Moore, 2003; Lusardi and Tufano, 2009; Lusardi and Scheresberg, 2013; Mottola. 2013). In Column (2) of Table 5, we include loan characteristics: the interest rate and a set of dummies for ARM, first-lien loans, prepayment penalty, interest-only loan, full documentation and subprime loans. Borrowers with lower financial literacy may not be good at controlling their debts lower than the affordable level, and they may choose larger loans (Stango and Zinman, 2009). Therefore in Column 3 of Table 5 we put in variables including the loan amount, LTV and the debt-to-income ratio.

The results indicate that the loan characteristics have statistically significant impacts on default. For instance, the borrowers who choose ARM are 7.4% more likely to default, consistent with the findings by Cunningham and Capone (2014). Borrowers who provide full documentation are about 17.2% less likely to default, as Jiang, Nelson, and Vytlacil (2014), and LaCour-Little and Yang (2013) find. Nevertheless, both the significance and the magnitude of the odds ratio of financial literacy do not change when we control for these loan characteristics, suggesting the effect of financial literacy on repayment behavior captured by the estimated odds ratio of the literacy dummy does not function through loan choice.

3.2.2. Education

The potential relation between education and repayment behavior may affect our causality inference between financial literacy and default. It is possible that people with better education are more likely to enter the financial industry, and higher education level makes them less likely to default. So the lower default rate of borrowers from the financial industry does not result from their better financial literacy but their better education. As discussed in Section 2.2, we proxy the borrower education by the co-borrower's years of schooling to address this issue.

The results are in Column (4) of Table 5. Education has a statistically significant impact on the borrowers' default, but the magnitude is very small, only 0.6%. Moreover, after we control for education, the effect of financial literacy on repayment behavior is still statistically significant and the magnitude of the odds ratio does not have any material change. Therefore, the relation between financial literacy and default we have identified is not the impact of education on repayment behavior.

3.2.3. Income

In Table 2 we find borrowers from the financial industry have higher average income. Higher income represents higher ability to repay debts. Thus, the financial-industry borrowers are less likely to default simply because they have higher income. We incorporate the natural logarithm of the combined monthly income to eliminate the effects of income. Column (5) of Table 5 presents the results. Income has no statistically significant effects on repayment behavior, suggesting that after considering debt-to-income ratio, the income *per se* is not an important predictor of default rate. More importantly, the estimates of financial literacy remain virtually unchanged. Borrowers from the financial industry are still about 14.1% less likely to default and the estimates are significant at 1% level. Thus, we can eliminate the possibility that the effect of financial literacy on default is from the correlation of financial literacy and income.

3.2.4. Local Real Estate Markets

Another concern is that in some areas, the finance industry may account for a larger portion of the local employment than in other areas because the finance industry has high spill-over effects and may cluster in some places. Real estate markets in different locations, which affect borrowers' repayment behavior, are partially separated and they are affected by local economic conditions and other geographic attributes. It is possible that the financial industry clusters in some areas where the real estate market condition is better, and borrowers from these areas have lower default rates. This negative correlation of the financial industry and default rates due to different real estate markets may contaminate our estimation.

In previous regression from Column (1) to Column (5) of Table 5 we have taken into account time-varying factors by allowing the standard errors to cluster at the MSA level.[10] To further address this issue, we control for MSA fixed effects in Column (6) of Table 5 to account for time-constant effects.[11] The impact of financial literacy on default is not affected by the incorporation of MSA fixed effects: the effect is still

[10] Throughout the paper, the standard errors are clustered at the MSA level unless otherwise stated.
[11] The number of observations of Column (6) is lower than those of others because some MSAs do not have defaults.

significant at 1% level and the magnitude of odds ratio is similar. Thus the impact of financial literacy on default is not related to the local real estate markets.

3.3. Heterogeneity of the Impact

We have displayed that better financial literacy reduces borrowers' default. In this section, we test how the effect differs depending on demographic and loan characteristics. We focus on lien status, ARM, borrower age, and gender. We divide the sample into different subsamples on these dimensions respectively, and compare the estimates across different subsamples.[12]

3.3.1. Lien Status

First, we divide our data into two subsamples, comprising the first-lien loans and the second-lien loans respectively. Then we conduct regression (1) using these two subsamples. Table 6 presents the results. For the second-lien loans (Column (1)), the effect of financial literacy on repayment behavior becomes insignificant, while the estimates remain significant for the first-lien loans (Column (2)). Additionally, the magnitude of the effect increases: the financial-industry borrowers are about 18.3% less likely to default as compared with the non-financial-industry borrowers. The effect is bigger than the one estimated using the full sample.

The intuition underlying the different estimates for the two subsamples is as follows. Many studies find that when borrowers encounter difficulties in repaying loans, they tend to default on first-lien loans, while maintaining repayment of their second-lien loans (Goodman et al., 2010; Jagtiani and Lang, 2010; Lee, Mayer, and Tracy, 2012). This finding is consistent with our estimated odds ratio of the first lien dummy in Column (6) of Table 5. The results indicate that financial literacy has greater impacts on the repayment of first-lien loans than that of second-lien loans. In Section 3.3.5, we further discuss this issue.

[12] Another way is to construct an interaction of literacy dummy and each of the demographic and loan characteristics variables. The features of our sample, however, do not allow us to use this method. For about 96% of the observations, the literacy dummy equals 0. Thus for the interaction of this literacy dummy to any other variable, at least 96% of the observations would be 0, and the interaction would be highly correlated with the literacy dummy, which leads multi-collinearity problem and make coefficient estimates insignificant. Therefore we do not use the interaction method.

3.3.2. ARM

Second, we test how the effect of financial literacy depends on the interest rate type. We divide the data into two subsamples, one comprising ARM and the other comprising fixed-rate mortgages (FRM), and estimate the impact of financial literacy using the two subsamples respectively. The results are in Table 7. The effect of financial literacy on repayment behavior remains significant for ARM, while the estimates become insignificant for FRM. Similarly, the magnitude of the effect is 19.4%, up from 14.4% using the full sample. One potential interpretation of the results is interest rates on ARM are floating and the monthly payment is difficult to predict. As a consequence, it requires better financial literacy to make a good repayment plan for ARM. Therefore, financial literacy has larger impacts on ARM borrowers than FRM borrowers.

3.3.3. Age

The relation between age and financial literacy has been widely studied. Theoretical research predicts that financial literacy follows a hump-shaped pattern during the life-cycle: Jappelli and Padula (2013) endogenize financial literacy in a multi-period life-cycle model, and find that households invest in financial literacy until retirement and after that financial literacy is falling. This finding is supported by empirical studies (Agarwal et al., 2007; Chen and Volpe, 1998; Lusardi, Mitchell, and Curto, 2010; Lusardi and Mitchell, 2011).

Our study also confirms this viewpoint. We divide our sample into three parts: borrowers younger than 35, between 35 and 50, and older than 50. Then we estimate regression (1) for the three subsamples respectively. Table 8 presents the results. For borrowers younger than 35 (Column (1)), the effect is significant and the magnitude of the effect is 19.7%, bigger than 14.4% for the full sample. The effect is even larger for borrowers older than 50 (Column (3)), which is 23.5%. For borrowers between 35 and 50, however, the estimate becomes insignificant. This result is consistent with the findings in previous studies. Young and old borrowers on average have lower financial literacy. The difference in financial literacy between borrowers from the financial industry and borrowers from the non-financial industry is bigger for young and old

borrowers than for middle-age borrowers. Therefore, the effect of financial literacy on default is greater for young and old people.

3.3.4. Gender

Borrower gender also affects the impact of financial literacy on default. We estimate equation (1) for male and female borrowers respectively and Table 9 displays the results. For female borrowers (Column (1)), the effect of financial literacy is insignificant. For male borrowers (Column (2)), this effect is still significant and its magnitude is 24.2%.

One possible interpretation of the results is that borrowing a mortgage to buy a house is an important decision. A borrower's action on the mortgage is influenced by other members in the family. Currently there is still a gender gap and males have higher social-economic status in family and society, so males are more influential in important decisions like financial budgeting. As a result, the improvement of financial literacy has bigger impacts on the repayment behavior of male borrowers than that of female borrowers.

3.3.5. Robustness Tests

There are several concerns for the analysis above. First, one may doubt that the estimates in some subsamples are insignificant merely because borrowers in those subsamples are more concentrated in the non-financial industry, so the literacy dummy does not have sufficient variation to estimate the effects. Table 10 lists the summary statistics of the literacy dummy for each subsample. We can see that the standard deviation of the literacy dummy, for those subsamples whose estimates are insignificant, are bigger or very close to those for the subsamples with significant estimates. For instance, the standard deviation of the literacy dummy for male borrowers is 0.027, while the one for female borrowers is 0.059. Therefore we can eliminate this possibility.

Another concern is that the assignment variables based on which we construct subsamples may correlated with each other. That correlation may make us obtain incorrect inferences from the results. For example, if most of the female borrowers in our sample are middle-age borrowers, then the insignificant estimate for the female

subsample results from age rather than gender. To solve this problem, we construct a correlation matrix for these variables in Table 11.[13] Almost all the correlation coefficients are smaller than 0.06. For the corresponding pairs, we can eliminate the possibility that one variable in a pair affects the estimates for subsamples constructed based on the other variable in the same pair.

The only exception is that the correlation coefficient between lien status and interest rate type is 0.6. We further estimate regression (1) using three subsamples: second-lien FRM, first-lien FRM, and first-lien ARM. Table 12 presents the results.[14] The impact of financial literacy is insignificant for the second-lien FRM (Column (1)). When we switch from the second lien FRM to the first lien FRM (Column (2)), the coefficient is insignificant. Then we switch from the first-lien FRM to the first-lien ARM, the coefficient becomes significant at 1% level. The result suggests that the difference in the impacts of financial literacy on default of first lien and second lien is from the interest rate type. Lien status has little impact on the effect of literacy on repayment behavior.

3.4. Strategic Default

A possible argument undermining our results is that the strategic default behavior of the high-literacy borrowers and the low-literacy borrowers is different, which leads to the default rates of the two types of borrowers. Strategic default occurs when the housing price is lower than the mortgage balance; the borrower's put option is in the money: the borrower can choose to foreclose the house rather than pay off the mortgage debt and save an amount equivalent to the difference between the loan balance and the housing price. To address this issue, we analyze the mark-to-market LTV of all delinquent mortgages. Mark-to-market LTV is the ratio of the loan balance to the market price of the house and reflects the borrowers' tendency to strategically default.

We extract all mortgages that have ever been delinquent, match their loan

[13] All the four variables are dummies. *First_lien* equals 1 if the loan is first lien and 0 otherwise; *ARM* equals 1 if the loan is ARM and 0 otherwise; *Age_d* equals 1 if borrower's age is between 35 and 50 and 0 otherwise; *Male* equals 1 if the borrower is male and 0 otherwise.

[14] The second-lien ARM subsample only has seven observations and is omitted.

characteristics information to their performance information of the months when they are delinquent, and get a panel dataset. First, we compare the mark-to-market LTV of high-literacy borrowers' loans and low-literacy borrowers' loans when they are delinquent. Table 13 shows the summary statistics of the mark-to-market LTV for the two groups of borrowers. The means for the two groups are similar and are far less than 100%, suggesting that the put option is out of the money. The pattern is confirmed in Figure 1 which plots the kernel density of the mark-to-market LTV of the two groups: for both groups, the mark-to-market LTV of almost all the loans are below 100%. Besides, the distributions of the mark-to-market LTV for the two groups are similar. Therefore there is little difference in the tendency to strategically default across the two groups.

To rigorously analyze the difference in the mark-to-market LTV, we apply equation (3) to the panel data, with the mark-to-market LTV as the dependent variable and using generalized least squares regression, to examine whether the mark-to-market LTV when a loan is delinquent is correlated to the financial literacy of the borrower. The results are in Table 14. The coefficient of the financial literacy dummy is insignificant, indicating that there is no statistically significant difference in the mark-to-market LTV between the two types of borrowers. Therefore, there is no difference in the tendency to strategically default.

3.5. Falsification

The key strategy of the paper is to assign higher financial literacy to people from the financial industry. In this section we conduct two falsifications: we change the way we assign higher financial literacy.

First, there are 4849 borrowers from the financial industry in our sample. Now we randomly choose 4849 observations in the sample and consider them high-literacy borrowers. Since the high financial literacy is randomly assigned, the high-literacy borrowers should not perform better than others. We call them pseudo-high-literacy borrowers. Then we repeat the regression (1) and (2) to see whether the pseudo-high-literacy borrowers have lower spreads and default rates. Since the assignment is

random, every time it repeats, the assigned observations change. To get a consistent result, we repeat the random assignment and regression thirty times. The results are in Table 15. For simplicity we only report the coefficients (for spread) or the odds ratio (for default) of the financial literacy dummy. In the thirty random assignments the pseudo-high-literacy borrowers do not have lower default rates; only in two cases, the pseudo-high-literacy borrowers have lower spreads. In combination, there is no case where the pseudo-high-literacy borrowers have lower spreads as well as lower default rates. Therefore, the pseudo-high-literacy borrowers do not perform better than the other borrowers.

Then we use an alternative strategy. The high financial literacy is assigned not to people from the financial industry but to other professions. We assign high financial literacy to engineers, and servicemen and police officers. Then we conduct regression (1) and (2). The results are in Table 16. The estimates of the pseudo financial literacy are not statistically significant in both spread regression (Panel A) and default regression (Panel B), for both the engineers, and the servicemen and police officers. The pseudo financial literacy has no impact on loan choices and defaults.

3.6. Alternative Delinquency Measures

The dependent variable we have used so far is a dummy variable indicating whether the borrower has defaulted or not. We also construct another two measures of mortgage delinquency as dependent variables. First, we define *delinq_ratio* as the fraction of time when the mortgage is in delinquency, i.e., the fraction of time when there are payments over due. Second, we define *missed_ratio* as the number of missed mortgage payments divided by the number of total payments. The difference between these two measures is that the first one only considers whether the mortgage is in delinquency in certain month, while the second one also considers the extent of delinquency (Gerardi, Goette, and Meier, 2010).

Table 17 presents the summary statistics for the two delinquency measures for borrowers from the financial industry and the non-financial industry respectively. The means of both *delinq_ratio* and *missed_ratio* are lower for borrowers from the financial industry, with similar standard deviations, suggesting that better financial

literacy may reduce the frequency of default.

Then we use the two delinquency measures as the dependent variable in (1) respectively to conduct generalized least-squares regression. Table 18 presents the results. Column (1) is the result for *delinq_ratio* and the coefficient of financial literacy is negative, suggesting that better financial literacy makes borrowers default less frequently by 9.0% (0.0026 divided by 0.029). The result is similar for *missed_ratio.* Therefore, the improvement of borrower repayment behavior due to higher financial literacy is robust when we use different delinquency measures.

4. Conclusion

We use a unique dataset including borrower, loan, and property characteristics, and analyze how subprime mortgage borrowers' financial literacy affects the default rates. We use borrowers' profession as the proxy of financial literacy. We find that borrowers working in the financial industry are 14.9% less likely to default than borrowers from the non-financial industry. The results support our hypothesis that better financial literacy reduces borrowers' default rates. We control for loan characteristics and find that the effect of financial literacy on default rates does not result from better loan choices; we also find the effect cannot be explained by education, a factor correlated with financial literacy, or geographic effects which may be correlated with the distribution of the financial industry. Additionally, we document the heterogeneity of the effect: it is bigger for first lien than second lien, male than female, and young and old borrowers than middle-age borrowers. We exclude the influence of strategic default. We construct two alternative delinquency measures as dependent variables and find the effect is robust.

This paper contributes to literature by using a unique dataset to overcome limitations in previous studies. It provides empirical evidence for the impact of financial literacy on borrower repayment behavior, identifies a source of the subprime loans delinquency, and fills in the research gaps in the literature of financial literacy.

Our results lead to several policy implications. We have documented that financial literacy has impact on borrowers' financial decisions, and we separated this

impact from the effects of financial literacy on borrowers' loan choice. The results indicate that it is insufficient to prevent borrowers from high-cost loans, such as the enactment of the anti-predatory lending law. In fact, Agarwal et al. (2014) argue there are some side effects of the anti-predatory lending policy such as market disruption. For these reasons, we need to provide financial education to change borrowers' repayment behavior and reduce defaults which have huge influences on the financial markets. Meanwhile, we have shown that the effect of financial literacy on default differs depending on loan and demographic characteristics. The heterogeneity of the effect tells us which group of people are the more important targets of financial education programs. For instance, we need to pay more attention to improving the financial literacy of young and old people because these two groups of people have lower literacy levels than middle-age people, and financial education has bigger impacts on them. Further investigation in this would generate helpful conclusions and we leave this issue for future research.

References

Agarwal, Sumit, Gene Amromin, Itzhak Ben-David, Souphala Chomsisengphet, and Douglas Evanoff, 2014, The Effectiveness of Mandatory Mortgage Counseling: Can One Dissuade Borrowers from Choosing Risky Mortgages?., *National Bureau of Economic Research Working Paper*.

Agarwal, Sumit, Gene Amromin, Itzhak Ben-david, Souphala Chomsisengphet, and Douglas D Evanoff, 2011, Financial Counseling , Financial Literacy , and Household Decision Making, *Financial Literacy: Implications for Retirement Security and the Financial Marketplace*.

Agarwal, Sumit, Gene Amromin, Itzhak Ben-David, Souphala Chomsisengphet, and Douglas D. Evanoff, 2010, Financial Literacy and Financial Planning: Evidence from India, *Available at SSRN 1728831*.

Agarwal, Sumit, Gene Amromin, Itzhak Ben-David, Souphala Chomsisengphet, and Douglas D. Evanoff, 2014, Predatory Lending and the Subprime Crisis, *Journal of Financial Economics* 113, 29–52.

Agarwal, Sumit, John C Driscoll, Xavier Gabaix, and David Laibson, 2009, The Age of Reason : Financial Decisions Over the Lifecycle, *Brookings Papers on Economic Activity*, 51–101.

Agarwal, Sumit, and Bhashkar Mazumder, 2013, Cognitive Abilities and Household Financial Decision Making, *American Economic Journal: Applied Economics* 5, 193–207.

Agnew, Julie R., Lisa R. Szykman, Stephen P. Utkus, and Jean a. Young, 2011, Trust, plan knowledge and 401(k) savings behavior, *Journal of Pension Economics and Finance* 11, 1–20.

Akerlof, George A., and Robert J. Shiller., 2010, Animal spirits: How human psychology drives the economy, and why it matters for global capitalism., *Princeton University Press*.

Amromin, Gene, Jennifer Huang, Clemens Sialm, and Edward Zhong., 2011, Complex Mortgages, *National Bureau of Economic Research Working Paper* No. 17315.

Annamaria Lusardi, and Olivia S Mitchell, 2009, How Ordinary Consumers Make Complex Economic Decisions: Financial Literacy and Retirement Readiness, *National Bureau of Economic Research Working Paper* No.15350.

Banks, James, and Zo ë Oldfield, 2007, Understanding Pensions: Cognitive Function, Numerical Ability and Retirement Saving, *Fiscal Studies* 28, 143–170.

Bateman, Hazel, Christine Eckert, John Geweke, Jordan Louviere, Susan Thorp, and Stephen Satchell, 2012, Financial Competence and Expectations Formation: Evidence from Australia*, *Economic Record* 88, 39–63.

Behrman, Jere R, Olivia S Mitchell, Cindy K Soo, and David Bravo, 2012, How Financial Literacy Affects Household Wealth Accumulation, *American Economic Review* 102, 300–304.

Boeri, Tito, and Luigi Guiso, 2008, "The subprime crisis: Greenspan's legacy.," *The first global financial crisis of the 21st century 73* 73, 37.

Bucks, Brian, and Karen Pence, 2008, Do borrowers know their mortgage terms?, *Journal of Urban Economics* 64, 218–233.

Chen, Haiyang, and Ronald P. Volpe, 1998, An Analysis of Personal Financial Literacy among College Students, *Financial services review* 7, 107–128.

Christelis, Dimitris, Tullio Jappelli, and Mario Padula, 2010, Cognitive abilities and portfolio choice, *European Economic Review* 54, 18–38.

Cole, Shawn, Thomas Sampson, and Bilal Zia, 2011, Prices or Knowledge ? What Drives Demand for Financial Services in Emerging Markets ?, 66, 1933–1967.

Cunningham, Donald F., and Charles A. Capone., 2014, The Relative Termination Experience of Adjustable to Fixed-Rate Mortgages, *The Journal of Finance* 45, 1687–1703.

Delavande, Adeline, Susann Rohwedder, and Robert J. Willis, 2008, Preparation for retirement, financial literacy and cognitive resources, *Michigan Retirement Research Center Research Paper* 2008-190.

Foundation, FINRA Investor Education, 2009, Financial Capability in the United States., *National Survey - Executive Summary* December 2.

Gerardi, Kristopher, Lorenz Goette, and Stephan Meier, 2010, Financial Literacy and Subprime Mortgage Delinquency: Evidence from a Survey Matched to Administrative Data, *Working Paper, Federal Reserve Bank of Atlanta*.

Goodman, Laurie S., Roger Ashworth, Brian Landy, and Ke Yin, 2010, Second liens: how important?, *The Journal of Fixed Income* 20, 1930.

Hastings, Justine S, and Olivia S Mitchell, 2011, How Financial Literacy and Impatience Shape Retirement Wealth and Investment Behaviors, *National Bureau of Economic Research Working Paper* No.16740.

Hsu, Joanne W, 2011, Aging and Strategic Learning: the Impact of Spousal Incentives on Financial Literacy, *Division of Research & Statistics and Monetary Affairs, Federal Reserve Board*.

Jagtiani, Julapa, and William W. Lang, 2010, Strategic default on first and second lien mortgages during the financial crisis. No. 11-3. 2010., *Federal Reserve Bank of Philadelphia* 3.

Jappelli, Tullio, and Mario Padula, 2013, Investment in financial literacy and saving decisions, *Journal of Banking & Finance* 37, 2779–2792.

Jiang, Wei, Ashlyn Aiko Nelson, and Edward Vytlacil, 2014, Liar's Loan? Effects of Origination Channel and Information Falsification on Mortgage Delinquency., *The Review of Economics and Statistics* 96, 1–18.

Jones, Tim, and G. Stacy Sirmans, 2014, Mortgage Default and Foreclosure : An Analysis of the Underlying Determinants , Associated Losses , and Valuation, *American Real Estate Society Working Paper*.

LaCour-Little, Michael, and Jing Yang, 2013, Taking the Lie out of Liar Loans: the Effect of Reduced Documentation on the Performance and Pricing of Alt-A and Subprime Mortgages, *Journal of Real Estate Research* 35, 507–553.

Lee, Donghoon, Christopher J Mayer, and Joseph Tracy, 2012, A New Look at Second Liens, *National Bureau of Economic Research* No. w18269.

Lusardi, Annamaria, 2011, Americans' financial capability, *National Bureau of Economic Research Working Paper* No.17103.

Lusardi, Annamaria, Pierre-Carl Michaud, and Olivia S. Mitchell, 2013, Optimal Financial Knowledge and Wealth Inequality, *National Bureau of Economic Research Working Paper* No. 18669.

Lusardi, Annamaria, and Olivia S Mitchell, 2013, The Economic Importance of Financial Literacy: Theory and Evidence, *National Bureau of Economic Research Working Paper* No.18952.

Lusardi, Annamaria, and Olivia S. Mitchell, 2007, Baby Boomer retirement security: The roles of planning, financial literacy, and housing wealth, *Journal of Monetary Economics* 54, 205–224.

Lusardi, Annamaria, and Olivia S. Mitchell, 2011, Financial literacy around the world: an overview, *Journal of Pension Economics and Finance* 10, 497–508.

Lusardi, Annamaria, Olivia S. Mitchell, and Vilsa Curto, 2010, Financial Literacy among the Young, *Journal of Consumer Affairs* 44, 358–380.

Lusardi, Annamaria, and Carlo de Bassa Scheresberg, 2013, Financial literacy and high-cost borrowing in the United States, *National Bureau of Economic Research Working Paper* No.18969.

Lusardi, Annamaria, and Peter Tufano, 2009, Debt literacy, financial experiences, and overindebtedness, *National Bureau of Economic Research Working Paper* No. 14808.

Mare, Robert D, 1991, Five decades of educational assortative mating., *American Sociological Review* 56, 15–32.

McArdle, John J., James P. Smith Smith, and Robert Wills, 2009, Cognition and Economic Outcomes in the Health and Retirement Survey, *National Bureau of Economic Research Working Paper* No. 15266.

Mian, Atif, and Amir Sufi, 2009, The Consequences of Mortgage Credit Expansion: Evidence from the US Mortgage Default Crisis., *The Quarterly Journal of Economics* No. 4, 1449–1496.

Moore, Danna L., 2003, Survey of financial literacy in Washington State: Knowledge, behavior, attitudes, and experiences, *Washington State Department of Financial Institutions*.

Mottola, Gary R, 2013, In Our Best Interest: Women, Financial Literacy, and Credit Card Behavior, *Numeracy* 6.

Rooij, Maarten van, Annamaria Lusardi, and Rob Alessie, 2011, Financial literacy and stock market participation, *Journal of Financial Economics* 101, 449–472.

Stango, Victor, and Jonathan Zinman, 2009, Exponential growth bias and household finance, *The Journal of Finance* 64, 2807–2849.

Table 1 Summary Statistics

This table presents the summary statistics of the full sample. We extract the performance information of loans and match this information to the origination data, creating a cross-sectional dataset. We restrict the value of the variables within reasonable boundaries as introduced in Section 2.3.

Variable	Obs	Mean	Std. Dev.	Min	Max
Default dummy	136619	0.111	0.314	0	1
Literacy dummy	136619	0.036	0.186	0	1
Interest rate	136619	9.658	1.596	3.000	23.472
Term	136619	352.993	32.143	120	480
Loan amount	136619	192250.5	140080.7	6350	1400000
Combined LTV	136619	87.995	13.900	4.546	121.818
Debt-to-income ratio	136619	28.027	9.617	1.001	86.670
ARM	136619	0.600	0.490	0	1
First lien	136619	0.803	0.398	0	1
Prepayment penalty	136619	0.708	0.455	0	1
Prepayment penalty years	136619	1.606	1.147	0	5
Interest only	136619	1.155	0.362	1	2
Full documentation	136619	0.706	0.456	0	1
Brokered Loan	136619	0.879	0.326	0	1
Subprime	136619	0.989	0.105	0	1
FICO score	136619	621.786	58.498	400	824
Purpose: refinance=1	136619	0.587	0.492	0	1
Loan age	136619	7.343	8.525	1	109
Owner occupied	136619	0.954	0.210	0	1
Co-borrower school year	136619	12.870	3.403	0	39
Combined income	136619	7607.403	4057.861	537	29999
Male	136619	0.713	0.453	0	1
Minority	136619	0.277	0.448	0	1
Borrower age	136619	41.543	10.548	18	79
Married	136619	0.954	0.210	0	1
Borrower residence years	136619	5.956	6.264	0	99
Self employed	136619	0.189	0.392	0	1
Citizen	136619	0.960	0.195	0	1

Table 2 Comparison between Borrowers from the Financial and the Non-financial Industry

This table compares the summary statistics of loans whose borrowers have different financial literacy. Panel A is for borrowers from the non-financial industry, while Panel B is for borrowers from the financial industry.

Panel A The non-financial industry

Variable	Obs	Mean	Std. Dev.	Min	Max
Default dummy	131738	0.112	0.315	0	1
Interest rate	131738	9.661	1.595	3	23.472
Term	131738	352.998	32.148	120	480
Loan amount	131738	191443.8	139404.5	9747	1400000
Combined LTV	131738	87.941	13.927	4.546	121.818
Debt-to-income ratio	131738	28.068	9.615	1.001	86.670
ARM	131738	0.599	0.490	0	1
First lien	131738	0.804	0.397	0	1
Prepayment penalty	131738	0.709	0.454	0	1
Prepayment penalty years	131738	1.609	1.146	0	5
Interest only	131738	1.154	0.361	1	2
Full documentation	131738	0.708	0.455	0	1
Brokered Loan	131738	0.879	0.326	0	1
Subprime	131738	0.989	0.105	0	1
FICO score	131738	621.480	58.450	400	821
Purpose: refinance=1	131738	0.588	0.492	0	1
Loan age	131738	7.329	8.492	1	109
Owner occupied	131738	0.955	0.207	0	1
Co-borrower school year	131738	12.854	3.396	0	39
Combined income	131738	7542.680	4000.033	537.21	29999
Male	131738	0.719	0.449	0	1
Minority	131738	0.277	0.447	0	1
Borrower age	131738	41.565	10.529	18	79
Married	131738	0.954	0.209	0	1
Borrower residence years	131738	5.971	6.268	0	99
Self employed	131738	0.187	0.390	0	1
Citizen	131738	0.960	0.196	0	1

Panel B The Financial Industry

Variable	Obs	Mean	Std. Dev.	Min	Max
Default dummy	4881	0.097	0.296	0	1
apr	4881	9.582	1.624	4.21	15.691
term	4881	352.870	32.011	120	480
Loan amount	4881	214022.1	155681.4	6350	1000000
Combined LTV	4881	89.452	13.061	8.293	104.957
Debt-to-income ratio	4881	26.939	9.603	1.692	70.538
ARM	4881	0.610	0.488	0	1
First lien	4881	0.781	0.413	0	1
Prepayment penalty	4881	0.684	0.465	0	1
Prepayment penalty years	4881	1.526	1.161	0	5
Interest only	4881	1.181	0.385	1	2
Full documentation	4881	0.654	0.476	0	1
Brokered Loan	4881	0.877	0.329	0	1
Subprime	4881	0.991	0.092	0	1
FICO score	4881	630.029	59.210	444	824
Purpose: refinance=1	4881	0.544	0.498	0	1
Loan age	4881	7.728	9.368	1	95
Owner occupied	4881	0.920	0.272	0	1
Co-borrower school year	4881	13.308	3.552	0	24
Combined income	4881	9354.283	5087.580	1086	29904.15
Male	4881	0.535	0.499	0	1
Minority	4881	0.286	0.452	0	1
Borrower age	4881	40.942	11.049	19	79
Married	4881	0.941	0.235	0	1
Borrower residence years	4881	5.546	6.134	0	54
Self employed	4881	0.240	0.427	0	1
Citizen	4881	0.974	0.158	0	1

Table 3 Loan Choice I

This table presents the regression of loan characteristics on financial literacy and other information using the full sample. In Column (1), the dependent variable is the spread of the loan and we use generalized least squares regression. In Column (2), the dependent variable is an indicator variable that equals 1 if the loan has prepayment penalty and 0 otherwise, and we use logistic regression. We report coefficients for Column (1) and odds ratio for Column (2) to measure the marginal effects. The controls variables comprise all controls in Table 5, except interest rate for Column (1) and prepayment penalty and prepayment penalty years for Column (2). Robust standard errors are clustered at the MSA level. Robust z-statistics are reported in the parentheses. *, **, and *** indicate the estimates are significant at 10%, 5%, and 1% levels respectively.

Dependent Variable	spread	prepayment penalty dummy
	(1)	(2)
Literacy Dummy (Coefficient/odds ratio)	-0.046***	0.789***
	(-3.35)	(-3.42)
Control Variables	Yes	Yes
State fixed effects	Yes	Yes
Constant	Yes	Yes
Property type fixed effects	Yes	Yes
Origination year fixed effects	Yes	Yes
Servicer fixed effects	Yes	Yes
R-squared/Pseudo R-squared	0.866	0.422
Observations	136619	133135

Table 4 Loan Choice II

This table presents the regression of spread and risk-free rate on the redefined financial literacy dummy: within the financial industry, we consider only senior staff, people with professional certificates, people whose jobs require high financial knowledge level, and people with loan-related jobs as high-literacy borrowers. Then we apply generalized least squares regression. The control variables comprise all controls in Table 5, except the interest rate. Robust standard errors are clustered at the MSA level. Robust z-statistics are reported in the parentheses. *, **, and *** indicate the estimates are significant at 10%, 5%, and 1% levels respectively.

Dependent Variable	spread	risk-free rate
	(1)	(2)
Redefined Literacy Dummy	-0.083***	-0.015**
	(-3.32)	(-2.14)
FICO score	-0.0096***	0
	(-34.51)	(1.04)
Combined LTV	0.0089***	-0.0004***
	(13.7)	(-3.97)
Control Variables	Yes	Yes
State fixed effects	Yes	Yes
Constant	Yes	Yes
Property type fixed effects	Yes	Yes
Origination year fixed effects	Yes	Yes
Servicer fixed effects	Yes	Yes
R-squared	0.866	0.607
Observations	136619	136619

Table 5 Baseline Results

This table presents the logistic regression of the default dummy on financial literacy and other information using the full sample. The dependent variable is an indicator variable that equals 1 if the borrower defaults at least once and 0 otherwise. We report odds ratio of each variable to measure the marginal effects. Robust standard errors are clustered at the MSA level. Robust z-statistics are reported in the parentheses. *, **, and *** indicate the estimates are significant at 10%, 5%, and 1% levels respectively.

Dependent Variable: Default dummy						
	(1)	(2)	(3)	(4)	(5)	(6)
Literacy dummy	0.870**	0.870**	0.874**	0.875**	0.859**	0.856**
	(-2.27)	(-2.22)	(-2.17)	(-2.15)	(-2.52)	(-2.55)
FICO score	0.987***	0.987***	0.986***	0.986***	0.986***	0.986***
	(-56.23)	(-52.32)	(-57.79)	(-57.88)	(-56.91)	(-56.02)
Purpose: refinance=1	0.660***	0.727***	0.871***	0.870***	0.901***	0.908***
	(-16.73)	(-11.75)	(-4.48)	(-4.49)	(-3.45)	(-3.12)
Loan age	1.150***	1.165***	1.166***	1.166***	1.167***	1.168***
	(41.79)	(36.48)	(36.23)	(36.27)	(36.18)	(36.02)
Owner occupied	0.828***	0.933	0.770***	0.769***	0.779***	0.776***
	(-3.57)	(-1.39)	(-5.24)	(-5.26)	(-4.75)	(-4.83)
Interest rate		1.220***	1.203***	1.204***	1.205***	1.209***
		(15.56)	(12.86)	(12.91)	(12.56)	(12.12)
Term		1.003***	1.002***	1.002***	1.002***	1.002***
		(5.29)	(4.21)	(4.19)	(3.96)	(3.76)
ARM		1.074*	1.004	1.003	0.986	1.001
		(1.65)	(0.08)	(0.07)	(-0.30)	(0.03)
First lien		0.984	1.264***	1.255***	1.319***	1.262***
		(-0.30)	(4.13)	(4.02)	(4.38)	(3.01)
Prepayment penalty		0.987	0.97	0.969	0.971	0.952
		(-0.38)	(-0.83)	(-0.86)	(-0.84)	(-1.19)
Interest only		1.039	0.988	0.988	0.984	0.988
		(1.09)	(-0.35)	(-0.37)	(-0.47)	(-0.39)
Full documentation		0.728***	0.672***	0.674***	0.663***	0.665***
		(-10.90)	(-14.27)	(-14.20)	(-13.86)	(-13.32)
Brokered loans		0.836***	0.825***	0.826***	0.818***	0.828***
		(-4.39)	(-4.49)	(-4.47)	(-4.74)	(-4.01)
Subprime		1.691	1.685	1.699	1.732	0.966
		(1.57)	(1.56)	(1.59)	(1.64)	(-0.06)
Log loan amount			0.949*	0.954*	0.933	0.956
			(-1.86)	(-1.69)	(-1.47)	(-0.82)
Combined LTV			1.018***	1.018***	1.016***	1.016***
			(18.56)	(18.51)	(15.76)	(14.46)
Debt-to-income ratio			1.014***	1.014***	1.014***	1.015***
			(8.98)	(8.94)	(6.38)	(6.73)

Table 5 (continued)

	(1)	(2)	(3)	(4)	(5)	(6)
Co-borrower school year				0.994**	0.993**	0.992***
				(-1.96)	(-2.42)	(-2.60)
Log combined income					1.035	1.033
					(0.63)	(0.59)
Male					0.957**	0.958*
					(-1.97)	(-1.84)
Minority					1.176***	1.138***
					(4.58)	(3.82)
Borrower age					0.992***	0.992***
					(-7.10)	(-6.81)
Married					1.103*	1.104*
					(1.85)	(1.83)
Borrower residence year					0.991***	0.991***
					(-4.31)	(-4.16)
Self employed					0.974	0.979
					(-0.94)	(-0.73)
Citizen					1.322***	1.325***
					(4.37)	(4.43)
MSA fixed effects	No	No	No	No	No	Yes
Constant	Yes	Yes	Yes	Yes	Yes	Yes
Property type fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Origination year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Servicer fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R-squared	0.213	0.225	0.229	0.229	0.231	0.236
Observations	136619	136619	136619	136619	136619	135440

Table 6 Lien Status

This table presents the logistic regression of the default dummy on financial literacy and other information. We divide our sample into two subsamples based on lien status and run the regression for each sample. Column (1) is for second lien and Column (2) is for first lien. We report odds ratio of each variable to measure the marginal effects. Robust standard errors are clustered at the MSA level. Robust z-statistics are reported in the parentheses. *, **, and *** indicate the estimates are significant at 10%, 5%, and 1% levels respectively.

Dependent variable: Default dummy		
	(1) second lien	(2) first lien
Literacy dummy	1.090 (0.66)	0.817*** (-3.27)
Control Variables	Yes	Yes
State fixed effects	Yes	Yes
Constant	Yes	Yes
Property type fixed effects	Yes	Yes
Origination year fixed effects	Yes	Yes
Servicer fixed effects	Yes	Yes
Pseudo R-squared	0.229	0.241
Observations	26200	108607

Table 7 Adjustable-Rate Mortgages

This table presents the logistic regression of the default dummy on financial literacy and other information. We divide our sample into two subsamples based on the interest rate type and run the regression for each sample. Column (1) is for FRM and Column (2) is for ARM. We report odds ratio of each variable to measure the marginal effects. Robust standard errors are clustered at the MSA level. Robust z-statistics are reported in the parentheses. *, **, and *** indicate the estimates are significant at 10%, 5%, and 1% levels respectively.

Dependent variable: Default dummy		
	(1)	(2)
	FRM	ARM
Literacy dummy	0.969	0.806***
	(-0.34)	(-3.12)
Control Variables	Yes	Yes
State fixed effects	Yes	Yes
Constant	Yes	Yes
Property type fixed effects	Yes	Yes
Origination year fixed effects	Yes	Yes
Servicer fixed effects	Yes	Yes
Pseudo R-squared	0.239	0.235
Observations	53178	81625

Table 8 Borrower Age

This table presents the logistic regression of the default dummy on financial literacy and other information. We divide our sample into three subsamples and run the logistic regression: borrowers younger than 35 (Column (1)), borrowers between 35 and 50 (Column (2)), and borrowers older than 50 (Column (3)). We report odds ratio of each variable to measure the marginal effects. Robust standard errors are clustered at the MSA level. Robust z-statistics are reported in the parentheses. *, **, and *** indicate the estimates are significant at 10%, 5%, and 1% levels respectively.

Dependent variable: Default dummy			
	(1)	(2)	(3)
	age <35	35<=age<=50	age>50
Literacy dummy	0.803**	0.932	0.765*
	(-1.96)	(-0.83)	(-1.76)
Control Variables	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes
Constant	Yes	Yes	Yes
Property type fixed effects	Yes	Yes	Yes
Origination year fixed effects	Yes	Yes	Yes
Servicer fixed effects	Yes	Yes	Yes
Pseudo R-squared	0.257	0.24	0.232
Observations	38272	68575	27271

Table 9 Gender

This table presents the logistic regression of the default dummy on financial literacy and other information. We divide our sample into two subsamples based on borrower gender and run the regression for each sample. Column (1) is for female and Column (2) is for male. We report odds ratio of each variable to measure the marginal effects. Robust standard errors are clustered at the MSA level. Robust z-statistics are reported in the parentheses. *, **, and *** indicate the estimates are significant at 10%, 5%, and 1% levels respectively.

Dependent variable: Default dummy		
	(1) female	(2) male
Literacy dummy	0.980 (-0.27)	0.758*** (-2.95)
Control Variables	Yes	Yes
State fixed effects	Yes	Yes
Constant	Yes	Yes
Property type fixed effects	Yes	Yes
Origination year fixed effects	Yes	Yes
Servicer fixed effects	Yes	Yes
Pseudo R-squared	0.236	0.241
Observations	37735	96869

Table 10 The Summary Statistics of Literacy Dummy

This table presents the summary statistics of the literacy dummy for the subsamples we use from Section 3.3.1 to Section 3.3.4. We divide the full sample into subsamples base on lien status, the interest rate type, borrower age and gender respectively, and compute the summary statistics for each subsample.

Variable: Literacy dummy					
Subsample	Obs	Mean	Std. Dev.	Min	Max
Second lien	26200	0.040	0.196	0	1
First lien	108607	0.035	0.183	0	1
FRM	53178	0.035	0.184	0	1
ARM	81625	0.036	0.187	0	1
Age <35	38272	0.041	0.198	0	1
35<=Age<=50	68575	0.032	0.177	0	1
50<Age	27271	0.037	0.189	0	1
Female	37735	0.059	0.235	0	1
Male	96869	0.027	0.161	0	1

Table 11 Correlation Matrix

This table presents the correlation matrix of those variables based on which the subsamples from Section 3.3.1 to Section 3.3.4 are generated. All the four variables are dummies. *First_lien* equals 1 if the loan is first lien and 0 otherwise; *ARM* equals 1 if the loan is ARM and 0 otherwise; *Age_d* equals 1 if borrower's age is between 35 and 50 and 0 otherwise; *Male* equals 1 if the borrower is male and 0 otherwise.

	First_lien	ARM	Age_d	Male
First_lien	1			
ARM	0.606	1		
Age_d	-0.055	-0.031	1	
Male	0.013	0.010	-0.026	1

Table 12 ARM and First Lien

This table presents the logistic regression of the default dummy on financial literacy and other information. We divide our sample into three subsamples based on the interest rate type and lien status, and run the regression for each subsample. Column (1) is for second-lien FRM; Column (2) is for first-lien FRM; Column (3) is for first-lien ARM. We report odds ratio of each variable to measure the marginal effects. Robust standard errors are clustered at the MSA level. Robust z-statistics are reported in the parentheses. *, **, and *** indicate the estimates are significant at 10%, 5%, and 1% levels respectively.

Dependent variable: Default dummy			
	(1)	(2)	(3)
	second-lien FRM	first-lien FRM	first-lien ARM
Literacy dummy	1.090	0.836	0.806***
	-0.66	(-1.23)	(-3.12)
Control Variables	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes
Constant	Yes	Yes	Yes
Property type fixed effects	Yes	Yes	Yes
Origination year fixed effects	Yes	Yes	Yes
Servicer fixed effects	Yes	Yes	Yes
Pseudo R-squared	0.229	0.278	0.235
Observations	26200	26145	81625

Table 13.Compare Mark-to-Market LTV

This table compares the summary statistics of the mark-to-market LTV of financial industry borrowers and non-financial industry borrowers when loans are delinquent. Since the loans can be delinquent in multiple periods, the dataset is a panel dataset.

Subsample	Obs	Mean	Std. Dev.	Min	Max
Non-financial industry	2018	63.900	23.533	0	97.315
Financial Industry	63897	67.139	21.040	0	104.679

Table 14 Strategic Default Incentives

This table presents the generalized least squares regression of the mark-to-market LTV of all delinquent mortgages on financial literacy and other information. Robust standard errors are clustered at the MSA level. Robust z-statistics are reported in the parentheses. *, **, and *** indicate the estimates are significant at 10%, 5%, and 1% levels respectively.

Dependent variable	Mark-to-Market LTV
	(1)
Literacy dummy	-0.032
	(-0.07)
Control Variables	Yes
State fixed effects	Yes
Constant	Yes
Property type fixed effects	Yes
Origination year fixed effects	Yes
Servicer fixed effects	Yes
Pseudo R-squared	0.947
Observations	65915

Table 15 Falsification I

We randomly assign high financial literacy to observations and run regression in equation (1) and (2), using default dummy and spread as dependent variable respectively. For simplicity we only report the coefficients (for spread) or the odds ratio (for default) of the financial literacy dummy. Robust z-statistics are reported in the parentheses. *, **, and *** indicate the estimates are significant at 10%, 5%, and 1% levels respectively.

	(1)	(2)	(3)	(4)	(5)
Spread (Coefficient)	-0.0012	0.0104	0.0024	0.0006	-0.0116
	(-0.10)	(0.98)	(0.21)	(0.05)	(-0.91)
Default dummy (Odds ratio)	0.965	1.057	1.120**	1.016	1.041
	(-0.59)	(0.97)	(2.36)	(0.35)	(0.78)

	(6)	(7)	(8)	(9)	(10)
Spread (Coefficient)	-0.0218*	0.0097	-0.01	-0.0061	0.0001
	(-1.78)	(0.64)	(-0.78)	(-0.49)	(0.01)
Default dummy (Odds ratio)	1.029	1.049	1.026	1.108*	0.971
	(0.46)	(0.89)	(0.57)	(1.85)	(-0.60)

	(11)	(12)	(13)	(14)	(15)
Spread (Coefficient)	0.0061	0.0054	-0.0173	0.0035	0.0006
	(0.49)	(0.43)	(-1.63)	(0.30)	(0.05)
Default dummy (Odds ratio)	0.969	1.013	1.032	0.959	0.956
	(-0.66)	(0.30)	(0.66)	(-0.87)	(-0.77)

	(16)	(17)	(18)	(19)	(20)
Spread (Coefficient)	-0.0026	0.0077	-0.0184*	0.0009	0.0156
	(-0.19)	(0.61)	(-1.71)	(0.07)	(1.28)
Default dummy (Odds ratio)	0.98	1	1.024	1.082	1.063
	(-0.44)	(-0.01)	(0.52)	(1.64)	(1.17)

	(21)	(22)	(23)	(24)	(25)
Spread (Coefficient)	0.0104	0.0039	0.0347***	-0.0035	-0.0186
	(0.79)	(0.27)	(3.35)	(-0.31)	(-1.52)
Default dummy (Odds ratio)	1.032	0.989	1.01	0.984	0.924
	(0.63)	(-0.20)	(0.20)	(-0.28)	(-1.50)

	(26)	(27)	(28)	(29)	(30)
Spread (Coefficient)	-0.019	0.0082	0.0121	-0.0036	0. 0029
	(-1.53)	(0.80)	(0.96)	(-0.31)	(0.27)
Default dummy (Odds ratio)	1.05	0.998	1.011	0.99	1.003
	(0.90)	(-0.03)	(0.19)	(-0.18)	(0.06)

Table 16 Falsification II

We assign high financial literacy to two professions: engineers, and servicemen and police officers. We report the coefficients (for the regression of spread in Panel A) or the odds ratio (for the regression of default in Panel B) of the financial literacy dummy. Robust z-statistics are reported in the parentheses. *, **, and *** indicate the estimates are significant at 10%, 5%, and 1% levels respectively.

Panel A Spread

Dependent variable: Spread		
	(1) Servicemen and Police Officers	(2) Engineers
Literacy dummy	0.008	-0.024
	(0.39)	(-1.26)
Control Variables	Yes	Yes
State fixed effects	Yes	Yes
Constant	Yes	Yes
Property type fixed effects	Yes	Yes
Origination year fixed effects	Yes	Yes
Servicer fixed effects	Yes	Yes
R-squared	0.866	0.866
Observations	136619	136619

Panel B Default

Dependent variable: Default Dummy		
	(1) Servicemen and Police Officers	(2) Engineers
Literacy dummy	1.120	0.976
	(1.08)	(-0.27)
Control Variables	Yes	Yes
State fixed effects	Yes	Yes
Constant	Yes	Yes
Property type fixed effects	Yes	Yes
Origination year fixed effects	Yes	Yes
Servicer fixed effects	Yes	Yes
R-squared	0.866	0.866
Observations	136619	136619

Table 17 Comparison between Borrowers from the Financial and the Non-Financial Industry: Continuous Measures

This table presents the summary statistics for the two delinquency measures for borrowers from the financial industry and the non-financial industry respectively. *delinq_ratio* is the fraction of time when the mortgage is in delinquency and *missed_ratio* is the number of missed mortgage payments divided by the number of total payments.

Variable	Industry	Obs	Mean	Std. Dev.	Min	Max
delinq_ratio	Non-financial industry	130591	0.030	0.106	0	1
	Financial Industry	4849	0.025	0.094	0	0.857
missed_ratio	Non-financial industry	130591	0.027	0.094	0	1
	Financial Industry	4849	0.023	0.086	0	0.857

Table 18 Alternative Delinquency Measures

This table presents the logistic regression of two alternative dependent variables on financial literacy and other information. We use the two other delinquency measures *delinq_ratio* and *missed_ratio* introduced in 3.4, and estimate the coefficients using generalized least-squares method. Robust standard errors are clustered at the MSA level. Robust z-statistics are reported in the parentheses. *, **, and *** indicate the estimates are significant at 10%, 5%, and 1% levels respectively.

Dependent variable	delinq_ratio	missed_ratio
	(1)	(2)
Mean Dependent variable	0.029	0.027
Literacy dummy	-0.0026**	-0.0021*
	(-2.03)	(-1.68)
Control Variables	Yes	Yes
State fixed effects	Yes	Yes
Constant	Yes	Yes
Property type fixed effects	Yes	Yes
Origination year fixed effects	Yes	Yes
Servicer fixed effects	Yes	Yes
Pseudo R-squared	0.351	0.330
Observations	136619	136619

Figure 1 Mark-to-Market LTV

This figure shows the kernel density of the mark-to-market LTV of financial industry borrowers and non-financial industry borrowers when loans are delinquent. The solid line is for the financial industry borrowers and the dashed line is for the non-financial industry borrowers.