Gerardi, K., Goette, L., & Meier, S. (2011, January 7). *Financial literacy and subprime mortgage delinquency: Evidence from a survey matched to administrative data* [Paper presentation]. American Economic Association 2011 Annual Meeting, Denver, CO, United States. https://www.aeaweb.org/conference/2011/retrieve.php?pdfid=348

Financial Literacy and Subprime Mortgage Delinquency:

Evidence from a Survey Matched to Administrative Data*

Kristopher Gerardi† FRB Atlanta

Lorenz Goette‡ University of Lausanne

Stephan Meier§ Columbia University

October 17, 2010

ABSTRACT

We measure several aspects of financial literacy and cognitive ability in a survey of subprime mortgage borrowers who took out loans in 2006 and 2007. These measures are matched to objective, administrative data on mortgage characteristics and payments. We find a large negative correlation between a particular aspect of financial literacy, numerical ability, and mortgage delinquency and default. The results are robust to controlling for a broad set of socio-demographic variables, and not driven by other aspects of cognitive ability. Furthermore, our results do not support the hypothesis that financial literacy impacts mortgage outcomes through the choice of risky mortgages.

*We thank Daniel Bergstresser, Chris Foote, Jeff Fuhrer, John Leahy, Robert Shiller and seminar audiences at FRB Boston, the ASSA meetings 2009, Cornell University, Harvard Business School, Universitat Autonoma de Barcelona, and Columbia University. We are especially grateful to the FRB Boston, where the authors were all employed at the time the survey was conducted, and which provided the funding for the survey. The views expressed in this paper are, however, those of the authors and not necessarily those of the Federal Reserve System or the Federal Reserve Bank of Boston.

†Federal Reserve Bank of Atlanta, Research Department, 1000 Peachtree Street N.E., Atlanta, GA 30309, USA. kristopher.gerardi@atl.frb.org.

‡University of Lausanne, Faculty of Business and Economics, Bâtiment Internef, 1015 Lausanne-Dorigny, Switzerland. lorenz.goette@unil.ch.

§Columbia University, Graduate School of Business, 3022 Broadway, New York, NY 10027, USA. sm3087@columbia.edu.

I. Introduction

While the subject of financial literacy has always been of first order importance to financial counselors and consumer advocates, it is now drawing the attention of more and more academic economists. Recent research has shown that many individuals have problems answering simple questions about basic financial principals (e.g., Lusardi and Mitchell, 2009; Banks and Oldfield, 2007; Lusardi and Tufano, 2008; McArdle et al., 2010) and routinely make systematic financial mistakes, such as underestimating interest rates from payment streams (Stango and Zinman, 2009a). It has been argued that with the deepening of financial markets, the degree of "financial illiteracy" has become an increasingly severe problem as a larger segment of the population now has to make important and complicated financial decisions (e.g., Bernanke, 2006). One case in point is the decision to purchase a home and the numerous types of mortgage instruments available to finance such a purchase. The expansion of mortgage credit in the early-to-mid 2000s in the United States has had a profound impact on real estate and financial markets. On a positive note, in a very short time, the expansion of mortgage credit broadened homeownership, particularly among individuals that had traditionally been shut out of credit markets. But, as we are now painfully aware, this increase in homeownership has not come without a significant cost. As house prices leveled off in 2006 and began to decline, a massive increase in subprime mortgage delinquencies, and an explosion of outright defaults occurred (e.g., Foote et al., 2008b; Mayer et al., 2009). This, in turn, led to a sharp drop in the value of mortgage-backed securities and to the worst financial and macroeconomic crisis since the Great Depression.

This paper investigates the impact of financial literacy on mortgage decisions and outcomes. In particular, it focuses on the relationship between financial literacy and mortgage delinquency and default within the setting of the recent U.S. subprime mortgage crisis. In light of the dramatic developments in the crisis, a debate has started over the sources of the increase in late mortgage payments and defaults, which precipitated the broad economic and financial crisis. Several papers discuss the role of credit supply changes, and in particular the potential role of relaxed underwriting standards in generating an expansion of mortgage credit (e.g., Gerardi et al., 2009; Mian and Sufi, 2009; Nadauld and Sherlund, 2009) and increased home equity leverage (e.g., Cooper, 2009). By contrast, this paper examines the impact of borrower behavior. Many market observers, including Akerlof and Shiller (2009), believe that departures from full rationality played an important role. In part, "irrational exuberance" (Shiller, 2005) — the belief that house prices will just keep rising — may have contributed to the boom and subsequent bust, but Akerlof and Shiller (2009) and others (e.g., Boeri and Guiso (2007)) argue that individuals' limited ability to make complicated financial decisions contributed importantly to the sharp rise in mortgage defaults.

This paper is the first, to our knowledge, to directly examine this hypothesis. The empirical results are based on a rich dataset that combines administrative data on subprime mortgages with survey data from a telephone interview with a sample of subprime mortgage borrowers. The survey sample is taken from a dataset on privately securitized subprime mortgages that the Federal Reserve Bank of Boston purchased from Corelogic (previously First American LoanPerformance). This dataset contains detailed information on mortgage terms and complete payment history streams as reported by mortgage servicers, which allows us to track the sample over time and follow the subsequent mortgage outcomes. The information on mortgage terms and performance are not based on self-reported information, which is important as research has shown that many borrowers are confused about their contractual terms (Bucks and Pence, 2008) and, in general, misreport information about debt and repayment (e.g., Karlan and Zinman, 2008). The survey collects three important pieces of information. The key set of questions measure several aspects of respondents' financial literacy, notably numerical ability and cognitive ability. A second set of questions measures risk and time preference parameters, and a third set of questions elicits detailed socio-demographic information.

The empirical results show a large and statistically significant negative correlation between financial literacy and measures of mortgage delinquency and default. The point estimates are remarkably robust, and quantitatively important: 20 percent of the borrowers in the bottom quartile of our financial literacy index have experienced foreclosure, compared to only 5 percent of those in the top quartile. Furthermore, borrowers in the bottom quartile of the index are behind on their mortgage payments 25 percent of the time, while those in the top quartile are behind approximately 10 percent of the time.

Financial literacy, or the understanding of basic financial principals, is composed of many different skills, and the survey design allows us to pin down the specific aspect of financial literacy that is important to mortgage outcomes. We measure a borrower's numerical ability, or the ability to perform simple mathematical calculations, as well as economic literacy (the understanding of economic principals like inflation) and cognitive skills using standard methods (Banks and Oldfield, 2007; Dohmen et al., 2009; Lusardi and Mitchell, 2009). Controlling for all three aspects shows that the correlation with mortgage default is highly specific to numerical ability. Thus, this analysis further contributes to the literature by delving deeper into the particular component of financial literacy that impacts mortgage decisions and outcomes and provides important insights for policy makers in order to provide a specific and tractable lever for interventions (which is discussed further in the concluding remarks).

While numercial ability is not randomized in this study, the rich dataset makes it possible to distinguish between some of the potential channels by which it could impact mortgage outcomes and to rule out a broad set of potential biases from omitted variables. The results are robust to the inclusion of controls for many socio-demographic variables, including income, income volatility, unemployment status and history, education and marital status, and the inclusion of parameters that capture risk aversion and time preferences. The Corelogic data also include a measure of a borrower's creditworthiness, the FICO score, at the time of the mortgage origination. Including FICO scores in the control set does not affect the correlation between financial literacy and delinquency, although it does significantly improve the explanatory power of the empirical model. While all possible explanations cannot be ruled out, the robustness of the empirical estimates leads us to conclude that limited financial literacy played a non-trivial role in the rise of subprime mortgage defaults.

In addition, a key contribution of this paper is an investigation of the manner by which numerical ability affects mortgage outcomes. The data allows us to test whether numerical ability impacts mortgage outcomes through its effect on housing and mortgage choices. The type of mortgage and the size and location of the home are important determinants of variation in default rates (e.g. Gerardi et al., 2009; Foote et al., 2008b; Mayer et al., 2009). For example, adjustable-rate mortgages have experienced significantly higher default rates compared to fixed-rate mortgages during the crisis. In addition, interest rate resets and interest-only payment periods are aspects of many subprime mortgage contracts that have received a large amount of scrutiny in the literature. Indirect evidence points to the possibility that cognitive limitations play an important role in the choice of a mortgage instrument. Evidence from micro data on mortgages shows that individuals are often confused about important contract terms of their mortgage (Bucks and Pence, 2008), that individuals with low cognitive ability overestimate the value of their home on home equity loan applications (Agarwal and Mazumder, 2010), and that individuals who were rated as confused by the interviewer were more likely to have adjustable-rate mortgages (Bergstresser and Beshears, 2009). As the data contain detailed information on mortgage characteristics, we can test whether numerical ability impacts mortgage outcomes through the particular choice of a mortgage instrument. The results show that numerical ability does not systematically predict riskier mortgage choices. Importantly, controlling for mortgage terms (interest rate, adjustable vs. fixed rate, cumulative loan-to-value (CLTV), debt-to-income (DTI) ratio, etc.) does not explain the link between numerical ability and the extent of mortgage delinquency. In addition, the data allow us to test for whether differences in prior experience with mortgage markets or different location choices can explain the link between numerical ability and mortgage outcomes. The estimation results do not support such a theory.

The finding that the link between numerical ability and mortgage default is not mediated by mortgage choice or initial differences in wealth, indicates that numerical ability likely impacts mortgage outcomes through another channel. One potential channel that is consistent with the empirical results is the inability of an individual to perform the simple mathematical calculations necessary to maintain a household budget or to calculate whether monthly mortgage payments are affordable over a long horizon. This interpretation is consonant with the picture that emerges from the survey evidence linking poor financial literacy to higher consumption, less saving, and out-of-control credit usage (Banks and Oldfield, 2007; Lusardi and Mitchell, 2009; Lusardi and Tufano, 2008). In the appendix we present a very simple, reduced-form model of how limited numerical ability can impede a household's ability to maintain an accurate budget and how this limitation could affect the decision to default on a mortgage.

The remainder of the paper is structured as follows: section II provides an overview of the data and the survey design; section III discusses the empirical methodology; section IV presents the basic results, and analyzes the link between financial literacy and mortgage choice and how that link affects mortgage outcomes; section V provides an interpretation of the results; and section VI contains the concluding remarks.

II. Design of Study

This section provides a detailed discussion of the sample and survey design. First, the mortgage data and the pool of mortgage borrowers that the survey sample is drawn from is described, and potential sample selection biases are discussed. Then, the details of the survey procedure and the questions of interest are presented.

A. Mortgage Data and Sampling

In order to obtain objective measures of mortgage delinquency and default, the survey sample is constructed from data that combines two micro-level mortgage datasets. The first is a loan-level dataset constructed and maintained by Corelogic (formerly FirstAmerican LoanPerformance). Corelogic collects information on individual mortgages that are used as collateral for non-agency, mortgage-backed securities (MBS) and sold to investors on the secondary mortgage market. The sample comes from data that the Federal Reserve Bank of Boston purchased in mid-2007, which covers Massachusetts, Connecticut, and Rhode Island from the late-1990s through March 2009. The dataset contains extensive loan-level information on mortgage characteristics, including interest rates (initial levels and changes over time), documentation levels, payment histories, loan-to-value ratios, and various other lending terms. It also contains some information regarding borrower characteristics, such as the borrower's credit score and debt-to-income ratio at origination (borrower's monthly debt payment divided by his or her monthly income). Finally, the Corelogic dataset identifies the type of MBS each loan was packaged into — subprime, Alt-A, or prime.[1]

The second source of data used in this study was supplied by The Warren Group, a private Boston firm that has been tracking real estate transactions in New England for more than a century. The Warren Group collects publicly available real estate transaction records that are filed at Registry of Deeds offices throughout New England, and have maintained an electronic database of these records for the past twenty years. The data includes the universe of purchase-money mortgages, refinance mortgages, home equity loans, home equity lines of credit (only information on capacities and no information on utilization rates), and purchase deeds (including foreclosure deeds) transacted in Massachusetts, Connecticut, and Rhode Island. Unlike the Corelogic data, this dataset contains the precise location of each property and the exact names of the buyers and sellers of each property as well as the names of the mortgage borrowers. These data make it possible to construct a history of mortgage transactions for a household in a given property. In other words, with the Warren Data it is possible to follow households in the same house across different mortgages. Since the data include information on all mortgage liens and the sale price for each property, it is also possible to construct a precise measure of the cumulative loan-to-value ratio at the time of purchase,[2] and to keep track of the total number of mortgages obtained by each homeowner.

The sample of first-lien mortgages contained in subprime MBS that were originated in 2006 and 2007 from the Corelogic dataset were matched to the Warren Group registry data. The match was based on the zip code of the property (Corelogic data contains only the identity of the zip code where the property is located), the date of mortgage origination, the amount of the mortgage, whether the mortgage was for purchase or refinance, and the identity of the institution that originated the mortgage. The match rate was approximately 45 percent, and left us with a sample of more than 74,000 mortgages.[3]

[1]The sample of prime loans in the Corelogic dataset consists of mortgages with values above the GSE (Government Sponsored Enterprise) conforming loan limits. This segment of the prime market is often referred to as jumbo-prime.

[2]The Corelogic data has only sporadic information on the presence of second liens, and thus does not allow for the construction of accurate cumulative loan-to-value ratios.

[3]The main issue that contributed to the low match rate was the inconsistent definition of dates between the two datesets. The date listed in Corelogic is the date of origination, while the date listed in the Warren Group data is the date that the mortgage document was recorded. It usually takes at least a few days for documents to be filed in the Registry of Deeds offices (sometimes a few weeks), and thus, these two dates do not match. Therefore, we were forced to use a date range in the matching algorithm, and consequently

Mortgages from this matched dataset were randomly selected to construct the sample of borrowers for the survey. Two different strategies were used to contact borrowers: 1) *Cold-calls* involved calling borrowers by phone, which was possible as each borrower's name and address is contained in the Warren Group data. This information was then used as an input into an internet search engine (USAPeopleSearch.com) to find each borrower's phone number(s). 2) *Mail-ins* involved mailing invitations to participate in the survey to the addresses listed in the Warren Group data.

The *Cold-call* strategy, entailed calling a total of 3,523 borrowers[4] in the summer of 2008 (June - August). The borrower was positively identified in approximately one-third of the cases (1,087).[5] In half of those cases it proved impossible to speak to the actual borrower, and thus, a response to the interview request was never received.[6] In 296 cases the borrower was contacted, but he or she refused to participate in the survey,[7] and in 253 cases the borrower was contacted and he or she agreed to participate in the survey. Based on these statistics, two participation rates are reported for the *Cold-calls*: Of the borrowers that were directly contacted, 46.1 percent agreed to participate in the survey; of the borrowers that were matched to a correct phone number, 10.5 percent agreed to participate.

The *Mail-in* strategy entailed mailing almost 5,000 invitation letters to borrowers for whom phone contact information could not be obtained. The invitation letter was one page (two-sided) and contained a brief description of the survey and the survey conductors (and was signed by the president of the Federal Reserve Bank of Boston). A small response card was also included that contained a question asking if the borrower would be interested in participating in the survey, and included space for borrowers who agreed to participate to list working phone numbers and the best times of day to call. In addition, a response envelope and postage was included. In the vast majority of cases (98.3 percent), a response was not received. For the cases in which a response was received, an attempt was made

often found cases of multiple mortgages of the same amount, originated in the same zip code, in a given date range. We were forced to throw out these cases of multiple matches. The identity of the originating institution often helped in these cases, but unfortunately the Corelgoic data contain only sparse information on this variable. The matched sample of mortgages appears to be quite representative of both the Corelogic and Warren Group datasets based on observable mortgage, property and borrower characteristics.

[4]Often multiple phone numbers were found for each borrower in the data, so the actual number of phone numbers called was much larger than the number of borrowers.

[5]For another one-third the phone line was not working. For the last one-third, a working line was reached, but it was not possible to verify that the phone number corresponded to the borrower in the data either because nobody picked up the phone or because an answering machine was reached (it was not possible to identify the borrower from the answering machine message, and furthermore the borrower never responded to a message that was left on the machine).

[6]In most of these cases either a message was left on an answering machine, or another member of the household answered, but the actual borrower was not available.

[7]This includes cases in which the borrower agreed to participate at a later date, but never followed through on that agreement.

to call the borrower to conduct the interview. Of the borrowers that were subsequently contacted, approximately 92 percent agreed to participate in the survey (70 percent of the borrowers for whom a correct address could be verified).

[Table 1 about here.]

Sample selection bias is always a serious concern in surveys such as this one. Since information about observable mortgage and borrower characteristics for *all* of the borrowers is available in the Corelogic and Warren Group datasets, it is possible to test whether there is sample selection on those observable characteristics. According to sample statistics listed in Table 1, there is little evidence of sample selection on observable characteristics. The table compares average characteristics between the respondents and non-respondents for both the *Cold-Calls* sample and the *Mail-In* sample. Only the mortgage amount for the phone sample and initial interest rate for the mailing is statistically significant at the 10 percent level. All the other differences are not statistically significant at even the 10 percent level. Importantly, information about the foreclosure rate of *all* individuals is also available, and there is no difference in the probability of foreclosure after the mailing went out between respondents and non-respondents. Furthermore, a more formal statistical test of sample selection is conducted, and the results support the findings from Table 1.[8]

While it does not appear that selection into the survey sample is an issue, the timing of the survey raises some important issues. The survey was conducted in the summer of 2008 between June and August, while the borrowers chosen for the survey obtained their mortgages in 2006 and 2007. August 2007 is the last month that a mortgage was originated in the survey sample, quite simply because the subprime mortgage market had completely shut down at that point and no new mortgages were originated. This means that the subprime borrowers taking the survey had been paying their respective mortgages for at least 10 months and up to 32 months (for mortgages originated in January 2006). In addition, one of the requirements imposed for inclusion into the sample was that each borrower not be in the foreclosure process at the time that the survey was conducted. This requirement was

[8]To test for potential sample selection bias on observables we estimate for each observable outcome measure k (see Table 1), the equation

$$y_i^k = \alpha_k + \gamma_k R_i + \beta_k CC_i + \epsilon_i^k \tag{1}$$

where α_k is the constant for outcome k, γ_k is the difference in the outcome if the individual was a respondent (and $R_i = 1$), and β_k is the difference in the outcome if individual i was a cold call (and $CC_i = 1$). Finally, ϵ_i^k is the residual for outcome k. The k equations in (1) are estimated by seemingly unrelated variables, thus allowing the residuals ϵ_i^k to be correlated across outcomes within individuals. The hypothesis $\gamma_k = 0$ is then tested for all k outcome measures. The p-value of the corresponding χ^2-test is $p = 0.51$, which suggests little to no evidence of selectivity into the survey on these 10 important variables.

made because of the increased difficulty in contacting borrowers in foreclosure. Many of those borrowers are either likely not living in the home anymore, or if they are, will likely refuse to answer phone calls or mail requests. Because of this design feature, the results in this study are not necessarily representative of all subprime mortgage borrowers. Many subprime borrowers defaulted on their loans and experienced foreclosure within the first year of origination. The average number of months to default for all subprime mortgages originated in 2006 and 2007 in the LP dataset for which the servicer has initiated foreclosure proceedings is slightly less than 18. More than one-quarter of the defaults occurred within one year of origination.

In the end, 339 individuals responded to the survey and constitute the main sample. The last two columns in Table 1 display summary statistics of the detailed information in the Corelogic data on respondents' mortgage characteristics. The statistics are consistent with what one might expect in a sample of subprime mortgage borrowers. The average FICO score of 632 is relatively low, the majority of borrowers have adjustable-rate mortgages (two-thirds), and the average debt-to-income ratio (ratio of the summation of all monthly debt payments to the monthly mortgage payment) of 0.42 is relatively high, which suggests that this group of borrowers is characterized by heavy debt burdens.

B. The Survey

The survey contains four important parts: 1) Measures of two aspects of individuals' financial literacy, numerical ability and basic economic literacy, and a measure of general cognitive ability. 2) Measures of time and risk preferences. 3) Questions about the details of the mortgage contract (we already know much of this information from the micro datasets) and the experience of shopping for the mortgage. 4) An extensive list of socio-demographic characteristics that complements information from the Corelogic and Warren Group datasets.

On average, the survey took about 20 minutes to complete, and individuals were compensated $20 for their participation.

B.1. Financial Literacy and Cognitive Ability

The first measure of financial literacy, which is the primary focus of this study, determines the proficiency of a respondent for solving basic mathematical calculations. Participants were asked the following five questions originally developed by Banks and Oldfield (2007). The questions are as follows:

1. *In a sale, a shop is selling all items at half price. Before the sale, a sofa costs $300.*

How much will it cost in the sale?

2. *If the chance of getting a disease is 10 per cent, how many people out of 1,000 would be expected to get the disease?*

3. *A second hand car dealer is selling a car for $6,000. This is two-thirds of what it cost new. How much did the car cost new?*

4. *If 5 people all have the winning numbers in the lottery and the prize is $2 million, how much will each of them get?*

5. *Let's say you have $200 in a savings account. The account earns ten per cent interest per year. How much will you have in the account at the end of two years?*

To construct an index of numerical ability, Banks and Oldfield (2007) suggest dividing individuals into four separate groups based on the responses to the five questions. A borrower is placed into the first group corresponding to the lowest level of numerical ability if he answers questions 1, 2, and 3 incorrectly *or* answers question 1 correctly, but answers questions 2, 3, and 4 incorrectly. The second group is made up of borrowers who answer at least one of the first four questions incorrectly (the outcome of the fifth question is not considered for the first or second groups). The third group contains borrowers who answered questions 1, 2, 3, and 4 correctly, but answered question 5 incorrectly. Finally, borrowers who answered all five questions correctly are placed into the fourth group. Table 2 shows the distribution of the numerical ability index in the survey sample as well as the distribution from the Banks and Oldfield paper. Approximately 16 percent of borrowers fall into the lowest group, 54 percent into the second group, 17 percent into the third group, and 13 percent into the highest group. Despite being characterized by a very different group of individuals, the distribution of the index in the Banks and Oldfield study is very similar to the distribution in the survey sample.

[Table 2 about here.]

It is important in this context to distinguish between the effects of financial literacy and the more general notion of cognitive ability. To do so, we use a verbal fluency measure first introduced by Lang et al. (2005). Participants were asked: *" In the next 90 seconds, name as many animals as you can think of. The time starts now."* The number of animals named is highly correlated with IQ (e.g. Lang et al., 2005). The reason for this is that intelligence is highly correlated with the ability to retrieve known information. As most people know hundreds, if not thousands of animals, the question reveals how easy it is to retrieve that information. Obviously, the ability to name animals in English also depends on individuals' English language skills, which is elicited separately (see below). In the economics literature, Dohmen et al. (2009) also use this question to measure cognitive ability. Figure 1 compares the distribution of responses in the survey sample to the distribution in their study, which used a representative sample of the German population. The shape of the distributions is very similar.[9]

[Figure 1 about here.]

In addition to the measure of financial literacy that focuses on respondents' numerical ability, we measure respondents' basic understanding of economic mechanisms using two questions from Lusardi and Mitchell (2009). Lusardi and Mitchell (2009) refer to these as "basic financial literacy" questions, but in our opinion they measure an individual's understanding of basic economic concepts, and thus we refer to them as questions that measure "economic literacy."

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

2. *Suppose that in the year 2020, your income has doubled and prices of all goods have doubled too. In 2020, how much will you be able to buy with your income? More than today, exactly the same as today, or less than today?*

In the survey sample, approximately 79 percent of borrowers answered the first question correctly, and 74 percent answered the second question correctly, while 60 percent answered both questions correctly. These results are very similar to those obtained by Lusardi and Mitchell (2009).

As an additional measure of cognitive ability, the average time it took the participants to respond to the Banks and Oldfield (2007) numerical ability questions was calculated. The time was measured from the moment the interviewer had finished reading the question to the moment an answer was given by the respondent. Table 3 displays the correlations between all of the measures of cognitive ability and financial literacy. There is a strong, positive correlation between the measures of financial literacy and cognitive ability as measured by the verbal fluency question. There is also a strong, negative correlation between the measures of financial literacy and verbal fluency with the response times to the numerical

[9] This is quite interesting in its own right, as a priori, we did not expect a sample of U.S. subprime mortgage borrowers to have a similar distribution of cognitive ability as a representative sample of German borrowers.

ability questions. Individuals who responded more quickly to these questions achieved higher scores.[10]

[Table 3 about here.]

A more formal factor analysis reveals one common factor among the five variables. Only the first eigenvalue is greater than one, while all others are almost exactly equal to zero. Finding one common component to different measures of intelligence is quite common, and has been found in many other studies (See, e.g., Flynn, 2007; Burks et al., 2009a).

B.2. Time and Risk Preferences

To measure time and risk preferences, respondents were asked to make a number of hypothetical choices that made it possible to calculate their discount factors and risk aversion parameters.

Similar to experimental measures of time preferences (see, e.g., Meier and Sprenger, 2010, 2009), individuals were asked to decide on a monetary amount that makes them indifferent between receiving the amount immediately versus waiting x months for a larger monetary amount. The answers to these questions allows for the calculation of individual discount factors. Individuals were asked to make such intertemporal trade-offs for the present versus both $x = 6$ months and $x = 12$ months. The two different time frames make it possible to construct a measure of whether individuals have dynamically inconsistent time preferences (e.g., Laibson, 1997). As can be seen in panel A in Table 4, which displays summary statistics of key variables from the survey, the average discount factor in our sample is 0.96 (over one month) and 80 percent of our sample exhibits dynamically consistent time preferences, similar to Meier and Sprenger (2010). In addition, the borrowers were asked to assess their own level of impatience on a 11-point scale from 0 corresponding to "very impatient" to 10 corresponding to "very patient." The measure of impatience that is based on the set of hypothetical choices is primarily used in the empirical work, but the results are largely unchanged if the subjective scale is used instead.

The measure of risk aversion also follows standard experimental strategies (e.g., Barsky et al., 1997). Participants were asked to hypothetically choose between a certain payoff and a 50-50 chance of receiving a good or bad payoff:

[10] A skeptic may argue that differences in the measure of cognitive ability instead may simply measure different styles in which individuals answer questions. Some may take the time to think about the question and then answer, while others may just blurt out the first thing that comes to their mind. The negative correlation between response times and the measures of financial literacy and cognitive ability provides some evidence against this interpretation, as it shows that individuals who struggled to answer in a timely manner, also were more likely to get the answer wrong.

Which would you prefer: A mortgage for which you paid \$1000 per month for the next thirty years, or a mortgage, in which, after two years the payment is either \$500 or \$1100 with equal chance?

If the participant accepted the uncertain lottery, the high payment of the uncertain mortgage was raised by increments of \$100. The payoff at which the participant switches to the safe mortgage is used as the measure of risk tolerance. The mean switching amount (see Table 4) was \$ 1184, revealing a substantial degree of risk aversion. In addition, participants were asked to assess their own level of risk tolerance on a scale from 0 to 10 as in Dohmen et al. (2005). As with the self-assessed impatience measure, the second risk measure does not require any numerical skills. Nevertheless, the risk measure based on the set of hypothetical choices (most related to experimental risk measures) is primarily used in the empirical work, but the results are robust to using the self-reported scale measure.

[Table 4 about here.]

B.3. Socio-demographics

The survey contains a number of detailed questions about socio-demographic characteristics and information about household income and employment status. Participants were asked about their race and ethnicity, gender, age, place of birth, amount of time spent in the United States, marital status, number of children, education level, and proficiency with the English language (scale from 0 corresponding to a "beginner" to 10 corresponding to a "native speaker"). In addition, questions were included to measure the amount of household income, the number of family members that contribute to household income, and the volatility of household income (on a three-point scale with 1 signifying that "it's been pretty stable"; 2 signifying "it has gone up and down a little over the last few years"; and 3 indicating that "it has gone up and down a lot over the last few years"). Finally, participants were asked about their current employment status and the number of times that they had been out of work over the previous five years.

Panel B in Table 4 presents summary statistics of the socio-demographic information. Approximately one-third of the respondents are minority (defined here as not white), and the split between males and females is about 50-50. The average age of respondents is 47 and their average household income is \$80 thousand, which is surprisingly high for a sample of subprime borrowers.[11] Almost 84% of respondents were born in the United States, and almost three-quarters have more than a high school education.

[11]The high average and huge standard deviation of income in the sample supports the idea that the subprime mortgage market was used by a diverse group of borrowers.

B.4. Mortgage Experience

As we discussed above it is possible to obtain information on respondents' previous experience in mortgage markets. The Warren Group data allows one to calculate the number of mortgages obtained since the home purchase (going back to January 1987), which allows one to calculate the number of mortgages taken out by each household before their current subprime loan.[12] This information is supplemented with additional proxies for borrowers' experience with mortgages and their search behavior prior to obtaining their current mortgage. The survey asks participants whether they were first-time homebuyers, whether they had taken a home buying class or had received counseling, if they obtained information about mortgage pricing before obtaining their loan, and if they had, how they obtained the information (internet, relative, friend, etc.). According to panel C of Table 4, more than half (55%) of the sample are first-time homebuyers, less than 10% took a home counseling class before purchasing their house, while 60% reported that they shopped around for a mortgage.

C. Measures of Mortgage Delinquency

Using the Corelogic mortgage performance data, three different measures of mortgage delinquency were created. All measures incorporate delinquency from the origination of the mortgage until March 2009 (the latest update provided by Corelogic).[13] The first measure of delinquency measures the fraction of time a borrower is behind by at least one mortgage payment. This measure captures the amount of time during which a household is unable or unwilling to meet the promised mortgage payments. For example, if a household missed its very first payment and made all future payments on time, this measure would consider that the household was behind in each period until that first payment is made.

The second measure of mortgage delinquency is the fraction of mortgage payments missed. This variable is an explicit measure of the extent of delinquency. For example, a borrower who has had a mortgage for 12 months and who has missed 6 payments would be assigned a value of 50 percent for this measure, while a borrower who has had the mortgage for the same amount of time, but who has only missed 3 payments, would be assigned a value of 25 percent.

The third measure is a dichotomous variable that takes a value of one if foreclosure proceedings have been initiated by the lender. Normally, foreclosure proceedings are initiated

[12] Mortgage information is only available for the current property.

[13] Restricting the time period until summer 2008 when we conducted the survey does not materially affect the results.

when a borrower is 120 days delinquent on their mortgage (or equivalently is 4 payments behind).[14]

[Table 5 about here.]

Table 5 contains information on the distributions of the three delinquency measures. The average borrower in the sample was behind on their payments 20 percent of the time, and had missed 11 percent of their payments. Half of the borrowers in the sample were delinquent more than 7 percent of the time and had missed more than 5 percent of their mortgage payments, while 10 percent of the borrowers were delinquent more than 60 percent of the time and had missed more than 30 percent of their payments. Almost 20 percent of the borrowers in the sample had been in the state of foreclosure at some point in their mortgage experience.

III. Empirical Specification

The primary empirical specification takes the following form:

$$D_i = \gamma NA_i + \mathbf{x_i}'\beta + \epsilon_i \tag{2}$$

where D_i corresponds to the first two measures of delinquency discussed above, the percent of time spent in delinquency and the percent of mortgage payments missed, for household i. The term NA_i represents the numerical ability group of household i, $\mathbf{x_i}$ represents a vector of control variables, and ϵ_i is the residual. The equation is estimated by ordinary least squares (OLS)[15], and potential heteroskedasticity in the standard errors is taken into account by estimating robust standard errors.

A probit specification is estimated for the third measure of delinquency, the initiation of foreclosure proceedings,

$$Pr[F_i = 1|NA_i, \mathbf{x_i}] = \Phi(\gamma NA_i + \mathbf{x_i}'\beta) \tag{3}$$

where F_i takes the value of one if foreclosure proceedings have been initiated on the borrower and zero otherwise, and $\Phi()$ is the cumulative distribution function of the standard normal distribution.

[14] One of the participation criteria was not being in foreclosure at the time of the survey. But, there are a few instances in which a borrower had been in foreclosure in the period before the survey was administered, but then had recovered by the time of the survey. These borrowers were included in the survey sample.

[15] The results are robust to using a Tobit specification instead of OLS (see Table A1 in the appendix).

We initially focus on numerical ability as the independent variable of interest because the strongest evidence from the previous literature comes from studies linking numerical ability to savings (Banks and Oldfield, 2007). In subsequent estimation, other measures of financial literacy that are not directly related to the ability to perform mathematical calculations, but are more related to the ability to comprehend financial concepts, as well as controls for general cognitive ability (Dohmen et al., 2009; Lusardi and Mitchell, 2009) are considered.

Control variables are included to avoid two types of potential omitted variable biases. First, a case whereby an omitted variable x has a causal impact on both mortgage delinquency and numerical ability, but its effect on delinquency, due to its omission, is captured by numerical ability. Second, the possibility that numerical ability does not have a direct effect on delinquency, but rather causes x, which in turn affects delinquency and default propensities. In this case, omitting x would lead to the improper conclusion that numerical ability affects delinquency directly, while it really is a determinant of some other variable that in turn affects delinquency. One of the strengths of this analysis is the expansive set of control variables available to deal with these potential biases and to thus narrow down the admissible set of interpretations.

The choice of the appropriate set of controls (besides basic demographic characteristics) is guided by the literature on models of mortgage default. Economic models of mortgage default emphasize the role of liquidity shocks and differences in household financial situations that make borrowers differentially vulnerable to those shocks (Gerardi et al., 2009; Pennington-Cross and Ho, 2010; Sherlund, 2008; Demyanyk and Van Hemert, 2010). If shocks or exposure to shocks are correlated with, or caused by, numerical ability, then they would need to be taken into account in order to avoid an omitted variable bias. For example, numerical ability may affect the financial situation of a household in general, putting them at an advantage from the very beginning stages of the mortgage contract. Thus, in order to understand the direct contribution of numerical ability on mortgage-related decisions, initial differences in financial situations would need to be controlled for in the estimation. To proxy for the financial condition of the borrower, we include income, income volatility, the number of dependents, employment status, unemployment history and education (and cognitive ability). In addition, credit scores (FICO) measured at the time of origination of the mortgage are included to proxy for differences in creditworthiness and the extent to which households may be borrowing constrained.[16]

Models of mortgage default also predict that preference parameters, such as time pref-

[16]The debt-to-income ratio at origination is also available in the Corelogic data, which measures the ratio of household debt (mortgage, credit card, auto, and education) to income. This variable is added to the control set in later specifications.

erences and risk preferences should affect the incidence of default (Foote et al., 2009). For example, more impatient individuals, who value consumption today much more than in the future, may be more likely to default on a mortgage, all else being equal. Furthermore, evidence from recent studies suggests that these preference parameters may be correlated with certain aspects of cognitive ability (Burks et al., 2009b; Dohmen et al., 2009). Thus, including measures of time preference and risk preference parameters is important in separating the direct impact of numerical ability on mortgage delinquency from a potential correlation with those preference parameters.

In a second step, we specifically investigate whether the type of mortgage chosen and the characteristics of the loan are influenced by variation in financial literacy and whether this influence explains the correlation between financial literacy and mortgage repayment behavior. The empirical mortgage literature also documents large differences across borrowers in the extent of delinquency and incidence of default that are due to differences in mortgage characteristics and house price movements. Gerardi et al. (2009), Foote et al. (2008b), Mayer et al. (2009) and Foote et al. (2009) show that there are large differences in default rates between fixed-rate and adjustable-rate mortgages, and that initial leverage (loan-to-value ratios) and local house price movements are strongly correlated with default rates. But, it is important to realize that the type and size of the mortgage contract, along with the location of a house are among the most important choices of any home purchase. Therefore, these differences should be regarded as differences in choices made by borrowers. The choices regarding mortgage characteristics could be directly influenced by financial literacy, as borrowers with limited financial literacy may be more likely to choose simpler contracts, or they may be more susceptible to predatory lending, which would suggest that they would be more likely to choose more complicated, opaque mortgage products. Thus, one of the goals in this study is to examine and document whether such a correlation exists, and if it does to see if it can explain some of the correlation between mortgage characteristics and mortgage default.

IV. Results on Financial Literacy and Delinquencies

A. Baseline Findings

Figure 2 displays the unconditional relationship between the numerical ability index and the three measures of delinquency. The bar graph in Panel A shows that there is a monotonically decreasing relationship between the percent of time spent behind on mortgage payments and numerical ability. Borrowers in the lowest numerical ability group on average spend almost

25 percent of the time in delinquency, while those in the highest group spend on average only 12 percent of the time in delinquency. In Panel B a similar relationship holds between the percent of missed mortgage payments and numerical ability. The lowest group has missed almost 15 percent of mortgage payments on average, while the highest group has missed only 6 percent of payments on average. The incidence of foreclosure also appears to be negatively related to numerical ability (Panel C). While there is only a small difference in the incidence of foreclosure between the first and second numerical ability group, the third group is characterized by a significantly lower incidence of foreclosure compared to the first two groups (15 percent versus more than 20 percent), while the fourth and highest group is characterized by a significantly lower incidence of foreclosure compared to the third group (7 percent versus 15 percent).

Table 6 displays the coefficient estimates from the linear regressions (columns (1) and (4)) and the estimated marginal effects from the probit model of foreclosure starts (column (7)). They indicate that, as suggested by the figure, the correlations between numerical ability and the delinquency measures are positive and statistically significant. As the figure already suggested, magnitudes are also quantitatively important: A borrower in the lowest category of the numerical ability index spends, on average, approximately 15 percent more time in delinquency than a borrower in the highest category. The differences in foreclosure rates across numerical ability groups are also very large. According to the estimates, the difference in foreclosure rates between the bottom quartile and top quartile is approximately 18 percentage points.[17]

[Figure 2 and Table 6 about here.]

In the remainder of the paper, we use the richness of our survey dataset and mortgage datasets to try to distinguish between alternative explanations in order to determine the channel through which numerical ability affects mortgage repayment behavior.

B. Socio-Economics, Preferences, and Household Financial Status

In columns (2), (5) and (8) of Table 6, the socio-economic variables and preference parameters that were collected in the survey are added to the regressions as control variables. The list of variables includes age, gender, ethnicity, education, marital status, the size of the household, time and risk preference parameters, labor market status over the previous five years, the household's income, and the subjective measure of income volatility. As can be

[17] The extent of delinquency and foreclosure is also estimated to be monotonically increasing in numerical ability when a specification that includes a separate dichotomous variable for each numerical ability group is employed (see Table A2 in the Appendix).

seen in the first row of Table 6, the inclusion of these control variables does not significantly alter the point estimates or the standard errors associated with the numerical ability index. Numerical ability remains significantly correlated with mortgage delinquency, and the point estimates remain large and virtually unchanged. The control set, however, does contain important predictors of delinquency, as can be seen in the increase in the R^2 from roughly 2 to 14 percent. In particular, variables related to labor market success, such as income and income stability, as well as the number of times out of work over the previous five years, have a significant impact on delinquency.

In columns (3), (6) and (9), controls are included for certain aspects of the household's financial situation at the time of mortgage origination. We include the FICO score, and dummy variables for whether the borrower is an investor (owner occupant as the reference group), as well as whether the mortgage is for a home purchase (refinance is the reference group left out of the regression). Again, the coefficient estimates are unaffected, and remain statistically significant in all specifications. The inclusion of these controls also significantly increases the R^2 of the regression from around 15 percent to approximately 25 percent. The FICO score, in particular, is an important determinant of delinquency and default. The fact that the correlation between numerical ability and delinquency does not change when the FICO score at origination is included is an important finding.[18] It implies that the measure of numerical ability is not just capturing the fact that borrowers who have defaulted on previous debts are more likely to default on their mortgage compared to borrowers with good credit histories. Therefore, initial creditworthiness, i.e. the ability to borrow to smooth out shocks, doesn't drive the effect of numerical ability.[19]

C. Financial Literacy: Numerical Ability, Economic Literacy or Cognitive Ability?

The next step in the analysis is to pin down the particular aspect of financial literacy that affects mortgage repayment behavior. In addition to the education variables, we include as control variables an additional aspect of financial literacy and a measure of cognitive ability that is unrelated to financial literacy. The measure of cognitive ability is a verbal IQ measure that is related to information processing, while the second measure of financial literacy includes two questions taken from Lusardi and Mitchell (2009) that are meant to

[18]Notice also that the inclusion of the FICO score renders most labor market controls that were significant in columns (2), (5), and (8), insignificant, with the exception of the volatility of income. Since the FICO score is constructed to be a catch-all predictor for delinquency, this is not entirely surprising.

[19]The estimated correlation between numerical ability and delinquency is not affected by the inclusion of debt-to-income ratios at origination, which capture other types of debt in addition to mortgage debt (see the discussion below and Table 9).

measure basic economic literacy. The response times to the numerical ability questions are also included as an explanatory variable.[20] Table 7 displays the results. In columns (1), (3) and (5), only the verbal IQ measure is included in the estimation. Its inclusion does not affect the magnitude or statistical significance of the estimated coefficient associated with financial literacy. The verbal IQ measure, conditional on the numerical ability measure, is not correlated with the first two measures of delinquency (percent of time behind, and percent of payments behind). However, it does enter significantly into the probit model for the initiation of the foreclosure process. An increase of one standard deviation in the verbal IQ measure (8 points), is associated with a 4.8 percentage point decrease in the foreclosure rate. An important difference between foreclosure and the other two delinquency measures is that foreclosure is initiated by the lender. One possible interpretation of this finding is that lenders may be less likely to foreclose on an intelligent person who is behind, and that this is picked up by our measure of IQ.

Columns (2), (4) and (6) display the results when the measures of economic literacy and the response times are included in the set of control variables. They are not correlated with any of the three measures of delinquency, and do not affect the point estimate of the numerical ability measure. These findings lead us to conclude that the correlation between financial literacy and mortgage repayment behavior is specific to borrowers' numerical ability. The addition of both a verbal IQ measure, the response time and a different aspect of financial literacy, economic proficiency, does not explain differences in mortgage delinquency and default, and does not affect the the correlation between the numerical ability index and mortgage delinquency.

[Table 7 about here.]

D. Mortgage Terms and Characteristics

An important channel through which financial literacy could affect mortgage delinquency is in leading individuals to obtain mortgages with unfavorable terms, because they may be more likely to make mistakes in assessing the financial consequences of a particular contract. By making use of the detailed information on mortgage and borrower characteristics in the administrative datasets, it is possible to directly examine this possibility. In a first step we determine whether numerical ability is correlated with the choices of various mortgage terms.

[20] One potential critique is that respondents who scored poorly on the numerical ability index do not really have low financial literacy, but rather are simply lazy and not interested in taking the time to carefully answer the questions. If this is the case then it could explain some of the correlation between numerical ability and mortgage delinquency. To also account for such an effect we include the response times to the questions in the estimation.

In a second step, we add two sets of control variables to the basic empirical specification to determine whether they help to explain the relationship between numerical ability and mortgage delinquency. The first set controls for the contract terms of the mortgage, such as the initial interest rate and whether the mortgage has a fixed or variable rate. The second control set includes the loan-to-value ratio (LTV) at origination, and the debt-to-income ratio (DTI) to examine whether individuals with poor numerical ability choose to take out loans that are significantly larger than those with higher literacy levels, and whether this explains some of their repayment problems.

Table 8 presents correlations between the numerical ability index and mortgage choices. Column (1) presents the estimated unconditional correlation between the numerical ability index and various mortgage contract terms. Column (2) shows how the correlation estimates change when the control variables included in Table 6 are added to the regressions (see Appendix Table A3 for the full results of the conditional models). The numerical ability index is unconditionally correlated with many of the mortgage and borrower attributes, but when the set of control variables is added, most of the correlations disappear. The lone exception is the initial interest rate, which is negatively correlated with numerical ability (individuals with higher ability have mortgages with lower interest rates on average).

Table 9 contains estimation results of equations 2 and 3 when mortgage and borrower characteristics are included in the set of control variables. Columns (1), (3) and (5) display the results when differences in contract terms are included. The control variables do not add to the explanatory power of the baseline specification and, consequently, leave the point estimate of the numerical ability index and its standard error, essentially unchanged. In columns (2), (4) and (6), LTV and DTI ratios at origination are added to the set of controls. The two variables are not correlated with the delinquency measures, but are correlated with the foreclosure measure. According to the estimates, a 10 percentage point increase in LTV at origination is associated with a 5.2 percentage point increase in the probability of foreclosure.[21] The DTI ratio is only weakly correlated with the probability of foreclosure.[22] But, more importantly, the inclusion of both variables does not affect the magnitude or statistical significance of the correlation between numerical ability and mortgage delinquency and default.

[Table 9 about here.]

[21] There are likely two explanations for this finding. First, all else equal, a higher LTV at origination, implies a worse equity position at each future date, and thus a higher probability of foreclosure (see, for example, Foote et al., 2008a). Second, there is likely a selection effect, whereby borrowers that are more likely to default, perhaps because they have less wealth, choose to produce lower down payments at the time of purchase.

[22] This is consistent with the finding in Foote et al. (2009) that DTI ratios at origination do not contain much predictive power for foreclosures.

E. Prior Experience in Mortgage Markets

Experience in mortgage markets may also impact mortgage repayment behavior. A borrower who has obtained numerous previous mortgages may have a better idea of the type of product that best fits their financial situation. In this section, we try to determine whether borrowers with poor numerical ability are less experienced with mortgages, and whether the extent of mortgage market experience has an independent effect on mortgage repayment behavior.[23] The number of previous mortgages obtained by the borrower is calculated from the Warren Group dataset, and is included in the control set. In addition an indicator for first-time homebuyers, as well as a number of variables collected in the survey pertaining to the amount of information the individual collected before signing the mortgage contract are included as control variables. Table 10 displays the results. The correlation between numerical ability and mortgage delinquency and default is not affected. Experience *per se* does not seem to have a strong effect on the extent of delinquency. There is, however, some evidence that individuals who purchased a house for the first time are more likely to experience foreclosure, though this effect is difficult to interpret.[24]

[Table 10 about here.]

F. Geographic Area and Mortgage Lenders

Finally, two additional channels through which numerical ability could indirectly affect mortgage repayment behavior are explored. The first is related to the decline in house prices. Declining house prices played an important role in the rise of foreclosures during the housing crisis that began in 2007 (e.g., Foote et al., 2008b,a; Gerardi et al., 2007). It is not readily apparent what effect declining house prices might have on the relationship between financial literacy and mortgage delinquency. It could be that the measure of numerical ability is correlated with some neighborhood characteristic like income or education that impacted mortgage default rates when house prices began to fall. For example, one of the stylized facts of the housing crisis is that house prices were more volatile (on both the upside and downside) in poorer neighborhoods with more subprime lending, and thus greater mortgage defaults and foreclosures. We try to address this issue by including a full set of town/city fixed effects into our specifications. The results are displayed in columns (1), (4) and (7) of Table 11 for each of the measures of delinquency, respectively. The inclusion leads

[23] Agarwal et al. (2008) show that borrowers learn to avoid making certain mistakes in the credit card market.

[24] It is possible that individuals who have purchased a house before have more assets, as they benefitted from increasing house prices. Since this is not the focus of the paper, we do not explore this topic in further detail.

to a large increase in the R^2, confirming that regional variation is important in explaining variation in mortgage delinquency, as found in many other studies (Foote et al., 2008b,a; Gerardi et al., 2007). However, with 175 town fixed effects, the large increase likely also reflects the fact that in many towns, we observe few borrowers. But, the correlation between numerical ability and delinquency remains significant, and for all three measures, the point estimate increases.[25]

[Table 11 about here.]

Another potentially important effect to control for is mortgage lender or servicer treatment effects. It could be, for example, that individuals with low numerical ability choose mortgage companies that provide poor support for their mortgage borrowers. For example, lenders and (and servicers to the extent that they differ from the lender) may differ in how diligent they are about reminding borrowers when payments are due, and such differences may explain variation in delinquency. Thus, in the remaining columns of Table 11, we add originator (42) and servicer (27) fixed effects to the baseline specification. The additional controls increase the R^2, but again leave the coefficient estimate associated with numerical ability unchanged.

V. Interpretation and Discussion

The results show a quantitatively large correlation between numerical ability and subprime mortgage delinquency. The correlation remains statistically significant and quantitatively unchanged when a nearly exhaustive set of variables is conditioned on, including measures of individual labor market outcomes, financial situations at the time of mortgage origination, and other socio-economic variables that could be correlated with numerical ability. Higher numerical ability may convey many advantages in life, such as better performance and higher productivity in the workplace, and a more successful experience in credit markets. However, the results show that the relationship between numerical ability and mortgage delinquency is not affected by these indirect effects. This result is similar to findings in other studies that show that better financial literacy is correlated with higher savings (Lusardi and Mitchell, 2009; Lusardi and Tufano, 2008; Banks and Oldfield, 2007).

The results of the analysis suggest that the specific component of financial literacy that is related to mortgage repayment behavior is numerical ability. The survey includes a

[25] We also estimated a specification in which we included the cumulative amount of house price appreciation experienced between the time the mortgage was originated and the time the survey was conducted. This controls for some of the cross-sectional dispersion in house prices that had developed over the course of the financial crisis. The results are robust to such a specification.

measure of economic literacy that is associated with outcomes in other studies (see, e.g., Lusardi and Mitchell, 2009), and a measure of cognitive ability that is more closely related to the ability to retrieve information quickly. However, these measures are not predictive of mortgage repayment behavior, and conditioning on these measures does not affect the estimated correlation between mortgage delinquency and numerical ability. Thus, the results suggest a correlation that is highly specific to the ability to perform simple mathematical computations.

In addition, the analysis attempts to distinguish various channels through which numerical ability might affect mortgage repayment behavior. One leading candidate is the possibility that low financial literacy could lead many borrowers to choose mortgages with riskier attributes, such as ARMs, mortgages with prepayment penalties, or mortgages that entailed higher leverage (i.e. higher LTVs and DTIs). However, these attributes do not significantly differ across the different numerical ability groups, and consequently conditioning on these mortgage and borrower characteristics does not alter the estimated correlation between numerical ability and mortgage delinquency. In generalizing this result, one needs to keep in mind the timing of the survey: by the time the survey was conducted, most borrowers had been paying their mortgage for at least one year. An important feature of the mortgage default crisis was that many defaults took place in the first 18 months of the mortgage tenure, and thus, this analysis does not address these early defaults.

The finding that numerical ability is associated with mortgage delinquency but not with initial mortgage choice is consonant with the intuition that limitations in numerical ability lead to mortgage delinquency through over-spending during the mortgage tenure, and not to prior behavior or to choices made at the very beginning. In appendix B, an analytical example is presented that makes this intuition more specific. A simple model is considered in which individuals make intertemporal consumption decisions, but have to initially chose a mortgage that takes the form of a consumption commitment in subsequent periods. Limitations in numerical ability is modeled as mistakes in adding up terms in the budget constraint. Poorer numerical ability means that individuals make larger mistakes in either direction.[26] Thus, half of the individuals will overestimate the house they can afford, while the other half will underestimate it. In the model it is never optimal to default, and thus a perfectly rational individual never defaults. However, the lower an individual's numerical ability, the higher the chance that he obtains a mortgage that is too big for his budget, which subsequently can lead to default. Thus, consistent with the

[26]The evidence from the association between numerical ability and savings in Banks and Oldfield (2007) could be interpreted as showing that limitations in numerical ability lead individuals to make systematic error and overestimate how much they can afford. However, our results, while compatible with this assumption, do not rely on it.

data, lower numerical ability in the model leads to a higher risk of mortgage default. In this example, the effects are driven by individuals running out of money, after having committed to the mortgage. However, consistent with the data, there is no relationship between the mean mortgage size across numerical ability groups because the errors are assumed to be symmetric (and the utility function is assumed to be quasi-linear in housing consumption). While the model is very simple and reduced form it does at least suggest how the results in this study could inform theories on numerical ability and household finance.

VI. Conclusion

This paper investigates whether subprime borrowers with limited financial literacy are more likely to be delinquent on their mortgage and more likely to default. The results have several implications for future research and applications. First, the results show that a normally unobservable characteristic/ability can explain part of the heterogeneity in default behavior. This finding provides insights to lending firms on designing contract terms and risk management strategies. Individuals who have difficulties dealing with numbers seem to be riskier, controlling for usual indicators like FICO scores.

Second, the results suggest that more intensive financial education could improve financial decisions later in life as shown by Agarwal et al. (2010) and Bernheim and Garrett (2003). However, much more research is needed to conclusively show that financial education has a causal and cost-effective impact. It is also important to remember that while our data show a strong and robust correlation that is highly specific and robust to a wide set of controls, it is not a setting in which financial literacy has been explicitly randomized. The next logical, but ambitious step, is to randomize numerical ability and then track the financial decisions of individuals with exogenously improved skills over time. The choice of mortgage terms is not found to be related to financial literacy, and thus, improved disclosure may not reduce the impact of financial literacy on mortgage repayment behavior. Whether the most unsophisticated consumers profit from mandatory disclosure crucially depends on its enforcement (Stango and Zinman, 2009b). In general, future research has to investigate whether numerical ability and disclosure are substitutes or whether they are more complementary. It is easy to imagine that individuals who cannot perform basic mathematical calculations, like the ones in this study, need extremely simple disclosure in order to avoid further confusion.

Finally, our results begin to inform theory about how to incorporate numerical ability into models of mortgage choice and mortgage default. In an extremely simple analytical example, we show that if limited numerical ability affects individual's perception of their budget constraint, we can generate some of the empirical results found in the paper: no association between numerical ability and initial house size but a higher probability of default for individuals with limited numerical ability. A logical and important next step would be to extend our example to a more complicated model and to see how it affects other household financial decisions like savings and consumption choices.

References

Agarwal, Sumit and Bhashkar Mazumder, "Cognitive Abilities and Household Financial Decision Making," *Working Paper*, 2010.

_ , Gene Amromin, Itzhak Ben-David, Souphala Chomsisengphet, and Douglas Evanoff, "Learning to Cope: Voluntary Financial Education Programs and the Housing Crisis," *American Economic Review, Papers and Proceedings*, 2010, *100* (2), Forthcoming.

_ , John Driscoll, Xavier Gabaix, and David Laibson, "Learning in the credit card market," *NBER Working paper*, 2008.

Akerlof, George and Robert Shiller, *Animal spirits: how human psychology drives the economy, and why it matters for global capitalism*, Princeton University Press, 2009.

Banks, J. and Z. Oldfield, "Understanding Pensions: Cognitive Function, Numerical Ability and Retirement Saving," *Fiscal Studies*, 2007, *28* (2).

Barsky, Robert B., F. Thomas Juster, Miles S. Kimball, and Matthew D. Shapiro, "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study," *The Quarterly Journal of Economics*, 1997, *112* (2), 537–579.

Bergstresser, Daniel and John Beshears, "Who selected Adjustable-Rate Mortgages? Evidence from Surveys of Consumer Finances," Working Paper, Harvard Business School 2009.

Bernanke, Ben, "Financial literacy," Testimony before the Committee on Banking, Housing, and Urban Affairs of the United States Senate on May 23 2006.

Bernheim, B. Douglas and Daniel M. Garrett, "The effects of financial education in the workplace: evidence from a survey of households," *Journal of Public Economics*, 2003, *87* (7-8), 1487–1519.

Boeri, T. and L. Guiso, "Subprime crisis: Greenspan's Legacy," *VoxEU.org* 2007.

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

Burks, Stephen, Jeffrey Carpenter, Lorenz Goette, and Aldo Rustichini, "Cognitive skills affect economic preferences, strategic behavior, and job attachment," *Proceedings of the National Academy of Science*, 2009, *106* (19), 7745–7750.

Burks, Stephen V., Jeffrey P. Carpenter, Lorenz Goette, and Aldo Rustichini, "Cognitive skills affect economic preferences, strategic behavior, and job attachment," *Proceedings of the National Academy of Sciences*, 2009, *106* (19), 7745–7751.

Cooper, Daniel, "Did Easy Credit Lead to Overspending? Home Equity Borrowing and Household Behavior in the Early 2000s," *FRB Boston Public Policy Discussion Paper No. 09-7*, 2009.

Demyanyk, Y. and O. Van Hemert, "Understanding the Subprime Mortgage Crisis," *The Review of Financial Studies*, 2010, p. Forthcoming.

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde, "Are Risk Aversion and Impatience Related to Cognitive Ability?," *American Economic Review*, 2009, p. forthcoming.

_ , _ , _ , _ , Juergen Schupp, and Gert G. Wagner, "Individual risk attitudes: New evidence from a large, representative, experimentally-validated survey," *Journal of the European Economic Association*, 2005, p. Forthcoming.

Flynn, James R., *What Is Intelligence?: Beyond the Flynn Effect*, Cambridge and New York: Cambridge University Press, 2007.

Foote, Christopher, Kristopher Gerardi, Lorenz Goette, and Paul S. Willen, "Reducing Foreclosures: No Easy Answers," *NBER Macroeconomics Annual*, 2009, p. Forthcoming.

Foote, Christopher L., Kristopher Gerardi, and Paul S. Willen, "Negative Equity and Foreclosure: Theory and Evidence," *Journal of Urban Economics*, 2008, *64* (2), 234–245.

_ , _ , Lorenz Goette, and Paul S. Willen, "Just the Facts: An Initital Analysis of the Subprime Crisis," *Journal of Housing Economics*, 2008, *17* (4), 1–24.

Gerardi, Kristopher, Adam Hale Shapiro, and Paul S. Willen, "Subprime Outcomes: Risky Mortgages, Homeownership Experiences, and Foreclosures," *Federal Reserve Bank of Boston Working Paper 07-15*, 2007.

_ , Adam Shapiro, and Paul Willen, "Decomposing the Foreclosure Crisis: House Price Depreciation versus Bad Underwriting," *Federal Reserve Bank of Atlanta Working Paper 2009-25*, 2009.

Karlan, Dean and Jonathan Zinman, "Lying About Borrowing," *Journal of the European Economic Association*, 2008, *6* (2-3), 510–521.

Laibson, David, "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, 1997, *112* (2), 443–477.

Lang, Frieder R., Dorothee Hahne, Stefanie Gymbel, Stefan Schroepper, and Katharina Lutsch, "Erfassung des kognitiven Leisungspotenzials und der "Big Five" mit Computer-Assisted-Personal-Interviewing (CAPI): Zur Reliabilitaet und Validitaet zweier ultrakurzer Tests und des BFI-S," *DIW Research Notes, Berlin* 2005.

Lusardi, Annamaria and Olivia Mitchell, "How Ordinary Consumers Make Complex Economic Decisions: Financial Literacy and Retirement Readiness," *NBER Working Paper No. 15350*, 2009.

_ **and Peter Tufano**, "Debt Literacy, Financial Experience, and Overindebteness," Working Paper, Harvard University 2008.

Mayer, Christopher, Karen Pence, and S.M. Sherlund, "The rise in mortgage defaults," *Journal of Economic Perspectives*, 2009, *23* (1), 27–50.

McArdle, John J., James P. Smith, and Robert Willis, "Cognition and Economic Outcomes in the Health and Retirement Survey," Technical Report 15266, NBER Working Paper 2010.

Meier, Stephan and Charles Sprenger, "Stability of Time Preferences," *Working Paper*, 2009.

_ **and** _ , "Present-Biased Preferences and Credit Card Borrowing," *American Economic Journal: Applied Economics*, 2010, *2* (1), 193–210.

Mian, Atif and Amir Sufi, "The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis," *Quarterly Journal of Economics*, 2009, (124), 1449–1496.

Nadauld, Taylor D. and Shane M. Sherlund, "The Impact of Secondary Mortgage Market Activity on the Expansion of Subprime Credit," Working Paper, Brigham Young University 2009.

Pennington-Cross, Anthony and Giang Ho, "The Termination of Subprime Hybrid and Fixed Rate Mortgages ," *Real Estate Economics*, 2010, p. Forthcoming.

Sherlund, Shane, "The Past, Present, and Future of Subprime Mortgages," *Finance and Economics Discussion Series Federal Reserve Board 2008-63*, 2008.

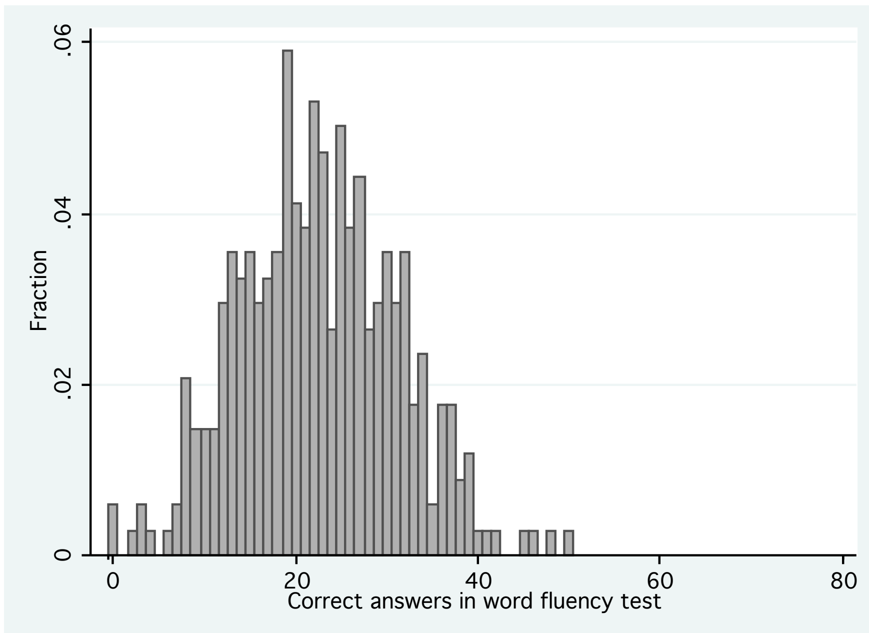
Shiller, Robert, *Irrational Exuberance*, Princeton University Press, 2005.

Stango, Victor and Jonathan Zinman, "Exponential Growth Bias and Household Finance," *Journal of Finance*, 2009, *64* (6), 2807–2849.

_ **and** _ , "Fuzzy Math, Disclosure Regulation, and Market Outcomes: Evidence from Truth-in-Lending Reform," *Review of Financial Studies*, 2009, p. Forthcoming.

Figure 1. Distribution of Verbal IQ Scores

Panel A: Distribution in this study



Panel B: Distribution in Dohmen et al. (2009)

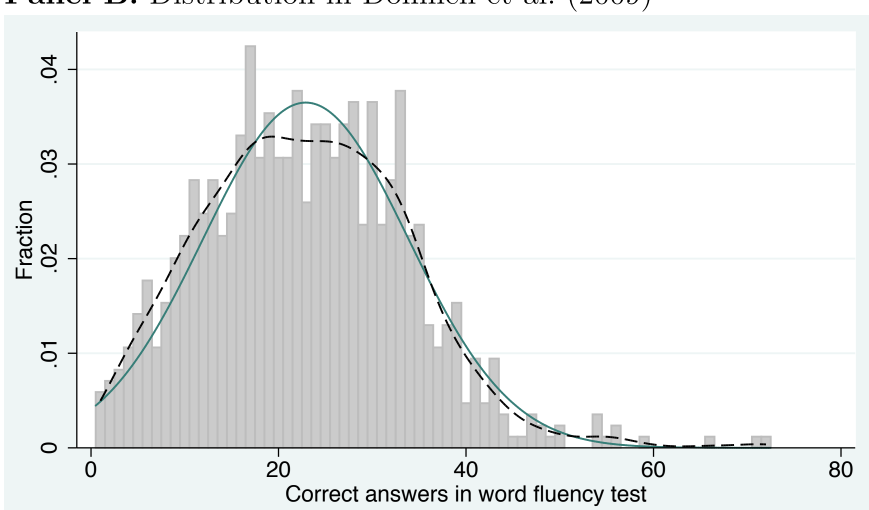


Figure 2. Delinquency and Numerical Ability Histograms

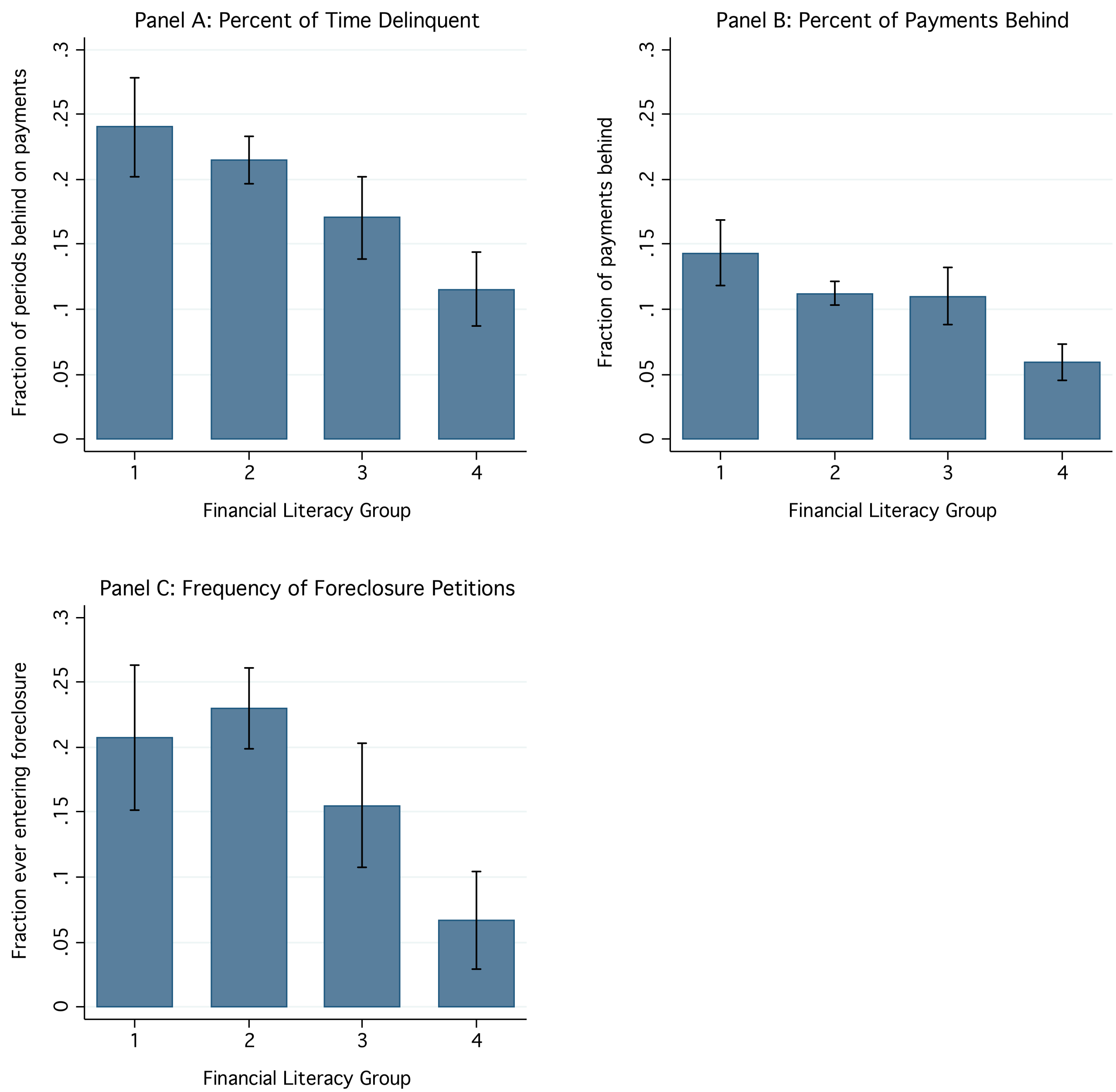


Table 1
Summary Statistics of Mortgage Data

	t-test of Differences NR vs. R				Summary Statistics	
	Cold Calls		Mail-Ins		Respondents	
	NR-R	*p*-value	*NR-R*	*p*-value	Means	Std. Dev.
FICO Score	-4.99	0.201	6.025	0.340	632	61.4
Fixed-Rate Mortage (=1)	-0.010	0.760	-0.026	0.523	0.339	.
Interest-Only (=1)	-0.011	0.516	0.020	0.483	0.083	.
Balloon Payment (=1)	-0.014	0.613	-0.035	0.490	0.257	.
Refinance (=1)	0.021	0.517	0.002	0.973	0.555	.
Loan-to-Value Ratio	0.814	0.324	0.147	0.899	78.6	12.7
Amount of Mortgage ($ thousands)	-14.5	0.068	20.6	0.119	245	130
Initial interest rate	0.118	0.123	0.243	0.053	7.94	1.19
Debt-to-Income Ratio	0.437	0.473	-1.48	0.136	42.0	8.18
Full-Doc Status (=1)	-0.018	0.532	0.060	0.245	0.699	.
Foreclosure after mailing went out (=1)	0.019	0.326	-0.004	0.901	0.094	.

Notes: The first part of the table shows differences of various mortgage characteristics between R and NR (NR-R) and *p*-values of *t*-tests of whether the differences are statistically significant. The information is based on 3,615 observations for the Cold Calls and 4,995 observations for the Mail-Ins. The information about *Debt-to-Income Ratio* is missing for a few observations in the Warren Group data. To compare *Foreclosure after mailing went out* we focus on individuals who were never in foreclosure between the origination and the date we contacted them. For some of the borrowers who were "current" on their mortgage when we contacted them, a foreclosure petition had been filed before and they may have already been in the process of moving out. The last two columns show summary statistics of all respondents (N= 339).

Table 2
Distribution of Numerical Ability Index

	Numerical Ability Group			
	1	2	3	4
This study:	15.6%	53.9%	17.1%	13.3%
Banks and Oldfield (2007):	16.2%	46.6%	26.8%	11.1%

Table 3
Correlation Between Measures of Cognitive Ability

	NA group	Verbal IQ measure	Savings scenario	Inflation scenario
Verbal IQ measure	0.356 (0.000)	1		
Savings scenario correct (DV)	0.236 (0.000)	0.153 (0.005)	1	
Inflation scenario correct (DV)	0.273 (0.000)	0.251 (0.000)	0.093 (0.087)	1
Reaction time in NA questions	-0.279 (0.000)	-0.303 (0.000)	-0.157 (0.004)	-0.207 (0.000)

Notes: $N = 339$. p-values in parentheses. A factor analysis performed on these correlations reveals one common factor ($\lambda = 1.17$), while all other eigenvalues are less than 0.005.

Table 4
Summary Statistics of Survey Questions

	Dummy?	Mean	Std. Dev.
Panel A: Time and Risk Preferences			
Discount Factor	No	0.964	0.026
Present Bias	Yes	0.201	.
Risk tolerance	No	1184.971	157.996
Panel B: Socio-Demographics			
Asian	Yes	0.012	.
African American	Yes	0.192	.
Hispanic	Yes	0.074	.
Native American	Yes	0.027	.
Other Ethnicity	Yes	0.027	.
Male	Yes	0.487	.
Age	No	46.6	10.3
Born in USA	Yes	0.838	.
# of Years lived in US	No	43.1	13.7
Married	Yes	0.631	.
Separated	Yes	0.029	.
Divorced	Yes	0.112	.
Single	Yes	0.192	.
Widowed	Yes	0.035	.
# of Children	No	2.09	1.49
No High School	Yes	0.032	.
Some High School	Yes	0.044	.
High School Degree	Yes	0.180	.
Some College	Yes	0.333	.
College Degree	Yes	0.271	.
Higher Degree	Yes	0.139	.
Fluency in English	No	9.74	1.05
Income ($ thousands)	No	80.0	57.4
Income Volatility	No	1.87	0.791
Employment Status	Yes	0.841	.
# of Years out of Work	No	0.730	1.54
Panel C: Mortgage Characteristics / Search Behavior			
First-time Homebuyer	Yes	0.552	.
Home counseling	Yes	0.091	.
Shop around	Yes	0.602	.

Notes: Based on 339 observations.

Table 5
Distribution of Delinquency Measures

			Percentiles				
	Mean	Std. Dev.	10	25	50	75	90
Fraction of periods during which household is behind on at least one payment	0.198	0.247	0	0	0.077	0.367	0.621
Fraction of missed payments	0.110	0.143	0	0	0.056	0.167	0.304
Foreclosure	0.192	.	.	.	.	.	.

Notes: $N = 339$ observations.

Table 6: The Baseline Result

	Fraction of Time in Delinquency			Fraction of Payments Missed			Foreclosure Initiated (=1)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Numerical Ability Index	–0.043***	–0.038*	–0.052***	–0.024***	–0.024**	–0.032***	–0.059**	–0.067**	–0.085***
	(0.014)	(0.020)	(0.019)	(0.009)	(0.011)	(0.010)	(0.026)	(0.030)	(0.028)
Log (Household Income)		–0.056**	–0.025		–0.022	–0.006		–0.062	–0.043
		(0.027)	(0.028)		(0.017)	(0.017)		(0.046)	(0.042)
Volatility of HH Income		0.038**	0.039**		0.022**	0.023**		0.048*	0.043
		(0.018)	(0.016)		(0.010)	(0.009)		(0.029)	(0.026)
Employed (DV) (d)		–0.001	–0.003		–0.002	–0.002		0.034	0.044
		(0.046)	(0.045)		(0.026)	(0.026)		(0.056)	(0.046)
# of times out of work in past		0.018**	0.009		0.006	0.002		–0.020	–0.026*
		(0.008)	(0.008)		(0.004)	(0.004)		(0.016)	(0.015)
Age		–0.003	–0.003		–0.001	–0.001		0.000	–0.000
		(0.003)	(0.003)		(0.002)	(0.002)		(0.005)	(0.005)
Male (DV) (d)		0.017	0.049		0.019	0.035**		0.120**	0.146***
		(0.032)	(0.031)		(0.018)	(0.017)		(0.049)	(0.046)
Asian (DV)		–0.162**	–0.285***		–0.108**	–0.171***			
		(0.077)	(0.087)		(0.049)	(0.050)			
African American (DV) (d)		0.116***	0.098**		0.077**	0.068**		0.177**	0.153**
		(0.044)	(0.040)		(0.030)	(0.027)		(0.076)	(0.074)
Hispanic (DV) (d)		0.003	0.015		0.007	0.011		–0.010	–0.016
		(0.056)	(0.055)		(0.029)	(0.027)		(0.093)	(0.081)
Native American (DV) (d)		–0.050	–0.031		–0.031	–0.024		0.036	0.005
		(0.110)	(0.100)		(0.054)	(0.040)		(0.185)	(0.180)
Other Ethnicity (DV) (d)		0.108	0.100		0.082	0.078		0.147	0.166
		(0.137)	(0.106)		(0.097)	(0.085)		(0.220)	(0.222)
Born in USA (DV) (d)		–0.038	–0.058		–0.023	–0.030		–0.152	–0.159
		(0.072)	(0.079)		(0.044)	(0.046)		(0.165)	(0.165)
Years lived in US		0.004	0.005		0.003	0.003		0.004	0.004
		(0.003)	(0.003)		(0.002)	(0.002)		(0.005)	(0.005)
Fluency in English		0.002	0.007		0.005	0.008		0.018	0.019
		(0.017)	(0.015)		(0.008)	(0.008)		(0.029)	(0.026)
Some High School (DV) (d)		–0.106	–0.072		–0.035	–0.016		–0.001	0.017
		(0.096)	(0.099)		(0.052)	(0.051)		(0.182)	(0.177)
High School Degree (DV) (d)		–0.073	–0.057		–0.013	–0.004		0.111	0.123
		(0.074)	(0.075)		(0.044)	(0.043)		(0.202)	(0.196)
Some College (DV) (d)		–0.024	0.002		0.010	0.024		0.244	0.260

Table 6: (continued)

	Fraction of Time in Delinquency			Fraction of Payments Missed			Foreclosure Initiated (=1)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		(0.074)	(0.075)		(0.043)	(0.042)		(0.195)	(0.189)
College Degree (DV) (d)		0.022	0.042		0.025	0.038		0.230	0.257
		(0.078)	(0.078)		(0.044)	(0.042)		(0.209)	(0.208)
Professional Degree (DV) (d)		0.020	0.048		0.019	0.039		0.311	0.384
		(0.080)	(0.079)		(0.045)	(0.044)		(0.255)	(0.266)
Married (DV) (d)		–0.036	–0.048		–0.053	–0.057		–0.166	–0.139
		(0.082)	(0.061)		(0.049)	(0.038)		(0.132)	(0.127)
Separated (DV) (d)		0.005	–0.001		0.015	0.017		0.021	0.030
		(0.113)	(0.091)		(0.083)	(0.070)		(0.160)	(0.160)
Divorced (DV) (d)		–0.080	–0.106		–0.093*	–0.106***		–0.158***	–0.135***
		(0.086)	(0.065)		(0.050)	(0.038)		(0.044)	(0.036)
Single (DV) (d)		–0.049	–0.044		–0.066	–0.064		–0.128*	–0.110*
		(0.090)	(0.067)		(0.052)	(0.040)		(0.074)	(0.064)
Number of Children		0.008	0.001		0.000	–0.003		0.017	0.015
		(0.011)	(0.010)		(0.006)	(0.006)		(0.016)	(0.014)
Estimated δ		–0.643	–0.298		–0.416	–0.251		–0.941	–0.562
		(0.548)	(0.529)		(0.313)	(0.298)		(0.816)	(0.734)
Present-Biased (DV) (d)		–0.026	–0.023		–0.007	–0.004		–0.065	–0.059
		(0.034)	(0.033)		(0.019)	(0.018)		(0.046)	(0.039)
Risk preferences		0.000	0.000		0.000	0.000		0.000	0.000
		(0.000)	(0.000)		(0.000)	(0.000)		(0.000)	(0.000)
FICO Score / 10			–0.015***			–0.009***			–0.020***
			(0.002)			(0.001)			(0.004)
Home Purchase (DV) (d)			–0.036			–0.009			0.049
			(0.036)			(0.019)			(0.061)
Months since home purchased			–0.000			–0.000**			–0.000
			(0.000)			(0.000)			(0.000)
Originated in 2007 (DV) (d)			–0.050			–0.021			–0.032
			(0.035)			(0.019)			(0.044)
Investor (DV) (d)			–0.009			0.001			0.010
			(0.057)			(0.032)			(0.106)
Constant	0.296***	0.881	1.452***	0.164***	0.498	0.857***			
	(0.037)	(0.598)	(0.547)	(0.023)	(0.323)	(0.306)			

Table 6: (continued)

	Fraction of Time in Delinquency			Fraction of Payments Missed			Foreclosure Initiated (=1)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
R^2	0.023	0.153	0.268	0.021	0.153	0.265			
F-test of H_0: All coefficients are equal to zero	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$
N	339	322	322	339	322	322	339	318	318

Notes: Robust standard errors in columns (1) - (6). ***, **, * indicate significance at the 1, 5, 10 percent level, respectively. Regression coefficients are reported in columns (1) - (6). Marginal effects from probit model are reported in columns (7) - (9).

Table 7
Controlling for General Cognitive Skills and Economic Literacy

	Fraction of Time in Delinquency		Fraction of Payments Missed		Foreclosure Initiated (=1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Numerical Ability Index	– 0.047**	– 0.051***	– 0.031***	– 0.033***	– 0.065**	– 0.061**
	(0.019)	(0.019)	(0.011)	(0.011)	(0.027)	(0.028)
Verbal IQ measure	– 0.001	– 0.002	0.000	0.000	– 0.006**	– 0.006**
	(0.002)	(0.002)	(0.001)	(0.001)	(0.003)	(0.003)
Savings Scenario correct (DV)		0.002		– 0.000		– 0.051
		(0.036)		(0.021)		(0.058)
Inflation scenario correct (DV)		0.006		0.011		– 0.016
		(0.033)		(0.018)		(0.047)
Reaction time in NA questions		– 0.003		– 0.000		– 0.001
		(0.002)		(0.001)		(0.003)
Control variables?	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.262	0.268	0.242	0.244		
F-test of H_0: All coefficients are equal to zero.	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$
N	322	322	322	322	318	318

Notes: Regression coefficients are reported in columns (1) - (4). Marginal effects from probit models are reported in columns (5) - (6). Robust standard errors in parentheses in columns (1) - (4). All specifications contain the full set of control variables as in Table 6.

Level of significance: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$

Table 8
Correlation between Numerical Ability and Mortgage Attributes

	(1) Unconditional	(2) Conditional
Fixed-Rate Mortgage (DV)	0.018	0.029
	(0.029)	(0.038)
Initial Interest Rate	-0.119*	-0.148**
	(0.068)	(0.074)
Low-Doc Loan (DV)	-0.065**	-0.058
	(0.029)	(0.036)
Prepayment Penalty (DV)	0.017	-0.001
	(0.031)	(0.040)
Log (origination amount)	0.069**	-0.037
	(0.029)	(0.031)
Loan-to-Value ratio	0.003	-0.013
	(0.011)	(0.012)
Debt-to-Income ratio	-0.874*	-0.342
	(0.514)	(0.676)

Notes: The table displays the estimated correlation between the numerical ability index and various mortgage attributes from regressions in which the mortgage attribute is the dependent variable. The first column reports the regression coefficient without control variables. The second column reports the regression coefficient with all of the controls from Table 6 included.
Level of significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 9
Controlling for Mortgage Attributes

	Fraction of Time in Delinquency		Fraction of Payments Missed		Foreclosure Initiated (=1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Numerical Ability Index	– 0.042**	– 0.032*	– 0.028***	– 0.026**	– 0.073***	– 0.043**
	(0.019)	(0.020)	(0.011)	(0.011)	(0.028)	(0.023)
Fixed-Rate Mortgage (DV)	0.036	0.033	0.015	0.011	0.028	0.028
	(0.029)	(0.029)	(0.017)	(0.017)	(0.047)	(0.037)
Initial Interest Rate	0.022	0.035*	0.006	0.012	0.016	0.008
	(0.017)	(0.018)	(0.009)	(0.009)	(0.021)	(0.015)
Low-Doc Loan (DV)	0.029	– 0.001	0.011	– 0.008	0.029	0.005
	(0.032)	(0.035)	(0.019)	(0.020)	(0.044)	(0.034)
Prepayment Penalty (DV)	– 0.005	0.032	0.008	0.031*	0.033	0.056
	(0.027)	(0.030)	(0.016)	(0.017)	(0.042)	(0.037)
Log (origination amount)		0.115***		0.070***		0.093**
		(0.037)		(0.023)		(0.044)
Loan-to-value ratio		– 0.009		0.018		0.517***
		(0.088)		(0.052)		(0.143)
Debt-to-income ratio		0.002		0.001		0.003*
		(0.002)		(0.001)		(0.002)
Control variables?	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.302	0.335	0.288	0.331		
F-test of H_0: All coefficients are equal to zero.	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$
N	315	288	315	288	311	286

Notes: Regression coefficients are reported in columns (1) - (4). Marginal effects from probit models are reported in columns (5) - (6). Robust standard errors in parentheses in columns (1) - (4). All specifications contain the full set of control variables as in Table 6.
Level of significance: $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 10
Controlling for Previous Mortgage Market Experience

	Fraction of Time in Delinquency	Fraction of Payments Missed	Foreclosure Initiated (=1)
Numerical Ability Index	– 0.049***	– 0.030***	– 0.075***
	(0.019)	(0.010)	(0.028)
Number of prev. mortgages	0.001	0.005	0.019
	(0.009)	(0.005)	(0.014)
First home purchase (DV)	0.046	0.023	0.089**
	(0.031)	(0.017)	(0.043)
Shopped around before getting mortgage (DV)	0.029	0.017	0.017
	(0.029)	(0.016)	(0.040)
Sought counceling for home buyers (DV)	– 0.026	– 0.003	– 0.068
	(0.050)	(0.029)	(0.053)
Attended home owner classes (DV)	– 0.009	– 0.010	0.127
	(0.047)	(0.024)	(0.102)
Control variables?	Yes	Yes	Yes
R^2	0.270	0.255	
F-test of H_0: All coefficients are equal to zero.	$p < 0.01$	$p < 0.01$	$p < 0.01$
N	322	322	318

Notes: Regression coefficients are reported in columns (1) and (2). Marginal effects from probit models are reported in column (3). Robust standard errors in parentheses in columns (1) and (2). All specifications contain the full set of control variables as in Table 6.
Level of significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 11
Including Town, Servicer, and Originator Fixed Effects

	Fraction of Time in Delinquency			Fraction of Payments Missed			Foreclosure Initiated (=1)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Numerical Ability Index	– 0.081**	– 0.055***	– 0.044***	– 0.045**	– 0.037***	– 0.026***	– 0.105*	– 0.106***	– 0.065**
	(0.033)	(0.020)	(0.015)	(0.022)	(0.011)	(0.009)	(0.056)	(0.032)	(0.025)
Town Fixed Effects?	Yes	No	No	Yes	No	No	Yes	No	No
Originator Effects?	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Servicer Effects?	No	No	Yes	No	No	Yes	No	No	Yes
Control Variables?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.735	0.361	0.350	0.690	0.358	0.337	0.668	0.318	0.298
F-test of H_0: All coefficients are equal to zero	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$
N	319	307	293	319	307	293	319	307	293

Notes: Regression coefficients are reported in columns (1) - (6). Marginal effects from probit models are reported in columns (7) - (9). Robust standard errors in parentheses in columns (1) - (6). All specifications contain the full set of control variables as in Table 6.
Level of significance: $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

A. Appendix Tables and Figures (NOT FOR PUBLICATION)

Table A1: The Baseline Result in Tobit Models

	Fraction of Time in Delinquency		Fraction of Payments Missed	
	(1)	(2)	(3)	(4)
Numerical Ability Index	–0.063***	–0.068***	–0.035***	–0.042***
	(0.023)	(0.026)	(0.013)	(0.015)
Log (Household Income)		–0.066		–0.028
		(0.042)		(0.024)
Volatility of HH Income		0.054**		0.031**
		(0.025)		(0.014)
Employed (DV)		0.025		0.015
		(0.054)		(0.032)
# of times out of work in past		0.016		0.006
		(0.012)		(0.007)
Age		–0.006		–0.003
		(0.005)		(0.003)
Male (DV)		0.067		0.047*
		(0.041)		(0.024)
Asian (DV)		–0.590***		–0.353***
		(0.227)		(0.132)
African American (DV)		0.143***		0.096***
		(0.050)		(0.029)
Hispanic (DV)		0.037		0.021
		(0.078)		(0.045)
Native American (DV)		–0.044		–0.034
		(0.128)		(0.074)
Other Ethnicity (DV)		0.113		0.095
		(0.140)		(0.080)
Born in USA (DV)		–0.042		–0.020
		(0.114)		(0.066)
Years lived in US		0.007		0.004
		(0.005)		(0.003)
Fluency in English		0.006		0.008

Table A1: (continued)

	Fraction of Time in Delinquency		Fraction of Payments Missed	
	(1)	(2)	(3)	(4)
		(0.021)		(0.012)
Some High School (DV)		–0.154		–0.061
		(0.133)		(0.077)
High School Degree (DV)		–0.119		–0.038
		(0.112)		(0.065)
Some College (DV)		–0.046		–0.003
		(0.108)		(0.063)
College Degree (DV)		0.028		0.030
		(0.110)		(0.064)
Professional Degree (DV)		0.030		0.026
		(0.118)		(0.068)
Married (DV)		–0.051		–0.060
		(0.104)		(0.060)
Separated (DV)		–0.028		0.009
		(0.143)		(0.082)
Divorced (DV)		–0.169		–0.147**
		(0.112)		(0.065)
Single (DV)		–0.038		–0.064
		(0.111)		(0.064)
Number of Children		0.013		0.003
		(0.014)		(0.008)
Estimated δ		–0.474		–0.385
		(0.706)		(0.409)
Present-Biased (DV)		–0.042		–0.013
		(0.046)		(0.027)
Risk preferences		0.000		0.000
		(0.000)		(0.000)
FICO Score / 10		–0.024***		–0.014***
		(0.003)		(0.002)
Home Purchase (DV)		–0.081		–0.033
		(0.054)		(0.031)
Months since home purchased		–0.001*		–0.001**
		(0.000)		(0.000)

Table A1: (continued)

	Fraction of Time in Delinquency		Fraction of Payments Missed	
	(1)	(2)	(3)	(4)
Originated in 2007 (DV)		–0.066		–0.028
		(0.047)		(0.027)
Investor (DV)		0.024		0.022
		(0.097)		(0.056)
Constant	0.249***	2.265***	0.137***	1.342***
	(0.056)	(0.779)	(0.032)	(0.451)
sigma				
Constant	0.350***	0.292***	0.202***	0.169***
	(0.018)	(0.015)	(0.011)	(0.009)
N	339	322	339	322

Notes: Coefficients of tobit models. Robust standard errors in columns (1) - (4). σ is the estimated standard deviation of the residual.

Level of significance: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$

Table A2
Dummy Variables of NA Categories Instead of Linear Term

	Fraction of Time in Delinquency		Fraction of Payments Missed		Foreclosure Initiated (=1)	
	(1)	(2)	(3)	(4)	(5)	(6)
NA Index = 2 (DV)	− 0.025	− 0.037	− 0.031	− 0.040	0.020	− 0.016
	(0.042)	(0.042)	(0.027)	(0.027)	(0.058)	(0.054)
NA Index = 3 (DV)	− 0.070	− 0.108**	− 0.033	− 0.056	− 0.050	− 0.084*
	(0.050)	(0.053)	(0.033)	(0.035)	(0.065)	(0.048)
NA Index = 4 (DV)	− 0.125***	− 0.142**	− 0.084***	− 0.104***	− 0.142***	− 0.145***
	(0.048)	(0.060)	(0.029)	(0.034)	(0.052)	(0.032)
Control variables?	No	Yes	No	Yes	No	Yes
F-Test: all coefficients of NA are zero	$p = 0.01$	$p = 0.04$	$p < 0.01$	$p = 0.01$	$p = 0.07$	$p = 0.02$
F-Test: Relationship is linear	$p = 0.86$	$p = 0.81$	$p = 0.6$	$p = 0.7$	$p = 0.28$	$p = 0.31$
R^2	0.024	0.261	0.026	0.247		
N	339	322	339	322	339	318

Notes: Regression coefficients are reported in columns (1) - (4). Marginal effects from probit models are reported in columns (5) - (6). Robust standard errors in parentheses in columns (1) - (4). All specifications contain the full set of control variables as in Table 6.
Level of significance: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$

Table A3: Full Results of Correlation Between NA and Mortgage Terms

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Numerical Ability Index	0.029	–0.148**	–0.058	–0.001	–0.037	–0.013	–0.342
	(0.038)	(0.074)	(0.036)	(0.040)	(0.031)	(0.012)	(0.676)
Log (Household Income)	–0.016	–0.078	–0.068	0.019	0.287***	–0.016	–1.436
	(0.062)	(0.119)	(0.058)	(0.064)	(0.051)	(0.019)	(1.111)
Volatility of HH Income	–0.059	–0.126*	0.071**	–0.036	0.065**	–0.021*	0.600
	(0.037)	(0.071)	(0.035)	(0.038)	(0.030)	(0.012)	(0.653)
Employed (DV)	–0.059	–0.113	–0.041	0.039	0.102	0.058**	2.775*
	(0.082)	(0.157)	(0.076)	(0.085)	(0.067)	(0.026)	(1.441)
# of times out of work in past	0.021	0.037	0.002	0.017	–0.030*	–0.005	–0.226
	(0.019)	(0.036)	(0.018)	(0.020)	(0.015)	(0.006)	(0.331)
Age	0.002	0.004	0.008	0.019***	–0.007	–0.002	0.113
	(0.007)	(0.014)	(0.007)	(0.007)	(0.006)	(0.002)	(0.128)
Male (DV)	–0.038	0.121	–0.035	0.041	0.088*	0.020	1.349
	(0.062)	(0.120)	(0.058)	(0.064)	(0.051)	(0.020)	(1.101)
Asian (DV)	0.062	–0.104	–0.086	–0.101	–0.234	–0.184**	–4.391
	(0.267)	(0.513)	(0.249)	(0.275)	(0.218)	(0.082)	(7.203)
African American (DV)	–0.067	–0.006	–0.025	–0.031	0.088	0.017	0.764
	(0.076)	(0.147)	(0.071)	(0.079)	(0.062)	(0.024)	(1.352)
Hispanic (DV)	0.117	–0.444**	–0.153	0.114	–0.057	0.012	–1.028
	(0.116)	(0.223)	(0.108)	(0.120)	(0.095)	(0.037)	(2.142)
Native American (DV)	0.129	0.058	0.376**	–0.041	0.074	0.062	5.966*
	(0.190)	(0.364)	(0.177)	(0.195)	(0.155)	(0.059)	(3.475)
Other Ethnicity (DV)	0.166	0.112	–0.100	–0.005	–0.134	0.019	–2.391
	(0.205)	(0.394)	(0.191)	(0.211)	(0.167)	(0.063)	(3.577)
Born in USA (DV)	–0.016	0.063	0.073	0.304*	–0.233*	0.013	1.171
	(0.159)	(0.305)	(0.148)	(0.169)	(0.129)	(0.049)	(2.805)
Years lived in US	0.005	–0.009	–0.007	–0.013*	0.008	0.002	–0.046

Table A3: (continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	(0.007)	(0.013)	(0.006)	(0.007)	(0.005)	(0.002)	(0.122)
Fluency in English	–0.014	–0.009	–0.013	0.004	0.009	0.007	–0.118
	(0.033)	(0.063)	(0.030)	(0.034)	(0.027)	(0.010)	(0.571)
Some High School (DV)	–0.098	0.011	–0.051	–0.088	–0.128	–0.164**	–1.684
	(0.206)	(0.397)	(0.193)	(0.213)	(0.168)	(0.067)	(3.936)
High School Degree (DV)	–0.016	0.049	–0.092	–0.336*	–0.240*	–0.081	–3.981
	(0.177)	(0.340)	(0.165)	(0.182)	(0.144)	(0.058)	(3.391)
Some College (DV)	–0.039	0.249	0.027	–0.204	–0.203	–0.084	–3.113
	(0.171)	(0.329)	(0.160)	(0.177)	(0.140)	(0.056)	(3.310)
College Degree (DV)	–0.033	0.042	0.039	–0.218	–0.102	–0.073	–3.628
	(0.175)	(0.335)	(0.163)	(0.180)	(0.142)	(0.057)	(3.357)
Professional Degree (DV)	–0.124	0.037	–0.058	–0.153	0.060	–0.092	–5.745
	(0.185)	(0.356)	(0.173)	(0.191)	(0.151)	(0.060)	(3.522)
Married (DV)	0.135	0.199	0.098	–0.006	0.052	0.019	–0.084
	(0.164)	(0.316)	(0.153)	(0.169)	(0.134)	(0.051)	(3.012)
Separated (DV)	0.168	0.442	0.022	–0.256	–0.206	0.060	0.543
	(0.220)	(0.423)	(0.206)	(0.227)	(0.180)	(0.068)	(3.978)
Divorced (DV)	0.082	0.044	–0.008	0.237	0.018	–0.001	1.681
	(0.174)	(0.334)	(0.163)	(0.180)	(0.142)	(0.054)	(3.165)
Single (DV)	0.103	0.295	0.080	–0.072	–0.086	0.039	2.148
	(0.175)	(0.336)	(0.163)	(0.180)	(0.143)	(0.054)	(3.204)
Number of Children	–0.001	0.004	0.038*	–0.043*	0.004	0.006	0.311
	(0.021)	(0.041)	(0.020)	(0.022)	(0.017)	(0.007)	(0.380)
Estimated δ	–0.229	4.351**	2.902***	–0.899	1.516*	0.601*	20.499
	(1.065)	(2.051)	(0.995)	(1.107)	(0.869)	(0.333)	(18.769)
Present-Biased (DV)	–0.075	–0.025	–0.091	–0.070	–0.062	0.012	–2.056*
	(0.068)	(0.132)	(0.064)	(0.071)	(0.056)	(0.022)	(1.223)
Risk preferences	0.000	0.000	0.000	–0.000	0.000	0.000*	0.001

Table A3: (continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)
FICO Score / 10	0.007	–0.108***	0.013***	0.000	0.008*	0.003	0.005
	(0.005)	(0.009)	(0.005)	(0.005)	(0.004)	(0.002)	(0.087)
Home Purchase (DV)	0.069	0.421***	0.035	0.145*	–0.293***	–0.028	–1.882
	(0.081)	(0.156)	(0.076)	(0.084)	(0.066)	(0.026)	(1.436)
Months since home purchased	0.001*	–0.002	–0.000	–0.000	–0.000	–0.001***	–0.026**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.012)
Originated in 2007 (DV)	–0.093	0.288**	–0.075	0.053	–0.012	–0.045**	–1.520
	(0.072)	(0.138)	(0.067)	(0.075)	(0.059)	(0.022)	(1.273)
Investor (DV)	0.159	0.739***	0.307**	0.169	–0.079	–0.124***	–5.856**
	(0.144)	(0.277)	(0.135)	(0.149)	(0.118)	(0.045)	(2.673)
Constant	–0.195	11.026***	–2.348**	1.065	9.251***	0.139	25.388
	(1.176)	(2.266)	(1.099)	(1.224)	(0.960)	(0.369)	(20.765)
R^2	0.092	0.398	0.194	0.133	0.385	0.279	0.131
N	322	321	322	316	322	312	303

Notes: Coefficients of OLS models. Standard errors in parenthesis. Dependent variable: (1) Fixed-Rate Mortgage (DV); (2) Initial Interest Rate; (3) Low-Doc Loan (DV); (4) Prepayment Penalty (DV); (5) Log (origination amount); (6) Loan-to-Value ratio; (7) Debt-to-Income ratio.
Level of significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

B. An illustrative model of imperfect numerical abilities

This model looks at the consumption aspect of housing/mortgage choices.

1. In period 0, individuals have to choose the size of their house, x. We assume full leverage so that the size of the mortgage is exactly equal to the size of the house, and thus x represents both the size of the house and the size of the mortgage.[27] We assume that individuals can either purchase a house of size x, which generates a utility flow of ϕx, or it can rent a house and derive zero utility.[28]

2. We assume that the only way the individual can change its house size is by defaulting, in which case it is banished to the rental market for the rest of its life. Thus, in periods 1 to T the individuals cannot change the size of their house (without defaulting). Instead, they have to decide how much to consume in each period subject to their intertemporal budget constraint.

We assume for notational convenience that there is no consumption in period 0. Also for simplicity, we ignore discounting and and set the interest rate to zero.

The individual has an income m at his disposal in every period, and capital markets are perfect so that only the present-value budget constraint applies. Utility from consumption in period t is given by a concave function $u(c_t)$ that is separable from the utility of owning the house ϕx.

A. The rational benchmark

We solve the model by backward induction.

A.1. Consumption Smoothing:

For a given size house x chosen in period 0, the individual will choose a consumption stream such that it maximizes intertemporal utility. Because there is no discounting and no interest, it is optimal to have a constant consumption across all periods.

[27]In our data the median subprime borrower had a cumulative loan-to-value ratio (cumulative in the sense that it includes all mortgage liens) at origination of 100%.

[28]Assuming that renting yields zero utility is without loss of generality, as the critical assumption is that houses in the rental market are smaller than owner-occupied houses, so that rental housing will yield less utility than owner-occupied housing. Alternatively, we could assume that rental houses are the same size as owner-occupied houses, but that individuals simply derive additional utility from owning their own home.

Define the available assets for consumption starting in period 1 as $a_1 = T(m - x)$. Then, $a_{t+1} = a_t - c_t$. Optimal consumption follows

$$c_t = c_s = \frac{a_1}{T} = \frac{a_{s-1} - c_{s-1}}{T - s + 1} = m - x \tag{4}$$

A.2. House/Mortgage Choice:

The individual solves the following maximization problem:

$$\max_x \sum_{t=0}^{T} u(m - x) + \phi x$$

For an interior solution ($x > 0$), the optimal choice of x is given by

$$u'(m - x) = \phi$$

so that $x^* = m - u'^{-1}(\phi)$ and $c^* = u'^{-1}(\phi)$. If $u'(m) \geq \phi$ then we have a corner solution where $x^* = 0$. Notice that because we assume quasi-linear utility in housing, any additional income will be allocated toward a larger house.

We assume that at the beginning of each period the individual can decide to default on his mortgage and transition to the rental market. At the beginning of each period the individual compares the present value of continuing to live in his current house and paying the current mortgage, x^*, versus defaulting, and moving into rental housing. We define the present value of owning and renting at time t as:

$$\begin{aligned} V_t^{own} &= \textstyle\sum_{j=0}^{T-t}[u(c^*) + \phi x^*] &= (T - t + 1)[u(c^*) + \phi x^*] \\ V_t^{rent} &= \textstyle\sum_{j=0}^{T-t}[u(m)] &= (T - t + 1)u(m) \end{aligned}$$

The individual will choose to default when $V_t^{own} < V_t^{rent}$. Note that in this model, a fully rational individual that chooses to own in period 0 will never choose to default on his mortgage in any subsequent period t, as along the optimal path defaulting would yield per-period utility $u(m) < \max_x u(m - x) + \phi x$.

B. Limited Numerical Ability

We assume that individuals with imperfect numerical ability misperceive their budget constraint. Instead of income m, they believe that they have income $m + \eta$ at their disposal in each future period. This is one way to interpret an individual's inability to perform simple mathematical calculations like addition and multiplication, which they were required to do in the survey questions that measured numerical ability. We will explore the impact of η on an individual's consumption stream. We assume that each individual receives a draw of η that applies to all periods.

We assume that individuals with poor numerical ability receive a draw from a symmetric distribution with a higher variance. Specifically, we assume that larger values of η are more likely for individuals with poor numerical ability. This can be represented as

$$\frac{\partial \Pr(|\eta| > y)}{\partial NA} < 0$$

for all y for increasing levels of numerical ability NA. We assume that $E(\eta|NA) = 0$, for all levels of numerical ability, so that limitations do not bias the perception of income in any way across individuals.

In this model, the individual maximizes utility given his expectation of income (i.e., given his η). Each period, he will have to revise his plans, as he notices a difference η between the predicted cash at hand $\tilde{a}_{t|t-1}$ and the actual cash at hand in period t, a_t. We assume for simplicity that the individual is constantly surprised about this and that there is no learning about η.

B.1. Consumption Smoothing.

Given a mortgage of size x in period 1, the individual believes that he has lifetime resources $a_1 + (T-1)\eta = T(m-x) + (T-1)\eta$ at his disposal. He plans to smooth consumption given his perceived wealth $\tilde{a}_1 = a_1 + (T-1)\eta$ over T periods:

$$c_1 = \frac{\tilde{a}_1}{T} = (m - x) + \frac{T-1}{T}\eta = u'^{-1}(\phi)$$

In period 1, the individual thinks that next period (and in all future periods) he is going to consume the same amount:

$$\tilde{c}_{2|1} = \tilde{c}_{3|1} = ... = \tilde{c}_{T|1} = c_1$$

However, in period 2, the individual has to revise his belief about the assets available to him to:

$$\tilde{a}_2 = (T-1)(m-x) + (T-2)\eta - [c_1 - (m-x)].$$

where he realizes that his income in period 1 fell short of his expectations by η, and as a result he consumed more than actual cash-on-hand $(m-x)$. Thus, he adjusts his lifetime wealth estimate by the amount of over-consumption, $c_1 - (m-x)$.

His consumption plan in period 2 is now to consume

$$c_2 = \frac{\tilde{a}_2}{T-1} = u'^{-1}(\phi) - \frac{\eta}{T-1} = c_1 - \frac{\eta}{T-1}$$

in each period. It is easy to see that for positive η:

$$c_2 - c_1 = \frac{-\eta}{T-1} < 0$$

In general, it is easy to show that the individual will choose consumption in period $t > 1$ according to

$$c_t = u'^{-1}(\phi) - \sum_{j=2}^{t} \frac{\eta}{T-j+1}$$

thinking that he can maintain this consumption level for the rest of his life because he misperceives the income in future periods by η. In particular, he thinks that in period $t+1$, he will consume

$$\tilde{c}_{t+1|t} = c_t$$

However, in period $t+1$, he will adjust his consumption downward to what he now thinks he will maintain for the rest of the periods:

$$c_{t+1} = u'^{-1}(\phi) - \sum_{j=2}^{t+1} \frac{\eta}{T-j+1}$$

It follows that for positive η:

$$\triangle c_{t+1} = c_{t+1} - c_t = \frac{-\eta}{T-t} < 0$$

The model has a number of features

- If individuals make errors that lead them to overestimate their income in the future relative to their consumption ($\eta > 0$), then consumption will be steadily declining over time. Conversely, if individuals make the opposite error, their consumption will be constantly growing over time.

- If $E(\eta|NA) = 0$, then it is impossible to identify limitations in numerical ability by looking at changes in consumption alone, since on average,

$$E(c_{t+1} - c_t|NA) = \frac{-E(\eta|NA)}{T-t} = 0$$

  Thus, it is, in general, impossible to distinguish individuals with limited numerical ability from perfectly rational individuals by looking at changes in consumption. The only prediction is about the variance of consumption.

- Notice also that this last property is very different from saying that these are simply errors that "cancel out." In terms of welfare, they don't cancel out. Individuals with $\eta > 0$ consistently consume too much and have to cut consumption dramatically over time, which reduces their welfare. Individuals with $\eta < 0$ consistently consume too little and save too much, lowering their welfare as well.

B.2. House/Mortgage Choice

Because the individual now thinks that he has income $m + \eta$ at his disposal in each period, the interior solution becomes

$$x^* = m + \frac{T-1}{T}\eta - u'^{-1}(\phi)$$

The individual aims for the same consumption level as a perfectly rational consumer given by $u'(m+\eta-x) = \phi$. For example, if $\eta > 0$ the individual will commit to a bigger mortgage than a perfectly rational individual, but start out in period 1 with the same consumption level c^* given by $c^* = u'^{-1}(\phi)$.

B.3. Implications for Mortgage Default.

We focus on the case in which the individual chooses to own in period 0 (i.e. $x^* > 0$), which is the only interesting case. The present value of owning in period $t > 1$ is given by:

$$V_t^{own} = (T - t + 1)u\left(\frac{\tilde{a}_t}{T - t + 1}\right) + (T - t + 1)\phi x^*$$

The present value of defaulting and renting in period $t > 1$ is given by:

$$V_t^{rent} = (T - t + 1)u\left(\frac{\tilde{a}_t}{T - t + 1} + x^*\right)$$

where $\tilde{a}_t = m - x^* + \frac{T-t}{T-t+1}\eta - \frac{(t-1)(T-1)}{T(T-t+1)}\eta + \sum_{j=2}^{t-1}\left(\frac{\eta}{T-j+1}\right)$ is perceived remaining lifetime wealth in period t net of mortgage payments x^*.

So the individual will choose to default in period t if $V_t^{rent} > V_t^{own}$:

$$\phi x^* < u\left(\frac{\tilde{a}_t}{T - t + 1} + x^*\right) - u\left(\frac{\tilde{a}_t}{T - t + 1}\right) \tag{5}$$

Assuming that $u(0) = -\infty$, equation (5) guarantees the existence of a reservation consumption level $r(x)$ for each possible choice of mortgage x, where the individual will be indifferent between defaulting on the mortgage and keeping the house:

$$u(r(x) + x) = u(r(x)) + \phi x.$$

That is, if per-period assets $\tilde{a}_t/(T - t - 1)$ fall below $r(x)$, the individual is better off defaulting on the mortgage, and consuming what remains of his assets. Moreover, the reference level of consumption is increasing in x: Differentiating with respect to x yields

$$r'(x) = \frac{u'(r(x) + x) - \phi}{u'(r(x)) - u'(r(x) + x))} > 0 \tag{6}$$

The denominator is greater than zero because marginal utility is decreasing in consumption. The numerator is greater than zero because $r + x < c$ (otherwise, the individual would not default!).

There are a number of important implications for mortgage defaults from this simple model. First, individuals with $\eta < 0$, will never default, as they under-predict their wealth.

They will simply have a growing consumption path. By contrast, individuals with $\eta > 0$ will start out at consumption level c^* and will realize that this is not sustainable. They will have a decreasing consumption (and a huge mortgage). This creates an incentive to default.

Thus, individuals with a bigger mortgage will default at a higher consumption level.[29] Taken together, the model provides us with a number of interesting results. First, this implies that individuals with poor numerical ability will be more likely to default. Why? Individuals with poor numerical ability are more likely to have a large and positive η. As a consequence, they will get a larger mortgage, which also raises $r(x)$. But because all individuals start at c^*, those with a large η also experience a quicker decline in consumption and will hit the consumption default threshold, $r(x)$, more quickly.[30] Intuitively, what is needed to observe limitations in numerical ability impacting behavior are choices that affect multiple periods, like the purchase of a home. Combined with the need to continuously revise consumption downward, this produces the result that defaults become more likely with limited numerical ability.

Second, conditional on loan size, numerical ability should still be predictive of default. Because individuals with poor numerical ability are more likely to have a decreasing consumption profile, they should hit $r(x)$ quicker, conditional on x. This corresponds to the result in the paper that conditional on loan size, numerical ability remains a significant predictor of default.

[29] There are two forces that produce this result: First, a bigger mortgage gives a higher utility. Thus, in order to be indifferent between renting and owning, consumption needs to be higher. This is reinforced by the high value of x, which increases the marginal utility of consumption if the individual keeps his mortgage.

[30] The fact that all individuals start at c^* follows from quasi-linearity of the utility function. If individuals with $\eta > 0$ also chose a higher initial consumption level, the argument would break down. But remember that this is an exercise to show that limitations in numerical abilities *can* produce the pattern we find in the data. It's not supposed to be a general characterization of behavior with limited numerical abilities.