Women are Left Behind: Social Norms, Math Skills and Mortgage Outcomes*

Athena Tsouderou†, Şelale Tüzel‡

December 2024

Abstract

Analyzing a near-universe sample of U.S. conventional mortgages originated and securitized in 2018-2019, we find that single women pay higher mortgage interest rates and are less likely to refinance when rates decline compared to single men, particularly in high-income, highly educated neighborhoods. This gender gap partly reflects differences in math skills and financial literacy, which tend to widen with socioeconomic status. These findings align with prevailing gender stereotypes and societal norms, contributing to disparities in financial sophistication that result in less favorable mortgage outcomes for women.

Keywords: Gender gap, mortgages, gender stereotypes, math skills, financial literacy.
JEL Classification: G21, G51, G53, D14, J16, I24.

---

*We thank Kenneth Ahern, Stefania Albanesi, Marieke Bos (discussant), Tom Chang, Vidhi Chhaochharia, Cary Frydman, Pedro Gete, Andra Ghent, Marco Giacoletti, Larry Harris, Erica Jiang, Yigitcan Karabulut, Amir Kermani, Xiang Li, Quinn Maingi, Bianca Minuth (discussant), Jordan Nickerson, Rodney Ramcharan, Miao Ben Zhang, participants at the Boca Finance and Real Estate Conference, Empirical Financial Intermediation Research Network Workshop, Financial Management & Accounting Research Conference, and seminar participants at the University of Miami and USC for their constructive and insightful comments.

†Herbert Business School, University of Miami, email: atsouderou@miami.edu.

‡Marshall School of Business, University of Southern California, email: tuzel@usc.edu.

1 Introduction

A substantial body of literature highlights gender disparities across various domains, including financial decision-making, investment returns, negotiation, and preferences for risk and competition.[1] Much of this research focuses on average outcome differences between men and women, prompting questions about whether inherent gender traits or societal attitudes contribute to these disparities. A key challenge is distinguishing whether observed differences arise from innate traits (nature) or environmental and societal factors (nurture) (Croson and Gneezy, 2009). Moreover, focusing solely on average (unconditional) disparities may overlook significant variations within the gender gap, where heterogeneity can be greater than the overall difference. Thus, the absence of an unconditional gap does not necessarily imply a lack of meaningful disparities within the data.

In this paper, we investigate gender disparities in mortgage outcomes, a critical financial decision with significant implications for household wealth. Beyond examining unconditional gender gaps, we analyze how these disparities vary across the socioeconomic spectrum. Our findings reveal that, within the same Census tract and time period, female borrowers pay an interest rate premium of approximately 3 basis points and a 2-basis-point increase in the rate spread (APR minus the prime offer rate) at origination. Women also incur net loan costs 16 basis points higher than men, including points, fees, and lender credits as a fraction of the loan amount. These disparities are more pronounced in higher-income, higher-education, and higher-socioeconomic-status neighborhoods; for example, in Census tracts with incomes one standard deviation above the mean, the rate spread gap increases by 80% after controlling for observable loan and borrower characteristics such as income, credit score, loan-to-value ratio, debt-to-income ratio, race, ethnicity, and age.

Female borrowers also refinance less frequently than males, particularly when market interest rates fall below their contracted mortgage rate. On average, while the prepayment probability is 2% per month, this rate is over 10% lower for female borrowers, with the disparity doubling in higher-income regions. Notably, while men in affluent areas exhibit higher prepayment rates, the refinancing behavior of women remains largely unchanged with tract income after accounting for loan and borrower characteristics. These findings underscore persistent gender gaps in mortgage costs and refinancing behavior, exacerbated in affluent regions, and highlight the role of socioeconomic factors and education in shaping these disparities.

[1] See Sunden and Surette (1998), Babcock and Laschever (2003), Gneezy et al. (2003), Niederle and Vesterlund (2007), Bertrand (2011), Andersen et al. (2021), and Goldsmith-Pinkham and Shue (2023) for examples of gender gaps in these domains.

Linking initial mortgage terms to refinancing behavior, we estimate that, over a 10-year period, female borrowers in high-income, highly educated Census tracts pay an additional mortgage cost equivalent to approximately 1% of the loan value. In lower-income, lower-education areas, this excess cost is about 0.5%. These estimates account for risk-based pricing differences relative to single men with similar loans in the same Census tract and time period. For the average mortgage amount of $231,000 in our sample, this translates to roughly $2,200 in excess payments for women in high-education neighborhoods and $1,240 in lower-education areas.

Our results go against the conventional wisdom that associates higher socioeconomic status with greater financial sophistication and anticipates a reduction in gender gaps, rather than an increase. However, regions characterized by higher socioeconomic status actually exhibit unexpected gender gaps. Reardon et al. (2019) documents that while, on average, male and female students score similarly on math tests, boys tend to outperform girls in math in socioeconomically privileged school districts. Although the study does not establish causal relationships, it suggests that gender disparities in math proficiency may be shaped by gender stereotypes and parental resources. These resources often promote investment in activities aligned with traditional gender roles, which tend to favor math-related pursuits for boys, but not for girls. Consistent with this view, Dossi et al. (2021) show that math gender gaps among children are most prevalent in affluent white families that exhibit biased family gender norms. In our analysis, we identify similar gender gaps in math skills among adults, using a novel measure of gender disparity that considers the relevance of math skills across occupations. This new measure reveals that, much like children's math scores, math skills are more relevant in jobs held by women in low-income, low-education Census tracts. In high-income, high-education tracts, however, this pattern reverses, with math skills being more critical in jobs held by men.

There is a compelling rationale to suggest a connection between gender disparities in math skills and gender gaps in mortgage outcomes. Several recent studies have highlighted the positive effects of enhanced math education on both labor market prospects (Goodman, 2019) and financial decision-making (Brown et al., 2016). Research by Cole et al. (2016) underscores the greater significance of high school math education in enhancing financial outcomes compared to personal finance courses. Additionally, Carpena et al. (2011) found that a financial education program that does not specifically address numeracy has little impact on an individual's ability to make accurate financial decisions.[2] Gerardi et al. (2013) finds that numerical ability was a strong predictor of mortgage defaults, even after controlling for cognitive ability, general

[2]Recent research on financial education differentiates the impact of business education from that of general quantitative skills on advanced financial outcomes. Altmejd et al. (2024) shows that college enrollment in business and economics programs has a causal positive effect on financial outcomes, supporting the notion that financial education and math skills are complementary.

economics knowledge, and a broad set of sociodemographic variables. Lusardi (2012) notes the widespread lack of numerical ability, particularly among certain demographic groups such as women, and highlights the strong correlation between numeracy and financial literacy and decision-making. Finally, Adams et al. (2023) found that the math gender gap predicts the proportion of women in investment professions across countries and U.S. states, underscoring the broader implications of math skills for gender gaps in financial domains.

Motivated by this body of evidence, we examine whether gender gaps in math skills help explain differences in mortgage outcomes between men and women. Our findings reveal that in areas where boys outperform girls in math in school, women tend to secure loans with significantly higher interest rates and loan costs and are less likely to refinance, especially when falling interest rates increase the value of refinancing. This disparity persists even after controlling for gender gaps in reading (English language arts, ELA) test scores. We obtain similar results when substituting childhood math scores with math skill requirements based on adults' occupations. While our study does not establish a causal link between gender disparities in math skills and mortgage outcomes, it highlights a plausible pathway through which these effects may emerge, particularly in socioeconomically advantaged areas.

Mortgages are complex financial products, encompassing a range of upfront and recurring costs for borrowers. Upfront costs include points and fees, which come in various forms like administrative fees, origination fees, funding fees, document preparation fees, discount points, and more. Additionally, borrowers pay various third-party fees, including appraisal fees, credit report fees, and various charges related to title and escrow. These fees, paid to various parties such as brokers, lenders, and title and escrow companies, can often be negotiated but are complex, especially for first-time borrowers. To add to this, lenders might offer credits against these costs, introducing an additional factor for consideration. Beyond these upfront costs, borrowers are also committing to making regular interest payments. The decision-making process involves weighing various expenses while considering factors such as the expected duration of the loan, liquidity needs, and individual time preferences, allowing borrowers to trade off upfront costs for the recurring cost, interest rate. Similar to Bhutta and Hizmo (2021), we examine both dimensions of mortgage costs — interest rates and upfront fees — along with the rate spread, derived from the loan's APR and accounting for both interest rates and upfront expenses, revealing similar gender gaps across all metrics.

Recent studies have shown that borrowers frequently overpay on origination fees due to limited knowledge of the mortgage market and inadequate loan shopping practices, resulting in substantial losses (Woodward and Hall, 2012; Agarwal et al., 2017). Studies by Bhutta et al. (2024) and Malliaris et al. (2022) underscore the tendency of less financially savvy homeowners to significantly overpay for mortgages, emphasizing the importance of informed shopping and negotiation behaviors. Expanding on these findings, using data from the National Survey of Mortgage Originations (NSMO), we identify significant differences in the loan shopping behavior between single female and male borrowers: Female borrowers typically initiate the mortgage process with less market information, explore fewer lender and loan options, and are less inclined to verify the reasonableness of the offered mortgage terms through additional sources.

Using NSMO survey data, we assess the financial literacy of single male and female borrowers and investigate how it varies across the socioeconomic spectrum based on their responses to questions about various mortgage concepts and features. Our analysis reveals that male borrowers, on average, exhibit higher levels of financial sophistication, with sophistication levels increasing across all borrowers as income and education levels rise, consistent with existing literature findings (Lusardi and Mitchell, 2014). However, we observe a notable divergence in the income and education gradients of financial sophistication between males and females.[3] While comprehension of simpler concepts such as the "consequences of missing mortgage payments" or the "difference between fixed- and adjustable-rate mortgages" increases similarly with education and income for both genders, for more complex concepts that involve a deeper understanding of mathematical relationships, like the "relationship between discount points and interest rate" or the "difference between mortgage interest rate and APR," the education and income gradients are considerably flatter for women, with gender gaps widening as education and income levels rise. This widening gap is significant and aligns with our core finding of increasing gender disparities in mortgage outcomes in more affluent areas, as understanding these advanced literacy concepts is essential for making informed mortgage decisions.

Beyond selecting the right mortgage, the timing and cost of refinancing are crucial for most borrowers (Campbell, 2006). The decision to refinance is complex, requiring careful assessment of costs versus benefits and necessitating proactive action from borrowers. Research indicates that borrowers often make suboptimal refinancing decisions, either by neglecting to refinance when rates drop or by choosing suboptimal rates and timing. Keys et al. (2016) found that 20% of unconstrained borrowers with outstanding U.S. mortgages in 2010 failed to refinance during rate declines, losing substantial savings. Agarwal et al. (2016) reported that 57% of U.S. borrowers refinance suboptimally, with the largest mistakes made by less financially sophisticated borrowers. Bajo and Barbi (2018) found that 87% of Italian mortgage

[3] The most commonly used financial literacy questions, developed by Annamaria Lusardi and Olivia Mitchell, are the "Big Three" and the "Big Five." The "Big Three" questions test understanding of interest rates, inflation, and risk diversification, while the "Big Five" also include bond pricing and mortgage payments. These questions gauge basic financial literacy, focusing on fundamental principles for everyday financial decisions. The NSMO survey, specifically targeting existing mortgage borrowers, questions understanding of mortgage concepts and features.

borrowers missed optimal refinancing opportunities post-policy change, particularly among the less financially literate. Extending these findings, we demonstrate that single women engage in mortgage refinancing less frequently than single men, particularly during periods of declining interest rates. These results persist even after including an extensive array of controls and fixed effects.

Our study adds to the broader investigation into gender disparities within financial outcomes, particularly in the context of housing investments. Research by Goldsmith-Pinkham and Shue (2023) highlights that single women in the U.S. tend to earn lower returns on housing investments compared to their male counterparts. They attribute approximately half of this discrepancy to timing in the housing market and the selection of neighborhoods. The other half of the gap is explained by women listing their houses at lower prices and obtaining less favorable negotiation outcomes than men. Conversely, Andersen et al. (2021) presents findings from Denmark showing no significant gender differences in negotiation outcomes for house purchases once adjustments are made for the distinct preferences in house characteristics between single men and single women. This finding aligns with our observation of small average gender differences in mortgage outcomes across men and women, despite notable variations across the socioeconomic spectrum. In our analysis, we employ rigorous controls for location and timing (using Census tract $\times$ origination year and month fixed effects) and compare loans issued at the same time for properties in close proximity, after adjusting for all loan characteristics that could influence pricing. This approach enables us to isolate the influence of gender from other factors that may lead to mortgage cost disparities, allowing us to focus on the residual gender gap within different neighborhoods.

Our research also relates to studies investigating gender disparities in borrowing practices, particularly in the context of business loans. Previous work primarily examines differences in business loan acquisition between female and male entrepreneurs. For example, in an analysis of loans from a major Italian bank, Bellucci et al. (2010) discovered that while women encounter more difficulties in accessing credit, they do not face higher interest rates compared to their male counterparts. Conversely, Alesina et al. (2013), utilizing data from the Italian credit registry, found that women are subject to higher rates for business loans, demonstrating that these discrepancies are not solely attributable to gender differences in credit history or the choice of banking institution. In the U.S. mortgage market, Cheng et al. (2011) documented distinct lender selection strategies between genders. Their findings reveal that approximately 40% of women base their lender choice on recommendations, while 20% prioritize the lowest cost. These preferences are inverted among men. Such differing approaches to lender selection significantly contribute to the disparity in mortgage rates paid by men and women.

Additionally, there is an active literature examining racial disparities in mortgage lending, with recent contributions from studies such as Conklin et al. (2023); Giacoletti et al. (2021); Li (2023). Previous research consistently shows that racial minorities experience higher interest rates across various mortgage types. Ghent et al. (2014) document that racial minorities face elevated interest rates for subprime, private-label, securitized mortgages, while Bartlett et al. (2022) find similar disparities in interest rates for GSE-securitized and FHA-insured mortgages. Furthermore, Ambrose et al. (2021) highlight increased broker fees on subprime mortgages for racial minorities, and Bhutta and Hizmo (2021) observe that racial minorities encounter higher interest rates for FHA-insured home purchase loans, although these disparities are partially offset by differences in discount points. Holding barriers to refinancing constant, Gerardi et al. (2023) find that racial minorities refinance less than White borrowers.

In summary, our paper makes a pioneering contribution to the literature by exploring cross-sectional variations in gender disparities, with a specific focus on mortgage outcomes. We find that gender gaps in mortgage outcomes systematically vary across communities along the socioeconomic spectrum and are closely tied to gender gaps in local math skills, both among children and adults. These findings indicate that gender gaps in mortgage outcomes are not merely the result of innate differences between men and women but are significantly shaped by local factors, including social norms and attitudes.

2 Data

The primary analysis in this paper draws on a novel dataset that we developed by merging multiple publicly available sources. Our core dataset combines the Home Mortgage Disclosure Act (HMDA) dataset with loan-level origination and performance data from Fannie Mae and Freddie Mac. Additionally, it incorporates student math and reading test scores from the Stanford Education Data Archive (SEDA) and occupational math skill importance, constructed using data from O*NET, the Bureau of Labor Statistics (BLS), and the American Community Survey (ACS). We supplement this dataset with tract-level income, education and household structure data from the ACS. Additionally, we utilize the public use file of the National Survey of Mortgage Originations (NSMO) to explore potential mechanisms underlying our main findings. We provide a description of each component of our core dataset and the merging procedures in the following sections, with additional information available in the Appendix.

2.1 HMDA Dataset

The HMDA requires financial institutions to publicly disclose loan-level information about mortgage applications and originations. The disclosures contain demographic information, such as the applicant's gender and race, and essential details such as the applicant's income, the loan amount, lender name, and the Census tract of the property. Overall, this dataset includes 90% of mortgage originations in the US and has been used extensively in previous studies. The HMDA data serves multiple purposes, one of which is to provide information for evaluating whether financial institutions are effectively addressing the housing needs of all members within their communities.

In response to the 2010 Dodd-Frank Wall Street Reform and Consumer Protection Act (Dodd-Frank Act) and the 2015 definitive rule from the Consumer Financial Protection Bureau (CFPB), the HMDA data was augmented to incorporate additional details on mortgage loans with the purpose to enhance understanding of the mortgage market. The implementation and commencement of data collection under the revised HMDA data framework began on January 1, 2018.

This expanded dataset provides comprehensive information on loan pricing, covering aspects such as interest rates, discount points, lender credits, points and fees, and negative amortization. Furthermore, it incorporates factors related to applicant creditworthiness, including debt-to-income ratio and loan-to-value (LTV) ratio, as well as specifics about the financed property, such as whether it involves manufactured housing with or without land. These newly added factors are commonly used by lenders in their loan decision-making processes and are essential controls for our study.

Utilizing the expanded HMDA dataset, our sample includes mortgages originated on or after January 1, 2018. We conclude our sample period for HMDA loans on December 31, 2019, avoiding the highly volatile and extraordinary years that ensued due to COVID-19 policies affecting the mortgage market. Our final HMDA dataset consists of all conventional first-lien mortgage originations for single-family properties between January 1, 2018, and December 31, 2019.

2.2 GSE Datasets

The expanded HMDA data provides deeper insights into how borrowers and lenders interact in the mortgage market by offering more detailed information on mortgage pricing. However, a key limitation is the absence of borrower credit scores in the publicly available data. To address this, we merge HMDA loans with loan-level datasets from Fannie Mae and Freddie Mac (the GSEs), which include borrower credit scores. The merged dataset not only incorporates credit scores but also provides precise mortgage origination month and year, more granular debt-to-income ratios, and, crucially for refinancing analysis, the monthly performance of loans. These additional variables from the GSE data significantly enhance the depth and accuracy of our analysis.

Since there is no unique identifier to connect the loans in the different datasets, we develop a very thorough algorithm to merge them. Our algorithm matches the loans based on the lender, year of origination, loan amount, interest rate, LTV ratio, property type, occupancy status, and loan purpose. The geographical matching links the Census tract in the HMDA data with the 3-digit zip code in the GSE data. We only keep loans that have a unique match between the datasets. This combined dataset contains approved loans securitized by the GSEs between 2018 and 2020.[4]

2.3 Additional Datasets

We obtain data on median household income, educational attainment, and household structure at the Census tract level from the U.S. Census Bureau's American Community Survey (ACS), the leading demographics survey conducted by the U.S. Census Bureau. We use the 5-year estimates ending in 2017. Educational attainment is measured as the percentage of the population aged 25 and older with a bachelor's degree or higher. For household structure, we calculate the share of households headed by single women with children in each tract. This is defined as the ratio of single female-headed households with children ("Female householder, no husband present, with own children under 18") to the total number of single female-headed households. Similarly, we calculate the share of households headed by single men with children in each tract.

We source data on standardized test scores from the Stanford Education Data Archive (SEDA; see Reardon et al., 2021), a project under the Educational Opportunity Project at Stanford University. We use data on the gender gap in math and reading (English language arts, ELA) scores at the school district level for grades 3 through 8, spanning from 2009 to 2018. In each district, we calculate the average of the cohort-standardized gender gap. To link the test scores with HMDA loan data, we match school districts to the Census tracts they

[4]The algorithm outlined in this study, which combines the public HMDA dataset with the public GSE datasets (explained in detail in Section A1 in the Appendix), offers a notable methodological contribution to the field. Since both datasets are publicly accessible, the code used to replicate this merging process can assist many researchers in addressing limitations present in the HMDA data, following a similar approach.

cover. While most tracts align with a single district, some sparsely populated tracts may span multiple districts. In these cases, we ensure an accurate match by requiring that at least 50% of the tract's land area falls within a single district to avoid multiple matches. This process uses the 2019 School District Geographic Relationship Files from the National Center for Education Statistics.

In addition to standardized test scores, the SEDA dataset provides socioeconomic variables at the school-district level. We incorporate the socioeconomic status (SES) composite variable, which captures the average characteristics of parents in public schools. This composite includes median household income, the share of adults with a bachelor's degree or higher, child poverty rates, unemployment rates, the proportion of households receiving food stamps, and the percentage of single mother–headed households, and is used by Reardon et al. (2019) to examine cross-sectional differences in gender achievement gaps across U.S. school districts.

We develop a novel measure of the math skill gender gap by combining the occupational composition of males and females in each Census tract with the math skill importance assigned to each occupation. We obtain data on the mathematical skills and knowledge required for nearly one thousand occupations from the Occupational Information Network (O*NET). O*NET is a comprehensive database that provides detailed information about U.S. occupations, relying on large-scale surveys of workers, employers, and experts. For each occupation, O*NET reports a score for "math skill importance," from 1 to 5, which measures the degree to which mathematical abilities are required to perform the tasks associated with the specific occupation.[5]

To calculate the gender gap in occupational math skill importance within each Census tract, we first compute the weighted average math skill importance for major occupation groups. This is done by aggregating math skill importance across detailed occupations, weighted by 2017 national employment figures from the Bureau of Labor Statistics' Occupational Employment and Wage Statistics (OEWS). We then calculate the weighted average math skill importance separately for males and females in each Census tract, using gender-specific employment numbers from the 2017 ACS 5-year estimates for each major occupational group. Finally, the gender gap is measured as the difference in math skill importance scores between males and females. A similar approach is used to calculate the occupational "economics and accounting knowledge" gap, based on the corresponding O*NET scores.

In addition to the core datasets described above, we use data from the National Survey

[5]We present the analysis for math skill importance, and replicate it using alternative measures from O*NET for math skill level, math knowledge level and math knowledge importance. These alternative metrics yield nearly identical results.

of Mortgage Originations (NSMO) to explore the mechanisms behind our findings. This nationally representative survey of newly originated residential mortgages in the United States is jointly conducted by the Federal Housing Finance Agency (FHFA) and the Consumer Financial Protection Bureau (CFPB). The survey captures borrowers' experiences obtaining a mortgage, their perceptions and understanding of the mortgage market, and importantly, links these responses to administrative data such as borrower and loan characteristics. Our analysis utilizes data from waves 1-30 of the survey, covering the period from the first quarter of 2014 to the second quarter of 2021.

2.4 Summary Statistics

Table 1 summarizes the key variables in our study for mortgages given to single borrowers, while Tables 2 and 3 present the key variables for low- and high-income tracts and low- and high-education tracts, respectively, as well as for male and female borrowers separately. We winsorize all continuous variables employed in the empirical analysis at the 0.5% level to limit the potential effects of outliers.

Overall, 42% of single borrowers in our sample are female. There is no robust relationship between the share of single female borrowers and socioeconomic status: Female borrowers constitute a slightly larger percentage of single borrowers in lower-income tracts (43% compared to 40%) but a slightly lower share in lower-education tracts (41.5% compared to 42.2%).[6] Unsurprisingly, we find a strong relationship between tract income and education: sorting by tract income yields nearly the same distribution in tract education as sorting by education, and vice versa. This is further supported by a 0.72 pairwise correlation between tract income and education, detailed alongside other local socioeconomic variables in Table A1 in the Appendix.

Gender gaps in student math and reading test scores are reported as the difference between the average cohort-standardized male score and female score, expressed in standard deviations.[7] Similar to Reardon (2019), we find significant gender gaps in reading scores, favoring females by 0.24 standard deviations. The gender gaps in reading scores do not vary systematically with the income and education levels of the district. In contrast, the gender gaps in math scores are small but exhibit systematic variation with tract income and education. Average math gender gaps are -0.02 (favoring females) in both low-income and low-education tracts, whereas they are 0.002 (favoring males) in high-income and high-education tracts. While these differences appear small,

[6]Figure A1 in the Appendix shows the geographical distribution of the share of single female borrowers in the US counties.

[7]For example, a gender gap of 0.5 indicates that the average male student scored approximately one half of a standard deviation higher than the average female student in that same cohort.

the distribution of math gaps across low- and high-income and education tracts, as shown in Figure 1, demonstrates a clear relationship between gender math gaps and these socioeconomic variables. Similarly, occupational math skill gender gaps, also in standard deviations, show a 0.12 to 0.14 gap favoring women in low-income and low-education tracts, reversing to a 0.02 to 0.04 gap favoring men in high-income and high-education tracts. The distribution of these skill gaps across socioeconomic levels, as depicted in Figure 2, reveals pronounced differences in gap distribution.

The average mortgage rate in our sample is 4.4%, and the average rate spread, which is the difference between the covered loan's APR and the average prime offer rate for a comparable transaction as of the date the interest rate is set, is 0.5%. Mortgage rates in low-income and education tracts exceed those in high-income and education tracts by approximately 15 basis points, and rate spreads are higher by about 20 basis points. We observe gender gaps in raw averages, which are more pronounced in high-income and education tracts. Average mortgage rates for females are 2 basis points higher than for males in low-income and education tracts, and 5 and 6 basis points higher in high-income and education tracts. Gender differences in rate spreads are smaller, with a difference of 2 and 3 basis points in high-income and education tracts and no difference in low-income and education tracts (rounded to two significant digits). Loan costs average 1.9% of the loan amount, which are approximately 0.7% higher in low-income and education neighborhoods. Females pay 12 and 14 basis points more in net loan costs compared to males in low-income and education neighborhoods, with this spread rising to 18 and 19 basis points in high-income and education neighborhoods.

The average credit score of borrowers is 748. The difference in credit scores between borrowers in low-income and high-income neighborhoods is small, about 7 points, whereas it is slightly larger, about 11 points, when comparing low-education and high-education tracts. There is no difference in the credit scores of male versus female borrowers. In both low- and high-income and education tracts, male borrowers have higher incomes and loan amounts, slightly higher loan-to-value ratios, and lower debt-to-income ratios compared to female borrowers.

White borrowers make up approximately 70% of our sample.[8] The proportion of White borrowers is slightly higher among single female borrowers, at 70.8% in high-income tracts and 72.9% in high-education tracts, compared to 67.9% and 69.5%, respectively, for single male borrowers.

[8]The "White" category includes borrowers with "White" as their reported race in HMDA, excluding those who indicated an ethnicity of "Hispanic or Latino." The remaining racial and ethnic breakdowns are Hispanic or Latino (8.7%), Asian (7.7%), Black (5.7%), American Indian (0.5%), Hawaiian (0.3%), Two or more minority races (0.1%), and Race not available (7.4%).

To understand the household structure of single borrowers, specifically whether they have children, we supplement the loan-level data with ACS data on single-adult households in the relevant Census tracts, as household structure details aren't available at the loan level. According to these combined data, 51.5% of households led by single females have children, with this proportion slightly decreasing in more affluent neighborhoods. Among single-male-led households, 47.6% have children, and this share increases slightly in more affluent neighborhoods. We revisit these statistics in our supplementary analyses and robustness checks.

Appendix Tables A2 and A3 expand upon Tables 1-3 by summarizing key variables for all mortgages in the HMDA-GSE merged dataset, including both single and joint mortgages. Table A2 presents summary statistics for single female borrowers (23.2% of loans), single male borrowers (32.2%), female primary borrowers in joint mortgages (11.7%), and male primary borrowers in joint mortgages (32.9%). The primary borrower is typically the individual who completes and submits the mortgage application, likely leading decision-making throughout the process.

While our analysis primarily focuses on single borrowers, comparisons of joint applications led by male versus female primary borrowers reveal consistent trends. For instance, among joint mortgages, the average mortgage rate for female primary borrowers is 4 basis points higher than for male primary borrowers, mirroring the raw gender gap observed among single borrowers. Additionally, the raw gender gap in the average rate spread is 5 basis points for joint mortgages, compared to 2 basis points for single borrowers. On average, joint mortgages feature higher loan amounts but lower debt-to-income ratios than those of single borrowers. Table A3 further details these statistics for each of the four borrower groups, segmented by low- and high-income tracts as well as low- and high-education tracts.

3 Gender Gaps in Mortgage Costs

Our empirical strategy to estimate the gender gaps in mortgage costs utilizes ordinary least squares (OLS) regressions of different measures of mortgage costs on an indicator for female borrowers. Our rich dataset allows us to control for Census tract $\times$ origination year and month fixed effects, meaning the estimated effect of female borrowers is within a Census tract in the same origination year and month. Moreover, we include numerous controls, as we describe below, to isolate the effect of gender. To obtain our first set of results for the unconditional gender effect we estimate the following regression:

$$Y_{i,j,c,t} = \beta_1 Female_i + \Gamma Z_{i,j} + a_c \times a_t + \epsilon_{i,j,c,t}, \tag{1}$$

where $Y_{i,j,c,t}$ represents the outcome of interest, for borrower i, with loan characteristics j, in Census tract c, and loan origination month t. The outcomes of interest are the mortgage interest rate and the rate spread, reported in HMDA and defined as the difference between the covered loan's APR and the average prime offer rate for a comparable transaction on the date the interest rate is set. Additional outcomes of interest are the total loan cost, comprising total points and fees, net of lender credits, as a fraction of the loan amount at origination, the lender credits as a fraction of the loan amount at origination and the net loan cost, which is the difference between the total loan cost and the lender credits as a fraction of the loan amount at origination.

$Female_i$ is an indicator variable that takes the value of 1 if the borrower is female and 0 if the borrower is male. $\Gamma Z_{i,j}$ is a rich set of borrower and loan characteristics that could affect the pricing of the mortgage. The pricing controls include dummies for nine loan-to-value (LTV) categories, nine credit score categories, manufactured homes, investment properties, high balance mortgages (that conform with high-cost area loan limits), borrowers with subordinate mortgages in addition to the first lien, loan purpose (home purchase, refinancing, or cash-out refinancing), and the number of units (1 to 4). These controls follow the loan-level price adjustment grids of Fannie Mae and the corresponding grids of Freddie Mac, which determine the credit-risk pricing adjustments to the guarantee fee that lenders must pay the GSEs for guaranteeing the mortgage.[9] Additional controls include dummies for eight borrower race and ethnicity categories, income deciles, loan amount deciles, debt-to-income deciles, dummies for three loan term intervals, and dummies for seven borrower age intervals. $a_c \times a_t$ denotes Census tract $\times$ origination year and month fixed effects. The estimation allows for clustering of standard errors at the Census tract level.[10]

Table 4 presents the results of the estimation of equation (1). The columns progressively include more controls for each mortgage outcome. Columns (1) to (3) present the estimates for the dependent variable of the mortgage interest rate. We find that mortgages to single female borrowers have an interest rate that is 2.8 basis points higher than that for single male borrowers within the same Census tracts, originated in the same year and month. The gender

[9] Bartlett et al. (2022) argue that the pricing grid accounts for all legitimate business necessity variables for GSE loans, and any pricing differences between loans beyond this grid cannot reflect differential credit risk, potentially indicating discrimination.

[10] Double clustering of standard errors at the Census tract and origination year-month level yields similar results when estimating equation (1).

gap increases to 4.2 basis points when we add the pricing controls.[11] This indicates that factors determining the borrower's credit risk and the mortgage pricing by the GSEs, such as the LTV ratio and credit score, do not explain the gender gap in the mortgage rate. On the contrary, controlling for these pricing factors actually increases the gender gap.

We find evidence that part of this gender gap in the mortgage rate is explained by other factors, which theoretically should not affect pricing beyond the credit-risk pricing adjustments. When we include additional controls such as borrower race, income, age and loan amount, the gender gap in mortgage rates is reduced to 0.8 basis points.

Columns (4) to (6) present the estimates for the dependent variable of rate spread. We observe that mortgages issued to single female borrowers have a rate spread that is 1.7 basis points higher than those issued to single male borrowers. This gap widens to 2.4 basis points after accounting for pricing controls, representing 4.8% of the average rate spread and 5.2% of its standard deviation. However, when additional controls are introduced, the gap narrows to 0.6 basis points.

Columns (7) to (9) present the estimates for the dependent variable of net loan cost. We find that mortgages issued to single female borrowers carry a net loan cost that is 17 basis points higher than those for single male borrowers. This difference decreases to 12 basis points after accounting for pricing controls, and further reverses to -0.6 basis points with the inclusion of additional controls.

In columns (10) and (11), we separately analyze the two components of net loan costs—fees and lender credits—presenting the results with all controls for brevity. This decomposition reveals that, on average, single female borrowers pay lower upfront fees than single male borrowers at the mortgage origination, but they also receive fewer lender credits after accounting for all controls. Overall, our findings indicate modest but statistically significant gender gaps favoring men in terms of interest rates, rate spreads, and mortgage credits.

3.1 Local Socioeconomic Factors and Gender Gaps

We next analyze how local socioeconomic factors are related to the gender gap in mortgage costs. As documented in Tables 2 and 3, the raw gender gaps in mortgage costs vary with the income and education levels of the local area. We now examine the relationship between these

[11] To gauge the magnitude of this effect, we can compare it to recent estimates of racial gaps in mortgage rates by Bartlett et al. (2022), which use a similar pricing grid to control for credit risk. In their estimates with Census tract x year fixed effects, reported in Table 5, they find that Black and Latinx borrowers pay approximately 2 basis points more for purchase GSE loans and 0.8 basis points more for refinance GSE loans.

gender gaps and various measures of income and education, along with a composite measure of socioeconomic status, using the following regression analysis:

$$Y_{i,j,c,t} = \beta_1 Female_i + \beta_2 Female_i \times \textit{Local socioeconomic indicator}_c + \Gamma Z_{i,j} + a_c \times a_t + \epsilon_{i,j,c,t}, \quad (2)$$

where all variables are defined as before. *Local socioeconomic indicator*$_c$ is the median household income or the level of educational attainment in Census tract c, which is measured as the share of the tract population over 25 years old with a bachelor's degree or higher, from the American Community Survey (ACS).[12] The linear effect of these local factors on the variables of interest is absorbed by the fixed effects. To facilitate the interpretation of results, these local factors are standardised to have a mean of zero and standard deviation of one. The estimation allows for clustering of standard errors at the Census tract level.[13]

Table 5 presents the results of estimating equation (2), which includes the interaction between the female indicator and the median tract income. As previously shown, the unconditional gender gap in mortgage rates, after controlling for all variables, stands at 0.8 basis points, while the gap in rate spreads is 0.6 basis points. Columns (3) and (6) of Table 5 reveal that in areas with median incomes one standard deviation above the mean, the mortgage rate gap widens by 50%, and the rate spread gap increases by 83%. Column (9) indicates that in these higher-income areas, the net loan cost gap increases by 0.8 basis points, effectively offsetting the unconditional gender gap of -0.6 basis points in net loan costs. This increase in net loan cost for women in affluent areas is driven by both a relative rise in fees and a reduction in lender credits.

Table 6 documents the cross-sectional variation in gender gaps based on local educational attainment. Focusing on the results with the full set of controls (Columns (3) and (6)), we find that in areas with a one standard deviation higher share of college graduates, the gender gap in mortgage rates increases by 63%, and the gender gap in rate spreads more than doubles. Additionally, Columns (9)-(11) demonstrate that in these highly educated areas, the gender gap in net loan costs widens by 0.9 basis points, driven by both higher fees and lower lender credits for female borrowers.

[12] These ACS estimates are based on data collected over a 5-year period, ending in 2017, and therefore describe the average characteristics for that 5-year time period. The multiyear estimates have an advantage of increased reliability for small or less populated geographical units, in our case, Census tracts. The median tract income and educational attainment level do not change dramatically from year to year, and our results are robust to using a different reference year. By using the 2017, 5-year estimates, we measure these factors prior to the origination of mortgages in our sample.

[13] Double clustering of standard errors at the Census tract and origination year-month level yields similar results when estimating equation (2).

These findings are further supported by additional analyses that incorporate local socioeconomic variables to measure the affluence of an area. Specifically, Table A4 in the Appendix presents results using a composite measure of socioeconomic status (SES).[14] Table A5 examines the ratio of Census tract median family income to the Metropolitan Statistical Area (MSA) median family income, providing a measure of relative affluence within metropolitan areas. Across all measures of affluence, the results consistently show that areas with higher socioeconomic levels exhibit larger gender gaps in mortgage pricing and costs.

The results remain robust when the sample is restricted to White borrowers, addressing potential concerns that observed gender gaps might be influenced by correlations between gender and race. Tables A6 and A7 in the Appendix show that cross-sectional variations in gender gaps by tract income and educational attainment persist within the White borrower subset. Notably, White borrowers constitute the largest racial group in our sample, accounting for 69.5% of single borrowers.

Similarly, the results are not driven by a specific cohort of borrowers or individuals at a particular life stage; they are consistently observed across all age groups. Tables A8 to A10 in the Appendix report findings for 10-year age cohorts, demonstrating that gender gaps persist across all age ranges. The only notable pattern is that gender gaps in net loan costs increase steadily with age after controlling for pricing variables, however, additional controls eliminate those differences across cohorts.

Moreover, the results hold when the sample is restricted to either home purchase or refinance mortgages. Tables A11 and A12 in the Appendix show that cross-sectional variations in gender gaps by tract income and educational attainment persist within the subset of purchase loans. Similarly, Tables A13 and A14 confirm the robustness of these findings within refinance loans.

We confirm that the effects are not driven by female borrowers, particularly in affluent communities, sorting into higher-cost lenders with universally unfavorable terms. Mortgage costs remain higher for females even when borrowing from the same set of lenders. To investigate the role of lenders, we repeat the earlier analysis presented in Tables 4 to 6 and Appendix Tables A4 and A5, with the addition of lender fixed effects in the regression equations (1) and (2). We present the consolidated results in Table A15 and A16 of the Appendix.

[14] SES comes from the Stanford Education Data Archive (SEDA) database and is the average socioeconomic status between the years 2009 to 2018. SES is a composite of the median household income, the share of adults with a bachelor's degree or higher, the poverty rate of 5- to 17-year-olds, the unemployment rate, the proportion of households receiving food stamps, and the proportion of single mother–headed households. Since SES is measured at the school district level, c denotes the school district level in equation (2) in this analysis, and $a_c \times a_t$ denotes school district $\times$ origination year and month fixed effects. In the specification that includes SES we allow for clustering of standard errors at the school district level.

Table A15 shows that female borrowers pay higher mortgage rates and rate spreads, even after accounting for fixed differences among lenders. The magnitudes of the effects remain almost the same, implying that the differences in borrowing rates are not necessarily due to female borrowers sorting into different lenders. Interestingly, Table A16 demonstrates that the previously observed effect of female borrowers having lower average net loan costs after controlling for loan and borrower characteristics diminishes and often becomes statistically insignificant once lender fixed effects are also included. However, the finding that female borrowers face higher net loan costs in more affluent areas persists—and in some cases, intensifies—across various socioeconomic indicators after accounting for lender fixed effects.

Furthermore, a large part of the variation in the outcomes is controlled for by the fixed effects, especially because the majority of the lenders in the dataset are very small lenders. Specifically, more than 50% of the lenders in our sample originate only one mortgage per Census tract per month. Including lender fixed effects effectively removes those loans from our sample, leaving loans from larger, more established lenders. Finding larger gender gaps within loans issued by these lenders, even with extensive controls, indicates that the gaps are indeed widespread in the data.

3.2 Gender Gaps in Default

Could the gender gaps in mortgage costs documented in Section 3 be due to gender differences in default propensities not accounted for by pricing controls and additional loan or borrower-level controls? From the lender's perspective, this is unlikely, as these loans are securitized. Additionally, from the MBS investors' point of view, this is not likely, since these loans are guaranteed against default by the GSEs, and the pricing controls account for the pricing grid that GSEs impose when setting their guarantee fees. Therefore, neither the lenders nor the investors in these mortgage-backed securities are exposed to differential default risk from female borrowers.

Are the GSEs, by guaranteeing these mortgages, exposed to a higher default risk from female borrowers? To investigate this, we replicate the analysis in Section 4 by replacing the prepayment hazard rate with the default hazard rate in equation (3). Consistent with the literature, we define default as being 90 days or more past due. In our sample, the average default rate is 0.22 percent per month. Table 7 presents the results.

Interestingly, we find that female borrowers actually have a lower default hazard rate compared to male borrowers, suggesting that lending to female borrowers is less risky. The default gender gap is substantial, representing 21 percent of the average monthly default rate for all single borrowers. This estimate remains nearly unchanged when accounting for pricing controls and additional loan and borrower characteristics. Our results are consistent with Gerardi et al. (2023), who report lower default rates for female borrowers in GSE loans. Additionally, Delis et al. (2022) find that female borrowers are less likely than male borrowers to default on small business loans.

Furthermore, interactions between the female indicator and local income and education variables show that the default gap does not decrease with the income or education level of the area. In one specification, we find that the female default rate is even lower in higher-income tracts; however, this finding is not robust, and the other specifications do not yield a significant estimate for the interaction terms.

4 Gender Gaps in Refinancing

Thus far, we have focused on the differences in mortgage terms for loans originated by male and female borrowers within conventional loans securitized by Fannie Mae and Freddie Mac. Our analysis revealed statistically significant gender gaps, particularly in more affluent areas, even after accounting for an extensive set of controls and lender fixed effects. The observed effects are comparable to estimates of racial gaps reported in the literature, and both are relatively small due to the standardization in this market.

However, mortgage decisions extend beyond selecting the best mortgage at origination. One critical decision is the opportunity to refinance mortgages, which can lead to substantial savings when interest rates decline. Households that fail to refinance their mortgages during these periods miss out on these potential savings. Therefore, our next step is to study prepayment outcomes to understand the gender gaps in refinancing behavior.[15]

Our analysis is conducted on a panel dataset at a monthly frequency, with the unit of observation being a loan-month. To reduce the computational burden, we utilize a 15 percent random sample of our merged HMDA-GSE dataset. A loan is considered prepaid when it is voluntarily paid off in full before the maturity date. In our sample average prepayment rate is

[15]Note that our data does not allow us to distinguish between prepayment due to refinancing or selling the property. If males buy and sell more frequently, part of the gap in prepayment may be due to differences in the frequency of sales. Goldsmith-Pinkham and Shue (2023) report that the average holding period for single males is 5.34 years compared to 5.76 years for single females, implying an slightly higher sales hazard rate for males in their sample.

2.08 percent per month.[16]

Figure 3 plots Kaplan–Meier estimates of the prepayment hazard rates by gender for low- and high-income tracts, as well as low- and high-education tracts. These are unconditional, average monthly rates as a function of months since origination. Specifically, the Kaplan–Meier estimates are calculated as the share of loans that are prematurely paid off in a given month, out of the number of loans that reached that month in our sample without being terminated or censored.

Comparisons of the top-left and top-right panels reveal several findings. Borrowers in lower-income tracts refinance at lower rates and tend to refinance later than those in higher-income tracts. During the peak of the refinancing waves, which occurred 15 to 30 months following origination, the difference in refinance rates between the two groups exceeds 1%. Turning our attention to gender differences, we observe little gender gap in refinancing in lower-income tracts, whereas the gender gap hovers around 0.5% in higher-income tracts. The bottom panels provide the same comparison for low- and high-education tracts and yield very similar results.

In Appendix Figure A2, we plot the hazard rates over calendar time. While overall patterns persist, a striking observation is that the male refinancing rate in high-income and high-education tracts peaked at over 6% in March 2021. In contrast, female refinancing in those areas remained around 5%, and both male and female refinancing rates were around 4% in low-income and low-education tracts.

To estimate how gender and the local socioeconomic development are related to mortgage prepayments, we estimate the following linear probability model:

$$\begin{aligned} Prepayment_{i,j,c,t,\tau} = {} & \beta_1 Female_i + \beta_2 Female_i \times \textit{Local socioeconomic indicator}_c \\ & + \Gamma Z_{i,j,t,\tau} + a_d + a_t + \epsilon_{i,j,c,t} \end{aligned} \tag{3}$$

where $Prepayment_{i,j,c,t,\tau}$ is a dummy variable that takes the value of 1 if a mortgage from borrower i, with mortgage characteristics j, in Census tract c, originated in month t is voluntarily prepaid in full in month τ, either due to refinancing or selling the property, and zero otherwise. Once a loan becomes prepaid, it is eliminated from the database. Local socioeconomic indicators include the median household income and educational attainment in Census tract c. Additional results with SES and tract/MSA median income are provided in the Appendix Tables A17 and A18. The long array of controls $Z_{i,j,t,\tau}$ includes all borrower and loan char-

[16] The average is calculated from the pooled loan-month observations from January 2018 to September 2023. Each loan contributes an observation for each month it remains active, with the prepayment indicator set to zero for every month except the final month, where it is set to one.

acteristics from equation (1) as well as the property value at origination and the loan age in months from origination t to the current date τ. a_d denotes county fixed effects and a_t denotes origination year-month fixed effects. The unit of analysis is the loan-month and the standard errors are double clustered at the county and origination year-month level.

Table 8 presents the results of the estimation of equation (3). Column (1) reports estimates from a specification which includes origination year-month and county fixed effects, but no additional borrower or loan-level controls. We find that female borrowers refinance at a rate that is 23 basis points lower than male borrowers on average. This refinancing gender gap represents more than 10 percent of the average monthly refinance rate of all single borrowers. Columns (2) and (3) add interactions with Census tract income and education. We find that gender gaps grow by approximately 50% for a 1 standard deviation increase in the area's socioeconomic variables.

Table 8 columns (4) to (6) present results from a specification that includes all pricing controls and additional loan and borrower controls introduced earlier for mortgage costs, along with loan age and property value, reducing the gender gap in the monthly refinancing hazard to 4 to 5 basis points. However, in areas with a 1 standard deviation higher income or education, the gender gap doubles. While we observe a higher overall refinancing hazard rate in higher-income areas, as demonstrated in Figure 3, the increase in the gender gap in high-income areas means that the prepayment intensity among women remains nearly unchanged with tract income after adjusting for loan and individual characteristics.[17]

Thus far, we have compared the overall prepayment hazard rates without distinguishing whether the refinancing is optimal. Next, we introduce a "rate drop" indicator to identify periods when refinancing is likely optimal—specifically, when the prevailing market mortgage rate falls below the rate at origination. This indicator is set to 1 if the 30-year fixed-rate mortgage rate, as reported by the Freddie Mac Primary Mortgage Market Survey (PMMS) at the time of origination, exceeds the market rate in a given month.

In Table 9, we include interactions of the rate drop dummy with the female indicator and area socioeconomic variables.[18] We also create an alternative rate drop dummy that takes the value of 1 if the contract mortgage rate exceeds the market rate in each month. The results using this measure, presented in Appendix Tables A21 and A22, show no material change in the findings. Both sets of results show that gender gaps in refinancing grow during periods of rate

[17] Column (3) of Appendix Table A19 presents the full specification of column (2) of Table 8 with linear terms, where the Female × Tract Income coefficient cancels out 64% of the Tract Income coefficient for female borrowers.

[18] Appendix Tables A19 and A20 present the full specification of the estimations in Tables 8 and 9 that include Census tract income and education.

declines, with this trend being particularly pronounced in areas with higher income, education, and math gaps. Notably, when comparing Tables 8 and 9 with all controls, we find that the gender gaps are approximately four times larger during rate drops. In areas with a 1 standard deviation higher income or education females have a monthly prepayment hazard rate that is 0.3% lower than males when it is most advantageous to refinance a loan, resulting in significant missed savings in future loan payments.

5 Gender Gaps in Overall Mortgage Costs

So far, we have shown that single female borrowers receive less favorable mortgage terms compared to single males, and that these disparities widen in higher-income, highly educated areas. We also found that female borrowers are less likely to refinance, particularly when interest rates decline, with the gender gap similarly increasing in affluent regions. If borrowers who receive unfavorable terms are also less likely to refinance, the combined cost of unfavorable terms and limited refinancing is likely greater than the impact of each factor alone.

To assess this combined effect, we link the initial mortgage terms to prepayment behavior and track each loan until it closes (due to prepayment or default) or until September 2023, whichever occurs first. Using this data, we calculate a hypothetical 10-year overall mortgage cost, which includes the origination costs and interest payments on the original loan amount over a 10-year period. If the original loan is not prepaid by September 2023, we assume it continues for the full 10 years. For loans that are paid off before September 2023, we estimate a new interest rate by subtracting the difference in market rates between the origination and prepayment month from the original rate, and we assume the borrower incurs refinancing costs equal to the original loan costs (fees minus lender credits). Given the typically small principal amortization in the early years of a long-term mortgage, we assume the loan balance remains unchanged.

We find that 68% of mortgages in our sample were paid off by September 2023, largely driven by the historically low mortgage rates following the COVID pandemic. On average, market interest rates fell by 0.81% from origination to prepayment for the mortgages in our sample, though the benefits of refinancing were partially offset by incurring new origination costs. Table 10 presents the results of estimating equations (1) and (2) using overall mortgage cost, expressed as a percentage of the loan amount, as the dependent variable, with tract income and education as the socioeconomic variables. Appendix Table A23 provides comparable results using SES and the tract-to-MSA median income ratio.

Our findings indicate that female borrowers' 10-year overall mortgage costs exceed those of male borrowers by approximately 0.8% of the loan value. Controlling for risk-based pricing factors does not reduce these disparities, though additional non-pricing controls substantially narrow the gap. The gaps vary considerably across different areas: female borrowers in high-income, highly educated Census tracts pay roughly 1% more of the loan amount, compared to about 0.5% in lower income and education areas, after accounting for risk-based pricing differences. For the average mortgage of $231,000, this equates to excess costs of $2,200 for women in high-education neighborhoods and $1,240 in lower-education areas. Even after all controls are applied, narrowing the average gender gap to below 0.05%, women in high-income, high-education Census tracts still pay 0.15% more than male borrowers.

6 Inspecting the Mechanism

The results thus far show that, on average, women—particularly those in more affluent, higher-income, and higher-education communities—face higher interest rates, rate spreads, and up-front fees at mortgage origination compared to men. They are also less likely than their male counterparts to refinance when interest rates decline. This combination—paying more for similar loans and refinancing less—suggests that women, particularly in affluent communities, may behave like less-informed borrowers. This could reflect less effective mortgage shopping, reduced search effort, or difficulties in evaluating loan terms. For instance, less-informed applicants may focus primarily on the mortgage rate, which is more salient, while overlooking fees and the trade-offs between rates and points/fees. The situation is further compounded by failing to refinance during periods of declining interest rates. Refinancing requires proactive decision-making and careful evaluation of its costs and benefits. A lack of attention or understanding could result in inaccurate loan comparisons and missed opportunities to refinance optimally.

6.1 Gender Gaps in Financial Literacy

The literature on financial literacy has documented that, on average, females have lower financial literacy (Lusardi and Mitchell, 2014; Fonseca et al., 2012) and exhibit different search behaviors (Malliaris et al., 2022). To further explore how financial literacy interacts with gender and socioeconomic factors, and how these factors influence mortgage search behavior, we use data from the National Survey of Mortgage Originations (NSMO). This survey gathers information on borrowers' experiences obtaining a mortgage, their perceptions, and their understanding of the mortgage market.

We begin by assessing how knowledgeable borrowers are about the mortgage process and the important economic decisions they made, by evaluating their answers to the following questions:[19] "How well could you explain to someone the... Process of taking out a mortgage / Difference between a fixed- and an adjustable-rate mortgage / Difference between a prime and subprime loan / Difference between a mortgage's interest rate and its APR / Amortization of a loan / Consequences of not making required mortgage payments / Relationship between discount points and interest rate."[20] Respondents chose from a three-point scale: "Not at all" (coded as 1) to "Very" (coded as 3).

While each question addresses an important feature of mortgages, we note that borrowers find some questions easier than others. For example, fewer than 10% of borrowers report difficulty explaining the process of taking out a mortgage, the difference between fixed- and adjustable-rate mortgages, and the consequences of not making required mortgage payments. However, proficiency drops significantly, and the dispersion in answers increases for questions requiring an understanding of quantitative features and trade-offs, such as loan amortization, the difference between the interest rate and APR, and the relationship between discount points and the interest rate. Given the quantitative nature of these latter concepts, which involve a deeper understanding of mathematical relationships and financial analysis, and the clear split in responses, we classify the first three questions as addressing "Basic Literacy" and the remaining four questions as addressing "Advanced Literacy."

To estimate how mortgage literacy is related to gender and income level, we estimate the following regression:

$$L_{i,s} = \beta_1 Female_i + \sum_k \beta_{2,k} Income_{i,k} + \sum_k \beta_{3,k} Female_i \times Income_{i,k} + \Gamma Z_i + a_s + \epsilon_{i,s}, \quad (4)$$

where $L_{i,s}$ represents the mortgage literacy level of borrower i, in survey wave s. Utilizing the rich NSMO questionnaire, we construct various measures of mortgage literacy, as outlined earlier. $Female_i$ is an indicator variable that takes the value of 1 if the borrower is female and 0 if the borrower is male. $Income_{i,k}$ represents the income level k of borrower i. To isolate the effects of gender and income, we control for ΓZ_i, a comprehensive set of borrower characteristics, which include dummies for four education levels, five race and ethnicity categories, six age groups, and nine credit score categories. Moreover, we include survey wave fixed effects a_s, which are

[19] Section A3 in the Appendix provides detailed information on the survey questions and the distribution of responses.

[20] In this block, there are two additional questions—one asking the difference between the lender's and owner's title insurance and the other asking the reason payments into an escrow account can change. Since these questions pertain to technical details rather than addressing important decisions and trade-offs, we excluded them from the analysis.

equivalent to year-quarter fixed effects. In all specifications, we use the provided survey weights from NSMO, which are based on sampling weights and non-response adjustments. The sample includes single borrowers of conventional loans, to be consistent with the analyses throughout the paper.

Table 11 reports the coefficient estimates of equation (4). Table 12 presents the results of a similar specification to equation (4) that includes the interaction of gender with education level instead of income level and controls additionally for income, with dummies for five income levels. Our results confirm several salient observations made previously in the literature. Female borrowers express lower levels of literacy on all questions, particularly when answering the more advanced literacy questions. Furthermore, literacy increases with the income and education level of the borrowers.

For basic literacy, gender gaps are relatively small, and literacy increases similarly for male and female borrowers as their income and education levels rise. This is most evident in the basic but important question of whether borrowers can explain the consequences of not making a required mortgage payment, where we observe the smallest gender gap, and a shallow gradient. In the other basic literacy questions, and in the basic literacy composite we construct, the returns to education and income are not significantly different for males and females.

However, we identify significant differences in the income and education gradients of male and female borrowers based on their answers to advanced literacy concepts that require a more nuanced understanding of trade-offs and quantitative features of mortgages. We find negative, growing, and highly significant estimates for the interaction terms of the female indicator with income and education level in Tables 11 and 12. This result is also evident in the estimated levels of advanced literacy (V1) by income and education, plotted in Figure 4. While the estimated level of advanced literacy grows with income for both males and females, the increasing gap, particularly in the highest income bracket, is noteworthy. The divergence in the return to education is more striking: for men, advanced literacy increases monotonically with education, whereas for women, literacy remains unchanged beyond some college education, even slightly declining for those with a postgraduate degree, leading to a large gender gap among highly educated borrowers.[21] These findings shed light on the growing gender gaps in mortgage borrowing costs and refinancing in more socioeconomically advantaged communities, which also

[21]The gender differences in the gradient of financial literacy with education and income have not been emphasized in the literature before. In a study of determinants of the gender gap in financial literacy, Fonseca et al. (2012) considers the interactions of various covariates with gender and identifies a similar effect for education using RAND American Life Panel Data, but not for income, and does not emphasize this result. Most of the financial literacy literature thus far focuses on a single and basic measure of literacy. Our results show that the depth and context of financial literacy matter.

exhibit larger gender gaps in math skills, as documented thus far.[22]

6.2 Social Norms and Math Gender Stereotypes

Why does financial literacy that requires mathematical reasoning diverge more across genders as income and education levels rise? Recent work by Breda et al. (2020) offers compelling insights. In their investigation of why gender segregation across occupations is more pronounced in more egalitarian and developed countries –a phenomenon known as the "gender-equality paradox" (Charles and Bradley, 2002; Stoet and Geary, 2018)– they demonstrate that the stereotype associating math with males is stronger in these more egalitarian and developed countries, and within the United States, among higher-income households.[23] Interestingly, they also show that math-gender stereotypes are *negatively* correlated with the acceptance of traditional gender roles, such as the belief that being a housewife is fulfilling or that attending university is less important for women.

Building on these findings, Breda et al. (2020) propose several mechanisms underlying the stronger presence of math-gender stereotypes in higher-income households. They suggest that greater financial flexibility may lead to increased internalization of these stereotypes, which subsequently shape individual choices. Additionally, high-income parents are more likely to invest in stereotypical activities and take a more active role in guiding their children's educational decisions. These increasing math-gender stereotypes in higher-income households align with findings from Penner and Paret (2008), Lubienski et al. (2013), and Reardon et al. (2019), who report that math gender achievement gaps are greatest among students with high parental education and socioeconomic status, as well as in socioeconomically advantaged school districts. We reproduce this result in Figure 1.

Crucially, to the extent that these achievement gaps reflect the math-gender stereotypes prevalent in the community and their influence on shaping outcomes, they may also serve as a proxy for the current math skill gap among adults in the same community. A substantial

[22] One concern with self-reported literacy measures is the potential difference in confidence levels between males and females. Some of the gender differences we see may result from a lack of confidence in mortgage knowledge rather than a lack of actual knowledge, as pointed out by Bucher-Koenen et al. (2012) and demonstrated in different financial contexts by Barber and Odean (2001). Reuben et al. (2014) show that men tend to overstate their performance in math-related pursuits, while women generally underreport it, reinforcing stereotypes about women and mathematics and contributing to biases against women. While gender differences in confidence can affect the baseline estimate for the female indicator, unless the confidence difference grows with the level of education or income, we would not expect it to generate the interaction effects we observe in the data.

[23] Specifically, Breda et al. (2020) measure the stereotype associating math with males based on differences between boys' and girls' beliefs that "doing well in math is completely up to them" and that "their parents think that math is important for their career," conditional on their math ability.

body of research (e.g., Parsons et al. (1982); Eccles (1994)) shows that parents' choices for their children often mirror their own beliefs and skill sets, suggesting that the stereotypes driving children's achievement gaps can similarly inform the skill distribution among adults. Our findings on the occupational math skill gender gap align with this view: we observe parallel gender gaps in math skills among both students and adults. Notably, these gaps shift from female- to male-favoring as community income and education levels rise, consistent with social norms and gender stereotypes influencing both adult occupational and student educational outcomes, thereby propagating gender gaps in the community.

6.3 Mortgage Shopping Behavior

Next, we investigate gender differences in mortgage shopping behavior at various critical stages of the process and whether these differences relate to our estimated financial literacy based on answers to other survey questions from NSMO. We consider initial familiarity with the mortgage market (When you began the process of getting this mortgage, how familiar were you with each of the following?), whether the borrowers seriously considered more than one lender before applying (How many different mortgage lenders/brokers did you seriously consider before choosing where to apply for this mortgage?), whether the borrowers were presented with multiple loan types or features (During the application process, were you told about mortgages with any of the following?), and whether they performed due diligence on the offered mortgage terms through additional resources (In the process of getting this mortgage from your mortgage lender/broker, did you check other sources to confirm that the terms of this mortgage were reasonable?).

Table 13 presents the results of our analysis. At every stage of the mortgage process, female borrowers perform a weaker search than their male counterparts. They start the process with less information, consider fewer lenders, fewer loan options, and tend to conduct due diligence on the offered terms at a lower rate.[24] Furthermore, these differences in search behavior are

[24] Another way to analyze mortgage search behavior is by examining the distribution of mortgage application outcomes in HMDA data. Beyond loans that are originated or denied, applications can also be withdrawn or not accepted by borrowers. Borrower-initiated withdrawals can signal mortgage shopping behavior (Jiang et al., 2024). Table A24 in the Appendix shows application outcomes for male and female borrowers, expressed as a percentage of loans originated, across Census tracts with varying income and education levels. On average, for every four approved mortgages, one application is withdrawn. Withdrawal rates are slightly higher for male borrowers (24.6% vs. 23.4% for females), with the gender gap slightly more pronounced in high-income, high-education tracts. While withdrawals likely indicate mortgage shopping, the underlying motivation—whether to improve the chances of loan approval or to secure better terms—remains unclear. Notably, loan denial rates exhibit substantial variation, differing by over 10 percentage points between high- and low-income/education areas. In contrast, withdrawal rates show a much smaller difference of about 1 percentage point. This suggests that borrowers in low-income, low-education areas may submit multiple applications to increase their approval

significantly related to financial literacy estimates. Including our financial literacy estimates in the regressions alongside the female indicator significantly reduces the coefficient for the female indicator for most search variables, while the coefficient for literacy remains unchanged. This suggests that lower financial literacy among females is a significant factor contributing to their suboptimal mortgage shopping behavior.

Complementing our earlier analysis for financial literacy, Tables A25 and A26 in the Appendix highlight significant disparities in the income and education gradients between male and female borrowers concerning their initial familiarity with specific mortgage concepts. The results indicate that the gender gap in initial familiarity with concepts such as interest rates, mortgage types, and down payments widens as income and education levels increase. This gap is particularly pronounced among borrowers in the highest income brackets and those with a college education or higher.

While documented differences in search behavior are likely contributors to the gender gaps, they are unlikely to be the only factors. Malliaris et al. (2022), using NSMO data, finds that performing a competitive search is not a substitute for financial sophistication. Considering the nature of the advanced literacy questions we examine and the differences we document, this finding is not surprising. For example, fewer than 30% of borrowers in the NSMO dataset state that they know the relationship between discount points and interest rates well, and a third of them state that they do not know it at all. Even if potential borrowers perform a thorough search and compare different loan options, they can make suboptimal choices due to not fully understanding the trade-offs inherent in mortgage options. Therefore, gender gaps in financial literacy, and the deep understanding of mathematical relationships it requires, are likely to contribute to gender gaps in mortgage outcomes through multiple channels, including suboptimal search and an inability to make complex mortgage decisions correctly.

6.4 Alternative Explanations

In addition to gender gaps in financial literacy and the persistent stereotype that "math is not for girls," we consider another potential mechanism that could explain why single women in more affluent communities tend to pay more for mortgages: they may have less time to shop for mortgage options. As noted earlier, this search process includes learning about mortgage products, comparing lenders and loan types, and thoroughly evaluating the terms offered. A common assumption is that single women might be more time-constrained than single men due

odds, whereas those in high-income, high-education areas are more likely focused on shopping for better loan terms.

to caregiving responsibilities, such as taking care of young children. However, our data do not support this hypothesis.

To investigate this, we merged tract-level data on household structure from the 2017 American Community Survey (ACS) 5-year estimates, with our primary mortgage dataset. Table 2 shows that in high-income Census tracts, 50.4% of households headed by single women have children (compared to 48.2% for those headed by single men). In low-income tracts, this percentage increases to 52.6% for single female heads of household (versus 47.0% for single male heads). Similarly, in high-education Census tracts, 51.0% of single female-headed households have children (compared to 47.7% for single male-headed households), while in low-education tracts, the percentage is 52.0% (versus 47.5% for single males). This indicates that a higher proportion of single female-headed households with children is found in less affluent areas, contrary to what we might expect if caregiving responsibilities were driving the higher mortgage costs in affluent areas. If caregiving were the main factor, we would expect larger gender gaps in mortgage costs in low-income and low-education areas.

Regression analyses that estimate equation (2) by replacing the local socioeconomic indicator with the share of households with children among single female households, or the differential between male and female single-headed households with children, yield non-significant results, further supporting that caregiving responsibilities do not explain the observed disparities.

7 Math Skills and Gender Gaps in Mortgage Outcomes

The literature on gender gaps in math education (Reardon et al., 2019) demonstrates that while unconditional gender differences in math proficiency are small, outcomes vary substantially across school districts and are closely tied to the socioeconomic characteristics of the areas. Figure 1 illustrates this relationship. Our analysis extends this insight to adults using a novel measure of math skills based on the occupational relevance of math, revealing even larger disparities across the socioeconomic spectrum, as shown in Figure 2. The consistent alignment of math gender gaps between school-age children and adults in the same regions strongly suggests that math-gender attitudes are deeply rooted, persist across generations, and systematically vary with socioeconomic conditions, as discussed in Section 6.2.

The links from math education and numeracy to financial outcomes, established by a substantial body of prior work, make it highly plausible that the larger gender gaps in math attitudes and skills observed in affluent areas significantly contribute to the relationship between socioeconomic status (income, education) and gender disparities in mortgage outcomes.

To explore this, we replicate our earlier analysis of mortgage rates, costs, refinancing behavior, and total mortgage expenses, substituting local measures of income or education with local gender gaps in math skills.

Table 14 presents the regression results for the 10-year overall mortgage cost, which incorporates both mortgage origination and refinancing effects, as defined in Section 5. The regressions incorporate interactions between the female indicator and the local gender gap in student math scores, as well as the local gender gap in occupational math skills. Comparing these results to those in Table 10, we find that the effect of a one standard deviation increase in the local math skill gender gap is comparable to the effect of a one standard deviation increase in local median income or education levels.[25]

Appendix Tables A27 to A29 present detailed mortgage cost results, incorporating interactions with math skills. The coefficients for the interactions of math skills with the female indicator closely align with the baseline results for the interactions of female with income and education, as reported in Tables 5 and 6. Similarly, Tables A30 and A31 provide results for refinancing outcomes in relation to math skills, similar to Tables 8 and 9. These findings reinforce the perspective that social norms that result in differences in gender attitudes toward math may play a pivotal role in explaining cross-sectional variations in gender gaps in mortgage outcomes.

8 Conclusions

We uncover a gender gap in key mortgage outcomes within the U.S. with significant implications for household wealth. As one of the most substantial financial decisions a household can make, securing a mortgage shapes long-term financial well-being. In addition to assessing unconditional gender gaps in outcomes like interest rates, loan costs, and refinancing activity, we analyze how these disparities vary across the socioeconomic spectrum.

Using data on the near universe of conventional, conforming mortgages in the United States reported in HMDA between 2018 and 2019, we find that females pay higher mortgage rates, rate spreads, and net loan costs at origination, and have a lower propensity to refinance when it is optimal to do so. Furthermore, these disparities widen in higher-income and higher-socioeconomic status neighborhoods, which also display a male-favoring gap in math test scores.

[25]Since student math score data is available at the school district level rather than the Census tract level, we adjust the regressions accordingly. Specifically, we replace the Census tract × origination year-month fixed effects with school district × origination year-month fixed effects. Additionally, we cluster standard errors at the school district level to account for the geographical unit at which math scores are reported.

In addition to the gender gap being connected to socioeconomic variables, we find that math scores in childhood account for part of the differences in mortgage outcomes between men and women. Our findings reveal that in areas where boys demonstrate superior math performance, females tend to secure loans at substantially higher interest rates and loan costs, and tend to refinance less, aligning with gender stereotypes associated with female strengths and weaknesses. This disparity in mortgage outcomes is not related to differences in reading (English Language Arts) performance. The math scores have a robust and independent effect, different from overall test performance.

We find similar patterns among adults using a novel measure of math skills based on the occupational relevance of math, with even larger disparities observed across the socioeconomic spectrum. These results are consistent with our findings based on NSMO survey responses, where we demonstrate that the male-female gap in advanced financial literacy, particularly in areas requiring a deeper understanding of mathematical relationships and analysis, grows with income and education.

The mechanism we propose is that gender gaps in mortgage outcomes stem from differences in financial literacy requiring mathematical reasoning, which are influenced by social norms and gender stereotypes. These norms and stereotypes lead to gender differences in attitudes toward math, particularly in socioeconomically advantaged areas with resources that promote math-related pursuits for boys more than for girls. Although we do not provide causal evidence directly linking this channel to our observed mortgage outcomes, our study builds on decades of research in psychology, sociology, education, and finance that interconnect these elements.

While our study focuses on gender gaps in mortgage outcomes, which are crucial for household portfolios and wealth accumulation due to their central role in household finances, the implications likely extend beyond the mortgage market. Understanding gender disparities in key financial outcomes and their connection to gender stereotypes regarding math is essential. This highlights the need for targeted financial education and societal change to bridge these gaps, making it an important area for future research and policy intervention.

References

Adams, Renee B, Brad M Barber, and Terrance Odean (2023). The math gender gap and women's career outcomes. *Journal of Investment Management 21*(2), 21–49.

Agarwal, Sumit, Itzhak Ben-David, and Vincent Yao (2017). Systematic mistakes in the mortgage market and lack of financial sophistication. *Journal of Financial Economics 123*(1), 42–58.

Agarwal, Sumit, Richard J Rosen, and Vincent Yao (2016). Why do borrowers make mortgage refinancing mistakes? *Management Science 62*(12), 3494–3509.

Alesina, Alberto F, Francesca Lotti, and Paolo Emilio Mistrulli (2013). Do women pay more for credit? Evidence from Italy. *Journal of the European Economic Association 11*(Supplement 1), 45–66.

Altmejd, Adam, Thomas Jansson, and Yigitcan Karabulut (2024). Business education and portfolio returns. Working Paper.

Ambrose, Brent W, James N Conklin, and Luis A Lopez (2021). Does borrower and broker race affect the cost of mortgage credit? *The Review of Financial Studies 34*(2), 790–826.

Andersen, Steffen, Julie Marx, Kasper Meisner Nielsen, and Lise Vesterlund (2021). Gender differences in negotiation: Evidence from real estate transactions. *The Economic Journal 131*(638), 2304–2332.

Babcock, Linda and Sara Laschever (2003). *Women Don't Ask: Negotiation and the Gender Divide*. Princeton University Press.

Bajo, Emanuele and Massimiliano Barbi (2018). Financial illiteracy and mortgage refinancing decisions. *Journal of Banking & Finance 94*, 279–296.

Barber, Brad M. and Terrance Odean (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics 116*(1), 261–292.

Bartlett, Robert, Adair Morse, Richard Stanton, and Nancy Wallace (2022). Consumer-lending discrimination in the FinTech era. *Journal of Financial Economics 143*(1), 30–56.

Bellucci, Andrea, Alexander Borisov, and Alberto Zazzaro (2010). Does gender matter in bank–firm relationships? Evidence from small business lending. *Journal of Banking & Finance 34*(12), 2968–2984.

Bertrand, Marianne (2011). New perspectives on gender. In *Handbook of Labor Economics*, Volume 4, pp. 1543–1590. Elsevier.

Bhutta, Neil, Andreas Fuster, and Aurel Hizmo (2024). Paying too much? Borrower sophistication and overpayment in the US mortgage market.

Bhutta, Neil and Aurel Hizmo (2021). Do minorities pay more for mortgages? *The Review of Financial Studies 34*(2), 763–789.

Breda, Thomas, Elyes. Jouini, Clotilde Napp, and Georgia Thebault (2020). Gender stereotypes can explain the gender-equality paradox. *Proceedings of the National Academy of Sciences 117*(52), 31063–31069.

Brown, Meta, John Grigsby, Wilbert van der Klaauw, Jaya Wen, and Basit Zafar (2016). Financial education and the debt behavior of the young. *The Review of Financial Studies 29*(9), 2490–2522.

Bucher-Koenen, Tabea, Annamaria Lusardi, Rob Alessie, and Maarten van Rooij (2012). How financially literate are women? Some new perspectives on the gender gap. *Network for Studies on Pensions, Aging and Retirement Panel Paper 31*.

Campbell, John Y (2006). Household finance. *The Journal of Finance 61*(4), 1553–1604.

Carpena, Fenella, Shawn Cole, Jeremy Shapiro, and Bilal Zia (2011). Unpacking the causal chain of financial literacy. World Bank Policy Research Working Paper 5798, World Bank, Washington, DC.

Charles, Maria and Karen Bradley (2002). Equal but separate? A cross-national study of sex segregation in higher education. *American Sociological Review 67*, 573–599.

Cheng, Ping, Zhenguo Lin, and Yingchun Liu (2011). Do women pay more for mortgages? *The Journal of Real Estate Finance and Economics 43*, 423–440.

Cole, Shawn, Anna Paulson, and Gauri Kartini Shastry (2016). High school curriculum and financial outcomes: The impact of mandated personal finance and mathematics courses. *Journal of Human Resources 51*(3), 656–698.

Conklin, James, Kristopher Gerardi, and Lauren Lambie-Hanson (2023). Can everyone tap into the housing piggy bank? Racial disparities in access to home equity. FRB of Philadelphia Working Paper.

Croson, Rachel and Uri Gneezy (2009). Gender differences in preferences. *Journal of Economic Literature 47*(2), 448–474.

Delis, Manthos D, Iftekhar Hasan, Maria Iosifidi, and Steven Ongena (2022). Gender, credit, and firm outcomes. *Journal of Financial and Quantitative Analysis 57*(1), 359–389.

Dossi, Gaia, David Figlio, Paola Giuliano, and Paola Sapienza (2021). The family origin of the math gender gap is a white affluent phenomenon. *American Economic Review, Papers and Proceedings 111*, 179–183.

Eccles, Jacquelynne S. (1994). Understanding women's educational and occupational choices: Applying the eccles et al. model of achievement-related choices. *Psychology of Women Quarterly 18*(4), 585–609.

Fonseca, Raquel, Kathleen Mullen, Gema Zammaro, and Julie Zissimopoulos (2012). What explains the gender gap in financial literacy? The role of household decision making. *Journal of Consumer Affairs 46*(1), 90–106.

Gerardi, Kristopher, Lorenz Goette, and Stephan Meier (2013). Numerical ability predicts mortgage default. *Proceedings of the National Academy of Sciences 110*(28), 11267–11271.

Gerardi, Kristopher, Paul S Willen, and David Hao Zhang (2023). Mortgage prepayment, race, and monetary policy. *Journal of Financial Economics 147*(3), 498–524.

Ghent, Andra C, Ruben Hernandez-Murillo, and Michael T Owyang (2014). Differences in subprime loan pricing across races and neighborhoods. *Regional Science and Urban Economics 48*, 199–215.

Giacoletti, Marco, Rawley Z Heimer, and Edison Yu (2021). Using high-frequency evaluations to estimate discrimination: Evidence from mortgage loan officers. Working Paper.

Gneezy, Uri, Muriel Niederle, and Aldo Rustichini (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics 118*(3), 1049–1074.

Goldsmith-Pinkham, Paul and Kelly Shue (2023). The gender gap in housing returns. *The Journal of Finance 78*(2), 1097–1145.

Goodman, Joshua (2019). The labor of division: Returns to compulsory high school math coursework. *Journal of Labor Economics 37*(4), 1141–1182.

Jiang, Erica, Gregor Matvos, Amit Seru, and Lulu Wang (2024). Minority specialized lenders. Working Paper.

Keys, Benjamin J, Devin G Pope, and Jaren C Pope (2016). Failure to refinance. *Journal of Financial Economics 122*(3), 482–499.

Li, Xiang (2023). Does the disclosure of consumer complaints reduce racial disparities in the mortgage lending market? Working Paper.

Lubienski, Sarah T., Joseph P. Robinson, Claire C. Crane, and Colleen M. Ganley (2013). Girls' and boys' mathematics achievement, affect, and experiences: Findings from the ECLS-K. *Journal for Research in Mathematics Education 44*, 634–645.

Lusardi, Annamaria (2012). Numeracy, financial literacy, and financial decision-making. *Numeracy 5*(1), Article 2.

Lusardi, Annamaria and Olivia S Mitchell (2014). The economic importance of financial literacy: Theory and evidence. *Journal of Economic Literature 52*(1), 5–44.

Malliaris, Steven, Daniel A. Rettl, and Ruchi Singh (2022). Is competition a cure for confusion? Evidence from the residential mortgage market. *Real Estate Economics 50*(1), 206–246.

Niederle, Muriel and Lise Vesterlund (2007). Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics 122*(3), 1067–1101.

Parsons, Jacquelynne Eccles, Terry F. Adler, and Caroline M. Kaczala (1982). Socialization of achievement attitudes and beliefs: Parental influences. *Child Development 53*(2), 310–321.

Penner, Andrew M. and Marcel Paret (2008). Gender differences in mathematics achievement: Exploring the early grades and the extremes. *Social Science Research 37*, 239–253.

Reardon, Sean F, Erin M Fahle, Demetra Kalogrides, Anne Podolsky, and Rosalía C Zárate (2019). Gender achievement gaps in U.S. school districts. *American Educational Research Journal 56*(6), 2474–2508.

Reardon, Sean F, Andrew D Ho, Benjamin R Shear, Erin M Fahle, Demetra Kalogrides, H Jang, and Belen Chavez (2021). Stanford education data archive technical documentation version 4.1 June 2021. Retrieved from https://edopportunity.org/.

Reuben, Ernesto, Paola Sapienza, and Luigi Zingales (2014). How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences 111*(12), 4403–4408.

Stoet, Gijsbert and David C. Geary (2018). The gender-equality paradox in science, technology, engineering, and mathematics education. *Psychological Science 29*(4), 581–593.

Sunden, Annika E and Brian J Surette (1998). Gender differences in the allocation of assets in retirement savings plans. *The American Economic Review 88*(2), 207–211.

Woodward, Susan E and Robert E Hall (2012). Diagnosing consumer confusion and sub-optimal shopping effort: Theory and mortgage-market evidence. *American Economic Review 102*(7), 3249–3276.

Figures

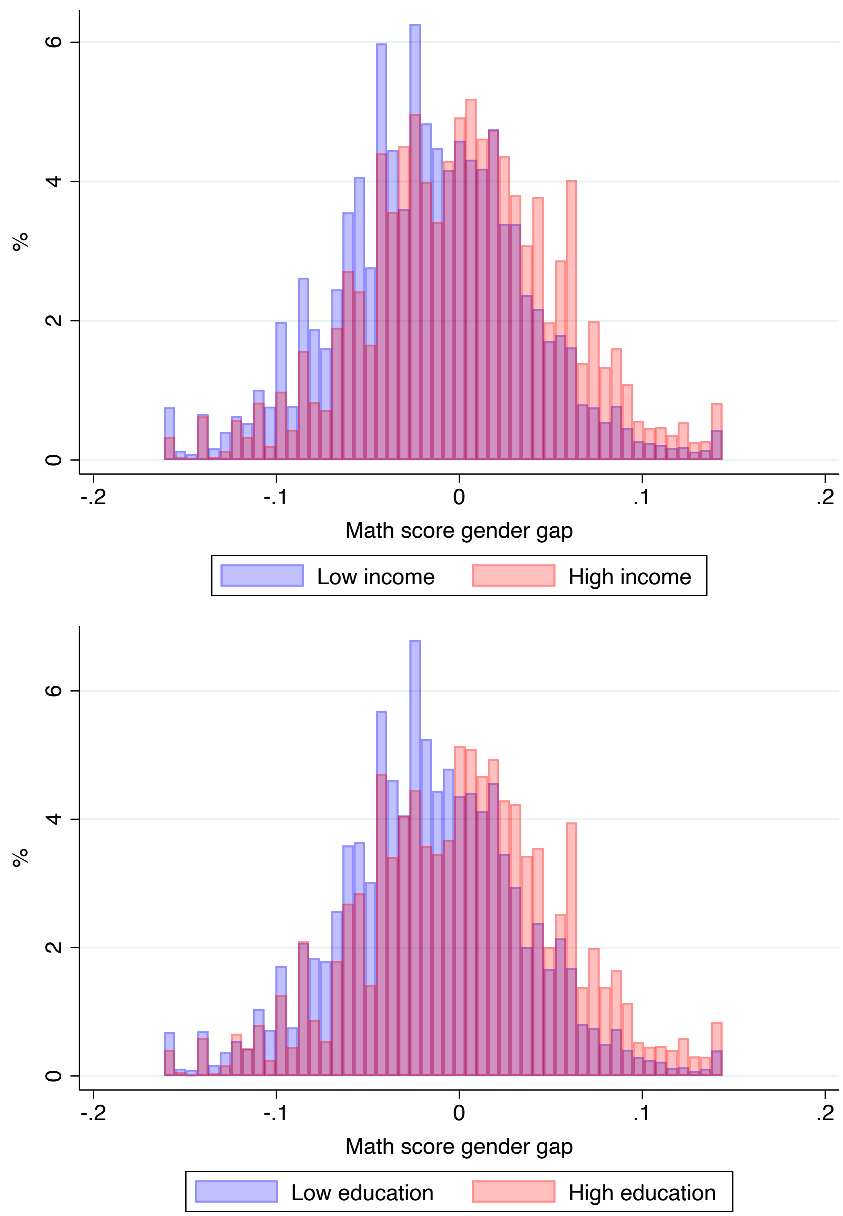

Figure 1. Gender Gap in Student Math Test Scores by Tract Income and Education. The histograms display the distribution of the student math score gender gap across the borrowers in our sample. The top figure shows the math score gap for low- and high-income Census tracts, while the bottom figure shows the math score gap for low- and high-education Census tracts. Construction of the math score gap is described in Table 1.

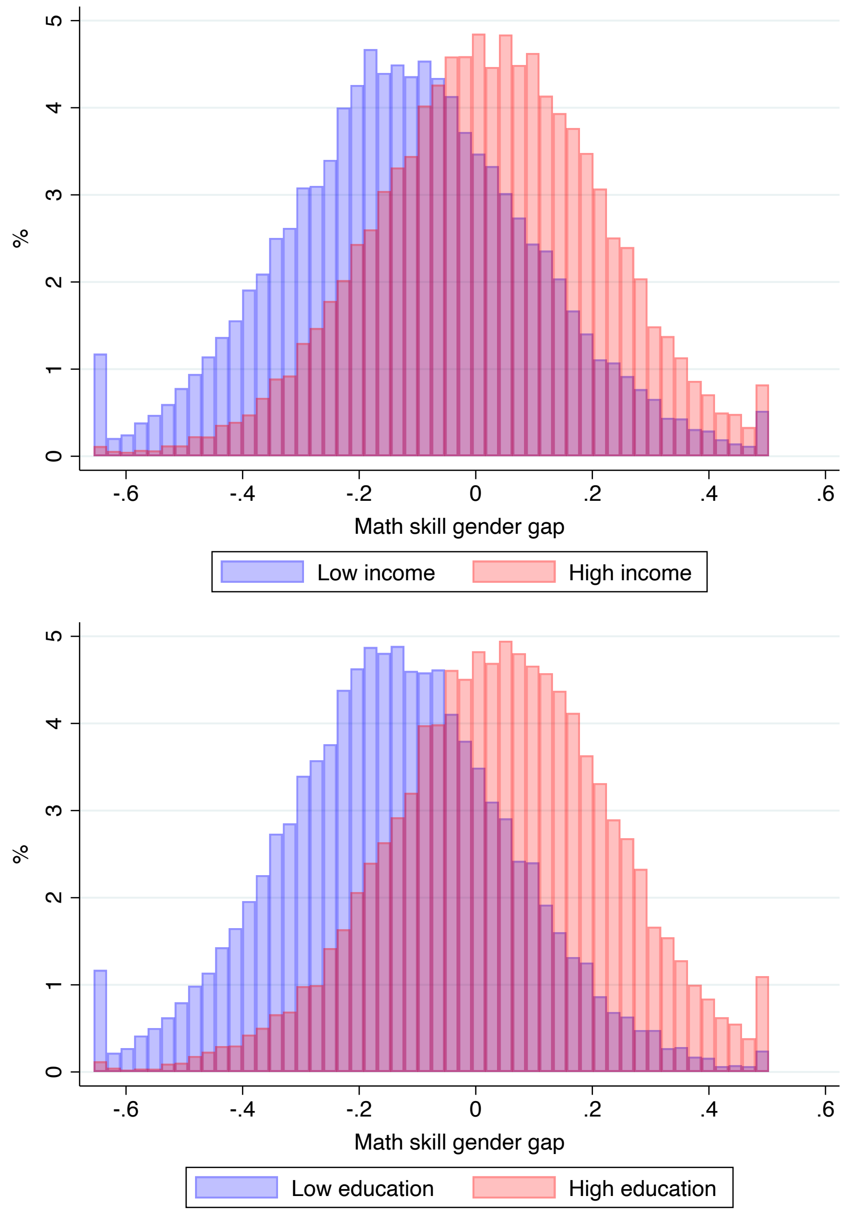

Figure 2. Gender Gap in Occupational Math Skill Importance by Tract Income and Education. The histograms display the distribution of the occupational math skill gender gap across the borrowers in our sample. The top figure shows the math skill gap for low- and high-income Census tracts, while the bottom figure shows the math skill gap for low- and high-education Census tracts. Construction of the math skill gap is described in Section 2.3.

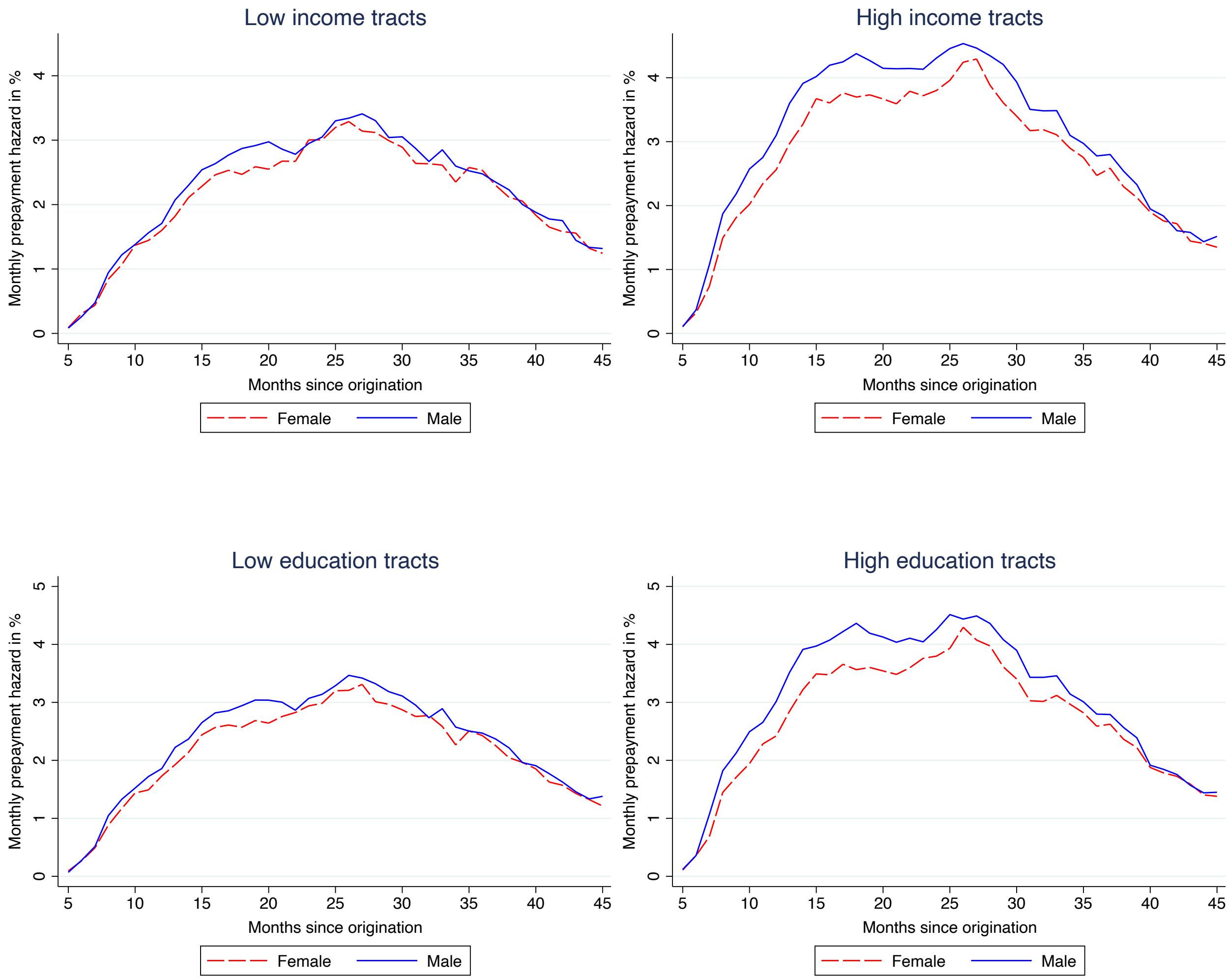

Figure 3. Kaplan–Meier Unconditional Prepayment Hazard Rates by Gender, Income, and Education. This figure plots the unconditional Kaplan–Meier monthly hazard rates for mortgage prepayment as a function of months since origination, separated by gender. The top-left panel shows the hazard rates for Census tracts with incomes below the median tract income, while the top-right panel shows the rates for Census tracts with incomes above the median tract income. The bottom-left panel displays the hazard rates for Census tracts with educational attainment below the median, and the bottom-right panel shows the rates for tracts with educational attainment above the median. Educational attainment is measured as the share of the tract population over 25 years old with a bachelor's degree. The Kaplan–Meier estimate of the hazard function is calculated as the number of prepayments by each gender in each month, divided by the number of outstanding mortgages for each gender in each month. These hazard rates are unconditional of loan and borrower characteristics. The underlying data come from a 15% random sample of loans originated between 2018 and 2019 from the HMDA-GSE database.

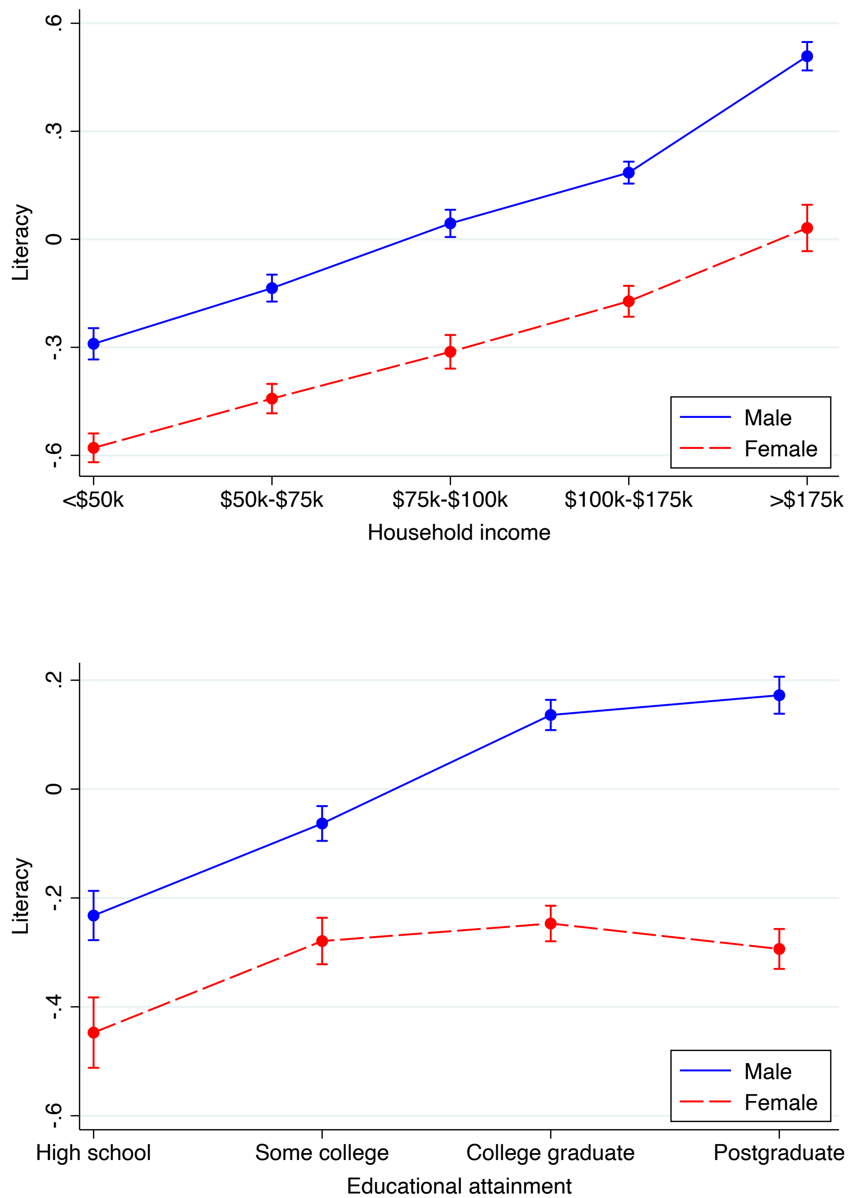

Figure 4. Financial Literacy by Gender, Income, and Education. The top figure displays the estimated financial literacy (Advanced Literacy V1) levels for male and female single mortgage borrowers across various income levels, with 95% confidence intervals. These estimates are derived from the regression model presented in Table 11, column (5). The bottom figure illustrates the estimated financial literacy levels across different educational attainment levels, based on the regression model presented in Table 12, column (5).

Tables

Table 1. Descriptive Statistics

	Observations	Mean	SD	Min	Max
Female	2,261,818	0.42	0.49	0.00	1.00
Mortgage rate (%)	2,261,818	4.40	0.60	2.25	7.00
Rate spread (%)	2,208,531	0.50	0.46	-0.39	2.23
Net loan cost (%)	2,261,818	1.93	1.36	-0.77	7.20
Total fees (%)	2,261,818	2.10	1.32	0.00	7.37
Total credits (%)	2,261,818	0.16	0.40	0.00	2.58
Tract income ($000)	2,261,725	72.5	28.4	5.65	250
Tract education (BA degree %)	2,261,793	35.6	17.6	0.00	96.3
Socioeconomic status (SES)	2,207,317	0.00	1.00	-5.97	3.09
Tract to MSA median income	2,261,818	114	38.6	35.0	257
Student math score gender gap	2,196,846	-0.01	0.05	-0.16	0.14
Student reading score gender gap	2,196,956	-0.24	0.06	-0.41	-0.10
Occupation math skill gender gap	2,261,779	-0.05	0.22	-0.66	0.50
Credit score	2,260,505	748	46.0	309	850
Loan-to-value ratio (%)	2,261,818	76.8	17.0	2.00	389
High balance loan	2,261,818	5×10^{-5}	0.01	0.00	1.00
Investment property	2,261,818	0.03	0.18	0.00	1.00
Manufactured home	2,261,818	0.01	0.09	0.00	1.00
Subordinate loan	2,261,818	0.02	0.14	0.00	1.00
Borrower income ($000)	2,143,893	89.5	60.6	17.0	455
Loan amount ($000)	2,261,818	231	123	5.00	1395
Debt-to-income ratio (%)	2,256,441	36.6	9.23	1.00	53.0
Loan term (months)	2,261,818	334	61.9	60.0	360
Percentage White (%)	2,261,818	69.5	46.0	0.00	100
Share of single female HH with kids (%)	2,258,817	51.5	19.5	0.00	100
Share of single male HH with kids (%)	2,226,149	47.6	28.6	0.00	100

This table presents descriptive statistics for the key variables in the study. Loan cost variables are expressed as a percentage of the loan amount. Rate spread, loan costs, gender gaps, tract/MSA income, and borrower income are winsorized at the top and bottom 0.5%. The gender gaps, tract income, tract education, tract/MSA income, and SES are standardized to have a mean of zero and a standard deviation of one in regressions. Source: HMDA data for conventional loans with a single borrower originated in 2018–2019, matched with loans from Fannie Mae and Freddie Mac, and merged with school district test scores and socioeconomic characteristics from SEDA, math skill importance levels from O*NET, and tract income, education and household structure information from ACS.

Table 2. Descriptive Statistics by Tract Income and Gender

	Low income tracts	High income tracts
Female	0.43	0.40
Tract income ($000)	50.7	94.3
Tract education (BA degree %)	25.4	45.8
Socioeconomic status	-0.49	0.48
Tract to MSA median income	91.1	137
Student math score gender gap	-0.02	0.002
Student reading score gender gap	-0.23	-0.24
Occupation math skill gender gap	-0.12	0.02
Share of single female HH with kids (%)	52.6	50.4
Share of single male HH with kids (%)	47.0	48.2

	Low income tracts		High income tracts	
	Female	Male	Female	Male
Mortgage rate (%)	4.49	4.47	4.35	4.30
Rate spread (%)	0.60	0.60	0.41	0.39
Net loan cost (%)	2.35	2.23	1.70	1.52
Total fees (%)	2.51	2.39	1.85	1.68
Total credits (%)	0.16	0.16	0.16	0.17
Credit score	745	744	752	752
Loan-to-value ratio (%)	77.0	79.2	73.4	76.5
High balance loan	10^{-4}	10^{-4}	2×10^{-5}	2×10^{-5}
Investment property	0.03	0.05	0.02	0.03
Manufactured home	0.01	0.01	0.003	0.003
Subordinate loan	0.02	0.02	0.02	0.02
Borrower income ($000)	66.9	82.3	90.2	112
Loan amount ($000)	176	192	258	290
Debt-to-income ratio (%)	37.1	35.8	37.6	36.2
Loan term (months)	336	332	337	331
Percentage White (%)	70.1	69.8	70.8	67.9

This table presents the descriptive statistics of the key variables in the study for low- and high-income tracts, and by gender. The top panel shows the percentage of female borrowers and area statistics for the variables at the tract level. Low-income tracts are defined as those with tract income in 2017 below the median, while high-income tracts are defined as those with tract income above the median. The bottom panel shows loan and borrower statistics for male and female borrowers in low- and high-income tracts. The variables are as described in Table 1. N = 2,261,818.

Table 3. Descriptive Statistics by Tract Education and Gender

	Low education tracts	High education tracts
Female	0.41	0.42
Tract income ($000)	56.3	88.7
Tract education (BA degree %)	21.3	50.0
Socioeconomic status	-0.39	0.39
Tract to MSA median income	93.5	135
Student math score gender gap	-0.02	0.002
Student reading score gender gap	-0.24	-0.24
Occupation math skill gender gap	-0.14	0.04
Share of single female HH with kids (%)	52.0	51.0
Share of single male HH with kids (%)	47.5	47.7

	Low education tracts		High education tracts	
	Female	Male	Female	Male
Mortgage rate (%)	4.49	4.47	4.35	4.29
Rate spread (%)	0.60	0.60	0.41	0.38
Net loan cost (%)	2.34	2.20	1.74	1.53
Total fees (%)	2.50	2.36	1.89	1.70
Total credits (%)	0.16	0.16	0.16	0.17
Credit score	743	743	754	754
Loan-to-value ratio (%)	76.6	79.0	73.9	76.6
High balance loan	6×10^{-5}	8×10^{-5}	3×10^{-5}	2×10^{-5}
Investment property	0.03	0.05	0.02	0.03
Manufactured home	0.01	0.02	0.002	0.002
Subordinate loan	0.02	0.02	0.02	0.02
Borrower income ($000)	66.2	80.8	89.9	115
Loan amount ($000)	182	200	249	285
Debt-to-income ratio (%)	37.5	36.3	37.1	35.8
Loan term (months)	336	332	337	331
Percentage White (%)	67.9	68.2	72.9	69.5

This table presents the descriptive statistics of the key variables in the study for low- and high-education tracts, and by gender. The top panel shows the percentage of female borrowers and area statistics for the variables at the tract level. Low-education tracts are defined as those where the share of the tract population over 25 years old with a bachelor's degree or higher is below the median, while high-education tracts are defined as those where this share is above the median. The bottom panel shows loan and borrower statistics for male and female borrowers in low- and high-education tracts. The variables are as described in Table 1. N = 2,261,818.

Table 4. Borrower Gender and Loan Cost

	Mortgage rate			Rate spread		
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.028***	0.042***	0.008***	0.017***	0.024***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Tract × Orig. yr-m FE	Yes	Yes	Yes	Yes	Yes	Yes
Pricing controls	No	Yes	Yes	No	Yes	Yes
Additional controls	No	No	Yes	No	No	Yes
Observations	1,826,143	1,826,143	1,703,815	1,773,253	1,773,253	1,653,500
R-squared	0.674	0.763	0.831	0.404	0.615	0.661

	Net loan cost			Total fees	Total credits
	(7)	(8)	(9)	(10)	(11)
Female	0.166***	0.123***	-0.005***	-0.009***	-0.004***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Tract × Orig. yr-m FE	Yes	Yes	Yes	Yes	Yes
Pricing controls	No	Yes	Yes	Yes	Yes
Additional controls	No	No	Yes	Yes	Yes
Observations	1,826,143	1,826,143	1,703,815	1,703,815	1,703,815
R-squared	0.453	0.510	0.622	0.638	0.328

Standard errors, clustered at the Census tract level, are in parentheses. All models include Census tract × origination year and month fixed effects. The pricing controls follow the loan-level price adjustment grids of Fannie Mae and the corresponding grids of Freddie Mac called credit fees. These grids determine the credit-risk pricing adjustments to the guarantee fee that lenders must pay the GSEs for guaranteeing the mortgage. The pricing controls include dummies for nine loan-to-value categories, dummies for nine credit score categories, dummies for manufactured homes, investment properties, high balance mortgages, borrowers who have a subordinate mortgage in addition to the first lien, loan purpose (home purchase, refinancing or cash-out refinancing) and for the number of units (1 to 4). The additional controls include dummies for eight borrower race and ethnicity categories, income deciles, loan amount deciles, debt-to-income deciles, dummies for three loan term ranges and dummies for seven borrower age groups. ***$p<0.01$.

Table 5. Borrower Gender, Census Tract Income and Loan Cost

	Mortgage rate			Rate spread		
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.027***	0.041***	0.008***	0.016***	0.024***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female × Tract income	0.013***	0.008***	0.004***	0.012***	0.007***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Tract × Orig. yr-m FE	Yes	Yes	Yes	Yes	Yes	Yes
Pricing controls	No	Yes	Yes	No	Yes	Yes
Additional controls	No	No	Yes	No	No	Yes
Observations	1,826,129	1,826,129	1,703,801	1,773,239	1,773,239	1,653,486
R-squared	0.674	0.763	0.831	0.404	0.615	0.661

	Net loan cost			Total fees	Total credits
	(7)	(8)	(9)	(10)	(11)
Female	0.164***	0.121***	-0.006***	-0.010***	-0.004***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Female × Tract income	0.021***	0.013***	0.008***	0.005**	-0.003***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Tract × Orig. yr-m FE	Yes	Yes	Yes	Yes	Yes
Pricing controls	No	Yes	Yes	Yes	Yes
Additional controls	No	No	Yes	Yes	Yes
Observations	1,826,129	1,826,129	1,703,801	1,703,801	1,703,801
R-squared	0.453	0.510	0.622	0.638	0.328

Standard errors, clustered at the Census tract level, are in parentheses. Tract income is the median tract income in 2017, standardized to have a mean of zero and a standard deviation of one. Pricing controls, additional controls and fixed effects are as in Table 4. **p<0.05; ***p<0.01.

Table 6. Borrower Gender, Census Tract Education and Loan Cost

	Mortgage rate			Rate spread		
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.027***	0.040***	0.008***	0.015***	0.023***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female × Tract education	0.014***	0.011***	0.005***	0.014***	0.010***	0.007***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Tract × Orig. yr-m FE	Yes	Yes	Yes	Yes	Yes	Yes
Pricing controls	No	Yes	Yes	No	Yes	Yes
Additional controls	No	No	Yes	No	No	Yes
Observations	1,826,137	1,826,137	1,703,809	1,773,247	1,773,247	1,653,494
R-squared	0.674	0.763	0.831	0.404	0.615	0.661

	Net loan cost			Total fees	Total credits
	(7)	(8)	(9)	(10)	(11)
Female	0.164***	0.120***	-0.006***	-0.010***	-0.004***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Female × Tract education	0.023***	0.023***	0.009***	0.005***	-0.004***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Tract × Orig. yr-m FE	Yes	Yes	Yes	Yes	Yes
Pricing controls	No	Yes	Yes	Yes	Yes
Additional controls	No	No	Yes	Yes	Yes
Observations	1,826,137	1,826,137	1,703,809	1,703,809	1,703,809
R-squared	0.453	0.510	0.622	0.638	0.328

Standard errors, clustered at the Census tract level, are in parentheses. Tract education is the share of the tract population over 25 years old with a bachelor's degree or higher in 2017, standardized to have a mean of zero and a standard deviation of one. Pricing controls, additional controls and fixed effects are as in Table 4. ***p<0.01.

Table 7. Mortgage Default, Borrower Gender and Local Factors

	90-Day Delinquency					
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.451***	-0.459***	-0.448***	-0.457***	-0.456***	-0.447***
	(0.043)	(0.042)	(0.043)	(0.036)	(0.036)	(0.037)
Female × Tract income		-0.046**			0.025	
		(0.018)			(0.023)	
Female × Tract education			-0.027			0.043
			(0.026)			(0.030)
Origination year-month FE	Yes	Yes	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Pricing controls	No	No	No	Yes	Yes	Yes
Additional controls	No	No	No	Yes	Yes	Yes
Observations	9,986,448	9,986,082	9,986,366	9,980,437	9,980,071	9,980,355
R-squared	0.001	0.001	0.001	0.002	0.002	0.002

Standard errors, clustered at the county and origination year-month level, are in parentheses. Coefficients and standard errors are multiplied by 1000. The unit of observation is a loan-month, and the sample contains 320,939 loans. All models include origination year-month fixed effects and county fixed effects. The pricing controls include dummies for nine loan-to-value categories, dummies for nine credit score categories, dummies for manufactured homes, investment properties, high balance mortgages, borrowers who have a subordinate mortgage in addition to the first lien, loan purpose (home purchase, refinancing or cash-out refinancing) and for the number of units (1 to 4). The additional controls include property value, loan age, dummies for eight borrower race and ethnicity categories, income deciles, loan amount deciles, debt-to-income deciles, dummies for three loan term at origination ranges and dummies for seven borrower age groups. In addition to the interaction terms, all models include the linear terms. **$p<0.05$; ***$p<0.01$.

Table 8. Mortgage Prepayment, Borrower Gender and Local Factors

	Prepayment					
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.228***	-0.211***	-0.230***	-0.042***	-0.042***	-0.048***
	(0.017)	(0.015)	(0.016)	(0.011)	(0.011)	(0.011)
Female × Tract income		-0.098***			-0.065***	
		(0.012)			(0.012)	
Female × Tract education			-0.095***			-0.055***
			(0.014)			(0.013)
Origination year-month FE	Yes	Yes	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Pricing controls	No	No	No	Yes	Yes	Yes
Additional controls	No	No	No	Yes	Yes	Yes
Observations	10,450,011	10,449,605	10,449,889	10,443,544	10,443,138	10,443,422
R-squared	0.0025	0.0029	0.0029	0.0051	0.0051	0.0051

Standard errors, clustered at the county and origination year-month level, are in parentheses. Coefficients and standard errors are multiplied by 100. All models include origination year and month fixed effects and county fixed effects. The pricing controls include dummies for nine loan-to-value categories, dummies for nine credit score categories, dummies for manufactured homes, investment properties, high balance mortgages, borrowers who have a subordinate mortgage in addition to the first lien, loan purpose (home purchase, refinancing or cash-out refinancing) and for the number of units (1 to 4). The additional controls include property value, loan age, dummies for eight borrower race and ethnicity categories, income deciles, loan amount deciles, debt-to-income deciles, dummies for three loan term at origination ranges and dummies for seven borrower age groups. In addition to the interaction terms, all models include the linear terms. ***$p<0.01$.

Table 9. Mortgage Prepayment, Rate Decline and Gender

	Prepayment					
	(1)	(2)	(3)	(4)	(5)	(6)
Female × Rate drop	-0.230***	-0.203***	-0.246***	-0.186***	-0.164***	-0.203***
	(0.023)	(0.020)	(0.021)	(0.024)	(0.020)	(0.022)
Female × Tract income × Rate drop		-0.141***			-0.127***	
		(0.018)			(0.017)	
Female × Tract education × Rate drop			-0.136***			-0.126***
			(0.019)			(0.018)
Origination year-month FE	Yes	Yes	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Pricing controls	No	No	No	Yes	Yes	Yes
Additional controls	No	No	No	Yes	Yes	Yes
Observations	10,450,011	10,449,605	10,449,889	10,443,544	10,443,138	10,443,422
R-squared	0.0059	0.0065	0.0065	0.0102	0.0105	0.0104

Standard errors, clustered at the county and origination year-month level, are in parentheses. Coefficients and standard errors are multiplied by 100. Controls and fixed effects are as in Table 8. In addition to the triple interactions all models include the linear terms and the pairwise interactions of the variables in the triple interaction. ***p<0.01.

Table 10. Overall Mortgage Cost, Borrower Gender and Local Factors

	Overall Cost								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female	0.769***	0.744***	0.744***	0.761***	0.744***	0.740***	0.045***	0.034***	0.034***
	(0.014)	(0.014)	(0.014)	(0.013)	(0.013)	(0.013)	(0.012)	(0.012)	(0.012)
Female × Tract income		0.239***			0.169***			0.113***	
		(0.014)			(0.013)			(0.012)	
Female × Tract education			0.234***			0.208***			0.115***
			(0.014)			(0.013)			(0.012)
Tract × Orig. yr-m FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pricing controls	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Additional controls	No	No	No	No	No	No	Yes	Yes	Yes
Observations	1,826,143	1,826,129	1,826,137	1,826,143	1,826,129	1,826,137	1,703,815	1,703,801	1,703,809
R-squared	0.440	0.440	0.440	0.521	0.521	0.522	0.587	0.587	0.587

Standard errors, clustered at the Census tract level, are in parentheses. Overall cost is expressed as a percentage of the loan amount and is calculated for 10 years after origination. The overall cost includes origination cost, interest, and additional cost if the loan gets refinanced within this interval. Tract income is the median tract income in 2017, standardized to have a mean of zero and a standard deviation of one. Tract education is the share of the tract population over 25 years old with a bachelor's degree or higher in 2017, standardized to have a mean of zero and a standard deviation of one. Pricing controls, additional controls and fixed effects are as in Table 4. ***p<0.01.

Table 11. Mortgage Literacy, Borrower Gender, and Income

	Knowledge Topic									
	Subprime	APR	Amortization	Points	Advanced V1	Advanced V2	Process	Fixed rate	Payment	Basic
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Female	-0.292***	-0.250***	-0.199***	-0.244***	-0.289***	-0.276***	-0.123***	-0.248***	0.020	-0.143***
	(0.029)	(0.030)	(0.030)	(0.040)	(0.029)	(0.039)	(0.031)	(0.032)	(0.032)	(0.031)
Income 50-75k	0.119***	0.128***	0.147***	0.126***	0.155***	0.185***	0.090***	0.096***	0.121***	0.127***
	(0.029)	(0.030)	(0.029)	(0.039)	(0.029)	(0.038)	(0.030)	(0.031)	(0.031)	(0.031)
Income 75-100k	0.287***	0.268***	0.299***	0.293***	0.334***	0.370***	0.210***	0.215***	0.211***	0.261***
	(0.030)	(0.030)	(0.030)	(0.039)	(0.029)	(0.038)	(0.031)	(0.032)	(0.032)	(0.031)
Income 100-175k	0.359***	0.390***	0.464***	0.517***	0.475***	0.553***	0.337***	0.286***	0.244***	0.355***
	(0.028)	(0.029)	(0.028)	(0.037)	(0.028)	(0.037)	(0.029)	(0.030)	(0.030)	(0.030)
Income >175k	0.733***	0.558***	0.742***	0.810***	0.799***	0.881***	0.535***	0.426***	0.373***	0.546***
	(0.031)	(0.032)	(0.031)	(0.041)	(0.031)	(0.040)	(0.033)	(0.034)	(0.034)	(0.033)
Female × Income 50-75k	-0.005	-0.019	-0.022	0.041	-0.018	-0.011	0.051	0.047	-0.065	0.012
	(0.041)	(0.042)	(0.041)	(0.054)	(0.040)	(0.053)	(0.043)	(0.044)	(0.044)	(0.043)
Female × Income 75-100k	-0.074*	-0.057	-0.042	-0.011	-0.068	-0.079	-0.020	0.039	-0.110**	-0.039
	(0.043)	(0.044)	(0.043)	(0.056)	(0.042)	(0.055)	(0.045)	(0.046)	(0.046)	(0.045)
Female × Income 100-175k	-0.012	-0.088**	-0.076*	-0.096*	-0.068*	-0.104**	-0.046	0.109**	0.007	0.030
	(0.040)	(0.041)	(0.040)	(0.052)	(0.039)	(0.051)	(0.042)	(0.043)	(0.043)	(0.042)
Female × Income >175k	-0.162***	-0.150***	-0.168***	-0.134**	-0.188***	-0.187***	-0.079	0.062	-0.078	-0.039
	(0.048)	(0.049)	(0.048)	(0.061)	(0.048)	(0.061)	(0.050)	(0.052)	(0.052)	(0.051)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	21,177	21,177	21,177	12,654	21,177	12,654	21,177	21,177	21,177	21,177
R-squared	0.111	0.077	0.132	0.135	0.140	0.159	0.058	0.083	0.036	0.075

This table presents the coefficient estimates obtained from regressions of mortgage literacy on female indicator, indicators for income groups, and their interactions. Regressions include survey wave fixed effects, dummy variable controls for four education levels, five race and ethnicity categories, six age groups and nine credit score categories, and use survey weights. Advanced literacy V1 is the sum of answers to prime vs. subprime, interest rate vs. APR and amortization questions. Advanced literacy V2 also include answers to points versus interest rates question, which starts in wave 11. Basic literacy is the sum of the answers to mortgage process, fixed vs. variable rate, consequences of nor making payments questions. All dependent variables are standardized to have a mean of zero and a standard deviation of one. Source: National Survey of Mortgage Originations for conventional loans with a single borrower originated between 2014 and 2021Q2 (30 waves). *p<0.10; **p<0.05; ***p<0.01.

Table 12. Mortgage Literacy, Borrower Gender and Education

	Knowledge Topic									
	Subprime	APR	Amortization	Points	Advanced V1	Advanced V2	Process	Fixed rate	Payment	Basic
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Female	-0.213***	-0.214***	-0.127***	-0.187***	-0.215***	-0.234***	-0.130***	-0.185***	-0.080*	-0.162***
	(0.040)	(0.041)	(0.040)	(0.053)	(0.040)	(0.052)	(0.042)	(0.043)	(0.043)	(0.042)
Some college	0.113***	0.110***	0.205***	0.178***	0.169***	0.178***	0.0856***	0.234***	0.127***	0.185***
	(0.028)	(0.029)	(0.028)	(0.037)	(0.028)	(0.036)	(0.030)	(0.030)	(0.030)	(0.030)
College graduate	0.256***	0.217***	0.458***	0.344***	0.368***	0.375***	0.249***	0.344***	0.127***	0.294***
	(0.028)	(0.028)	(0.028)	(0.036)	(0.027)	(0.035)	(0.029)	(0.030)	(0.030)	(0.029)
Postgraduate	0.265***	0.251***	0.507***	0.357***	0.405***	0.402***	0.226***	0.358***	0.128***	0.292***
	(0.030)	(0.031)	(0.030)	(0.039)	(0.030)	(0.038)	(0.031)	(0.032)	(0.032)	(0.032)
Female × Some college	-0.030	-0.003	0.029	0.031	-0.001	0.038	0.086*	-0.021	0.053	0.048
	(0.048)	(0.049)	(0.048)	(0.063)	(0.047)	(0.062)	(0.050)	(0.052)	(0.052)	(0.051)
Female × College graduate	-0.165***	-0.092**	-0.169***	-0.108*	-0.168***	-0.141**	-0.027	-0.032	0.097**	0.018
	(0.045)	(0.047)	(0.046)	(0.059)	(0.045)	(0.059)	(0.048)	(0.049)	(0.049)	(0.048)
Female × Postgraduate	-0.179***	-0.207***	-0.253***	-0.215***	-0.251***	-0.248***	-0.063	0.025	0.028	-0.003
	(0.047)	(0.049)	(0.047)	(0.062)	(0.047)	(0.061)	(0.049)	(0.051)	(0.051)	(0.050)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	21,177	21,177	21,177	12,654	21,177	12,654	21,177	21,177	21,177	21,177
R-squared	0.112	0.078	0.134	0.136	0.142	0.161	0.058	0.083	0.035	0.075

This table presents the coefficient estimates obtained from regressions of mortgage literacy on female indicator, indicators for education levels, and their interactions. Regressions include survey wave fixed effects, dummy variable controls for five income groups, five race and ethnicity categories, six age groups and nine credit score categories, and use survey weights. Advanced literacy V1 is the sum of answers to prime vs. subprime, interest rate vs. APR and amortization questions. Advanced literacy V2 also include answers to points versus interest rates question, which starts in wave 11. Basic literacy is the sum of the answers to mortgage process, fixed vs. variable rate, consequences of nor making payments questions. All dependent variables are standardized to have a mean of zero and a standard deviation of one. Source: National Survey of Mortgage Originations for conventional loans with a single borrower originated between 2014 and 2021Q2 (30 waves). *p<0.10; **p<0.05; ***p<0.01.

Table 13. Mortgage Shopping Behavior, Gender and Financial Literacy

	Initial Familiarity			Lender Search		
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.161***		-0.013	-0.063***		-0.051***
	(0.014)		(0.013)	(0.007)		(0.007)
Financial literacy		0.425***	0.424***		0.039***	0.035***
		(0.007)	(0.007)		(0.004)	(0.004)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	21,177	21,177	21,177	21,177	21,177	21,177
R-squared	0.140	0.276	0.276	0.028	0.029	0.032
	Loan Search			Confirm Terms		
	(7)	(8)	(9)	(10)	(11)	(12)
Female	-0.126***		-0.077***	-0.028***		-0.014*
	(0.015)		(0.015)	(0.008)		(0.008)
Financial literacy		0.146***	0.139***		0.044***	0.043***
		(0.007)	(0.007)		(0.004)	(0.004)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	21,177	21,177	21,177	15,741	15,741	15,741
R-squared	0.033	0.048	0.049	0.020	0.027	0.028

This table presents the coefficient estimates obtained from regressions of mortgage search behaviors on female indicator and financial literacy. Regressions include survey wave fixed effects, dummy variable controls for five income groups, four education levels, five race and ethnicity categories, six age groups, and nine credit score categories, and use survey weights. The financial literacy variable is what we defined previously as Advanced Literacy V1. Source: National Survey of Mortgage Originations for conventional loans with a single borrower originated between 2014 and 2021Q2 (30 waves). *p<0.10; ***p<0.01.

Table 14. Overall Mortgage Cost, Borrower Gender and Math Skills

	Overall Cost					
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.807***	0.845***	0.014	0.757***	0.752***	0.040***
	(0.024)	(0.020)	(0.011)	(0.014)	(0.013)	(0.012)
Female × Student math score gap	0.183***	0.148***	0.041***			
	(0.021)	(0.018)	(0.012)			
Female × Occupation math skill gap				0.177***	0.147***	0.083***
				(0.014)	(0.013)	(0.012)
Location × Orig. yr-m FE	Yes	Yes	Yes	Yes	Yes	Yes
Pricing controls	No	Yes	Yes	No	Yes	Yes
Additional controls	No	No	Yes	No	No	Yes
Observations	2,159,375	2,159,375	2,041,530	1,826,137	1,826,137	1,703,809
R-squared	0.218	0.346	0.452	0.440	0.521	0.587

Standard errors in parentheses, clustered at the schools district level in models (1) to (3) and at the Census tract level in models (4) to (6). Overall cost is expressed as a percentage of the loan amount and is calculated for 10 years after origination. The overall cost includes origination cost, interest, and additional cost if the loan gets refinanced within this interval. Student math score gap is the difference between the average male score and the average female score in standardized math tests from grades 3 to 8 in each school district, from the SEDA dataset. Occupation math skill gap is the difference between the average level of math skill importance in the occupations held by males and females within each Census tract. This average math skill level is calculated by weighting each occupation's math skill importance, from the O*NET dataset, by the 2017 employment figures for males and females in major occupation groups within each tract. In the regressions, the math score gap and the math skill gap is standardized to have a mean of zero and a standard deviation of one. Models (1) to (3) include school district × origination year-month fixed effects. Models (4) to (6) include Census tract × origination year-month fixed effects. Pricing controls and additional controls are as in Table 4. ***p<0.01.

APPENDIX

A1 HMDA-GSE Matching Procedure

Below we describe the steps to match the mortgages from Fannie Mae and Freddie Mac, the Government Sponsored Enterprises (GSEs), to the HMDA database.

A1.1 HMDA Data Processing

We begin with all loans in the HMDA 2018 and 2019 files. We keep loans secured by a first lien, for single family homes (1 to 4 units), either site-built or manufactured. We remove reverse mortgages, open-end lines of credit, loans with negative amortization, interest-only, loans with balloon payments and other non-amortizing features. We remove the loans for which the Census tract, the interest rate, the total loan cost or the loan term is missing. Finally, we keep the loans that were originated (action taken = 1) or were purchased (action taken = 6). The total number of mortgages in this raw database is 15,592,356, with 3,272,471 being purchased loans.

A1.1.1 Matching purchased loans with originated loans

Purchased loans are loans that were purchased by a financial institution after originated by another financial institution. These include closed mortgage loans where the purchasing entity has not reviewed the application or made a credit decision prior to purchasing. With the expansion of the HMDA reporting requirements, entities purchasing covered loans are required to collect, record and report on the new and modified data points for purchases that occur on or after January 1, 2018.

However, the entities covered by the extended requirements for loan purchases are not required to report certain data for purchased loans. For some types of data, such as the rate spread, a covered purchaser complies with the HMDA requirements by reporting "not applicable." For certain other types of data, such as the borrower's income, age, ethnicity, race and sex, a covered purchaser may, but is not required to, report such data, and can either report "not applicable" or provide the data if it so elects.

To link purchased loans to GSE loans where applicable, we first need to address the issue of missing key loan characteristics in the purchased loan data. Many critical fields required for both matching and future analysis, such as borrower gender, credit score, and rate spread, are often absent. Therefore, we must match purchased loans with their corresponding originated loans in the HMDA dataset to recover these missing data fields. These fields are essential for accurately conducting the HMDA-GSE matching process and for supporting our subsequent analysis.

We match the originated loans with the purchased loans on the following ten variables, imposing an exact match for all variables: Census tract, loan amount, interest rate, total loan cost, loan term, conforming loan limit, construction method, occupancy type, number of units and loan type. This results in 2,170,157 matched loans from the 3,272,471 purchased loans (66.3% match rate).

Not all purchased loans have a match, as some of them were probably originated in 2017, and sold in 2018, and we were only matching the loans originated in 2018 or 2019. Moreover, we impose very strict matching conditions. A slight difference in the reported loan amount or interest rate would prevent a match even if it were the same loan.

We do the following post-match tests to ensure correct matches. We drop matched loans if one of the following applies:

(a) The origination year is after the year of purchase

(b) The absolute difference between the property value of the originated and purchased loan is larger than $5,000

(c) The absolute difference between the borrower income of the originated and purchased loan is larger than $5,000

(d) One of the loans in a matched pair is reported as a high-cost mortgage and the other as not a high-cost mortgage

(e) The matched loans do not have the same loan purpose (home purchase, home improvement, refinancing or cash-out refinancing) when the loan purpose is different from "Other purpose" or "Not applicable."

The next step ensures that we only keep unique matches. Originated loans that match to multiple purchased loans might have been sold multiple times. If an originated loan matches to multiple purchased loans we keep the purchased loan that is sold to Fannie Mae or Freddie Mac, or the purchaser type is "Not applicable". After this cleaning procedure, 99.91% of the loans have a unique match. We drop the remaining 0.09% originated or purchased loans that match to more than one loan. In total, 1,922,181 originated loans uniquely match to purchased loans in HMDA between 2018 and 2019.

A1.1.2 Updating the HMDA dataset

Once we have the unique matches of the purchased loans to the originated loans, we replace the demographic variables of the purchased loans, which were mostly missing, with the demographic variables of the matched originated loans. These variables include the applicant and co-applicant ethnicity, race, sex, age, credit score type, and various loan characteristics, such as combined LTV, rate spread, total points and fees, origination charges, discount points, lender credits, prepayment penalty term, intro rate period, debt to income ratio, direct submission of application indicator, initially payable to institution indicator, automated underwriting system used and indicator for preapproval request.

The updated HMDA dataset contains the originated mortgages that were not sold to another financial institution before they were potentially sold to the GSEs. In addition, for those mortgages that were sold to another financial institution, the updated dataset replaces the originated loans with the matched purchased loans, since these will be matched in the next step to the loans in the GSE databases. The total number of unique mortgages in the updated HMDA dataset is 13,670,175.

A1.2 GSE Data Processing

We put together the mortgages acquired by Fannie Mae or Freddie Mac, from January 1st, 2018 to December 31st, 2020, from quarterly data files, publicly available on the websites of the GSEs. The mortgage database contains mortgage characteristics such as the month and year of origination, the interest rate, the 3-digit zip code and the identifier of the financial institution that originated the mortgage. In this initial database there are 8,195,899 mortgages from Fannie Mae and 6,988,080 from Freddie Mac.

A1.3 HMDA-GSE Matching

We create two separate files from the updated HMDA data. The first one, to be matched with loans from Fannie Mae, keeps the mortgages for which the purchaser type is Fannie Mae. The second one, to be matched with loans from Freddie Mac, keeps the mortgages for which the purchaser type is Freddie Mac. We also do a second round matching with the purchaser type being "Not applicable", which improves the percentage of GSE loans matched. Throughout the full process we impose strict matching criteria over multiple loan characteristics, and ensure that the matches are unique.

A1.3.1 Lenders

We match the loans on several dimensions. First, we match the financial institution in HMDA (legal entity identifier) that holds the loan to the financial institution in the GSE dataset (seller) that sold the loan. The matching of the financial institutions uses the HMDA lender file for the years 2018 and 2019. This crosswalk file, called "The Avery file", was constructed by Robert Avery and contains matching information for all lenders who have ever filed a HMDA report. The HMDA ID is matched to information in the FRB's NIC system and FFIEC or TFR Call Reports for each filing year.

We manually check the matching of lenders and adjust for mergers and acquisitions of lenders within the time frame of the loan originations. Since the seller identifiers in the GSE data are set to "Other sellers" for smaller financial institutions, we allow the matching algorithm to match "Other sellers" to a number of legal entity identifiers in HMDA.

A1.3.2 Geographical units

Second, we match the Census tract identifier in HMDA, with the 3-digit zip code in the GSE datasets. To do this, we utilize the 2018 crosswalk file from the U.S. Department of Housing and Urban Development (HUD). At this stage we keep all the observations, even if a Census tract belongs to more than one 3-digit zip code. Most of the duplicated HMDA loans that were created in the process of matching them to multiple zip codes are eliminated at later stages since we require matching on many other loan characteristics.

A1.3.3 Other loan characteristics

We make the following adjustments to the HMDA data, to improve potential matches with the GSE loans:

(a) We create a new combined LTV variable, to follow the GSE reporting rule: The result of the calculation of the LTV must be truncated (shortened) to two decimal places, then rounded up to the nearest whole percent. For example, 94.01% is converted to 95%, and 80.001% is converted to 80%.

(b) We create a new variable for the number of borrowers, that takes the values of 1, 2 or missing, based on the values of the applicant sex and co-applicant sex.

We match the mortgages in the updated HMDA dataset to the mortgages in the Fannie

Mae and Freddie Mac datasets on the following variables, in addition to the lender and geographical unit we described earlier, imposing an exact match for these five variables: interest rate, loan term at origination, combined LTV, number of borrowers and occupancy type (principal residence, second residence or investment property).

Next, we perform post-match tests and we require that the loans match on a number of additional characteristics. For these characteristics we allow some margins for the values in the HMDA and GSE databases not to be exactly equal.

(a) We drop loans for which the origination year in HMDA is after the GSE acquisition year.

(b) We drop loans for which the absolute difference between the loan amount in HMDA and the loan amount in the GSE database is larger than \$5,000

(c) We require the following matching of the loan purpose from the two databases: Loans with purpose in HMDA of 1 (home purchase) should match with the purpose from GSEs of "P" (purchase), loans with purpose in HMDA of 31 (refinancing) should match with the purpose from GSEs of "R" (refinance), and loans with purpose in HMDA of 32 (cash-out refinancing) should match with the purpose from GSEs of "C" (cash-out refinance). Other values of loan purpose from HMDA (2 - home improvement, 4 - other purpose or 5 - not applicable) are let to match with any of the three purposes from GSEs.

(d) We require the following matching of the origination channel used by the party that delivered the loan to the issuer. In the GSE databases the origination channel is either correspondent (a loan that is originated by a party other than a mortgage loan seller and is then sold to a mortgage loan seller), retail (a loan for which the mortgage loan seller takes the loan application and then processes, underwrites, funds, and sells the loan to the GSE) or broker. We drop loans if the channel is correspondent and the action taken in HMDA is 1 (loan originated). Also, we drop loans if the channel is retail and the action taken in HMDA is 6 (purchased loan).

(e) We require the following matching of the debt-to-income ratio (DTI) in HMDA that is reported either as an integer or as a bucket of values to the DTI in the GSE databases that is reported as an integer. The GSE reporting rule for DTI is as follows: The DTI value is rounded down to the nearest whole number if the decimal is less than 0.5, and rounded up if the decimal is 0.5 or greater.

DTI from GSEs (%)	DTI in HMDA
<20	“<20%”
20	“<20%” or “20%-<30%”
>20 and <30	“20%-<30%”
30	“20%-<30%” or “30%-<36%”
>30 and <36	“30%-<36%”
36	“30%-<36%” or “36” or “37”
37	“36” or “37” or “38”
38	“37” or “38” or “39”
...	...
48	“47” or “48” or “49”
49	“48” or “49” or “50%-60%”
50	“49” or “50%-60%”
>50 and <60	“50%-60%”
60 or 61	“50%-60%” or “>60%”
>61	“>60%”

(f) We require matching of the number of total units, 1 to 4, in the two databases.

(g) We keep only conventional loans and drop FHA, FSA/RHS and VA loans. After the matching and the previous additional checks more than 99.99% of the matched loans are conventional, which is a confirmation that the algorithm matches the HMDA loans to the loans acquired by Fannie and Freddie correctly.

After we implement the above matching algorithm, we require that each loan from HMDA is matched to a unique loan from either Fannie or Freddie and vice versa. Out of the matched loans from Fannie 96.95% are unique, and out of the matched loans from Freddie 97.76% are unique. Out of the HMDA loans matched to Fannie 95.95% are unique, and out of the HMDA loans matched to Freddie 97.24% are unique. We remove from the database the remaining loans that have multiple matches. The HMDA-GSE matched dataset contains 4,197,266 loans.

A2 Occupation Math Skill Gap: Data and Variable Construction

The occupation math skill gap measures the relative importance of math skills in the occupations held by men versus women within each Census tract. We develop this novel variable by combining the occupational composition of males and females in each Census tract with the math skill importance assigned to each occupation, using it as an indicator of gender disparities in the adult population's ability to apply mathematical skills to problem-solving. These skills are crucial for effective mortgage decisions, such as securing lower costs, evaluating trade-offs, or refinancing at optimal times. In this section, we describe the datasets and calculations used to construct the occupation math skill gap measure.

The two primary data sources for constructing this variable are the Occupational Information Network (O*NET), which provides detailed scores on the math skills and knowledge required for specific occupations, and the American Community Survey (ACS), which offers Census tract employment data by gender. Because O*NET scores are available for a highly granular set of occupations, while ACS data are at broader occupation levels, we use various crosswalks to connect these datasets.

O*NET is a comprehensive database that provides detailed information about U.S. occupations, relying on large-scale surveys of workers, employers, and experts. Specifically, we use the May 2017 database from the O*NET Database Releases Archive. Using the 2017 occupation classification ensures the maximum match with the 2017 occupation codes from the Bureau of Labor Statistics (BLS).

The O*NET database contains multiple files that quantify different characteristics and abilities required for each occupation. We use the files for "Skills" and "Knowledge", and focus on the importance and level of mathematics in each one. We describe the construction of the measure for math skill importance, and replicate the steps and analysis for math skill level, math knowledge level and math knowledge importance.

The O*NET "Skills" file contains a list of 964 occupations (8-digit occupation codes) and reports a "math skill importance" score for each one. This score ranges from from 1 to 5, and measures the degree to which mathematical problem solving abilities are required to perform the tasks associated with the specific occupation. For example, Mathematicians score 5/5 in math skill importance, Financial Analysts 3.9/5, Sales Managers 3/5 and Actors 1/5.

Using the occupation codes, we merge the O*NET dataset with the number of people employed in each occupation from the 2017 Occupational Employment and Wage Statistics (OEWS) Tables from BLS. The BLS dataset classifies occupations in various levels of detail, specifically in major, minor, broad and detailed groups. We keep the occupations in the most refined classification, which is the detailed groups. This results in 809 6-digit occupation codes.

We merge the occupation codes from O*NET with the occupation codes from BLS. Before merging, we process the 8-digit codes in O*NET to create 6-digit codes as follows: If the O*NET code ends in "00" we keep that occupation and remove any other occupation with the same 6-digit code (and different last two digits). We do this because the codes ending in digits other than "00" are subsets of the main occupation, e.g. 11-1011.00 is the code for Chief Executives and 11-1011.03 is the code for Chief Sustainability Officers. For occupations with the same 6-digit codes that do not have a representative occupation ending in "00" we allocate the average value of the math skill importance to the 6-digit code. After creating the 6-digit codes, we do a visual check of the occupation descriptions in both databases, and do manual adjustments to the codes, so that the same occupation description always corresponds to the same 6-digit occupation code in both databases. The merge of O*NET math skill importance dataset with the BLS employment data results in 765 unique 6-digit occupation codes.

Next, we obtain table S2401 with Census tract employment data by gender from the 2017 American Community Survey (ACS) 5-year estimates. This dataset contains descriptions of 25 occupations, which we match with major occupation groups in BLS that correspond to 2- or 3-digit codes. We merge the ACS data with the O*NET-BLS dataset, using the 25 major occupation group codes.

To calculate the gender gap in math skill importance in each Census tract we follow these steps: First, we calculate the weighted average of the math skill importance for the major occupation groups, using as weights the 2017 national employment figures from BLS. Second, we calculate the weighted average math skill importance for males and females in each Census tract, using as weights the employment numbers for each gender in the ACS. Finally, we calculate the gap as the math skill importance score of males minus females.

A3 National Survey of Mortgage Originations (NSMO) Survey Questions

We utilize responses from NSMO questions to construct measures of financial literacy and assess the thoroughness of the mortgage search conducted by borrowers. Our analysis focuses on single borrowers (where borrower_num=1) who are also the survey respondents (borrower_r=1).

A3.1 Financial Literacy

We assess the financial literacy of survey participants based on their responses to the survey block X56. Respondents chose their answers from a three-point scale: “Not at all” (coded as 1) to “Very” (coded as 3). The text and codes of the questions, along with the distribution of the responses, are as follows:

How well could you explain to someone the...	Code	Response (%)			Score	
		Not	Somewhat	Very	mean	std
Process of taking out a mortgage	x56a	3.8	48.6	47.5	2.41	0.57
Difference between a fixed- and an adjustable-rate mortgage	x56b	5.8	27.0	67.2	2.58	0.62
Consequences of not making required mortgage payments	x56f	6.6	26.7	66.8	2.58	0.62
Difference between a prime and subprime loan	x56c	42.8	35.4	21.7	1.74	0.77
Difference between a mortgage's interest rate and its APR	x56d	25.6	46.2	28.3	1.98	0.73
Amortization of a loan	x56e	27.3	34.1	38.7	2.05	0.81
Relationship between discount points and interest rate	x56h	34.2	37.2	28.6	1.89	0.79

We construct the Basic Literacy measure as the sum of scores from the first set of three questions (x56a, x56b, x56f). Advanced Literacy V1 is the sum of the first three questions in the second set (x56c, x56d, x56e), while Advanced Literacy V2 includes all questions in the second set (x56c, x56d, x56e, x56h). Advanced Literacy V2 is only available for waves 11-30, as question x56h was not included in waves 1-10.

A3.2 Mortgage Shopping Behavior

We evaluate mortgage shopping behavior in four stages: initial familiarity, lender search, loan options, and due diligence. The results are in Table 13.

We assess initial familiarity with the mortgage market based on responses to survey block X05, where respondents choose their answers from a three-point scale: "Not at all" (coded as 1) to "Very" (coded as 3):

When you began the process of getting this mortgage, how familiar were you (and any co-signers) with each of the following?	
	Code
The mortgage interest rates available at that time	x05a
The different types of mortgages available	x05b
The mortgage process	x05c
The down payment needed to qualify for a mortgage	x05d
The income needed to qualify for a mortgage	x05e
Your credit history or credit score	x05f
The money needed at closing	x05g

We construct the Initial Familiarity measure by summing the scores from all the questions in the survey block.

To understand the extent of the lender search, we consider the responses to question x11:

How many different mortgage lenders/brokers did you seriously consider before choosing where to apply for this mortgage?

We created a Lender Search indicator variable that equals 1 if the borrowers considered more than one lender, and 0 if they did not.

In order to assess whether the borrower is presented with different loan features and options, we used the survey block X23, which has the following questions:

During the application process were you told about mortgages with any of the following?	Code
An interest rate that is fixed for the life of the loan	x23a
An interest rate that could change over the life of the loan	x23b
A term of less than 30 years	x23c
A higher interest rate in return for lower closing costs	x23d
A lower interest rate in return for paying higher closing costs (discount points)	x23e
Interest-only monthly payments	x23f
An escrow account for taxes and/or homeowner insurance	x23g
A prepayment penalty (fee if the mortgage is paid off early)	x23h

We created a Loan Search variable by adding the number of "yes" responses to these questions.

Finally, we created a Confirm Terms indicator variable if the respondent answered "yes" to question x20h:

> In the process of getting this mortgage from your mortgage lender/broker, did you...
> Check other sources to confirm that the terms of this mortgage were reasonable.

For the additional analyses in Tables A25 and A26 we code the responses separately for each question from the X05 survey block. For example, to measure the familiarity with interest rates, we utilize the question x05a:

> When you began the process of getting this mortgage, how familiar were you (and any co-signers) with each of the following?...
> The mortgage interest rates available at that time.

We create a dummy variable that takes the value of 1 if the answer is "Very" and zero otherwise. We do the same for questions x05b to x05g.

Figures for the Appendix

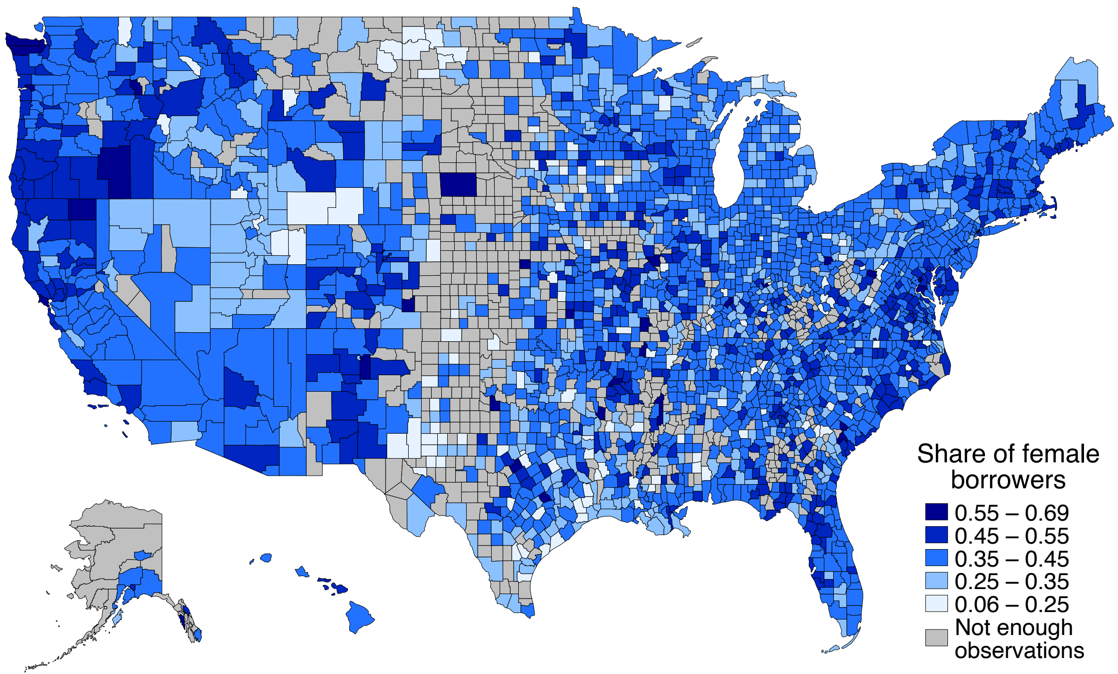

Figure A1. Geographical Distribution of Single Female versus Single Male Borrowers. The map illustrates the share of single female borrowers among all single borrowers across U.S. counties, derived from the merged HMDA-GSE loan dataset. The share is calculated for counties with at least 20 loan observations in the sample.

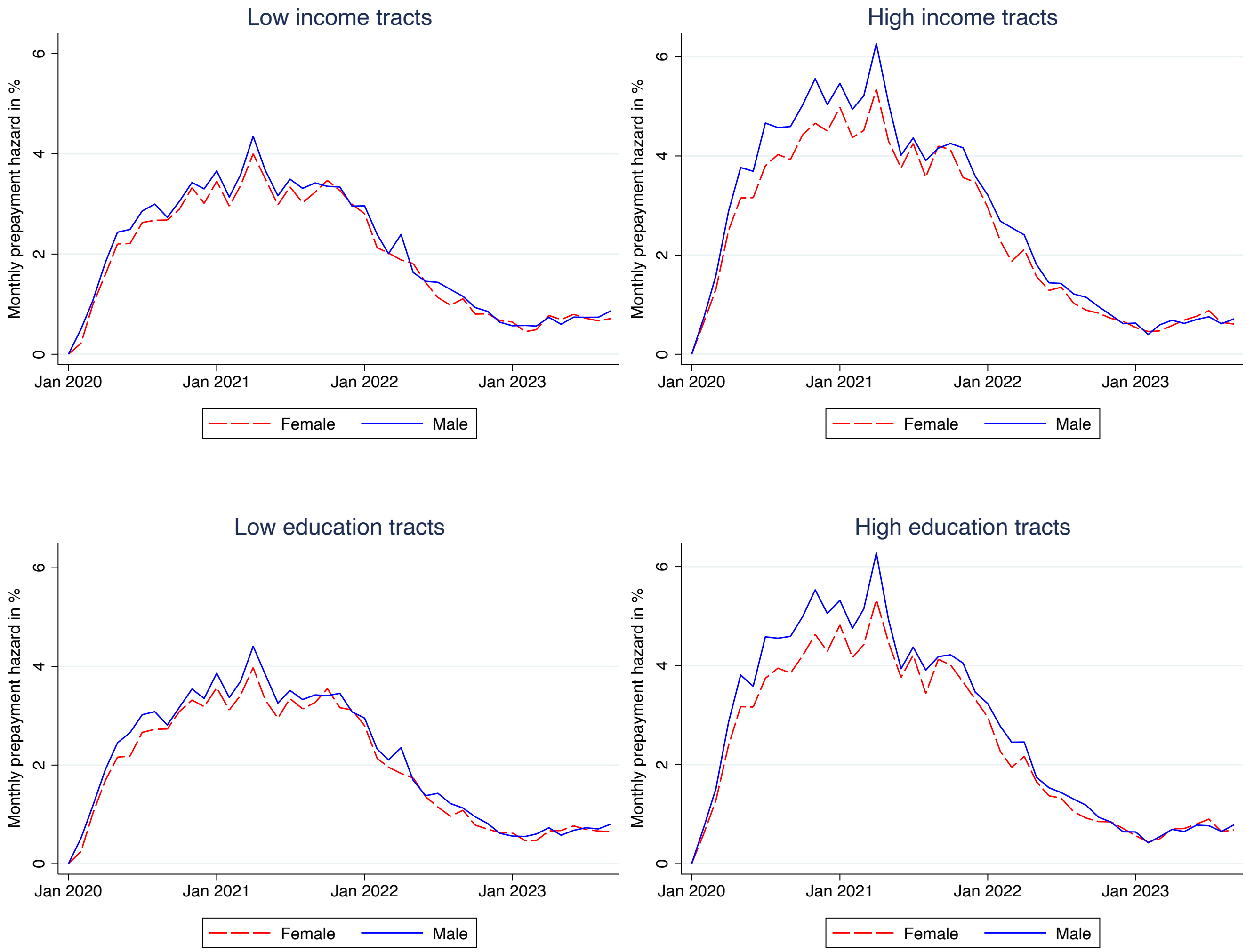

Figure A2. Kaplan–Meier Unconditional Prepayment Hazard Rates by Gender, Income, and Education Over Time. This figure plots the unconditional Kaplan–Meier monthly hazard rates for mortgage prepayment over time, separated by gender. The top-left panel shows the hazard rates for Census tracts with incomes below the median tract income, while the top-right panel shows the rates for Census tracts with incomes above the median tract income. The bottom-left panel displays the hazard rates for Census tracts with educational attainment below the median, and the bottom-right panel shows the rates for tracts with educational attainment above the median. Educational attainment is measured as the share of the tract population over 25 years old with a bachelor's degree. The Kaplan–Meier estimate of the hazard function is calculated as the number of prepayments by each gender in each month, divided by the number of outstanding mortgages for each gender in each month. These hazard rates are unconditional of loan and borrower characteristics. The underlying data come from a 15% random sample of loans originated between 2018 and 2019 from the HMDA-GSE database.

Tables for the Appendix

Table A1. Pairwise Correlations of Local Socioeconomic Indicators

	(1)	(2)	(3)	(4)	(5)	(6)
(1) Tract median income						
(2) Tract education (BA degree %)	0.716					
(3) Socioeconomic status (SES)	0.576	0.440				
(4) Tract to MSA median income	0.756	0.687	0.336			
(5) Student math score gender gap	0.242	0.193	0.517	0.180		
(6) Student reading score gender gap	-0.073	-0.056	-0.049	0.037	0.480	
(7) Occupation math skill gender gap	0.386	0.489	0.294	0.328	0.195	0.007

This table presents pairwise correlations of the socioeconomic variables and gender gaps in math and reading test scores and math skills that we use in the study. N = 2,194,370.

Table A2. Descriptive Statistics for Single and Joint Mortgages

	Single borrower		Primary borrower in joint mortgage	
	Female	Male	Female	Male
Percentage of borrowers (%)	23.2	32.2	11.7	32.9
Mortgage rate (%)	4.42	4.38	4.37	4.33
Rate spread (%)	0.51	0.49	0.47	0.42
Net loan cost (%)	2.03	1.86	1.68	1.62
Total fees (%)	2.19	2.03	1.83	1.77
Total credits (%)	0.16	0.16	0.15	0.15
Credit score	749	748	748	753
Loan-to-value ratio (%)	75.3	77.8	76.4	74.5
High balance loan	4×10^{-5}	5×10^{-5}	3×10^{-5}	2×10^{-5}
Investment property	0.03	0.04	0.02	0.03
Manufactured home	0.01	0.01	0.01	0.01
Subordinate loan	0.02	0.02	0.02	0.02
Borrower income ($000)	78.3	98.0	125	128
Loan amount ($000)	216	242	267	271
Debt-to-income ratio (%)	40.0	38.0	36.2	36.2
Loan term (months)	337	332	330	326
Percentage White (%)	70.4	68.8	74.0	76.6

This table presents descriptive statistics for all loans in the HMDA-GSE matched data, by borrower gender. For mortgages with two borrowers, statistics are shown separately for male and female primary borrowers, as HMDA records gender and race for both applicants and co-applicants. Loan cost variables are expressed as a percentage of the loan amount. Rate spread, loan costs, and borrower income are winsorized at the top and bottom 0.5% to minimize outliers. Source: HMDA data for conventional loans originated in 2018–2019, matched with loans from Fannie Mae and Freddie Mac. N = 4,092,384.

Table A3. Descriptive Statistics for Single and Joint Mortgages by Income and Education

	Single borrower		Primary borrower in joint mortgage		Single borrower		Primary borrower in joint mortgage	
	Female	Male	Female	Male	Female	Male	Female	Male
	Low income tracts				High income tracts			
Percentage of borrowers (%)	25.6	33.6	10.9	29.9	21.1	31.0	12.3	35.6
Mortgage rate (%)	4.49	4.47	4.45	4.42	4.35	4.30	4.31	4.27
Rate spread (%)	0.60	0.60	0.58	0.52	0.41	0.39	0.39	0.35
Net loan cost (%)	2.35	2.23	2.00	1.94	1.70	1.52	1.44	1.38
Total fees (%)	2.51	2.39	2.14	2.08	1.85	1.68	1.59	1.53
Total credits (%)	0.16	0.16	0.15	0.14	0.16	0.17	0.16	0.15
Credit score	745	744	744	750	752	752	751	756
Loan-to-value ratio (%)	77.0	79.2	77.8	75.9	73.4	76.5	75.2	73.4
Borrower income ($000)	66.9	82.3	109	112	90.2	112	137	140
Loan amount ($000)	176	192	216	219	258	290	306	310
Debt-to-income ratio (%)	37.1	35.8	36.6	36.7	37.6	36.2	35.9	35.8
Loan term (months)	336	332	330	325	337	331	331	327
Percentage White (%)	70.1	69.8	74.3	77.7	70.8	67.9	73.7	75.8
	Low education tracts				High education tracts			
Percentage of borrowers (%)	23.9	33.8	11.3	31.0	22.4	31.8	12.1	34.7
Mortgage rate (%)	4.49	4.47	4.45	4.41	4.35	4.29	4.30	4.26
Rate spread (%)	0.60	0.60	0.58	0.52	0.41	0.38	0.38	0.34
Net loan cost (%)	2.34	2.20	1.96	1.91	1.74	1.53	1.44	1.38
Total fees (%)	2.50	2.36	2.11	2.05	1.89	1.70	1.60	1.53
Total credits (%)	0.16	0.16	0.15	0.14	0.16	0.17	0.15	0.15
Credit score	743	743	741	748	754	754	753	758
Loan-to-value ratio (%)	76.6	79.0	77.7	75.8	73.9	76.6	75.3	73.3
Borrower income ($000)	66.2	80.8	107	109	89.9	115	140	143.6
Loan amount ($000)	182	200	227	229	249	285	301	306
Debt-to-income ratio (%)	37.5	36.3	37.1	37.2	37.1	35.8	35.4	35.3
Loan term (months)	336	332	330	326	337	331	331	326
Percentage White (%)	67.9	68.2	71.9	75.4	72.9	69.5	75.7	77.6

This table presents the descriptive statistics for all loans in the HMDA-GSE matched data for low- and high-income tracts, for low- and high-education tracts, and by gender. The definitions of low- and high-income tracts and low- and high-education tracts are as in Tables 2 and 3 respectively. The variables are as described in Table A2. N = 4,092,384.

Table A4. Borrower Gender, Local Socioeconomic Status and Loan Cost

	Mortgage rate			Rate spread		
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.028***	0.043***	0.007***	0.021***	0.028***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female × SES	0.015***	0.010***	0.003***	0.011***	0.007***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
District × Orig. yr-m FE	Yes	Yes	Yes	Yes	Yes	Yes
Pricing controls	No	Yes	Yes	No	Yes	Yes
Additional controls	No	No	Yes	No	No	Yes
Observations	2,165,397	2,165,397	2,047,174	2,112,697	2,112,697	1,996,706
R-squared	0.546	0.677	0.770	0.167	0.473	0.550

	Net loan cost			Total fees	Total credits
	(7)	(8)	(9)	(10)	(11)
Female	0.179***	0.147***	-0.010***	-0.015***	-0.005***
	(0.004)	(0.003)	(0.002)	(0.002)	(0.001)
Female × SES	0.042***	0.029***	0.010***	0.008***	-0.003***
	(0.004)	(0.003)	(0.002)	(0.002)	(0.001)
District × Orig. yr-m FE	Yes	Yes	Yes	Yes	Yes
Pricing controls	No	Yes	Yes	Yes	Yes
Additional controls	No	No	Yes	Yes	Yes
Observations	2,165,397	2,165,397	2,047,174	2,047,174	2,047,174
R-squared	0.219	0.299	0.496	0.521	0.092

Standard errors, clustered at the school district level, are in parentheses. SES is the socioeconomic status composite variable from SEDA dataset, standardized to have a mean of zero and a standard deviation of one. All models include school district × origination year and month fixed effects. Pricing controls and additional controls are as in Table 4. ***p<0.01.

Table A5. Borrower Gender, Tract to MSA Median Income and Loan Cost

	Mortgage rate			Rate spread		
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.027***	0.041***	0.008***	0.016***	0.023***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female × Tract/MSA income	0.014***	0.011***	0.005***	0.012***	0.009***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Tract × Orig. year-month FE	Yes	Yes	Yes	Yes	Yes	Yes
Pricing controls	No	Yes	Yes	No	Yes	Yes
Additional controls	No	No	Yes	No	No	Yes
Observations	1,826,143	1,826,143	1,703,815	1,773,253	1,773,253	1,653,500
R-squared	0.674	0.763	0.831	0.404	0.615	0.661

	Net loan cost			Total fees	Total credits
	(7)	(8)	(9)	(10)	(11)
Female	0.164***	0.120***	-0.005***	-0.009***	-0.004***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Female × Tract/MSA income	0.028***	0.025***	0.003*	0.001	-0.002**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Tract × Orig. year-month FE	Yes	Yes	Yes	Yes	Yes
Pricing controls	No	Yes	Yes	Yes	Yes
Additional controls	No	No	Yes	Yes	Yes
Observations	1,826,143	1,826,143	1,703,815	1,703,815	1,703,815
R-squared	0.453	0.510	0.622	0.638	0.328

Standard errors, clustered at the Census tract level, are in parentheses. Tract to MSA median income is from HMDA, standardized to have a mean of zero and a standard deviation of one. Pricing controls, additional controls and fixed effects are as in Table 4. *p<0.10; **p<0.05; ***p<0.01.

Table A6. Borrower Gender, Census Tract Income and Loan Cost: White Borrowers

	Mortgage rate			Rate spread		
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.023***	0.040***	0.008***	0.010***	0.023***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female × Tract income	0.007***	0.004***	0.003***	0.008***	0.004***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Tract × Orig. yr-m FE	Yes	Yes	Yes	Yes	Yes	Yes
Pricing controls	No	Yes	Yes	No	Yes	Yes
Additional controls	No	No	Yes	No	No	Yes
Observations	1,166,489	1,166,489	1,083,948	1,133,611	1,133,611	1,052,658
R-squared	0.696	0.779	0.841	0.414	0.623	0.671

	Net loan cost			Total fees	Total credits
	(7)	(8)	(9)	(10)	(11)
Female	0.169***	0.125***	-0.014***	-0.017***	-0.003***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.001)
Female × Tract income	0.013***	0.006**	0.002	0.001	-0.001
	(0.003)	(0.003)	(0.002)	(0.002)	(0.001)
Tract × Orig. yr-m FE	Yes	Yes	Yes	Yes	Yes
Pricing controls	No	Yes	Yes	Yes	Yes
Additional controls	No	No	Yes	Yes	Yes
Observations	1,166,489	1,166,489	1,083,948	1,083,948	1,083,948
R-squared	0.465	0.521	0.637	0.653	0.357

Standard errors, clustered at the Census tract level, are in parentheses. Tract income is the median tract income in 2017, standardized to have a mean of zero and a standard deviation of one. Pricing controls, additional controls and fixed effects are as in Table 4. The sample contains only White borrowers. **p<0.05; ***p<0.01.

Table A7. Borrower Gender, Census Tract Education and Loan Cost: White Borrowers

	Mortgage rate			Rate spread		
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.023***	0.040***	0.007***	0.009***	0.022***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female × Tract education	0.007***	0.006***	0.004***	0.009***	0.007***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Tract × Orig. yr-m FE	Yes	Yes	Yes	Yes	Yes	Yes
Pricing controls	No	Yes	Yes	No	Yes	Yes
Additional controls	No	No	Yes	No	No	Yes
Observations	1,166,491	1,166,491	1,083,950	1,133,613	1,133,613	1,052,660
R-squared	0.696	0.779	0.841	0.414	0.623	0.671

	Net loan cost			Total fees	Total credits
	(7)	(8)	(9)	(10)	(11)
Female	0.169***	0.124***	-0.014***	-0.017***	-0.003***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.001)
Female × Tract education	0.009**	0.010***	0.003	0.001	-0.002**
	(0.003)	(0.003)	(0.002)	(0.002)	(0.001)
Tract × Orig. yr-m FE	Yes	Yes	Yes	Yes	Yes
Pricing controls	No	Yes	Yes	Yes	Yes
Additional controls	No	No	Yes	Yes	Yes
Observations	1,166,491	1,166,491	1,083,950	1,083,950	1,083,950
R-squared	0.465	0.521	0.637	0.653	0.357

Standard errors, clustered at the Census tract level, are in parentheses. Tract education is the share of the tract population over 25 years old with a bachelor's degree or higher in 2017, standardized to have a mean of zero and a standard deviation of one. Pricing controls, additional controls and fixed effects are as in Table 4. The sample contains only White borrowers. **p<0.05; ***p<0.01.

Table A8. Borrower Gender and Loan Cost by Age Group

	Mortgage rate		Rate spread		Net loan cost	
	(1)	(2)	(3)	(4)	(5)	(6)
			Age ≤ 34			(N=302,412)
Female	0.025***	0.007***	0.011***	0.002	0.069***	-0.003
	(0.002)	(0.001)	(0.002)	(0.002)	(0.005)	(0.005)
R-squared	0.814	0.857	0.694	0.717	0.559	0.631
			Age 35-44			(N=268,901)
Female	0.044***	0.018***	0.028***	0.018***	0.098***	-0.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.005)	(0.005)
R-squared	0.811	0.865	0.687	0.717	0.573	0.652
			Age 45-54			(N=202,088)
Female	0.039***	0.006***	0.017***	0.005**	0.111***	-0.010*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.006)	(0.006)
R-squared	0.799	0.861	0.677	0.717	0.586	0.678
			Age 55-64			(N=123,630)
Female	0.033***	-0.001	0.016***	-0.000	0.144***	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.009)	(0.008)
R-squared	0.803	0.864	0.676	0.719	0.598	0.702
			Age ≥ 65			(N=57,640)
Female	0.041***	0.006*	0.015***	0.003	0.141***	-0.009
	(0.004)	(0.004)	(0.004)	(0.004)	(0.014)	(0.013)
R-squared	0.810	0.857	0.651	0.706	0.597	0.706
Tract × Orig. yr-m FE	Yes	Yes	Yes	Yes	Yes	Yes
Pricing controls	Yes	Yes	Yes	Yes	Yes	Yes
Additional controls	No	Yes	No	Yes	No	Yes

Standard errors, clustered at the Census tract level, are in parentheses. Pricing controls, additional controls and fixed effects are as in Table 4. *p<0.10; **p<0.05; ***p<0.01.

Table A9. Borrower Gender, Census Tract Income and Loan Cost by Age Group

	Mortgage rate		Rate spread		Net loan cost	
	(1)	(2)	(3)	(4)	(5)	(6)
			Age $\leq$ 34			(N=302,410)
Female	0.025***	0.007***	0.011***	0.002	0.069***	-0.003
	(0.002)	(0.001)	(0.002)	(0.002)	(0.005)	(0.005)
Female $\times$ Tract income	0.006***	0.005***	0.009***	0.006***	0.026***	0.012**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.006)	(0.005)
R-squared	0.814	0.857	0.694	0.717	0.559	0.631
			Age 35-44			(N=268,901)
Female	0.039***	0.016***	0.025***	0.015***	0.092***	-0.007
	(0.002)	(0.002)	(0.002)	(0.002)	(0.006)	(0.006)
Female $\times$ Tract income	0.013***	0.007***	0.008***	0.007***	0.017***	0.018**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.005)	(0.005)
R-squared	0.811	0.865	0.687	0.717	0.573	0.652
			Age 45-54			(N=202,084)
Female	0.036***	0.005**	0.015***	0.003	0.107***	-0.012*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.007)	(0.007)
Female $\times$ Tract income	0.011***	0.005**	0.005**	0.004**	0.012*	0.005
	(0.002)	(0.002)	(0.002)	(0.002)	(0.006)	(0.006)
R-squared	0.800	0.861	0.677	0.717	0.586	0.678
			Age 55-64			(N=123,630)
Female	0.033***	-0.001	0.015***	-0.000	0.141***	-0.002
	(0.003)	(0.003)	(0.003)	(0.003)	(0.009)	(0.008)
Female $\times$ Tract income	0.007**	0.001	0.004	0.001	0.038***	0.020**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.010)	(0.009)
R-squared	0.803	0.864	0.676	0.719	0.598	0.702
			Age $\geq$ 65			(N=57,640)
Female	0.041***	0.006	0.015***	0.003	0.148***	-0.006
	(0.004)	(0.004)	(0.004)	(0.004)	(0.014)	(0.013)
Female $\times$ Tract income	0.003	-0.002	0.005	-0.001	0.058***	0.021
	(0.005)	(0.004)	(0.005)	(0.004)	(0.017)	(0.015)
R-squared	0.810	0.857	0.651	0.706	0.598	0.706
Tract $\times$ Orig. yr-m FE	Yes	Yes	Yes	Yes	Yes	Yes
Pricing controls	Yes	Yes	Yes	Yes	Yes	Yes
Additional controls	No	Yes	No	Yes	No	Yes

Standard errors, clustered at the Census tract level, are in parentheses. Pricing controls, additional controls and fixed effects are as in Table 4. *p<0.10; **p<0.05; ***p<0.01.

Table A10. Borrower Gender, Census Tract Education and Loan Cost by Age Group

	Mortgage rate		Rate spread		Net loan cost	
	(1)	(2)	(3)	(4)	(5)	(6)
			Age $\leq$ 34			(N=302,412)
Female	0.024***	0.006***	0.009***	0.001	0.065***	-0.005
	(0.002)	(0.001)	(0.002)	(0.002)	(0.005)	(0.005)
Female × Tract education	0.009***	0.006***	0.010***	0.008***	0.028***	0.015***
	(0.002)	(0.001)	(0.002)	(0.002)	(0.005)	(0.005)
R-squared	0.814	0.857	0.694	0.717	0.559	0.631
			Age 35-44			(N=268,901)
Female	0.040***	0.016***	0.025***	0.015***	0.092***	-0.005
	(0.002)	(0.002)	(0.002)	(0.002)	(0.006)	(0.005)
Female × Tract education	0.015***	0.007***	0.012***	0.008***	0.022***	0.016***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.006)	(0.005)
R-squared	0.812	0.865	0.687	0.717	0.573	0.652
			Age 45-54			(N=202,086)
Female	0.036***	0.005**	0.015***	0.003	0.105***	-0.012*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.007)	(0.006)
Female × Tract education	0.015***	0.006***	0.009***	0.006***	0.026***	0.009
	(0.002)	(0.002)	(0.002)	(0.002)	(0.007)	(0.006)
R-squared	0.800	0.861	0.677	0.717	0.586	0.678
			Age 55-64			(N=123,630)
Female	0.033***	-0.001	0.015***	-0.000	0.141***	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.009)	(0.008)
Female × Tract education	0.008***	0.001	0.007**	-0.001	0.050***	0.004
	(0.003)	(0.003)	(0.003)	(0.003)	(0.010)	(0.009)
R-squared	0.803	0.864	0.676	0.719	0.598	0.702
			Age $\geq$ 65			(N=57,640)
Female	0.041***	0.006*	0.015***	0.003	0.141***	-0.009
	(0.004)	(0.004)	(0.004)	(0.004)	(0.014)	(0.013)
Female × Tract education	0.006	-0.003	0.004	-0.002	0.059***	0.029**
	(0.005)	(0.004)	(0.005)	(0.004)	(0.017)	(0.015)
R-squared	0.810	0.857	0.651	0.706	0.598	0.706
Tract × Orig. yr-m FE	Yes	Yes	Yes	Yes	Yes	Yes
Pricing controls	Yes	Yes	Yes	Yes	Yes	Yes
Additional controls	No	Yes	No	Yes	No	Yes

Standard errors, clustered at the Census tract level, are in parentheses. Pricing controls, additional controls and fixed effects are as in Table 4. *p<0.10; **p<0.05; ***p<0.01.

Table A11. Borrower Gender, Census Tract Income and Loan Cost: Purchase Loans

	Mortgage rate			Rate spread		
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.017***	0.033***	0.006***	0.006***	0.023***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female × Tract income	0.012***	0.009***	0.005***	0.013***	0.009***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Tract × Orig. yr-m FE	Yes	Yes	Yes	Yes	Yes	Yes
Pricing controls	No	Yes	Yes	No	Yes	Yes
Additional controls	No	No	Yes	No	No	Yes
Observations	927,924	927,924	857,996	898,507	898,507	830,066
R-squared	0.700	0.781	0.833	0.442	0.675	0.701

	Net loan cost			Total fees	Total credits
	(7)	(8)	(9)	(10)	(11)
Female	0.129***	0.106***	-0.001	-0.001	-0.000
	(0.003)	(0.003)	(0.003)	(0.003)	(0.001)
Female × Tract income	0.025***	0.018***	0.007**	0.006**	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.001)
Tract × Orig. yr-m FE	Yes	Yes	Yes	Yes	Yes
Pricing controls	No	Yes	Yes	Yes	Yes
Additional controls	No	No	Yes	Yes	Yes
Observations	927,924	927,924	857,996	857,996	857,996
R-squared	0.485	0.526	0.630	0.647	0.375

Standard errors, clustered at the Census tract level, are in parentheses. Tract income is the median tract income in 2017, standardized to have a mean of zero and a standard deviation of one. Pricing controls, additional controls and fixed effects are as in Table 4. The sample contains only purchase loans (59.5% of all loans). **p<0.05; ***p<0.01.

Table A12. Borrower Gender, Census Tract Education and Loan Cost: Purchase Loans

	Mortgage rate			Rate spread		
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.015***	0.032***	0.006***	0.004***	0.022***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female × Tract education	0.014***	0.013***	0.006***	0.015***	0.011***	0.008***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Tract × Orig. yr-m FE	Yes	Yes	Yes	Yes	Yes	Yes
Pricing controls	No	Yes	Yes	No	Yes	Yes
Additional controls	No	No	Yes	No	No	Yes
Observations	927,928	927,928	858,000	898,511	898,511	830,070
R-squared	0.700	0.781	0.833	0.443	0.675	0.701

	Net loan cost			Total fees	Total credits
	(7)	(8)	(9)	(10)	(11)
Female	0.128***	0.104***	-0.002	-0.002	-0.000
	(0.003)	(0.003)	(0.003)	(0.003)	(0.001)
Female × Tract education	0.023***	0.022***	0.007**	0.005**	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.001)
Tract × Orig. yr-m FE	Yes	Yes	Yes	Yes	Yes
Pricing controls	No	Yes	Yes	Yes	Yes
Additional controls	No	No	Yes	Yes	Yes
Observations	927,928	927,928	858,000	858,000	858,000
R-squared	0.485	0.526	0.630	0.647	0.375

Standard errors, clustered at the Census tract level, are in parentheses. Tract education is the share of the tract population over 25 years old with a bachelor's degree or higher in 2017, standardized to have a mean of zero and a standard deviation of one. Pricing controls, additional controls and fixed effects are as in Table 4. The sample contains only purchase loans (59.5% of all loans). **p<0.05; ***p<0.01.

Table A13. Borrower Gender, Census Tract Income and Loan Cost: Refinance Loans

	Mortgage rate			Rate spread		
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.044***	0.052***	0.010***	0.030***	0.026***	0.007***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female × Tract income	0.012***	0.008***	0.005***	0.009***	0.005***	0.005***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Tract × Orig. yr-m FE	Yes	Yes	Yes	Yes	Yes	Yes
Pricing controls	No	Yes	Yes	No	Yes	Yes
Additional controls	No	No	Yes	No	No	Yes
Observations	554,451	554,451	515,060	536,899	536,899	498,648
R-squared	0.682	0.773	0.849	0.472	0.625	0.685

	Net loan cost			Total fees	Total credits
	(7)	(8)	(9)	(10)	(11)
Female	0.192***	0.133***	-0.010***	-0.018***	-0.008***
	(0.004)	(0.004)	(0.004)	(0.003)	(0.001)
Female × Tract income	0.012***	0.006	0.006*	0.003	-0.004**
	(0.004)	(0.004)	(0.003)	(0.003)	(0.002)
Tract × Orig. yr-m FE	Yes	Yes	Yes	Yes	Yes
Pricing controls	No	Yes	Yes	Yes	Yes
Additional controls	No	No	Yes	Yes	Yes
Observations	554,451	554,451	515,060	515,060	515,060
R-squared	0.547	0.598	0.690	0.700	0.395

Standard errors, clustered at the Census tract level, are in parentheses. Tract income is the median tract income in 2017, standardized to have a mean of zero and a standard deviation of one. Pricing controls, additional controls and fixed effects are as in Table 4. The sample contains only refinance loans (40.5% of all loans). *p<0.10; **p<0.05; ***p<0.01.

Table A14. Borrower Gender, Census Tract Education and Loan Cost: Refinance Loans

	Mortgage rate			Rate spread		
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.045***	0.053***	0.010***	0.031***	0.026***	0.008***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female × Tract education	0.014***	0.011***	0.005***	0.012***	0.009***	0.006***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Tract × Orig. yr-m FE	Yes	Yes	Yes	Yes	Yes	Yes
Pricing controls	No	Yes	Yes	No	Yes	Yes
Additional controls	No	No	Yes	No	No	Yes
Observations	554,451	554,451	515,060	536,899	536,899	498,648
R-squared	0.682	0.773	0.849	0.472	0.625	0.685

	Net loan cost			Total fees	Total credits
	(7)	(8)	(9)	(10)	(11)
Female	0.192***	0.131***	-0.010***	-0.018***	-0.008***
	(0.004)	(0.004)	(0.004)	(0.003)	(0.001)
Female × Tract education	0.027***	0.025***	0.012***	0.006**	-0.006***
	(0.004)	(0.004)	(0.004)	(0.003)	(0.001)
Tract × Orig. yr-m FE	Yes	Yes	Yes	Yes	Yes
Pricing controls	No	Yes	Yes	Yes	Yes
Additional controls	No	No	Yes	Yes	Yes
Observations	554,451	554,451	515,060	515,060	515,060
R-squared	0.547	0.598	0.690	0.700	0.395

Standard errors, clustered at the Census tract level, are in parentheses. Tract education is the share of the tract population over 25 years old with a bachelor's degree or higher in 2017, standardized to have a mean of zero and a standard deviation of one. Pricing controls, additional controls and fixed effects are as in Table 4. The sample contains only refinance loans (40.5% of all loans). **p<0.05; ***p<0.01.

Table A15. Borrower Gender and Loan Rates, with Lender Fixed Effects

	Mortgage rate				
	(1)	(2)	(3)	(4)	(5)
Female	0.007***	0.007***	0.006***	0.006***	0.007***
	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)
Female × Tract income		0.004***			
		(0.001)			
Female × Tract education			0.005***		
			(0.001)		
Female × SES				0.003***	
				(0.001)	
Female × Tract/MSA income					0.004***
					(0.001)
Observations	1,703,654	1,703,640	1,703,648	2,047,044	1,703,654
R-squared	0.847	0.847	0.847	0.792	0.847

	Rate spread				
	(6)	(7)	(8)	(9)	(10)
Female	0.005***	0.004***	0.004***	0.003***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female × Tract income		0.005***			
		(0.001)			
Female × Tract education			0.006***		
			(0.001)		
Female × SES				0.003***	
				(0.001)	
Female × Tract/MSA income					0.005***
					(0.001)
Observations	1,653,343	1,653,329	1,653,337	1,996,582	1,653,343
R-squared	0.705	0.705	0.705	0.610	0.705

Standard errors in parentheses are clustered at the Census tract level, except in Models (4) and (9), in which they are clustered at the school district level. This table replicates the analysis from Tables 4 to 6, A4 and A5 for mortgage rates and rate spreads (Models (3) and (6) in each table, including tract or school district × origination year-month fixed effects, pricing controls and additional controls), with the addition of lender fixed effects. ***p<0.01.

Table A16. Borrower Gender and Loan Cost, with Lender Fixed Effects

	Net loan cost				
	(1)	(2)	(3)	(4)	(5)
Female	-0.002	-0.003*	-0.003*	-0.003	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Female × Tract income		0.008***			
		(0.002)			
Female × Tract education			0.010***		
			(0.002)		
Female × SES				0.008***	
				(0.002)	
Female × Tract/MSA income					0.005***
					(0.002)
Observations	1,703,654	1,703,640	1,703,648	1,665,880	1,703,654
R-squared	0.695	0.695	0.695	0.695	0.695

Standard errors in parentheses are clustered at the Census tract level, except in Model (4), in which they are clustered at the school district level. This table replicates the analysis from Tables 4 to 6, A4 and A5 for the net loan cost (Model (9) in each table, including tract or school district × origination year-month fixed effects, pricing controls and additional controls), with the addition of lender fixed effects. *p<0.10; ***p<0.01.

Table A17. Mortgage Prepayment, Borrower Gender and Local Factors

	Prepayment			
	(1)	(2)	(3)	(4)
Female	-0.225***	-0.218***	-0.043***	-0.045***
	(0.017)	(0.016)	(0.012)	(0.011)
Female × SES	-0.098***		-0.056***	
	(0.013)		(0.010)	
Female × Tract/MSA income		-0.079***		-0.034***
		(0.012)		(0.012)
Origination year-month FE	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes
Pricing controls	No	No	Yes	Yes
Additional controls	No	No	Yes	Yes
Observations	10,197,214	10,450,011	10,190,820	10,443,544
R-squared	0.0027	0.0029	0.0051	0.0051

Standard errors, clustered at the county and origination year-month levels, are in parentheses. Coefficients and standard errors are multiplied by 100. Controls and fixed effects are as in Table 8. In addition to the interaction terms, all models include the linear terms. ***p<0.01.

Table A18. Mortgage Prepayment, Rate Decline and Gender

	Prepayment			
	(1)	(2)	(3)	(4)
Female × Rate drop	-0.223***	-0.216***	-0.181***	-0.175***
	(0.020)	(0.021)	(0.021)	(0.022)
Female × SES × Rate drop	-0.150***		-0.135***	
	(0.020)		(0.019)	
Female × Tract/MSA inc. × Rate drop		-0.131***		-0.118***
		(0.020)		(0.019)
Origination year-month FE	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes
Pricing controls	No	No	Yes	Yes
Additional controls	No	No	Yes	Yes
Observations	10,197,214	10,450,011	10,190,820	10,443,544
R-squared	0.0062	0.0064	0.0104	0.0104

Standard errors, clustered at the county and origination year-month level, are in parentheses. Coefficients and standard errors are multiplied by 100. Controls and fixed effects are as in Table 8. In addition to the triple interactions all models include the linear terms and the pairwise interactions of the variables in the triple interaction. ***p<0.01.

Table A19. Mortgage Prepayment, Borrower Gender and Tract Income (Full Specification)

	Prepayment			
	(1)	(2)	(3)	(4)
Female	-0.0021***	-0.0005***	-0.0004***	0.0007***
	(0.0002)	(0.0001)	(0.0001)	(0.0001)
Tract income	0.0039***	-0.0006***	0.0010***	-0.0028***
	(0.0002)	(0.0002)	(0.0001)	(0.0002)
Female × Tract income	-0.0010***	0.0001	-0.0006***	0.0003*
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Rate drop		0.0195***		0.0260***
		(0.0008)		(0.0016)
Female × Rate drop		-0.0020***		-0.0016***
		(0.0002)		(0.0002)
Tract income × Rate drop		0.0059***		0.0053***
		(0.0003)		(0.0003)
Female × Tract income × Rate drop		-0.0014***		-0.0013***
		(0.0002)		(0.0002)
Constant	0.0217***	0.0078***	0.0163***	-0.0088***
	(0.00005)	(0.0006)	(0.0009)	(0.0022)
Pricing controls	No	No	Yes	Yes
Additional controls	No	No	Yes	Yes
Origination year-month FE	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes
Observations	10,449,605	10,449,605	10,443,138	10,443,138
R-squared	0.0029	0.0065	0.0051	0.0105

Standard errors, clustered at the county and origination year-month level, are in parentheses. All models include origination year-month fixed effects and county fixed effects. The pricing controls include dummies for nine loan-to-value categories, dummies for nine credit score categories, dummies for manufactured homes, investment properties, high balance mortgages, borrowers who have a subordinate mortgage in addition to the first lien, loan purpose (home purchase, refinancing or cash-out refinancing) and for the number of units (1 to 4). The additional controls include property value, loan age, dummies for eight borrower race and ethnicity categories, income deciles, loan amount deciles, debt-to-income deciles, dummies for three loan term at origination ranges and dummies for seven borrower age groups. *p<0.10; ***p<0.01.

Table A20. Mortgage Prepayment, Borrower Gender and Tract Education (Full Specification)

	Prepayment			
	(1)	(2)	(3)	(4)
Female	-0.0023***	-0.0004***	-0.0005***	0.0010***
	(0.0002)	(0.0001)	(0.0001)	(0.0001)
Tract education	0.0039***	0.000003	0.0009***	-0.0023***
	(0.0002)	(0.0002)	(0.0001)	(0.0002)
Female × Tract education	-0.0009***	0.0001	-0.0005***	0.0004**
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Rate drop		0.0197***		0.0262***
		(0.0008)		(0.0016)
Female × Rate drop		-0.0025***		-0.0020***
		(0.0002)		(0.0002)
Tract education × Rate drop		0.0051***		0.0046***
		(0.0003)		(0.0002)
Female × Tract education × Rate drop		-0.0014***		-0.0013***
		(0.0002)		(0.0002)
Constant	0.0218***	0.0078***	0.0163***	-0.0089***
	(0.0001)	(0.0006)	(0.0009)	(0.0022)
Pricing controls	No	No	Yes	Yes
Additional controls	No	No	Yes	Yes
Origination year-month FE	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes
Observations	10,449,889	10,449,889	10,443,422	10,443,422
R-squared	0.0029	0.0065	0.0051	0.0104

Standard errors, clustered at the county and origination year-month level, are in parentheses. All models include origination year-month fixed effects and county fixed effects. The pricing controls include dummies for nine loan-to-value categories, dummies for nine credit score categories, dummies for manufactured homes, investment properties, high balance mortgages, borrowers who have a subordinate mortgage in addition to the first lien, loan purpose (home purchase, refinancing or cash-out refinancing) and for the number of units (1 to 4). The additional controls include property value, loan age, dummies for eight borrower race and ethnicity categories, income deciles, loan amount deciles, debt-to-income deciles, dummies for three loan term at origination ranges and dummies for seven borrower age groups. $^{**}p<0.05$; $^{***}p<0.01$.

Table A21. Mortgage Prepayment, Rate Decline and Gender

	Prepayment				
	(1)	(2)	(3)	(4)	(5)
Female × Rate drop V2	-0.220***	-0.195***	-0.239***	-0.215***	-0.209***
	(0.023)	(0.020)	(0.021)	(0.020)	(0.022)
Female × Tract income × Rate drop V2		-0.129***			
		(0.020)			
Female × Tract education × Rate drop V2			-0.128***		
			(0.021)		
Female × SES × Rate drop V2				-0.132***	
				(0.020)	
Female × Tract/MSA income × Rate drop V2					-0.121***
					(0.020)
Origination year-month FE	Yes	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	10,450,011	10,449,605	10,449,889	10,197,214	10,450,011
R-squared	0.0052	0.0065	0.0058	0.0056	0.0057

Standard errors, clustered at the county and origination year-month level, are in parentheses. Coefficients and standard errors are multiplied by 100. All models include origination year and month fixed effects and county fixed effects. ***p<0.01.

Table A22. Mortgage Prepayment, Rate Decline and Gender, with Additional Controls

	Prepayment				
	(1)	(2)	(3)	(4)	(5)
Female × Rate drop V2	-0.160***	-0.140***	-0.172***	-0.155***	-0.150***
	(0.023)	(0.020)	(0.021)	(0.021)	(0.021)
Female × Tract income × Rate drop V2		-0.096***			
		(0.020)			
Female × Tract education × Rate drop V2			-0.097***		
			(0.019)		
Female × SES × Rate drop V2				-0.099***	
				(0.019)	
Female × Tract/MSA income × Rate drop V2					-0.088***
					(0.020)
Pricing controls	Yes	Yes	Yes	Yes	Yes
Additional controls	Yes	Yes	Yes	Yes	Yes
Origination year-month FE	Yes	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	10,443,544	10,443,138	10,443,422	10,190,820	10,443,544
R-squared	0.0103	0.0104	0.0104	0.0104	0.0103

Standard errors, clustered at the county and origination year-month level, are in parentheses. Coefficients and standard errors are multiplied by 100. Controls and fixed effects are as in Table 8. ***p<0.01.

Table A23. Overall Mortgage Cost, Borrower Gender and Local Factors

	Overall Cost					
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.805***	0.842***	0.014	0.748***	0.743***	0.038***
	(0.022)	(0.019)	(0.010)	(0.014)	(0.013)	(0.012)
Female × SES	0.296***	0.209***	0.092***			
	(0.020)	(0.017)	(0.012)			
Female × Tract/MSA income				0.240***	0.202***	0.087***
				(0.014)	(0.013)	(0.012)
Location × Orig. yr-m FE	Yes	Yes	Yes	Yes	Yes	Yes
Pricing controls	No	Yes	Yes	No	Yes	Yes
Additional controls	No	No	Yes	No	No	Yes
Observations	2,165,397	2,165,397	2,047,174	1,826,143	1,826,143	1,703,815
R-squared	0.219	0.346	0.453	0.440	0.522	0.587

Standard errors in parentheses, clustered at the schools district level in models (1) to (3) and at the Census tract level in models (4) to (6). Overall cost is expressed as a percentage of the loan amount and is calculated for 10 years after origination. The overall cost includes origination cost, interest, and additional cost if the loan gets refinanced within this interval. Models (1) to (3) include school district × origination year-month fixed effects. Models (4) to (6) include Census tract × origination year-month fixed effects. Pricing controls and additional controls are as in Table 4. ***p<0.01.

Table A24. Mortgage Application Outcomes

	Number of applications originated	Percentage of originated applications (%)					
		File closed for incompleteness	Application withdrawn by applicant	Application denied by lender	Application approved but not accepted	Preapproval request denied	Preapproval request approved but not accepted
Female	1,894,578	6.52	23.38	21.63	4.21	0.07	0.11
Male	2,796,214	7.33	24.57	20.39	4.35	0.06	0.10
Low income Census tracts							
Female	948,010	7.17	23.70	26.84	4.73	0.10	0.13
Male	1,306,195	7.96	24.60	25.22	4.86	0.09	0.11
High income Census tracts							
Female	946,568	5.87	23.07	16.41	3.70	0.03	0.09
Male	1,490,019	6.79	24.55	16.15	3.91	0.03	0.09
Low education Census tracts							
Female	906,238	7.65	23.99	28.37	4.86	0.10	0.12
Male	1,326,137	8.26	24.73	25.78	4.91	0.08	0.10
High education Census tracts							
Female	988,340	5.48	22.82	15.46	3.62	0.04	0.09
Male	1,470,077	6.49	24.43	15.53	3.85	0.03	0.09

This table presents the number of applications that resulted in mortgage originations and the different outcomes of mortgage applications, as a percentage of originated mortgages, by applicant gender and by Census tract income and education. The sample contains conventional mortgage applications by single applicants in 2018 and 2019 from HMDA. Low-income Census tracts are defined as those with tract income in 2017 below the median, while high-income Census tracts are defined as those with tract income above the median. Low-education Census tracts are defined as those where the share of the tract population over 25 years old with a bachelor's degree or higher is below the median, while high-education Census tracts are defined as those where this share is above the median.

Table A25. Familiarity with Mortgage Concepts and Income

	Familiarity with specific concepts at the beginning of the search						
	Interest rates	Mortgage types	Mortgage process	Down payment	Income needed	Credit history	Money at closing
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	-0.0720***	-0.0638***	-0.0269*	-0.0200	-0.0292*	-0.0056	-0.0122
	(0.0147)	(0.0149)	(0.0149)	(0.0149)	(0.0151)	(0.0130)	(0.0152)
Income 50-75k	0.0606***	0.0624***	0.0597***	0.0724***	0.0603***	0.0286**	0.0549***
	(0.0144)	(0.0146)	(0.0146)	(0.0146)	(0.0148)	(0.0127)	(0.0149)
Income 75-100k	0.1113***	0.1087***	0.1159***	0.1273***	0.1175***	0.0598***	0.0895***
	(0.0148)	(0.0150)	(0.0149)	(0.0150)	(0.0151)	(0.0130)	(0.0152)
Income 100-175k	0.1586***	0.1791***	0.1910***	0.1896***	0.1733***	0.0913***	0.1427***
	(0.0139)	(0.0141)	(0.0141)	(0.0141)	(0.0142)	(0.0123)	(0.0144)
Income >175k	0.2455***	0.3062***	0.2956***	0.2833***	0.2576***	0.1188***	0.2222***
	(0.0157)	(0.0159)	(0.0158)	(0.0159)	(0.0160)	(0.0138)	(0.0162)
Female × Income 50-75k	0.0041	-0.0027	0.0092	-0.0318	0.0004	0.0296*	-0.0302
	(0.0203)	(0.0207)	(0.0206)	(0.0207)	(0.0209)	(0.0180)	(0.0210)
Female × Income 75-100k	-0.0413*	-0.0180	0.0065	-0.0563***	-0.0305	0.0086	-0.0557**
	(0.0213)	(0.0216)	(0.0215)	(0.0216)	(0.0218)	(0.0188)	(0.0220)
Female × Income 100-175k	-0.0156	-0.0307	-0.0299	-0.0418**	-0.0081	0.0107	-0.0488**
	(0.0198)	(0.0202)	(0.0201)	(0.0202)	(0.0203)	(0.0175)	(0.0205)
Female × Income >175k	-0.0554**	-0.0766***	-0.0575**	-0.0523**	-0.0451*	-0.0125	-0.0158
	(0.0240)	(0.0244)	(0.0243)	(0.0244)	(0.0246)	(0.0212)	(0.0248)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	21,177	21,177	21,177	21,177	21,177	21,177	21,177
R-squared	0.1101	0.0852	0.1016	0.0817	0.0702	0.0555	0.0653

This table presents the coefficient estimates obtained from regressions of self-reported familiarity with mortgage concepts on female indicator, indicators for income groups, and their interactions. Regressions include survey wave fixed effects, dummy variable controls for four education levels, five race and ethnicity categories, six age groups and nine credit score categories, and use survey weights. Source: National Survey of Mortgage Originations for conventional loans with a single borrower originated between 2014 and 2021Q2 (30 waves). *p<0.10; **p<0.05; ***p<0.01.

Table A26. Familiarity with Mortgage Concepts and Education

	Familiarity with specific concepts at the beginning of the search						
	Interest rates	Mortgage types	Mortgage process	Down payment	Income needed	Credit history	Money at closing
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	-0.0416**	-0.0308	-0.0323	-0.0202	-0.0653***	-0.0126	-0.0595***
	(0.0200)	(0.0203)	(0.0202)	(0.0203)	(0.0205)	(0.0176)	(0.0206)
Some college	0.0476***	0.0514***	0.0229	0.0366**	-0.0152	0.0376***	0.0131
	(0.0141)	(0.0143)	(0.0143)	(0.0143)	(0.0144)	(0.0124)	(0.0146)
College graduate	0.0932***	0.0912***	0.0568***	0.0900***	0.0187	0.0414***	0.0080
	(0.0138)	(0.0140)	(0.0140)	(0.0140)	(0.0142)	(0.0122)	(0.0143)
Postgraduate	0.0860***	0.0838***	0.0388**	0.0815***	-0.0172	0.0350***	-0.0169
	(0.0149)	(0.0151)	(0.0151)	(0.0151)	(0.0153)	(0.0132)	(0.0154)
Female × Some college	-0.0156	-0.0301	0.0342	0.0022	0.0600**	0.0262	0.0385
	(0.0240)	(0.0243)	(0.0242)	(0.0243)	(0.0246)	(0.0212)	(0.0247)
Female × College graduate	-0.0612***	-0.0624***	-0.0081	-0.0430*	0.0163	0.0160	0.0144
	(0.0227)	(0.0231)	(0.0230)	(0.0231)	(0.0233)	(0.0201)	(0.0235)
Female × Postgraduate	-0.0789***	-0.0873***	-0.0413*	-0.0695***	0.0071	0.0147	0.0044
	(0.0236)	(0.0240)	(0.0239)	(0.0240)	(0.0242)	(0.0208)	(0.0244)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	21,177	21,177	21,177	21,177	21,177	21,177	21,177
R-squared	0.1104	0.0854	0.1018	0.0821	0.0704	0.0553	0.0651

This table presents the coefficient estimates obtained from regressions of self-reported familiarity with mortgage concepts on female indicator, indicators for income groups, and their interactions. Regressions include survey wave fixed effects, dummy variable controls for five income groups, five race and ethnicity categories, six age groups and nine credit score categories, and use survey weights. Source: National Survey of Mortgage Originations for conventional loans with a single borrower originated between 2014 and 2021Q2 (30 waves). *p<0.10; **p<0.05; ***p<0.01.

Table A27. Borrower Gender, Student Math Gap and Loan Cost

	Mortgage rate			Rate spread		
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.028***	0.043***	0.007***	0.021***	0.028***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female × Student math score gap	0.009***	0.007***	0.002***	0.005***	0.005***	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
District × Orig. year-month FE	Yes	Yes	Yes	Yes	Yes	Yes
Pricing controls	No	Yes	Yes	No	Yes	Yes
Additional controls	No	No	Yes	No	No	Yes
Observations	2,159,375	2,159,375	2,041,530	2,106,796	2,106,796	1,991,173
R-squared	0.545	0.677	0.770	0.167	0.473	0.550

	Net loan cost			Total fees	Total credits
	(7)	(8)	(9)	(10)	(11)
Female	0.180***	0.147***	-0.010***	-0.015***	-0.005***
	(0.004)	(0.003)	(0.002)	(0.002)	(0.001)
Female × Student math score gap	0.030***	0.022***	0.001	-0.001	-0.003***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.001)
District × Orig. year-month FE	Yes	Yes	Yes	Yes	Yes
Pricing controls	No	Yes	Yes	Yes	Yes
Additional controls	No	No	Yes	Yes	Yes
Observations	2,159,375	2,159,375	2,041,530	2,041,530	2,041,530
R-squared	0.216	0.297	0.495	0.520	0.091

Standard errors, clustered at the school district level, are in parentheses. Student math score gap is the difference between the average male score and the average female score in standardized math tests from grades 3 to 8 in each school district, from SEDA dataset, standardized to have a mean of zero and a standard deviation of one. All models include school district × origination year and month fixed effects. Pricing controls and additional controls are as in Table 4. ***p<0.01.

Table A28. Borrower Gender, Student Math and Reading Gap and Loan Cost

	Mortgage rate			Rate spread		
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.028***	0.043***	0.007***	0.021***	0.028***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female × Student math score gap	0.008***	0.007***	0.001**	0.003**	0.004***	0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female × Student reading score gap	0.001	0.001	0.002***	0.003**	0.003**	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
District × Orig. yr-m FE	Yes	Yes	Yes	Yes	Yes	Yes
Pricing controls	No	Yes	Yes	No	Yes	Yes
Additional controls	No	No	Yes	No	No	Yes
Observations	2,159,303	2,159,303	2,041,461	2,106,725	2,106,725	1,991,105
R-squared	0.545	0.677	0.770	0.167	0.473	0.550

	Net loan cost			Total fees	Total credits
	(7)	(8)	(9)	(10)	(11)
Female	0.180***	0.147***	-0.010***	-0.015***	-0.005***
	(0.004)	(0.003)	(0.002)	(0.002)	(0.001)
Female × Student math score gap	0.030***	0.019***	0.004	0.001	-0.003***
	(0.004)	(0.004)	(0.002)	(0.002)	(0.001)
Female × Student reading score gap	-0.001	0.005	-0.005**	-0.005**	0.000
	(0.004)	(0.004)	(0.002)	(0.002)	(0.001)
District × Orig. yr-m FE	Yes	Yes	Yes	Yes	Yes
Pricing controls	No	Yes	Yes	Yes	Yes
Additional controls	No	No	Yes	Yes	Yes
Observations	2,159,303	2,159,303	2,041,461	2,041,461	2,041,461
R-squared	0.216	0.297	0.495	0.520	0.091

Standard errors, clustered at the school district level, are in parentheses. Math and reading gaps are the differences between the average male score and the average female score in standardized math and reading tests from grades 3 to 8 in each school district, from SEDA dataset, standardized to have a mean of zero and a standard deviation of one. All models include school district × origination year and month fixed effects. Pricing controls and additional controls are as in Table 4. **p<0.05; ***p<0.01.

Table A29. Borrower Gender, Occupational Math Skills and Loan Cost

	Mortgage rate			Rate spread		
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.028***	0.041***	0.008***	0.016***	0.024***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female × Occupation math skill gap	0.009***	0.007***	0.003***	0.009***	0.006***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Tract × Origination year-month FE	Yes	Yes	Yes	Yes	Yes	Yes
Pricing controls	No	Yes	Yes	No	Yes	Yes
Additional controls	No	No	Yes	No	No	Yes
Observations	1,826,137	1,826,137	1,703,809	1,773,247	1,773,247	1,653,494
R-squared	0.674	0.763	0.831	0.404	0.615	0.661

	Net loan cost			Total fees	Total credits
	(7)	(8)	(9)	(10)	(11)
Female	0.165***	0.121***	-0.006***	-0.009***	-0.004***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Female × Occupation math skill gap	0.021***	0.018***	0.007***	0.004**	-0.003***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Tract × Origination year-month FE	Yes	Yes	Yes	Yes	Yes
Pricing controls	No	Yes	Yes	Yes	Yes
Additional controls	No	No	Yes	Yes	Yes
Observations	1,826,137	1,826,137	1,703,809	1,703,809	1,703,809
R-squared	0.453	0.510	0.622	0.638	0.328

Standard errors, clustered at the Census tract level, are in parentheses. Occupation math skill gap is the difference between the average level of math skill importance in the occupations held by males and females within each Census tract. This average math skill level is calculated by weighting each occupation's math skill importance, from the O*NET dataset, by the 2017 employment figures for males and females in major occupation groups within each tract. In the regressions, the math skill gap is standardized to have a mean of zero and a standard deviation of one. All models include Census tract × origination year and month fixed effects. Pricing controls and additional controls are as in Table 4. **p<0.05; ***p<0.01.

Table A30. Mortgage Prepayment, Borrower Gender and Math Skills

	Prepayment			
	(1)	(2)	(3)	(4)
Female	-0.228***	-0.228***	-0.043***	-0.043***
	(0.017)	(0.016)	(0.012)	(0.011)
Female × Student math score gap	-0.054***		-0.030**	
	(0.013)		(0.012)	
Female × Occupation math skill gap		-0.067***		-0.040***
		(0.012)		(0.010)
Origination year-month FE	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes
Pricing controls	No	No	Yes	Yes
Additional controls	No	No	Yes	Yes
Observations	10,141,903	10,449,889	10,135,533	10,443,422
R-squared	0.0026	0.0026	0.0051	0.0051

Standard errors, clustered at the county and origination year-month levels, are in parentheses. Coefficients and standard errors are multiplied by 100. All models include origination year and month fixed effects as well as county fixed effects. Models (3) and (4) also include pricing and additional controls as in Table 8. In addition to the interaction terms, all models include the linear terms. **$p<0.05$; ***$p<0.01$.

Table A31. Mortgage Prepayment, Rate Decline, Gender and Math Skills

	Prepayment			
	(1)	(2)	(3)	(4)
Female × Rate drop	-0.226***	-0.231***	-0.182***	-0.188***
	(0.022)	(0.022)	(0.023)	(0.022)
Female × Math score gap × Rate drop	-0.078***		-0.068***	
	(0.018)		(0.016)	
Female × Math skill gap × Rate drop		-0.111***		-0.102***
		(0.018)		(0.017)
Origination year-month FE	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes
Pricing controls	No	No	Yes	Yes
Additional controls	No	No	Yes	Yes
Observations	10,141,903	10,449,889	10,135,533	10,443,422
R-squared	0.0060	0.0060	0.0103	0.0103

Standard errors, clustered at the county and origination year-month level, are in parentheses. Coefficients and standard errors are multiplied by 100. All models include origination year and month fixed effects as well as county fixed effects. Models (3) and (4) also include pricing and additional controls as in Table 8. In addition to the triple interactions all models include the linear terms and the pairwise interactions of the variables in the triple interaction. ***p<0.01.