

Mortgage Prepayment, Race, and Monetary Policy

Kristopher Gerardi, Paul Willen, and David Hao Zhang *

January 26, 2021

Abstract

Over the period 2005 to 2020, Black borrowers with mortgages insured by Fannie Mae or Freddie Mac paid interest rates that were almost 50 basis points higher than those paid by Non-Hispanic white borrowers. We show that the main reason is that Non-Hispanic white borrowers are much more likely to exploit periods of falling interest rates by refinancing their mortgages or moving. Black and Hispanic white borrowers face challenges refinancing because, on average, they have lower credit scores, equity and income. But even holding those factors constant, Black and Hispanic white borrowers refinance less suggesting that other social factors are at play. Because they are more likely to exploit lower interest rates, white borrowers benefit more from monetary expansions. Policies that reduce barriers to refinancing for minority borrowers and alternative mortgage contract designs that more directly pass through interest rate declines to borrowers can reduce racial mortgage pricing inequality.

*We thank Manuel Adelino, David Berger, Neil Bhutta, Scott Frame, Andreas Fuster, Ed Glaeser, Lauren Lambie-Hanson, Joe Peek, Geoff Tootell, Joe Tracy, Larry Wall, Christina Wang, Jon Willis, and attendees of the 2020 Atlanta Fed/Princeton Bendheim Conference on Racial Justice and Finance for helpful comments. We especially thank Daniel Sexton for excellent research assistance. Kristopher Gerardi, kristopher.gerardi@atl.frb.org, is at the Federal Reserve Bank of Atlanta, 1000 Peachtree St., Atlanta GA. Paul Willen, paul.willen@atl.frb.org, is at the Federal Reserve Bank of Boston, 600 Atlantic Avenue, Boston MA. David Zhang, dzhang@hbs.edu, is at Harvard Business School, Soldiers Field Rd, Boston MA. The opinions expressed herein are those of the authors and do not represent the official positions of Black Knight Inc., Equifax, the Federal Reserve Bank of Atlanta, Federal Reserve Bank of Boston, or the Federal Reserve System. All remaining errors are our own.

1 Introduction

At the end of 2012, Black borrowers with mortgages insured by Fannie Mae or Freddie Mac (GSEs) paid interest rates that were approximately 60 basis points higher than those paid by Non-Hispanic white borrowers. This difference was not a new phenomenon, although the gap has waxed and waned over time, as depicted in Figure 1. What accounts for this gap? One explanation could be that loans to Black borrowers are riskier and lenders charge higher rates on riskier loans. Another explanation is that racially biased loan officers charge higher rates to Black borrowers. We show in this paper that neither of these explanations directly explains most of the gap. If we make the counterfactual assumption that all borrowers, regardless of race or perceived risk, receive the Freddie Mac Primary Mortgage Market Survey (PMMS) rate prevailing that quarter, we find that the gap shrinks by only about 15 percent. In other words, even if lenders completely ignored risk and race when they priced new loans, Black homeowners in 2012 would still have paid over 30 basis points more than their non-Hispanic white counterparts.

In this paper, we document large differences in prepayment behavior across racial groups and show that they generate the rate disparities discussed above. The quarterly hazard of prepayment due to refinance for a Black borrower with a loan from the GSEs is 0.75 percentage points lower than it is for a non-Hispanic white GSE borrower, which corresponds to approximately 44 percent of the average quarterly refinance probability for all borrowers with GSE loans in our sample (1.71 percentage points). For prepayments due to sale, the Black-white gap is -0.524 percentage points, which corresponds to approximately 55 percent of the average quarterly sale probability (0.96 percentage points). Given the trend decline in mortgage rates over the last 40 years, differences in prepayment speeds alone would lead to lower rates for non-Hispanic white borrowers. However, the problem is compounded by the fact that non-Hispanic white borrowers appear to respond much more strongly to fluctuations in interest rates. In 2006 and 2007, when the PMMS 30-year FRM rate averaged over 6 percent, which is higher than it had been since 2001, Black and non-Hispanic white borrowers refinanced at roughly the same rate. In 2009 and 2010, when the PMMS 30-year FRM rate fell to historic lows of under 5 percent, non-Hispanic white borrowers were almost twice as likely to refinance as Black borrowers.

What explains these differences in prepayment behavior across racial groups? Our rich data provide answers. We use the Credit Risk Insights Servicing McDash-Home Mortgage Disclosure Act (CRISM-HMDA) data set, a three-way match between administrative mortgage data from McDash, Home Mortgage Disclosure Act (HMDA) data collected by the Federal Reserve, and credit bureau data from Equifax. In contrast to data used in previous

work in this area, the CRISM-HMDA data set allows us to distinguish between mortgage prepayments due to sales and refinances, provides up-to-date information on borrower creditworthiness, and is nationally representative. We find that observable differences between Black and non-Hispanic white borrowers account for approximately 80 percent of the difference in refinance rates. The typical Black borrower has a lower credit score, lower income, and higher leverage. A Black borrower is also more likely to be female and less likely to have a co-borrower. All of those factors lead to lower refinance propensities, regardless of race. However, a small gap remains even after controlling for these factors in addition to extremely fine geographic-by-time fixed effects. Suppose we take two borrowers living in the same Zip code with the same credit score, income, and gender and we observe them in the same year and quarter. If one borrower is Black and the other is non-Hispanic white, we show that the Black borrower is 0.15 percentage points less likely to refinance.

Further insights come from looking at responses to refinance incentives through the course of a loan. Refinance opportunities emerge for two reasons: macroeconomic and idiosyncratic. The main macro reason to refinance is lower interest rates, which increase the incentive for borrowers to exercise the prepayment option. Idiosyncratic reasons stem from individual increases in creditworthiness such as a reduction in leverage from higher house prices or an increased credit score resulting from higher income and employment security. We show that in our sample of GSE mortgages, minority and non-Hispanic white borrowers respond similarly to idiosyncratic shocks. An 100 point increase in credit score leads to a 0.7 percent increase in the refinance probability, and the number is not significantly different across races. Therefore, we find that the refinance gap is due to non-Hispanic white borrowers responding much more strongly to macroeconomic shocks compared with minority borrowers. Yet, macroeconomic changes in interest rates are precisely the channel through which the interest rate reduction driven by monetary policy gets passed through to mortgage borrowers, which suggests that there is large heterogeneity by race in the mortgage refinancing channel of monetary policy.

The implications for monetary policy here are significant. Expansionary monetary policy by definition leads to lower interest rates and so, given the evidence we have presented, disproportionately benefits non-Hispanic white borrowers and exacerbates mortgage rate inequality. While mortgage rates have always played a role in Federal Reserve policy, policymakers explicitly targeted mortgage rates only in 2008. Quantitative Easing (QE1), initiated in November of that year, consisted of large scale asset purchases (LSAPs) of mortgage-backed securities (MBS). The announcement of the LSAPs on November 25, 2008, provides a good laboratory to study the interaction between monetary policy and mortgage rate inequality. We compare the six months before with the six months after the announcement

of QE1 and find that the quarterly refinance probability for non-Hispanic white borrowers increased by 3.2 percentage points (per quarter) compared with only 1 percentage point for Black borrowers. This led to differential effects on outstanding mortgage rates, with a 21 basis point drop for the average non-Hispanic white borrowers versus a 9 basis point drop for the average Black borrower in the six months following QE1.

The observation that minority borrowers have lower prepayment speeds also has implications for mortgage pricing. Slower prepayment speeds typically make mortgages more valuable to investors, which drives down rates. We show evidence that in a competitive market, lenders would offer *lower* rates to Black and Hispanic white households as compared with otherwise identical non-Hispanic white households. This makes the observation that Black borrowers tend to be charged a higher interest rate than observationally similar non-Hispanic white borrowers at origination less justifiable as being due to statistical discrimination.¹

Our research draws a distinction between the extensive and intensive margins of opportunity in credit markets. If we think of the intensive margin here as mortgage rates offered to Black and Non-Hispanic White borrowers at origination, we find the intensive margin does not contribute that much to rate disparities. A series of recent papers (Bartlett et al. (2019), Bhutta and Hizmo (2020), and Zhang and Willen (2020)) has also documented small differences in average rates between Non-Hispanic and minority borrowers, on the order of 2–8 basis points. However, the extensive margin, defined here as whether Black borrowers get new loans by refinancing, appears to be more important.

Our paper contributes to the literature on heterogeneity in monetary policy transmission in mortgage markets. Factors such as the type of mortgage contract (Calza, Monacelli, and Stracca (2013), Di Maggio, Kermani, Keys, Piskorski, Ramcharan, Seru, and Yao (2017)), house price growth (Beraja, Fuster, Hurst, and Vavra, 2018), renting versus owning a home (Cloyne, Ferreira, and Surico, 2019), borrower age (Wong, 2019), income (Agarwal, Chomsisengphet, Kiefer, Kiefer, and Medina, 2020), and lender concentration (Scharfstein and Sunderam (2017), Agarwal, Amromin, Chomsisengphet, Landvoigt, Piskorski, Seru, and Yao (2020)) have all been found to lead to differential pass-through of monetary policy through the mortgage market across households and regions. Our finding that Black and Hispanic white mortgagees benefit less from monetary policy is therefore complementary to these results.

Our paper is also related to the literature on racial differences in mortgage performance and their implications for pricing. Previous studies including Kelly (1995), Clapp, Goldberg,

¹Higher interest rates for Black borrowers at origination was found by Black and Schweitzer (1985), Boehm, Thistle, and Schlottmann (2006), Bocian, Ernst, and Li (2008), Kau, Keenan, and Munneke (2012), Ghent, Hernández-Murillo, and Owyang (2014), Cheng, Lin, and Liu (2015), Bartlett, Morse, Stanton, and Wallace (2019), and Zhang and Willen (2020).

Harding, and LaCour-Little (2001), Deng and Gabriel (2006), Firestone, Van Order, and Zorn (2007), and Kau, Fang, and Munneke (2019) document that minority borrowers prepay their mortgages at lower rates than non-Hispanic white borrowers. There are some important differences between our analysis and these papers, however. First, none is able to distinguish between prepayments caused by home sales and those caused by refinances. Second, these studies use relatively narrow mortgage samples from either small geographic areas, short time periods, or individual banks/lenders. Third, previous studies focus exclusively on the pricing implications of prepayment differences and do not establish their implications for disparities in outstanding mortgage rates and the effect of monetary policy in exacerbating those differences.

Finally, our paper is related to the literature documenting that many borrowers appear to exercise their prepayment option in a suboptimal manner. Recently, Keys, Pope, and Pope (2016) show that a significant fraction of financially unconstrained households (approximately 20 percent) do not refinance when it is optimal to do so. Johnson, Meier, and Toubia (2018) find that more than 50 percent of borrowers neglect to refinance in a setting with zero up-front monetary costs and substantial gains in monthly payment savings. Agarwal, Ben-David, and Yao (2017) find that many homebuyers appear to suffer from the sunk cost fallacy when deciding whether to refinance. Andersen, Campbell, Nielsen, and Ramadorai (2020) decompose the inertia in refinancing into time and state dependence, and find large heterogeneity in refinancing behavior by demographics in the Danish context, many of which (e.g. income, education, immigration status) can partially explain the racial differences in refinancing behavior.² Our paper focuses on documenting the large racial differences in refinancing and its implications for mortgage rate disparities, monetary policy, pricing, and mortgage contract design.

The rest of this paper is organized as follows. Section 2 details our data and summary statistics. Section 3 contains the empirical approach we use and our results on differential prepayment tendencies across racial groups. Section 4 explores the implications of the differences in prepayment for the interest rate gap and the pass-through of monetary policy. Section 5 describes the implications of our results for mortgage pricing. Section 6 concludes.

²Earlier papers that find evidence of borrowers failing to refinance when it is likely beneficial to do so include Campbell (2006), Chang and Yavas (2009), Deng and Quigley (2012), Green and LaCour-Little (1999), and Schwartz (2006).

2 Data and Summary Statistics

We use a novel data set that combines three sources of administrative data: Home Mortgage Disclosure Act (HMDA) data, Black Knight McDash mortgage servicing data (hereafter referred to as the McDash data), and credit bureau data from Equifax. The three data sources are linked together through two separate loan-level matches: a match between the HMDA and McDash databases, which we will refer to as the HMDA-McDash data set; and a match between the McDash and Equifax databases, which is referred to as CRISM (Equifax Credit Risk Insight Servicing McDash Database). We are then able to merge the two matched data sets, creating a final data set with information from all three sources, which we will refer to as the HMDA-McDash-CRISM data set. We will briefly describe each of the three sources of data below. We describe the details of the matching procedures in the Appendix (section A.1).³

The HMDA database provides information on approximately 90 percent of US mortgage originations (see National Mortgage Database, 2017). It has been frequently used in the literature to study issues around mortgage market discrimination.⁴ The database contains a limited amount of information on borrower and loan characteristics at the time of mortgage origination, such as loan amount, borrower income, and borrower race and ethnicity. However, it does not contain some of the important underwriting variables, such as borrower credit scores, LTV ratios, loan maturities, and mortgage rates. In addition, since HMDA does not contain any information on mortgage performance over time, it is impossible to use the database to study prepayment and/or default behavior.

The McDash data set is constructed using information from mortgage servicers, which are financial institutions that are responsible for collecting payments from borrowers. It covers 60 percent to 80 percent of the US mortgage market (depending on the year) and contains detailed information on the characteristics and performance of both purchase-money mortgages and refinance mortgages. For example, it includes information on borrower credit scores, LTV ratios, maturities, interest rates, documentation levels, and additional variables measured at the time of mortgage origination. Each loan is tracked at a monthly frequency from the month of origination until it is paid off voluntarily or involuntarily via the foreclosure process. The McDash database has been used by many papers in the literature to study

³We note that all information on borrower race and gender used in this analysis comes from the HMDA database and not from the CRISM database.

⁴Examples include Carr and Megbolugbe (1993), Schill and Wachter (1993), Schill and Wachter (1994), Munnell, Tootell, Browne, and McEneaney (1996), Tootell (1996), Avery, Beeson, and Calem (1997), Black, Collins, and Cyree (1997), Holloway (1998), Reibel (2000), Black, Robinson, and Schweitzer (2001), Cherian (2014), Hauptert (2019), Bartlett et al. (2019), Bhutta and Hizmo (2020), Zhang and Willen (2020).

questions around loan performance.⁵

Finally, the CRISM data set consists of an anonymous credit file match of McDash loans to credit bureau data from Equifax at the borrower level. The Equifax data are updated at a monthly frequency and include information on outstanding consumer loans and credit lines for the primary borrower as well as all co-borrowers associated with the McDash mortgage.⁶ The CRISM data set provides the borrower’s credit bureau information beginning six months before the McDash mortgage is originated and ending six months after the McDash mortgage is terminated.⁷ It contains fields that allow us to distinguish between mortgage prepayments that are due to the borrower refinancing versus prepayments that are due to the borrower selling the property and moving. This is a significant advantage, as one of the drawbacks of virtually all loan-level data sets is that it is impossible to distinguish between prepayments due to refinances and prepayments due to home sales.

We follow the methodology used by Lambie-Hanson and Reid (2018) to classify prepayments as either refinances or sales.⁸ Specifically, we categorize a prepayment as a refinance if the borrower’s address does not subsequently change and we observe new first mortgage debt being originated either just before or just after the time of the prepayment.⁹ We categorize a prepayment as a property sale and move if we observe the borrower’s address change within a six-month window of the prepayment date.¹⁰

In addition to allowing us to distinguish between prepayments due to refinances and sales, the CRISM data set provides updated information about borrower credit scores, which we use in some of our empirical specifications to proxy for liquidity shocks.¹¹

⁵Examples include Keys, Seru, and Vig (2012), Piskorski, Seru, and Vig (2010), Jiang, Nelson, and Vytlačil (2013), Bubb and Kaufman (2014), Jiang, Nelson, and Vytlačil (2014), Kaufman (2014), Ding (2017), Fuster, Goldsmith-Pinkham, Ramadorai, and Walther (2018), Adelino, Gerardi, and Hartman-Glaser (2019), Agarwal, Ambrose, and Yao (2020) and Berger, Milbradt, Tourre, and Vavra (2020).

⁶We keep only observations that pertain to the primary mortgage borrower to avoid double counting.

⁷The McDash data set provides only information about the timing of mortgage prepayment and whether the prepayment was voluntary or involuntary due to foreclosure or distressed sale, and it does not provide any further information after the month of prepayment.

⁸Lambie-Hanson and Reid (2018) use similar data to study differences in refinancing behavior between subprime and prime borrowers.

⁹The CRISM data set provides a field that tells us the most recent quarter in which the borrower’s first mortgage debt balance changed. We use this field to identify changes in a borrower’s first mortgage debt.

¹⁰There are two fields in the CRISM data set that provide information on individuals changing their mailing addresses, which we use to identify moves. First, there is a field updated monthly that lists the month of the most recent change in the individual’s mailing address. Second, there is a field that shows the current Zip code associated with the individual’s mailing address that is also updated monthly. We assume that a borrower moves when we see either the Zip code change *or* when we see that the individual’s address has changed within a six-month window of the termination date of the mortgage. Our results are robust to narrowing the window to three months.

¹¹There are numerous alternative credit score measures in CRISM. Our analysis below focuses on the Equifax Risk Score 3.0 that was introduced in 2005 and predicts the likelihood of a consumer becoming seriously delinquent on any debt account. However, we have verified that our results are not sensitive to

Our final HMDA-McDash-CRISM data set includes loans originated in the 2005–2015 (inclusive) period. The CRISM database begins in June 2005 but does include mortgages originated prior to 2005. However, the McDash database has poorer coverage of pre-2005 mortgage originations, and thus we include only originations on or after 2005 in our sample.¹² Our data on loan performance extends through June 2020. In order to focus on a homogeneous mortgage product, we limit the sample to 30-year, fully amortizing, fixed-rate mortgages (FRMs) that were insured (against default risk) by the federal government. Specifically, we include loans that were acquired and insured by the GSEs (Fannie Mae and Freddie Mac) as well as loans that were insured by the Federal Housing Administration (FHA).¹³ We impose some additional sample restrictions to address outliers and missing information on key underwriting variables. Table A.4 in the Appendix lists all of the restrictions and how they impact the size of our sample. Most of the sample restrictions are adopted from Fuster et al. (2018), which uses the McDash-HMDA matched database.¹⁴ Finally, we include loans that were originated to Asian, Black, and white borrowers. Since HMDA provides separate identifiers for race and ethnicity, we are also able to distinguish between Hispanic/Latino white borrowers and non-Hispanic white borrowers.¹⁵

Since most of our analysis is conducted on a panel data set at the quarterly frequency where the unit of observation is a loan-quarter, we work with a 7.5 percent random sample of the HMDA-McDash-CRISM data set to ease the computational burden.¹⁶ We also dis-

the particular credit score employed. For example, our results are virtually identical if we instead use FICO scores.

¹²In 2005 McDash added a large servicer to its database, which substantially increased its overall coverage of the mortgage market. In addition, the large servicer provided information only on its *active* loans as of January 2005, while providing no information on its historical loans that had terminated prior to 2005. This raises the possibility of attrition bias being an issue in the pre-2005 McDash sample as well as the pre-2005 McDash-HMDA merged database.

¹³GSE and FHA loans account for the vast majority of 30-year FRM originations during our sample period. Loans insured by the GSEs prior to September 2008, when they were placed in conservatorship, were not technically backed by the federal government. However, most market participants believed those loans to be implicitly guaranteed by the government.

¹⁴There are a few notable sample differences between that study and our current analysis. Fuster et al. (2018) focus on 2009–2013 loan originations and consider data on loan performance only through 2016. In addition, their paper includes loans with maturities of less than 30 years as well as loans held by portfolio lenders (banks) and loans that are privately securitized.

¹⁵The race codes in HMDA are (1) American Indian or Alaska Native, (2) Asian, (3) Black or African American, (4) Native Hawaiian or other Pacific Islander, (5) white, (6) information not provided by applicant in mail, internet, or telephone application, (7) not applicable. We exclude groups 1) and 4) due to low observation counts. We also exclude groups 6) and 7). The ethnicity codes in HMDA are (1) Hispanic or Latino, (2) not Hispanic or Latino, (3) information not provided by applicant in mail, internet, or telephone application, (4) not applicable. We classify borrowers in the first group as “Hispanic,” but we make the distinction only for white borrowers. We combine Hispanic and non-Hispanic Black borrowers into the single “Black” category.

¹⁶This was the maximum sample size that we were able to work with on our Unix cluster.

tinguish between the GSE and FHA loans in our sample and conduct our analysis on each group separately. The two loan types represent very different segments of the US mortgage market, as the FHA program typically focuses on more disadvantaged and riskier borrowers who have lower credit scores and lower down payments compared with the GSEs.

Tables 1 and 2 display summary statistics (means and standard deviations) for key observable variables in our sample of GSE and FHA loans, respectively. The top panel in each table displays mortgage and borrower characteristics at origination where the unit of observation is a loan (that is, one observation per loan), while the bottom panels display summary statistics of the time-varying variables included in our analysis where the unit of observation is a loan-quarter (that is, multiple observations per loan). In both tables we display statistics for the pooled sample of borrowers as well as separately for Black, Hispanic white, and non-Hispanic white borrowers.¹⁷ There are large differences across the racial/ethnic categories for many of the observable variables in both tables. Focusing on the GSE sample, for example, non-Hispanic white borrowers have significantly higher average credit scores and household incomes compared with Black and Hispanic white borrowers (752 versus 715 and 730 and \$97.6K versus \$81.6K and \$79.1K, respectively). Non-Hispanic white borrowers obtain significantly lower mortgage rates on average (5.18 versus 5.64 and 5.45, respectively), which is documented by several papers in the literature.¹⁸ Interestingly, Black borrowers are much more likely to be female (47.8 percent) compared with both Hispanic white (31.2 percent) and non-Hispanic white (28.4 percent) borrowers, while non-Hispanic white borrowers are much more likely to have a co-applicant on the mortgage (53.1 percent) compared with Black (27.8 percent) and Hispanic white (35.7 percent) borrowers. While we see similar discrepancies between the racial/ethnic groups in the FHA sample, the values of the group averages are quite different. For example, average credit scores and household income levels are significantly lower for all groups in the FHA sample compared with the GSE sample. In addition, LTV ratios are much higher in the FHA sample (93.6 percent versus 72.6 percent).

The bottom panel of Table 1 shows that the average prepayment rate due to refinancing is 1.71 percent per quarter in our GSE sample, while the average prepayment rate due to selling and moving is 0.96 percent per quarter. The average quarterly default rate is only 0.35 percent.¹⁹ The average refinance rate is slightly lower in the FHA sample (1.33 percent)

¹⁷Asian borrowers are included in the pooled sample, but due to space constraints we do not include separate statistics for them in the table. The characteristics of Asian borrowers look very similar to non-Hispanic white borrowers across most observable variables.

¹⁸See, for example, Black and Schweitzer (1985), Boehm et al. (2006), Bocian et al. (2008), Ghent et al. (2014), Cheng et al. (2015), Bartlett et al. (2019), Bhutta and Hizmo (2020), Zhang and Willen (2020).

¹⁹We use a serious delinquency (90 days or more past due) measure of default in our analysis to be consistent with the previous literature. We also employ an involuntary prepayment definition of default that

while the average sale hazard is virtually identical. The FHA default rate is more than twice as high (0.89 percent) as the GSE rate, which is unsurprising since the FHA program is characterized by mostly first-time homebuyers with low income and low credit scores. There are large differences in average refinance rates across racial/ethnic groups in both loan samples. In the GSE sample, non-Hispanic white borrowers refinance at an average rate of 1.74 percent per quarter compared to only 1.21 percent for Black and Hispanic white borrowers. There are similar differences between non-Hispanic white and Black refinance rates in the FHA sample (1.44 percent versus 0.89 percent). There are also fairly large differences across racial/ethnic groups in both quarterly default rates as well as quarterly sale rates in both mortgage samples.

The left panel in Figure 2 plots Kaplan-Meier estimates of the hazard rates of prepayment due to refinancing by racial/ethnic group. These are unconditional, average quarterly rates as a function of duration that account for right censoring.²⁰ The figure shows that the unconditional hazard estimates of refinancing for non-Hispanic white borrowers are approximately 1 to 1.5 percentage points higher than those for Black borrowers, and that difference is fairly constant over the first 10 years of the mortgage life cycle. Hispanic white borrowers also have considerably lower refinance hazards compared with non-Hispanic white borrowers, although the difference is not as large as it is for Black borrowers.

The right panel in Figure 2 displays the Kaplan-Meier estimates of the sale hazards by racial/ethnic group. Consistent with the summary statistics discussed above, the level of the sale hazards is significantly lower than those of the refinance hazards. However, similar to the refinance estimates, we see large gaps between the hazards for non-Hispanic white borrowers and our two minority borrower groups, as non-Hispanic white households are much more likely to sell and move each quarter compared with Black and Hispanic white households.

There are also significant differences in quarterly default rates across the racial/ethnic groups. Table 1 shows that in the GSE sample, Black borrowers are almost three times as likely to default as non-Hispanic white borrowers (0.30 percent versus 0.87 percent per quarter). Hispanic white borrowers are also characterized by relatively high default hazards (0.80 per per quarter). These differences are similar in the FHA sample.²¹

includes loans that terminated due to foreclosure (both auction sales and bank/REO sales) or pre-foreclosure distressed sales (that is, short sales). We discuss results using this measure below.

²⁰Specifically, the Kaplan-Meier estimates are calculated as follows: Assuming that hazards occur at discrete times t_j where $t_j = t_{0+j}$, $j = 1, 2, \dots, J$, if we define the number of loans that have reached time t_j without being terminated or censored as n_j , and the number of terminations due to refinancing at t_j as d_{pj} , then the Kaplan-Meier estimate of the hazard function is: $\lambda_p(t_j) = \frac{d_{pj}}{n_j}$.

²¹The Kaplan-Meier estimates for defaults are displayed in Figure A.4 in the Appendix.

3 Prepayment Results

In this section we present our main empirical results. We start by showing estimates of the gap between minority and non-Hispanic white households in voluntary prepayments due to both refinancing and selling. Next, we test for differences in default behavior across the racial/ethnic borrower groups. We then show that differences in refinancing propensities are primarily due to differences in the extent to which borrowers refinance when their prepayment options are in the money, which are in turn mostly explained by observables such as income, credit scores, and loan-to-value ratios. Finally, we provide evidence that monetary policy has exacerbated the gaps in refinance propensities.

3.1 Empirical Setup

We examine differences in mortgage prepayment behavior due to refinance and home sale as well as differences in the propensity to default across racial/ethnic groups. For the bulk of our analysis we will focus on linear probability models (LPMs) that are estimated at a quarterly frequency.²² While linear probability models have some notable drawbacks,²³ they allow us to work with relatively large sample sizes and easily incorporate multiple levels of fixed effects, including highly disaggregated geographic fixed effects. We also consider logit models and show that the estimated average marginal effects are very similar to the LPM coefficient estimates.

Our primary specifications take the following general form:

$$Outcome_{it} = \beta_1 * Black_i + \beta_2 * Hispanic_i + \beta_3 * Asian_i + \gamma * X_{ijt} + \nu_g + \mu_v + \epsilon_{it}, \quad (1)$$

where i indexes the individual mortgage and t indexes the year-quarter. We focus on three mortgage outcomes: the likelihood of voluntary prepayment due to refinance, prepayment due to home sale, and finally, the likelihood of default. Specifically, $Prepay_{it}^{refi}$ is an indicator variable that takes a value of 1 if loan i prepays due to the borrower refinancing in year-quarter t , and $Prepay_{it}^{sale}$ takes a value of 1 if loan i prepays due to the borrower selling the house and moving in year-quarter t . $Default_{it}$ is an analogous indicator variable that identifies when a loan defaults. Our focus will be on testing for differences in mortgage

²²Our data set provides only the year-quarter in which each mortgage was originated due to privacy concerns. We describe the data in detail below.

²³For example, Horrace and Oaxaca (2006) prove that the LPM can lead to biased and inconsistent estimates of structural parameters when the predicted values from the regression falls outside of the [0,1] interval. On the other hand, Jörn-Steffen Pischke notes that if marginal effects are of interest, the linear probability model will be a good approximation to the conditional expectation function: <http://www.mostlyharmlesseconometrics.com/2012/07/probit-better-than-lpm/>.

outcomes across the racial/ethnic borrower groups, which will include Black, Hispanic white, Asian, and non-Hispanic white borrowers. We specify indicator variables for each group in equation (1) with non-Hispanic white borrowers representing the omitted category. Thus, the β coefficients will tell us how much more or less likely Black, Hispanic white, and Asian borrowers are to prepay/default compared with non-Hispanic white borrowers. X_{it} is a vector of control variables that include numerous mortgage and borrower characteristics, which we describe in detail below. Most of the control variables are time-invariant, but a few vary at the quarterly frequency. In some specifications we will include geographic fixed effects, ν_g , typically at the state level or Zip code level, as well as vintage year-quarter fixed effects, μ_v . The standard errors are heteroskedasticity robust and are double clustered by county and year-quarter of origination.

Since the LPMs are estimated at a quarterly frequency, we are working in a hazard framework in which we model the likelihood of prepayment/default in year-quarter t conditional on the loan surviving through $t - 1$. For example, if a loan is active for three years, at which point it prepays due to the borrower refinancing into a new loan, it will contribute 12 observations, with the $Prepay_{it}^{refi}$ indicator taking a value of 0 for the first 11 observations and a value of 1 for the final observation. Hazard models are commonly employed in the mortgage literature due to their ability to account for right-censored data (that is, loans that neither prepay or default during the sample period and are either still active at the end of the sample or exit the data set prior to the end of the sample period for other reasons).²⁴

3.2 Prepayment due to Refinancing

We begin by estimating the LPM model in equation (1) for prepayment due to borrowers refinancing into new loans. Table 3 contains the results. Columns (1) through (6) report estimates for the GSE sample, while columns (7) through (10) show estimates for the FHA sample. In all columns, we have multiplied the dependent variable (refinance indicator) by 100 so that the coefficients can be interpreted in terms of percentage points. Column (1) reports estimates from our simplest specification, which includes vintage year-quarter fixed effects to control for unobservable changes in underwriting standards over time and a control for mortgage age (third-order polynomial).²⁵ Black (Hispanic white) borrowers refinance at a rate that is 0.75 (0.69) percentage point lower than non-Hispanic white borrowers on average, while Asian borrowers refinance at a rate that is 0.44 percentage point higher than

²⁴A nontrivial number of loans in our sample are transferred to different mortgage servicers before they terminate. If the new servicer is not a contributor to the database, the loan drops out and we do not know its final outcome. These servicing transfers make up a significant fraction of our right-censored observations.

²⁵We experimented with higher order polynomials as well as one-year bins for loan age, but the results did not materially change.

non-Hispanic white borrowers on average. These differences are all statistically significant as well as economically meaningful. The gap between Black and non-Hispanic white borrowers is approximately 44 percent of the average quarterly refinance hazard among all GSE loans (1.71 percentage points).

To examine the extent to which lower prepayment likelihood of minority borrowers can be explained by their observable characteristics, in column (2) of Table 3 we include controls for some basic underwriting characteristics at origination, such as the borrower’s credit score (Equifax risk score), LTV ratio, loan size, and indicator variables for loans that are refinances, less than full documentation of income/assets, and different property types (condominiums and 2 to 4 units).²⁶ In addition, we include an estimate for the borrower’s change in LTV over time, which we calculate by updating the mortgage balance based on the amortization schedule and the value of the property using the change in the county-level house price index since the quarter of origination. Finally, we add state fixed effects to the specification. The underwriting coefficient estimates are consistent with our expectations and with previous findings in the prepayment literature. Borrowers with higher credit scores and larger loan sizes refinance at faster rates. The differences in refinancing propensities between racial/ethnic groups decrease significantly with the addition of these controls. The difference between Black and non-Hispanic white borrowers drops by almost 50 percent, from 0.75 to −0.38 percentage point per quarter. The differences between non-Hispanic white borrowers and the other minority groups also decline (in absolute magnitude) with the addition of the underwriting controls. These results suggest that about half of the difference in refinance behavior can be attributed to differences in basic underwriting variables.

In column (3) we add more information about the borrower. First, we add three variables from the HMDA database: the borrower’s reported income at the time of loan origination, an indicator for female borrowers, and an indicator for the presence of a co-applicant. We do not display the estimates due to space constraints, but they can be found in Table A.5 in the Appendix. Borrowers with higher income are more likely to refinance, while female borrowers are slightly less likely to do so. Borrowers with a co-applicant are more likely to prepay. The differences across income categories (displayed in Table A.5) are economically large and comparable to the racial/ethnic group differences. We also control for three additional variables in column (3). We control for borrower age (second order polynomial), which we obtain from the CRISM data set. We control for the “moneyness” of the refinance option using a measure constructed by Deng, Quigley, and Van Order (2000) that compares the present discounted value of the remaining stream of mortgage payments discounted at the borrower’s current mortgage rate and the remaining stream discounted at the prevailing

²⁶We also include indicators for missing information about documentation and property type.

market rate. Specifically, the “Call Option” measure of Deng et al. (2000) is calculated as:

$$Call\ Option_{i,k} = \frac{V_{i,m} - V_{i,r}}{V_{i,m}}$$

where

$$V_{i,m} = \sum_{s=1}^{TM_i - k_i} \frac{P_i}{(1 + m_t)^s}$$

$$V_{i,r} = \sum_{s=1}^{TM_i - k_i} \frac{P_i}{(1 + r_i)^s}$$

and r_i is borrower i ’s mortgage rate, TM_i is the mortgage term, k_i is the age/seasoning of the mortgage, m_i is the prevailing market rate (the PMMS index), and P_i is the mortgage payment. The larger the value of the “Call Option,” the more the borrower would benefit from refinancing into a new loan with a lower rate and payment. The third variable, “SATO” (spread at origination), is the difference between the borrower’s mortgage rate and the value of the PMMS index in the year-quarter of origination. *SATO* is often included in prepayment models to proxy for unobserved constraints that may prevent a borrower from being able to obtain the prevailing market rate. Both *Call Option* and *SATO* are strong predictors of refinance propensities as a one standard deviation increase in “Call Option” (6.4 percentage points) is associated with a 1.97 percentage point increase in the refinance hazard, while a one standard deviation increase in *SATO* (0.41 percentage points) is associated with a -0.65 percentage point decrease in the refinance hazard. Finally, we specify credit score, LTV, and loan size in small, discrete bins, rather than as continuous variables in column (3), in order to allow for any non-linearities that might exist in their relationship with the propensity to refinance. The inclusion of all these additional controls and the more flexible functional forms has only a small effect on the prepayment gaps between racial/ethnic groups relative to basic underwriting variables.

Comparing the coefficients associated with the minority groups and the non-Hispanic white group in columns (1) and (3), we see that approximately 44 percent of the gap remains for Black borrowers, while two-thirds of the gap remains for Hispanic white borrowers. One possibility is that minority borrowers are more likely to experience adverse income or liquidity shocks that make it difficult to qualify for a new loan. While we do not have direct information on income or wealth over time, the CRISM data include updated information about borrower credit scores over the life of the mortgage. Since income and wealth shocks are correlated with the likelihood of debt repayment, updated credit scores should serve as a proxy for such shocks. In column (4) of Table 3 we use this information and include the

change in the borrower’s credit score between the current year-quarter and the quarter of origination. The change in the Risk Score is highly correlated with the likelihood of refinancing. A 100 point increase is associated with a 0.78 percentage point increase in the quarterly refinance hazard. The addition of the variable also has a significant impact on the difference in refinance propensities between Black borrowers and non-Hispanic white borrowers, as the gap declines by approximately 23 percent (0.075 percentage points). Therefore, evidence suggests that a majority of the refinancing gap between non-Hispanic white and minority borrowers can be attributed to differences in underwriting variables and time-varying credit scores. This in turn implies that for policy, addressing the heterogeneous refinancing behavior of borrowers by their characteristics in a race-neutral way, such as creating and providing outreach for streamlined refinancing programs, or promoting the use of adjustable rate mortgages (ARMs), could resolve most of the refinancing gap by race.

Next, we examine whether refinancing differences are more correlated with race or the neighborhoods that minorities live in. The specification reported in column (5) of Table 3 includes Zip code fixed effects, so that differences in refinance hazards between groups in column (5) are estimated using variation only within a fairly small geographic area. This specification has the virtue of accounting for many sources of time-invariant, unobserved heterogeneity, such as the demographic composition of the Zip code area as well as the average income/wealth of the area. Controlling for the Zip code significantly narrows the gap between the racial/ethnic groups. Both the Black and Hispanic white coefficients decline by more than one-third in absolute magnitude, from -0.255 to -0.148 , and -0.421 to -0.278 , respectively. Finally, in column (6) we add a full set of Zip-code-by-year-quarter fixed effects. This specification controls for time-varying, unobserved heterogeneity at the Zip code level, and thus accounts for local economic shocks as well as local house price dynamics.²⁷ The addition of Zip-code-by-year-quarter fixed effects has almost no effect on the gap in quarterly refinance hazards. Black (Hispanic white) borrowers refinance by approximately 0.15 (0.29) percentage points less per quarter compared with non-Hispanic white borrowers in the same year-quarter in the same Zip code, controlling for credit score, change in credit score, LTV, income, gender, and our additional underwriting variables. Comparing columns (1) and (6), controlling for all observable variables at the time of mortgage origination, in addition to the change in credit scores, LTV, and Zip code level shocks over time, we can explain approximately 80 percent of the gap between the refinance behaviors of Black and non-Hispanic white borrowers and about two-thirds of the gap between Hispanic and non-

²⁷There are almost 800,000 Zip-code-by-year-quarter fixed effects. A few thousand are dropped due to there being only a single observation. Since the specification also includes vintage year-quarter fixed effects, we are unable to include the third order polynomial for mortgage age.

Hispanic white borrowers. This again suggests that a race-neutral policy based on addressing refinancing gaps by neighborhood and borrower characteristics would resolve most of the gap in refinancing.

Columns (7) through (10) in Table 3 display results corresponding to four LPM specifications estimated on our sample of FHA loans. Column (7) is analogous to column (1) and includes only vintage effects and controls for mortgage age, while column (8) is the same specification displayed in column (2), which includes basic underwriting controls such as credit score and LTV. Columns (9) and (10) are the same specifications as columns (5) and (6) and include Zip code and Zip-code-by-year-quarter fixed effects, respectively. The differences in refinance hazards across the racial/ethnic groups in the FHA sample and the patterns across the different specifications are similar to what we found in the GSE sample. Notably, similar to the results that we obtained from the GSE sample, comparing columns (7) and (10), controlling for observable borrower and mortgage characteristics and geographic differences, explains a large fraction (about 73 percent) of the differences in refinance propensities between Black and non-Hispanic white borrowers.²⁸

In Table A.6 in the Appendix we show that the results in Table 3 are not sensitive to our choice of the LPM, which assumes that the refinance hazard is a linear function of the covariates. The table contains estimated average marginal effects from logit models corresponding to each specification in Table 3.²⁹ The average marginal effects associated with the logits in all specifications are very close to the corresponding LPM coefficients.

3.3 Prepayment due to Selling

In Table 4 we test for prepayment differences between non-Hispanic white and minority borrowers due to home sales rather than refinancing activity. Our dependent variable in the LPM regressions is an indicator that takes a value of 1 if mortgage i voluntarily prepays in year-quarter t and we see that the borrower has moved and changed addresses (and 0 otherwise). We multiply the sale indicator by 100 so that the coefficients can be interpreted in terms of percentage points. The table is structured identically to Table 3, as we estimate the exact same set of specifications.

Columns (1) and (7) show that there are large differences in the propensity to sell between minority and non-Hispanic white households, controlling for only vintage effects and the age of the loan in both the GSE and FHA samples. Black borrowers are approximately 0.52

²⁸Interestingly, this is not the case for Hispanic white borrowers, however. Observables can explain only about 20 percent of the gap in refinance behavior between Hispanic and non-Hispanic white borrowers in the FHA sample.

²⁹The exception is the specifications with Zip code and Zip-code-by-year-quarter fixed effects. Those specifications include too many fixed effects to include in a logit model.

(0.64) percentage points less likely to sell their homes in a given quarter compared with non-Hispanic white borrowers in the GSE (FHA) sample, which corresponds to about 54 percent (68 percent) of the quarterly sample average (0.96 and 0.94 percentage points, respectively). In contrast to our analysis of prepayment due to refinancing, adding detailed controls for borrower and mortgage characteristics in columns (2) and (8) does not have a large effect on the minority coefficients. The gap between sale hazards for Black borrowers and Hispanic white borrowers decreases (in absolute magnitude) by approximately 20 percent in the GSE sample and even less in the FHA sample.

The addition of the HMDA variables (income, gender, and co-applicant indicator), updated credit score information, our proxy for the incentive to refinance (*Call Option*), and geographic fixed effects (state and Zip code) does further attenuate the gaps between the sale propensities of the racial/ethnic groups. However, controlling for our detailed observable borrower and loan characteristics does not have as large of an effect on the differences in sale hazards as it did on the differences in refinance hazards that we see in Table 3. Comparing the simplest specification in column (1) with our most sophisticated specification in column (6), we can explain approximately one-third of the differences between sale hazards of minority and non-white Hispanic borrowers in our GSE loan sample. Comparing columns (7) and (1), we find very similar effects in our FHA sample.

3.4 Default

In this section we present results on differences in default hazards across racial/ethnic groups. We assume that borrowers default when they miss at least three payments (that is, 90-plus days past due), to be consistent with the recent mortgage default literature. Table 5 presents estimation results for the same LPM specifications in Tables 3 and 4, with one exception. We do not include a separate specification in which we add a control for changes in borrowers' credit scores.³⁰ Again, we multiply the default indicator by 100 so that the coefficients can be interpreted in terms of percentage points.

In column (1) we see large differences between the default hazards of minority borrowers compared with non-Hispanic white borrowers. Black borrowers with GSE loans are 0.44 percentage points more likely to default on their payments each quarter, which is more than 125 percent of the average default hazard in the GSE sample (0.35 percentage points). The addition of basic controls attenuates this difference, as the Black coefficient declines to 0.29 percentage points in column (2). Further controlling for our HMDA variables and Zip code

³⁰Since credit scores are likely to decline quickly when borrowers miss mortgage payments, it wouldn't be clear whether the changes in the scores are reflecting liquidity/income shocks that drive borrowers to default or, alternatively, whether the missing payments are causing the credit scores to decrease.

fixed effects reduces the coefficient to 0.15 percentage points. Comparing columns (1) and (5), we are able to explain almost 70 percent of the differences in Black versus non-Hispanic white default hazards by controlling for observable borrower and loan characteristics and highly disaggregated geographic-by-time fixed effects. The pattern is similar for the estimated differences between Hispanic white and non-Hispanic white borrowers.

The default patterns are largely similar for Black borrowers in the FHA sample, but they are different for Hispanic white borrowers. The gap for Hispanic white borrowers of 0.165 percentage points is much smaller in column (6) (only 17 percent of the FHA sample average), and it becomes statistically insignificant in column (9) when we add our controls and the Zip-code-by-year-quarter fixed effects.

These results are consistent with previous studies documenting that Black borrowers tend to have higher cumulative default probabilities than non-Hispanic white borrowers.³¹ However, it is important to note that they are quite sensitive to the definition of default that one employs. In Table A.9 in the Appendix we estimate the same specifications but use a default definition that is based on involuntary prepayments due to foreclosure or pre-foreclosure distressed sales (that is, short sales) rather than serious delinquency. The table shows that minority loans are significantly more likely to end in involuntary prepayment when we do not control for borrower and mortgage characteristics. However, when those controls are included (in both the GSE and FHA samples), minority loans are significantly *less* likely to involuntarily prepay. This pattern suggests that minority borrowers are more likely to miss payments, but are less likely to actually lose their homes to foreclosure.³²

3.5 Racial Differences in the Sensitivity of Refinancing to Mortgage Rates

In this section we dig a bit deeper into the results on refinance disparities that we documented in section 3.2. The most common reason for borrowers to refinance is to take advantage of lower market rates and save on interest payments. In Table 3 we found that the *Call Option* variable, which proxies for the “moneyness” of the prepayment option and is driven by movements in market rates relative to the borrower’s current rate, is an important predictor of the propensity to refinance. One possible explanation for the large disparities in refinancing behavior between our racial/ethnic groups is that minority borrowers are less likely or less

³¹See, for example, Canner, Gabriel, and Woolley (1991), Berkovec, Canner, Gabriel, and Hannan (1994), and Berkovec, Canner, Gabriel, and Hannan (1998)

³²One possibility is that minority households are more likely to obtain loan modifications and avoid foreclosure. We provide some evidence below that modifications appear to disproportionately impact the interest rates that minority borrowers pay on outstanding mortgages, which is consistent with such an interpretation.

able to refinance to take advantage of lower rates. We test this hypothesis by estimating a version of equation (1) in which we interact our race/ethnicity variables with *Call Option*:

$$Prepay_{it} = \beta * Black_i + \eta * Call\ Option_{it} + \delta * (Black_i * Call\ Option_{it}) + \gamma * X_{ijt} + \nu_g + \mu_v + \epsilon_{it}, \quad (2)$$

If differences in refinance behavior between Black/Hispanic white and non-Hispanic white borrowers are explained by differential sensitivities of minority borrowers to respond to declining rates, then we should expect to find $\delta < 0$, and we should also expect to see that the inclusion of the interaction term attenuates the estimate of β .

Before discussing the results from estimating equation (2), we present a simple binned scatter plot in Figure 3 that shows the unconditional relationship between the propensity to refinance and *Call Option* for each of our racial/ethnic groups. Specifically, in Figure 3 we group the *Call Option* variable into deciles (separately for each racial/ethnic group) and then plot the average value of *Call Option* against the average quarterly refinance rate within each decile. The chart shows that all borrowers are more likely to refinance when the *Call Option* variable increases in magnitude, which corresponds to the prepayment option being deeper in the money. However, the figure clearly shows that non-Hispanic white and Asian borrowers are much more likely to refinance compared with minority borrowers when their prepayment options are deeper in the money. When market rates are either higher or about the same as the borrowers' coupon, so that *Call Option* is negative or close to zero, all borrowers have a similarly low propensity to refinance. When market rates are lower relative to the rates on outstanding loans and *Call Option* becomes more positive, the refinance hazard for non-Hispanic white and Asian borrowers increases by more than a factor of five to approximately 5 percentage points. In contrast, minority borrowers' average refinance hazard approximately doubles.

These patterns are confirmed in Table 6, which displays the results from estimating equation (2) separately for GSE and FHA loans. We start by displaying results for the LPM model without any interactions in columns (1) and (5). These specifications closely correspond to the specifications in columns (5) and (9) in Table 3, which include all of our controls as well as Zip code fixed effects, but do not include Asian borrowers. In columns (2) and (6) we add the interactions between the Black and Hispanic white dummies and the *Call Option* variable. The addition of the *Call Option* interaction explains the entire discrepancy in refinance behavior between minority and non-Hispanic white borrowers in both samples. That is, differences in refinance propensities between minority GSE borrowers and non-Hispanic white GSE borrowers comes entirely from differences in the sensitivity of refinancing in response to interest rate movements. Both columns (2) and (6) show that Black and Hispanic white borrowers are significantly less likely to refinance in response to market rates declining and the prepayment option becoming more valuable. In the GSE sample, a one standard deviation increase in *Call Option* (6.40 percentage points) increases the likelihood of refinancing by 2.1 percentage points for non-Hispanic white borrowers but only 1.4 percentage points for Black and Hispanic white borrowers. While the qualitative patterns are

similar in the FHA sample, the differences are not as large. However, the differential sensitivity to the *Call Option* variable also explains all of the difference in refinance propensities between minority and non-Hispanic white borrowers in the FHA sample.³³

The change in a borrower’s credit score is another time-varying factor that we found to be a strong predictor of refinance behavior in Table 3 and that has an important effect on the estimated disparities in refinance hazards between minority and non-Hispanic white borrowers. Our contention is that changes in credit scores over time likely reflect liquidity/income shocks that are impacting a borrower’s ability to repay debt. In columns (3) and (7) we interact the change in credit score with the Black and Hispanic white dummies to see if there are heterogeneous effects across racial/ethnic groups in their propensity to refinance in response to credit score changes. In the GSE sample, we do not find any statistically significant differences. In contrast, minority FHA borrowers are statistically significantly less likely to refinance in response to credit score improvements compared with non-Hispanic white borrowers, though the difference is not as strong in percentage terms compared with the different sensitivity to the *Call Option* value.

It is possible that the effect of changes in credit scores on refinancing propensities depends on the original credit score level. For example, an increase of 50 points for a borrower with a very low initial credit score may not improve that borrower’s ability to refinance into a lower rate, but an increase of 50 points for a borrower with a score closer to the sample average may appreciably increase the likelihood that the borrower can qualify for a lower rate. Thus, in columns (4) and (8) we add triple interaction terms between our race/ethnicity dummies, the change in credit score, and the credit score level at the time of mortgage origination. The triple interaction terms are all positive and statistically significant, which suggests that minority borrowers with high initial credit scores are more likely to refinance for a given increase in their credit scores compared with non-Hispanic white borrowers.

3.6 The Effect of Monetary Policy on Refinance Gaps

In the previous section we found that minority borrowers respond significantly less to changes in market rates that make their prepayment options more valuable compared with non-Hispanic white borrowers. This suggests that expansionary monetary policy that lowers mortgage rates could exacerbate the refinancing disparities that we have documented. In this section we take a closer look at this issue.

Figure 4 displays unconditional, quarterly refinance rates for Black (solid black line) and non-Hispanic white (dashed red line) GSE loans in calendar time over the course of our sample period. The figure shows that the refinance gap is relatively small in the first few years of the sample

³³In the Appendix we show that these results are robust to an alternative measure of the moneyness of the prepayment option. Specifically, we use the more sophisticated measure derived by Agarwal, Driscoll, and Laibson (2013) that accounts for mobility, the volatility of interest rates, closing costs, and inflation. Those results can be found in Table A.10 and are consistent with the patterns in Table 6.

period, but then it increases dramatically beginning in early 2009, right about the time of the announcement of the Federal Reserve’s first large-scale asset purchase program (LSAP), which is commonly referred to as quantitative easing (QE1). The gap closes in late 2009/early 2010, but then grows again in the third quarter of 2010, which coincides with the first Federal Reserve discussions of the second LSAP, QE2.³⁴ Finally, the third increase in the refinance gap in the figure occurs around the time of the announcement of the Fed’s final LSAP, QE3, in the third quarter of 2012.³⁵

While Figure 4 is consistent with the hypothesis that the Federal Reserve’s unconventional monetary policies played an important role in generating large differences in refinancing behavior between minority and non-Hispanic white borrowers, it is not definitive. The post-crisis period was extremely turbulent, with many other policies and shocks impacting the mortgage market.³⁶ For that reason, we implement a more direct test for monetary policy effects on the gap between the refinance behaviors of minority and non-Hispanic white households. We focus exclusively on our GSE sample since we showed in the previous section that the racial gaps in refinance behavior among FHA borrowers are not as sensitive to fluctuations in market rates. We also explicitly focus on the first LSAP, QE1. Beraja et al. (2018) show that mortgage rates fell significantly and refinancing activity expanded considerably when QE1 was announced.³⁷ Furthermore, the paper argues that unlike later LSAPs, QE1 was unanticipated by mortgage borrowers and thus provides for a fairly clean source of identification for the monetary policy effects on refinancing behavior.

QE1 was announced by the Federal Reserve on November 25, 2008, and initially called for purchases of as much as \$500 billion in MBS guaranteed by the GSEs.³⁸ In March 2009, the Federal Reserve announced that it would expand the program by purchasing \$750 billion more in MBS. QE1 terminated at the end of the first quarter of 2010 with the Federal Reserve having purchased a total of \$1.25 trillion in MBS.³⁹

We test whether QE1 exacerbated the gap between the refinance rates for minority and Non-Hispanic White borrowers by estimating the following difference-in-differences regression, which is similar in spirit to the specification used in Beraja et al. (2018):⁴⁰

³⁴On August 27, 2010, Fed Chairman Ben Bernanke stated in his speech at the Jackson Hole monetary policy conference, “A first option for providing additional monetary accommodation if necessary, is to expand the Federal Reserve’s holdings of longer-term securities.”

³⁵QE3 was announced and initiated on September 13, 2012. It involved the Federal Reserve purchasing large amounts of both MBS and Treasury securities at a monthly frequency.

³⁶For example, the Home Affordable Refinance Program (HARP) was initiated by the Federal Housing Finance Agency in March 2009 and was reformed and expanded in December 2011.

³⁷Beraja et al. (2018) show that the large increase in mortgage originations following QE1 was entirely driven by refinancings rather than purchases.

³⁸It also announced purchases of as much as \$100 billion in debt obligations of Fannie Mae, Freddie Mac, Ginnie Mae, and the Federal Home Loan Banks.

³⁹See Fuster and Willen (2010) for further details about QE1 and its effect on the mortgage market.

⁴⁰See equation (1) and Table I in the paper. The focus of that paper is on regional differences in housing equity, rather than racial differences, causing regional differences in refinancing behavior.

$$Prepay_{it} = \beta * Black_i + \eta * postQE1_t + \delta * (Black_i * postQE1_t) + \gamma * X_{ijt} + \nu_g + \mu_v + \epsilon_{it}, \quad (3)$$

where *postQE1* is an indicator variable that equals 1 for the period after QE1 and 0 for the period before QE1 as well as the quarter in which QE1 was announced (2008:Q4).⁴¹ We consider two different sample windows around the QE1 announcement: a one-year window that consists of the two quarters before and after the announcement as well as a two-year window that consists of the 4 quarters before and after the announcement.

Table 7 displays the estimation results. In columns (1) through (3) we restrict the sample to a one-year window around QE1, and in columns (4) through (6) we expand the sample to a two-year window. For each window we estimate three specifications. First, we estimate an unconditional regression with no additional controls. Second, we estimate our preferred specification from above that includes all of our loan and borrower underwriting variables as well as Zip code and origination year-quarter fixed effects (the specification in column (5) in Table 3). Finally we estimate a specification that adds interaction terms between our *postQE1* dummy and credit scores as well as LTV ratios. This is a more flexible specification that allows QE1 to differentially impact borrowers with different credit scores and LTVs, and it is motivated by anecdotal evidence that suggests the refinancing boom that followed QE1 was driven mainly by borrowers with high credit scores and low LTVs.

The estimation results in Table 7 suggest that QE1 had a large effect on the racial gap in refinance propensities. According to column (1), Black borrowers were about 0.1 percentage point less likely to refinance in the six months prior to QE1 compared with non-Hispanic white borrowers, and the gap increases by an order of magnitude to approximately 2.3 percentage points after QE1. While refinance propensities for non-Hispanic white borrowers increased by 3.2 percentage points, an increase of approximately 520 percent of their rate prior to QE1 (0.6 percent points), Black borrowers increased their refinance rates by approximately 1.0 percentage point, an increase of approximately 200 percent of their pre-QE1 rate (0.5 percent points). Including our controls and fixed effects slightly changes the magnitudes, but the large effect of QE1 on refinance gaps remains. In column (2) Black and Hispanic white conditional prepayment rates are actually significantly higher than those of non-Hispanic white borrowers in the six months before QE1, but afterwards their rates fall more than 2.6 percentage points below the rates for non-Hispanic white borrowers.

In column (3) the addition of the interactions between the *postQE1* dummy and credit scores and LTVs slightly attenuates the gaps between refinances by minority and non-Hispanic white borrowers that emerged after QE1, but the differences remain large and statistically significant. The interactions with credit score, which are displayed in the table, are striking.⁴² High-credit-

⁴¹Since QE1 was announced at the end of November, refinances driven by QE1 would not show up until the beginning of 2009:Q1.

⁴²The interaction effects with LTV are much smaller and thus not shown due to space constraints.

score borrowers (Risk Score > 740) increased their refinance rates by more than 3.7 percentage points after QE1 compared with an increase of about 0.77 percentage points for low-credit-score borrowers (Risk Score ≤ 600). Since the refinance differences across credit score bins are small in the period before QE1, these findings are consistent with the claim that the refinancing boom from QE1 was disproportionately driven by borrowers with high credit scores.

Columns (4) through (6) show that expanding the window size to one year slightly changes the estimated magnitudes, but does not alter the main patterns. QE1 appears to have generated a much larger increase in refinancing behavior by non-Hispanic white borrowers compared with minority borrowers as well as high-credit-score borrowers compared with those with lower credit scores.

While the results in Table 7 strongly suggest that QE1 significantly exacerbated refinance disparities between minority and non-Hispanic white borrowers, there were other major policies enacted around the same time as QE1, which could confound inference from our difference-in-differences estimator. For example, the Home Affordable Refinance Program (HARP) and the Home Affordable Modification Program were both enacted in March 2009, and may have had an impact on refinancing disparities across racial/ethnic groups. To address this issue and increase our confidence that QE1 really drove the differential changes in refinancing behavior in the relevant window, we zero in on the day of the announcement. To do this, we use confidential HMDA data, which provide information on the exact day on which a borrower applied for a mortgage. Figure 5 shows that from November 24 to 25, refinance applications by non-Hispanic white borrowers increased from 15,000 to more than 30,000, an increase of over 100 percent. Over those same days, applications by Black borrowers increased from 1,800 to 2,100, a gain of a little over 15 percent. Black borrowers did make further gains over the next week, but overall, over the next few weeks, the maximum increase relative to November 24 was about 50 percent, whereas for non-Hispanic white borrowers the increase rarely fell below 100 percent.

4 Implications for Mortgage Rate Disparities

The literature on statistical discrimination in mortgage market pricing focuses almost exclusively on the flow of mortgage rates—the difference in rates obtained by minority and non-Hispanic white borrowers at the time of origination. In this section we show that the large differences across groups in prepayment behavior drives large disparities in the *stock* of mortgage rates across racial/ethnic groups—the difference in rates associated with outstanding mortgages. While there are certainly good reasons to focus on the *flow* of rates, as we will show, the disparities in the stock of rates are significantly larger than the flow differences. Furthermore, we will show that monetary policy appears to have driven disparities in the stock of rates while having little impact on flow disparities.

The top panel of Figure 6 displays the difference in the flow of average mortgage rates (solid red line) for Black and non-Hispanic white borrowers during our sample period and the difference in

the stock of average rates (solid blue line). The left panel pools together FHA and GSE loans, while the right panel focuses on only GSE mortgages. These graphs are very similar to Figure 1, with the only difference being that they are constructed using our estimation sample of loans originated during the 2005–2015 period. Figure 1 uses loans originated during the 1996–2015 period. In the initial quarter (2005:Q1), the two measures coincide since we do not include any loans originated prior to 2005. There is an initial gap of about 15 basis points. The flow gap fluctuates between 10 and 25 basis points over the first few years of the pooled sample before falling to zero in 2011 and remaining below 10 basis points through the end of the sample period. In the GSE sample, the flow gap falls from just over 30 basis points in 2008 to 10 basis points in 2010 and then fluctuates between 5 and 20 basis points for the remainder of the period.⁴³ In contrast to the gap in the flow of rates, the gap in the stock of mortgage rates rises substantially after 2008 in both graphs. In the pooled sample it peaks at about 35 basis points in 2013, while it climbs to almost 60 basis points in the GSE sample.

We include a third series in each panel (dotted blue line) that adjusts the gap in outstanding rates to account for loan modifications. As we discussed above, HAMP was introduced in early 2009 and provided loan modifications to many borrowers in distress. One of the common types of modifications was interest rate reductions. Our McDash data provide information on interest rate changes over time, which we use to adjust the gap in the stock of rates to account for modifications that reduced borrower rates.⁴⁴ Interestingly, modifications appear to have had a significant impact on the rate gaps. In both panels, we can see that the difference between the average outstanding rate for Black versus non-Hispanic white borrowers is significantly reduced when we account for rate-reducing modifications. This suggests that broad-based modification programs disproportionately affected minority borrowers and helped alleviate rate disparities in the aftermath of the crisis.

To isolate the disparities in the stock of rates that is due only to prepayment behavior (as opposed to differences in pricing at origination) in the bottom panel of Figure 6, instead of using actual interest rates paid by borrowers, we assume that every mortgage origination receives that quarter’s PMMS value. Thus, by construction, there are no disparities in the rate of mortgage flows for Black and non-Hispanic white borrowers, so that the disparities in the stock of rates are driven only by the differences in prepayment propensities. The bottom panel of Figure 6 shows that beginning in 2009, the tendency of Black borrowers to pay higher than market rates for longer than non-Hispanic white borrowers increases the rate gap by more than 35 basis points in the pooled

⁴³These are slightly larger differences compared with the results in Bartlett et al. (2019), who find differences between interest rates for minority and non-Hispanic white borrowers of 7.9 and 3.6 basis points for purchase and refinance 30-year FRMs originated between during the 2009–2015 period and insured by the GSEs. However, the gap in Figure 6 is unconditional while the differences documented in Bartlett et al. (2019) are conditional on credit scores and LTV ratios. In Appendix A.7 we repeat the exercise with Survey of Consumer Finances (SCF) data as a robustness check. Although the data are much more noisier due to a smaller sample size and an inability to control for the quarter of origination, we do find a similar pattern in that the rate difference by race is larger in the stock of mortgages than at origination for new mortgages.

⁴⁴Since our sample comprises only fixed-rate loans, any change in the interest rate must be due to a modification or measurement error.

sample and by almost 50 basis points in the GSE sample.

If we go back to Figure 1, where we have a longer time series that goes back to 2000, we can see the obvious correlation between refinance waves and the differences in the stock of rates. The gap spikes during the refinance wave in the early 2000s and then again during the 2009–2015 period when unconventional monetary policy, largely through the purchases of trillions of dollars in mortgage-backed securities (MBS), drove down mortgage rates and spurred another refinance boom.

We now look further into the role played by unconventional monetary policy in driving the large increase in the gap in outstanding mortgage rates that we see in Figure 6 by estimating a difference-in-differences specification that is similar to equation 3 above. Specifically we estimate the following regression:

$$R_{it}^M = \beta * Black_i + \eta * postQE1_t + \delta * (Black_i * postQE1_t) + \epsilon_{it}, \quad (4)$$

where the dependent variable, R_{it}^M is the current mortgage interest rate paid by borrower i (which is the same as the rate at origination, since all loans in our sample are fixed rate).

Table 8 displays the estimation results for three windows around the announcement of QE1: one year, two years, and four years. For each window we display two different specifications. In columns (1), (3), and (5) we estimate specifications with no additional controls, while in columns (2), (4), and (6) we add a set of vintage year-quarter fixed effects. Adding vintage year-quarter fixed effects means that only loans originated in the same year-quarter identify the QE1 coefficients, and thus, it eliminates all variation due to prepayment differences.

The unconditional regression estimates are consistent with Figure 6. Rates paid by non-Hispanic white borrowers drop significantly after QE1—21 basis points in the one-year window and 46 basis points in the four-year window. At the same time, rates paid by minority borrowers also decline, but by much smaller magnitudes. For the one-year window, average rates paid by black borrowers drop by 11.5 basis points after QE1 and by about 23 basis points in the four-year window. This causes the gap in outstanding rates to grow from 21 basis points in the two years before QE1 to 44 basis points in the two years after the policy.

The addition of vintage year-quarter fixed effects completely eliminates the positive post-QE1 estimates on mortgage rates for all borrowers. This confirms that it is loans originated in different periods that drive the unconditional results, which is consistent with differential refinancing behavior driving the large divergence in mortgage rates for minority and non-Hispanic white borrowers in the period after QE1.

5 Pricing Implications

Differential prepayment behavior of Black and Hispanic borrowers has significant implications for the pricing of mortgages. We focus on three aspects. First, lower prepayments mean that loans

to Black and Hispanic white borrowers are more valuable to lenders and investors. Second, as a result, equilibrium interest rates paid by Black and Hispanic white borrowers should be *lower* at origination than rates paid by otherwise identical non-Hispanic white borrowers, conditional on the GSEs and FHA’s pricing of default insurance. Third, lower prepayment rates mean that the GSEs and FHA’s cost of providing default insurance could be higher for Black and Hispanic white borrowers even when the hazard of default is the same as it is for comparable non-Hispanic white borrowers.

Consider a mortgage with an initial balance S_0 . Assume that time is continuous and the loan has constant prepayment and default hazards, λ_p and λ_d , respectively. The interest rate in the economy is r , the note rate on the mortgages is m , and the lender pays a guarantee fee g to insure timely repayment of principal and interest. The value of this loan is

$$V = \int_0^\infty e^{-rt} S_t (m - g + \lambda_p + \lambda_d) dt.$$

We assume that the hazards are exponential so $S_t = S_0 e^{-(\lambda_p + \lambda_d)t}$, implying that:

$$V - S_0 = \frac{m - g - r}{r + \lambda_p + \lambda_d} \quad (5)$$

We follow industry practice and refer to the left-hand side of equation (5) as the gain-on-sale of a mortgage. Two key insights emerge from equation (5). First, gain-on-sale is positive if and only if the flow income from the loan $m - g - r$ is positive. In the top part of Figure A.3 in the Appendix, we use MBS market prices for Fannie Mae and Freddie Mac loans to compute $V - S_0$ for different pools of loans. The line labelled “TBA” is for low-risk mortgages with a note rate equal to the Freddie Mac PMMS rate for a 30-year FRM. The figure shows that $V - S_0$ is always positive and, in the later years of our sample, substantial, which in turn implies that the flow income from the loan, $m - g - r$, is always positive. Second, equation (5) shows that a reduction in λ_p , the prepayment speed, reduces gain-on-sale if $m - g - r$ is positive. These two facts imply that for the typical loan, a reduction in the prepayment rate should increase the value of the mortgage to lenders and investors.

We quantify the extent to which lower prepayment speeds increase the value of minority borrowers’ mortgages in two ways: (1) by comparison with low-balance mortgages, and (2) by looking at actual traded MBS prices. It is well known in the industry some borrowers are less likely to prepay, and mortgages with lower prepayment risk tend to trade in their own specified pools or “spec” pools because they are more valuable than the typical, To-Be Announced (TBA) pools that are only differentiated by issuer and coupon. We can use pricing information from spec pools to obtain a rough estimate of the rate premium that Black borrowers might obtain if lenders took into account their lower prepayment speeds.

It is widely recognized that low balance mortgages are less likely to prepay. An economic explanation is that because some costs of refinancing are fixed, but the benefits of refinancing are

proportional to the balance of the loan. We compare the Black and Hispanic borrower’s prepayment speeds with those of low balance mortgages. In Table A.12 in the Appendix we combine refinances and home sales into a single prepayment variable (since investors do not care about the reason for voluntary prepayment) and regress this prepayment variable on our race dummies (column (1)) and then separately on our indicator variables for loan amount (column (2)). The difference in quarterly prepayment hazards for Black and non-Hispanic white borrowers is approximately 1.63 percentage points. Column (2) in Table A.12 shows that this is very similar to the prepayment gap between loans that are below \$85,000 and those that are above \$175,000 (1.70 percentage points).⁴⁵ Thus, we will focus on spec pools that consist of loans with original balances lower than \$85,000. The gain-on-sale premium for pools of loans in these spec pools is typically between 50 and 100 basis points.⁴⁶ Therefore, by comparison with low balance mortgages, we estimate that Black borrowers’ mortgages are between 50 to 100 basis points more valuable than those of non-Hispanic white borrowers’ on the secondary market.

An alternative way of estimating the premium on Black borrowers’ mortgages is to look at the actual secondary market prices paid by investors and dealers. While the race of the borrower is not typically disclosed to investors, the addresses associated with loans are, and investors may be able to proxy for race by looking at the racial compositions of neighborhoods. We investigate whether spec pools with more mortgages from Black census tracts receive higher premiums. To do so, we merge data on loan-level collateral of spec pools from eMBS with the 2018-2019 HMDA. We then regress transaction prices in the TRACE MBS data by the percentage of Black borrowers in a census tract along with a series of control variables to see whether pools with loans from more heavily Black census tracts receive higher prices. Table A.11 shows the results. Column 1 of Table A.11 shows that, looking at all trades with MBS value greater than or equal to \$85k, an 1% increase in the Black presence in a census tract is associated with a 1 basis point increase in secondary market value of the spec pool, implying that Black borrowers’ mortgages are 100 basis points more valuable than those of non-Hispanic white borrowers’. Column 2 of Table A.11 shows that the implied difference in value grows to 141 basis points if we look at trades with MBS value greater than or equal to 1 million. The 100-141 basis points estimate is in the same magnitude as the 50-100 basis points gain on sale estimate we obtained from comparison with low balance mortgages. The difference between them may be attributable to the difference in sample period (the MBS results are only for more recent vintages).

How does this affect borrowers? To get some sense of how rates paid by minority borrowers would change if lenders took into account lower prepayment speeds, we can look at the low-balance mortgages. Assuming that a lender wants to maintain a constant gain-on-sale across all loans, we can then ask what the rate reduction on loans to Black borrowers would need to be to ensure that outcome. If the 50-100 basis point gain on sale differences were fully passed through to Black

⁴⁵The omitted/reference group in the regression consists of mortgages that are greater than \$175,000 so that the coefficients in the table should be interpreted as relative comparisons with that group.

⁴⁶Compare the lines labeled “Low-balance spec pool” and “TBA” in Figure A.3 in the Appendix.

borrowers, they would typically pay 5 to 15 basis points less than they currently do.⁴⁷

In our sample, mortgages are either insured by Fannie Mae and Freddie Mac or guaranteed by FHA. Because of a quirk in the way Fannie Mae and Freddie Mac price insurance, the higher prepayment speeds of non-Hispanic white borrowers may make them more attractive to insure. Fannie and Freddie receive income from the flow of mortgage insurance payments g and from a one-time fee called an *LLPA*. Using our assumptions from above, a mortgage insurance contract is worth

$$I = LLPA + \int_0^\infty e^{-rt} S_t (g - \lambda_d LGD) dt = LLPA + S_0 \left[\frac{g - LGD\lambda_d}{r + \lambda_p + \lambda_d} \right],$$

where $LGD \cdot S_t$ is the loss suffered by the lender on a loan that defaults. Suppose the lender chooses *LLPA* and g for a given pool of loans in which all borrowers have the same λ_d . It is easy to see that unless $g = LGD\lambda_d$, the value of the insurance contract I depends on the prepayment speed. If $LGD\lambda_d > g$, then higher prepayment speeds will make insurance contracts more valuable. The issue is that Fannie and Freddie set g independently of risk characteristics and adjust the *LLPA* to account for LTV and FICO score. Because they have higher unconditional default hazards, λ_d , $g - \lambda_d LGD$ is more likely to be negative for Black and Hispanic white borrowers. Thus, I will be lower for a Black or Hispanic white borrower when compared with an otherwise identical non-Hispanic white borrower who has a higher λ_p .

6 Conclusion

In this paper we have shown that minority borrowers refinance their fixed-rate mortgages at a significantly lower rate compared with non-Hispanic white borrowers, and that expansionary monetary policy appears to have exacerbated these differences. In turn, the large differences in refinance propensities have resulted in significant disparities in the average interest rate that minority borrowers pay on the stock of outstanding mortgages compared with their non-Hispanic white counterparts. These differences in the stock of rates are much larger in magnitude than the corresponding differences in the rates paid on newly originated loans.

To be clear, our analysis does not suggest that policies that drive down mortgages rates are harmful to minority borrowers. To the contrary, minority borrowers do benefit from lower mortgage rates. However, our analysis suggests that they benefit much less than white borrowers.

Our research leads to two important questions. First, why do Black and Hispanic white borrowers refinance less frequently? In particular, why are they so much less responsive to variation in interest rates. As we have shown, observable differences across borrowers can explain about 80 percent of the difference, but a nontrivial gap remains. The remaining gap could be explained by numerous factors that are omitted from our analysis including different levels of education and/or financial literacy, differential exposure to negative income/employment shocks that may inhibit

⁴⁷Figure A.3 shows that there are periods, such as early 2009 and late 2010, when they would pay substantially less (~ 30 bps).

the ability to refinance into low rates and that are not reflected in updated credit scores, or even heterogeneous social networks, which have been shown to be important transmitters of information about refinancing opportunities (Maturana and Nickerson (2019)).

The second question is, what can policymakers do to reduce racial differences? The prepayable, fixed-rate mortgage plays a central role in the story. Many commentators argue that the FRM offers the best of both worlds. Essentially, the prepayment option enables the borrower to take advantage of falling rates while providing insurance against rising rates. But the value of this option, in the real world, depends on both the willingness and ability of borrowers to exercise the option. The data show systematic variation across racial groups in refinancing and moving propensities, and thus, in a sense, the value of the option.

How could a policymaker enable Black and Hispanic white borrowers to exploit rate reductions more effectively? One way would be to expand the use of adjustable-rate mortgages (ARMs). The United States is almost unique in its reliance on FRMs. In many countries, the mortgage ecosystem is largely populated with ARMs, and those countries enjoy high home-ownership rates and have foreclosure problems that are no worse than in the United States. Another would be to encourage the mortgage industry to develop products that combine the benefits of FRMs and ARMs. For example, for many years, market participants have discussed “ratchet” mortgages, which adjust down but not up. These alternative mortgage contract designs may lead to a more equitable distributional impact of monetary policy. Finally, complementary, race-neutral policies that make it easier and less costly to refinance such as streamlined refinancing programs may also be effective in closing these rate disparities.

References

- Adelino, M., K. Gerardi, and B. Hartman-Glaser (2019). Are lemons sold first? dynamic signaling in the mortgage market. *Journal of Financial Economics* 132(1), 1 – 25.
- Agarwal, S., B. W. Ambrose, and V. W. Yao (2020). Lender steering in residential mortgage markets. *Real Estate Economics* 48(2), 446–475.
- Agarwal, S., G. Amromin, S. Chomsisengphet, T. Landvoigt, T. Piskorski, A. Seru, and V. Yao (2020). Mortgage refinancing, consumer spending, and competition: Evidence from the home affordable refinancing program. *Working paper*.
- Agarwal, S., I. Ben-David, and V. Yao (2017). Systematic mistakes in the mortgage market and lack of financial sophistication. *Journal of Financial Economics* 123(1), 42–58.
- Agarwal, S., S. Chomsisengphet, H. Kiefer, L. C. Kiefer, and P. C. Medina (2020). *Working paper*.
- Agarwal, S., J. C. Driscoll, and D. I. Laibson (2013). Optimal mortgage refinancing: A closed-form solution. *Journal of Money, Credit and Banking* 45(4), 591–622.
- Andersen, S., J. Y. Campbell, K. M. Nielsen, and T. Ramadorai (2020, October). Sources of inaction in household finance: Evidence from the danish mortgage market. *American Economic Review* 110(10), 3184–3230.
- Avery, R. B., P. E. Beeson, and P. S. Calem (1997). Using hmda data as a regulatory screen for fair lending compliance. *Journal of Financial Services Research* 11, 9–12.
- Bartlett, R., A. Morse, R. Stanton, and N. Wallace (2019). Consumer-lending discrimination in the fintech era. *Working paper*.
- Beraja, M., A. Fuster, E. Hurst, and J. Vavra (2018). Regional Heterogeneity and the Refinancing Channel of Monetary Policy*. *The Quarterly Journal of Economics* 134(1), 109–183.
- Berger, D. W., K. Milbradt, F. Tourre, and J. Vavra (2020). Mortgage prepayment and path-dependent effects of monetary policy. *Working paper*.
- Berkovec, J. A., G. B. Canner, S. A. Gabriel, and T. H. Hannan (1994). Race, redlining, and residential mortgage loan performance. *The Journal of Real Estate Finance and Economics* 9, 263–294.
- Berkovec, J. A., G. B. Canner, S. A. Gabriel, and T. H. Hannan (1998). Discrimination, competition, and loan performance in fha mortgage lending. *The Review of Economics and Statistics* 80(2), 241–250.
- Bhutta, N. and A. Hizmo (2020). Do minorities pay more for mortgages?
- Black, H. A., M. C. Collins, and K. B. Cyree (1997). Do black-owned banks discriminate against black borrowers? *Journal of Financial Services Research* 11, 189–204.

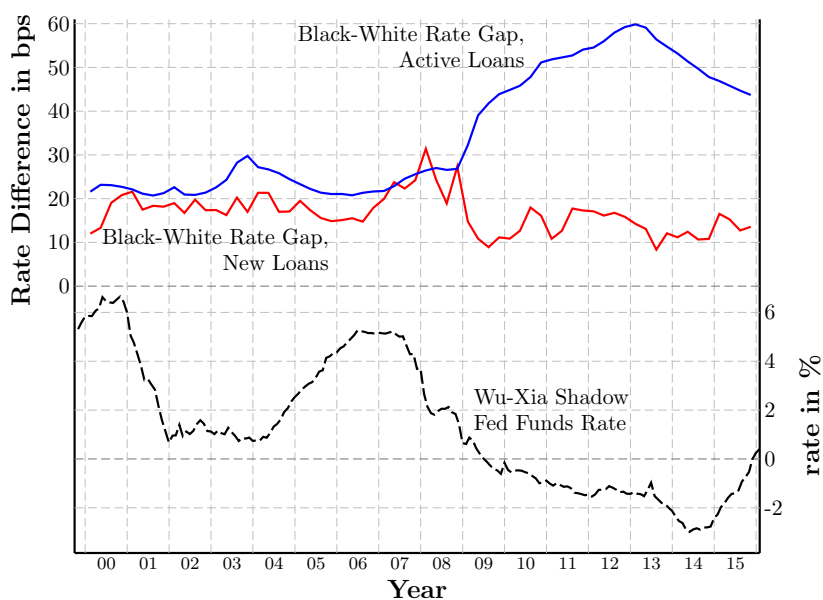
- Black, H. A., B. L. Robinson, and R. L. Schweitzer (2001). Comparing lending decisions of minority-owned and white-owned banks: Is there discrimination in mortgage lending? *Review of Financial Economics* 10(1), 23 – 39.
- Black, H. A. and R. L. Schweitzer (1985). A canonical analysis of mortgage lending terms: Testing for lending discrimination at a commercial bank. *Urban Studies* 22(1), 13–19.
- Bocian, D. G., K. S. Ernst, and W. Li (2008). Race, ethnicity and subprime home loan pricing. *Journal of Economics and Business* 60(1), 110 – 124. Financing Community Reinvestment and Development.
- Boehm, T. P., P. D. Thistle, and A. Schlottmann (2006). Rates and race: An analysis of racial disparities in mortgage rates. *Housing Policy Debate* 17(1), 109–149.
- Bubb, R. and A. Kaufman (2014). Securitization and moral hazard: Evidence from credit score cutoff rules. *Journal of Monetary Economics* 63, 1–18.
- Calza, A., T. Monacelli, and L. Stracca (2013). Housing finance and monetary policy. *Journal of the European Economic Association* 11(s1), 101–122.
- Campbell, J. Y. (2006). Household finance. *The journal of finance* 61(4), 1553–1604.
- Canner, G. B., S. A. Gabriel, and J. M. Woolley (1991). Race, default risk and mortgage lending: A study of the fha and conventional loan markets. *Southern Economic Journal* 58(1), 249–262.
- Carr, J. H. and I. F. Megbolugbe (1993). The federal reserve bank of boston study on mortgage lending revisited. *Journal of Housing Research* 4(2), 277–313.
- Chang, Y. and A. Yavas (2009). Do borrowers make rational choices on points and refinancing? *Real Estate Economics* 37(4), 635–658.
- Cheng, P., Z. Lin, and Y. Liu (2015). Racial discrepancy in mortgage interest rates. *The Journal of Real Estate Finance and Economics* 51(1), 101–120.
- Cherian, M. (2014). Race in the mortgage market: An empirical investigation using hmda data. *Race, Gender Class* 21(1/2), 48–63.
- Clapp, J. M., G. M. Goldberg, J. P. Harding, and M. LaCour-Little (2001). Movers and shuckers: interdependent prepayment decisions. *Real estate economics* 29(3), 411–450.
- Cloyne, J., C. Ferreira, and P. Surico (2019, 01). Monetary Policy when Households have Debt: New Evidence on the Transmission Mechanism. *The Review of Economic Studies* 87(1), 102–129.
- Deng, Y. and S. Gabriel (2006). Risk-based pricing and the enhancement of mortgage credit availability among underserved and higher credit-risk populations. *Journal of money, Credit and Banking*, 1431–1460.

- Deng, Y. and J. M. Quigley (2012). Woodhead behavior and the pricing of residential mortgages. *NUS Institute of Real Estate Studies Working Paper Series IRES2012-025*.
- Deng, Y., J. M. Quigley, and R. Van Order (2000). Mortgage terminations, heterogeneity and the exercise of mortgage options. *Econometrica* 68(2), 275–307.
- Di Maggio, M., A. Kermani, B. J. Keys, T. Piskorski, R. Ramcharan, A. Seru, and V. Yao (2017). Interest rate pass-through: Mortgage rates, household consumption, and voluntary deleveraging. *American Economic Review* 107(11), 3550–88.
- Ding, L. (2017). Borrower credit access and credit performance after loan modifications. *Empirical Economics* 52(3), 977–1005.
- Firestone, S., R. Van Order, and P. Zorn (2007). The performance of low-income and minority mortgages. *Real Estate Economics* 35(4), 479–504.
- Fuster, A., P. Goldsmith-Pinkham, T. Ramadorai, and A. Walther (2018). Predictably unequal? the effects of machine learning on credit markets. *Working paper*.
- Fuster, A. and P. Willen (2010). \$1.25 trillion is still real money: Some facts about the effects of the federal reserve’s mortgage market investments. *FRB of Boston Public Policy Discussion Paper* (10-4).
- Ghent, A. C., R. Hernández-Murillo, and M. T. Owyang (2014). Differences in subprime loan pricing across races and neighborhoods. *Regional Science and Urban Economics* 48, 199 – 215.
- Green, R. K. and M. LaCour-Little (1999). Some truths about ostriches: Who doesn’t prepay their mortgages and why they don’t. *Journal of Housing Economics* 8(3), 233–248.
- Hauptert, T. (2019). Racial patterns in mortgage lending outcomes during and after the subprime boom. *Housing Policy Debate* 29(6), 947–976.
- Holloway, S. R. (1998). Exploring the neighborhood contingency of race discrimination in mortgage lending in columbus, ohio. *Annals of the Association of American Geographers* 88(2), 252–276.
- Horrace, W. C. and R. L. Oaxaca (2006). Results on the bias and inconsistency of ordinary least squares for the linear probability model. *Economics Letters* 90(3), 321 – 327.
- Jiang, W., A. A. Nelson, and E. Vytlačil (2013, 11). Securitization and Loan Performance: Ex Ante and Ex Post Relations in the Mortgage Market. *The Review of Financial Studies* 27(2), 454–483.
- Jiang, W., A. A. Nelson, and E. Vytlačil (2014). Liar’s loan? effects of origination channel and information falsification on mortgage delinquency. *The Review of Economics and Statistics* 96(1), 1–18.
- Johnson, E. J., S. Meier, and O. Toubia (2018, 05). What’s the Catch? Suspicion of Bank Motives and Sluggish Refinancing. *The Review of Financial Studies* 32(2), 467–495.

- Kau, J. B., L. Fang, and H. J. Munneke (2019). An unintended consequence of mortgage financing regulation—a racial disparity. *The Journal of Real Estate Finance and Economics* 59(4), 549–588.
- Kau, J. B., D. C. Keenan, and H. J. Munneke (2012). Racial discrimination and mortgage lending. *The Journal of Real Estate Finance and Economics* 45(2), 289–304.
- Kaufman, A. (2014). The influence of fannie and freddie on mortgage loan terms. *Real Estate Economics* 42(2), 472–496.
- Kelly, A. (1995). Racial and ethnic disparities in mortgage prepayment. *Journal of Housing Economics* 4(4), 350–372.
- Keys, B. J., D. G. Pope, and J. C. Pope (2016). Failure to refinance. *Journal of Financial Economics* 122(3), 482–499.
- Keys, B. J., A. Seru, and V. Vig (2012, 05). Lender Screening and the Role of Securitization: Evidence from Prime and Subprime Mortgage Markets. *The Review of Financial Studies* 25(7), 2071–2108.
- Lambie-Hanson, L. and C. Reid (2018). Stuck in subprime? examining the barriers to refinancing mortgage debt. *Housing Policy Debate* 28(5), 770–796.
- Maturana, G. and J. Nickerson (2019). Teachers teaching teachers: The role of workplace peer effects in financial decisions. *The Review of Financial Studies* 32(10), 3920–3957.
- Munnell, A. H., G. M. B. Tootell, L. E. Browne, and J. McEneaney (1996). Mortgage lending in boston: Interpreting hmda data. *The American Economic Review* 86(1), 25–53.
- Piskorski, T., A. Seru, and V. Vig (2010). Securitization and distressed loan renegotiation: Evidence from the subprime mortgage crisis. *Journal of Financial Economics* 97(3), 369 – 397. The 2007-8 financial crisis: Lessons from corporate finance.
- Reibel, M. (2000). Geographic variation in mortgage discrimination: Evidence from los angeles. *Urban Geography* 21(1), 45–60.
- Scharfstein, D. and A. Sunderam (2017). Market power in mortgage lending and the transmission of monetary policy. *Working paper*.
- Schill, M. H. and S. M. Wachter (1993). A tale of two cities: Racial and ethnic geographic disparities in home mortgage lending in boston and philadelphia. *Journal of Housing Research* 4(2), 245–275.
- Schill, M. H. and S. M. Wachter (1994). Borrower and neighborhood racial and income characteristics and financial institution mortgage application screening. *The Journal of Real Estate Finance and Economics* 9, 223–239.
- Schwartz, A. (2006). Household refinancing behavior in fixed rate mortgages. *unpublished paper, Harvard University*.

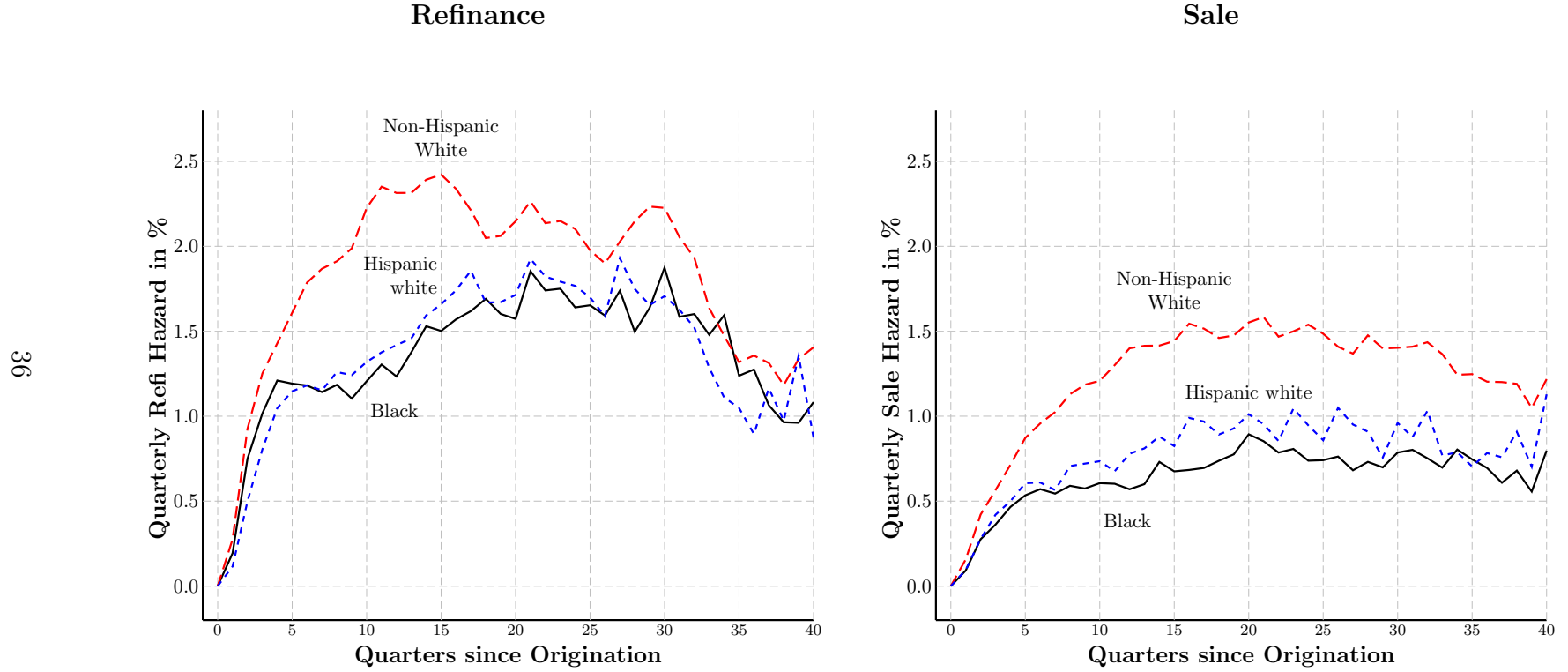
- Tootell, G. M. B. (1996, 11). Redlining in Boston: Do Mortgage Lenders Discriminate Against Neighborhoods?*. *The Quarterly Journal of Economics* 111(4), 1049–1079.
- Wong, A. (2019). Refinancing and the transmission of monetary policy to consumption. *Working Paper*.
- Zhang, D. H. and P. S. Willen (2020). Do lenders still discriminate? a robust methodology for detecting differences in menus. *Working Paper*.

Figure 1: Rates on outstanding mortgages: Black versus non-Hispanic white Borrowers for mortgages originated from 1996–2015



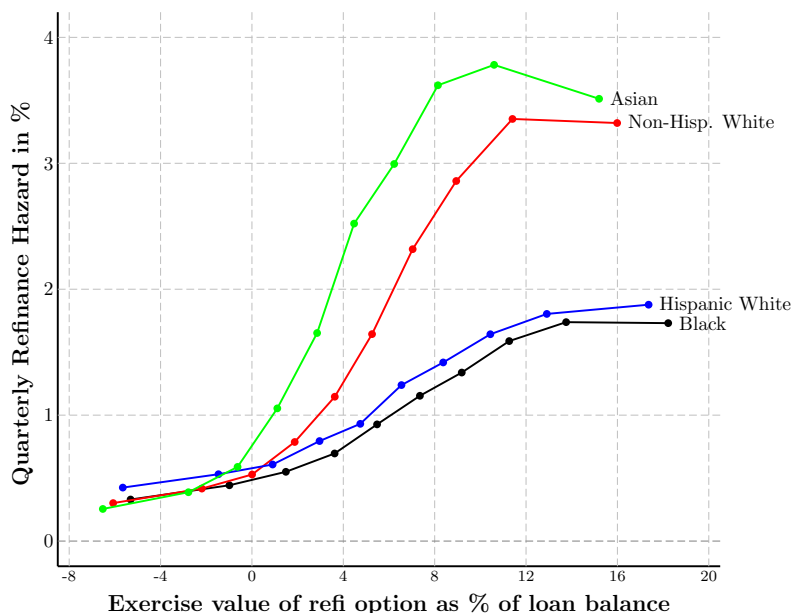
Notes: This figure displays the rate gap for Black and non-Hispanic white borrowers with 30-year FRMs. New Loans are originated in the quarter and active loans are all outstanding loans. Data to compute the rate gaps come from the HMDA-McDash database. The Wu-Xia Shadow Fed Funds rate comes from <https://www.frbatlanta.org/cqer/research/wu-xia-shadow-federal-funds-rate>.

Figure 2: Kaplan Meier unconditional refinance and sale hazard rates



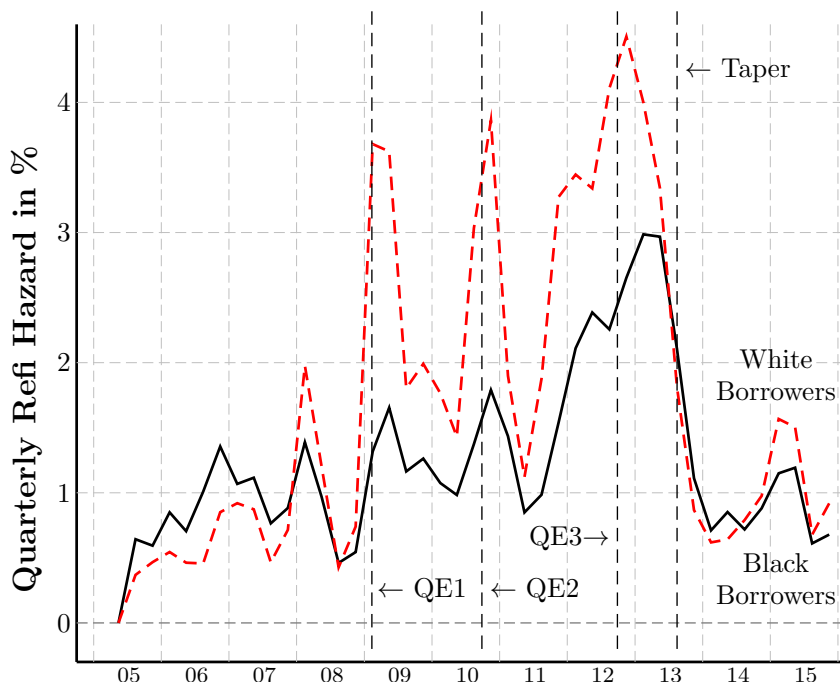
Notes: This figure displays the Kaplan-Meier hazard estimates of refinance and home sale broken down by racial/ethnic groups. The Kaplan-Meier estimate of the hazard function is: $\lambda_p(t_j) = \frac{d_{pj}}{n_j}$, where the number of loans that have reached time t_j without being terminated or censored is given by n_j , and the number of terminations due to prepayment at t_j is given by d_{pj} . The underlying data come from the HMDA-McDash-CRISM database.

Figure 3: Responses to gain from exercising the refinance option.



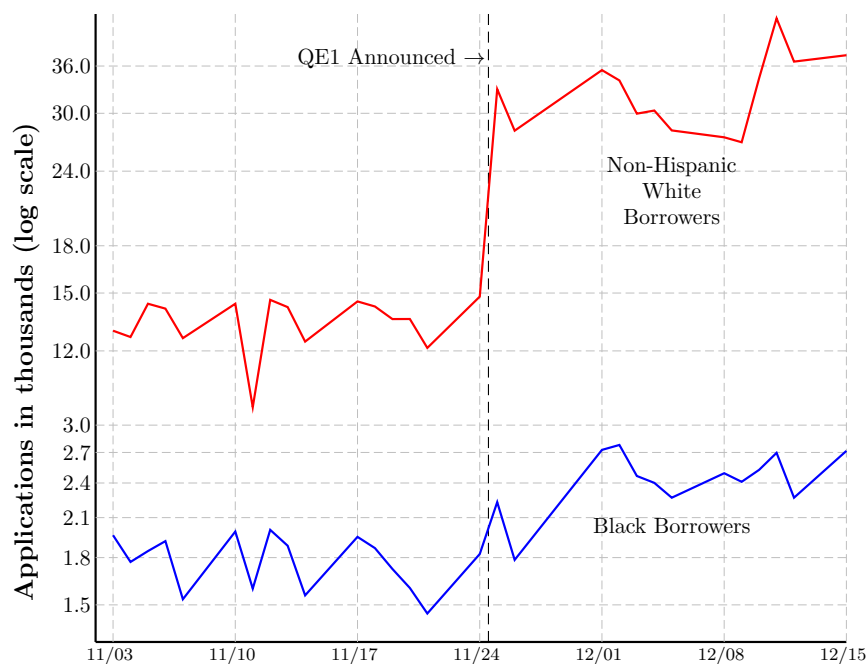
Notes: This figure shows a binned scatter plot of the hazard of refinance as a function of the gain from exercising the refinance option as calculated in Deng et al. (2000).

Figure 4: Unconditional quarterly refinance hazards for Black and non-Hispanic white borrowers.



Notes: Hazard is defined as the percentage of matched HMDA-McDash-CRISM loans at the beginning of a quarter that are refinanced by the end of the quarter. Events are QE1, announcement of original LSAP in November 2008; QE2, Bernanke's August 2010 speech suggesting an expansion of LSAPs; QE3, FOMC vote to buy \$40b bonds per month in September 2012; Taper, Bernanke's 2013 FOMC press conference suggesting that FOMC would wind down purchases of MBS.

Figure 5: Event study of the announcement of first quantitative easing (QE1) on November 25, 2008.



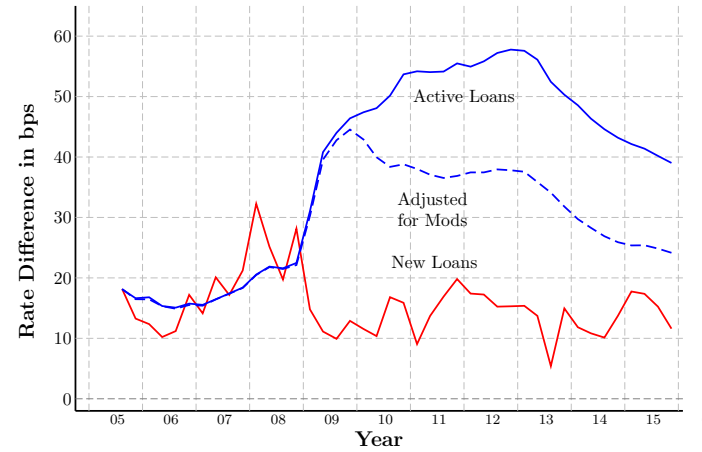
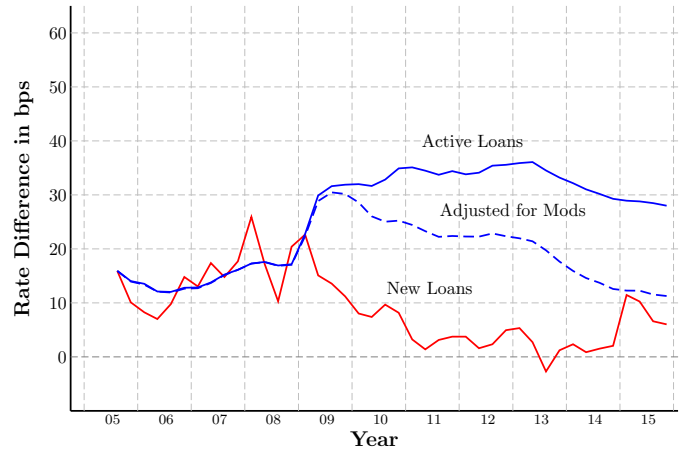
Notes: This figure shows the ratio of Black versus non-Hispanic white refinance applications normalized to the day before the announcement of QE1. The data come from the confidential Home Mortgage Disclosure Act (cHMDA) files.

Figure 6: Gap between interest rates for Black and non-Hispanic white borrowers for mortgages originated from 2005–2015

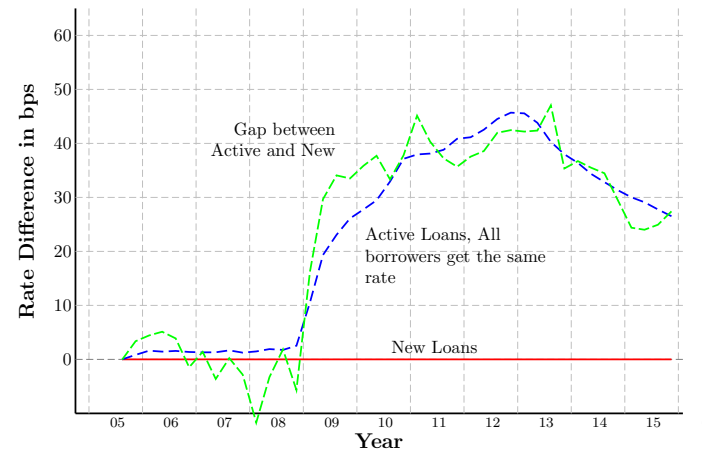
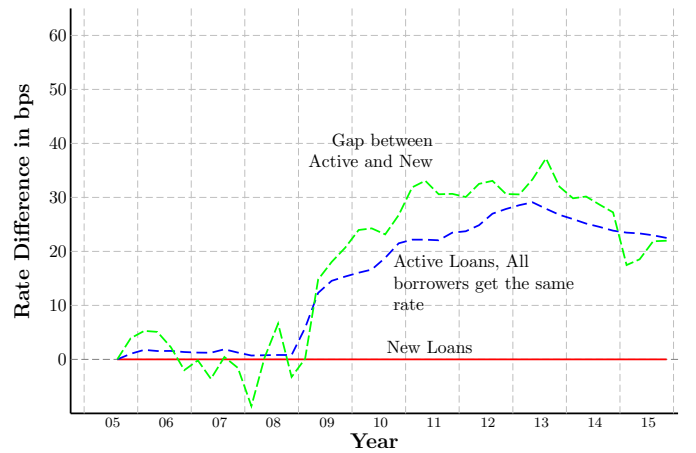
1. GSE and FHA Loans

2. GSE Loans Only

A. With Actual Rates



B. Assuming all borrowers receive average quarterly rate at origination



Notes: This figure displays the difference between the average interest rate paid by a Black versus a non-Hispanic white borrower. “New Loans” are loans originated in the quarter. “Active Loans” are all loans outstanding in the quarter (including new loans.) The top two panels shows the actual rates reported in the HMDA-McDash-Equifax database. In the bottom panels, we isolate the effect of refinances by assigning every borrower the average FHLMC Primary Mortgage Market Survey rate for the quarter in which their loan was originated.

Table 1: Summary Statistics: GSE Sample

Panel A: Fixed Characteristics								
	All		Black		Hispanic White		Non-Hispanic White	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Equifax Risk Score (100s points)	7.50	0.59	7.15	0.72	7.30	0.63	7.52	0.57
LTV (%)	72.6	15.9	75.6	15.4	74.0	15.9	72.5	16.0
Loan Amount (\$100k)	2.12	1.13	1.84	1.00	1.98	1.03	2.10	1.11
Interest Rate (ppts)	5.20	1.02	5.64	1.09	5.45	1.06	5.18	1.01
Income (\$1k)	97.6	64.0	81.6	51.9	79.1	51.5	98.6	64.7
Borrower Age (years)	46.3	13.4	48.4	13.2	45.2	12.8	46.5	13.5
Refinance (d)	0.538	0.499	0.588	0.492	0.513	0.500	0.543	0.498
Condo (d)	0.140	0.347	0.149	0.356	0.141	0.348	0.134	0.340
2-4 Family (d)	0.018	0.133	0.039	0.194	0.040	0.197	0.015	0.121
Low Documentation (d)	0.308	0.462	0.325	0.468	0.308	0.462	0.309	0.462
Non-Occupant Owner (d)	0.140	0.347	0.160	0.367	0.144	0.351	0.137	0.344
Female (d)	0.294	0.455	0.478	0.500	0.312	0.463	0.284	0.451
Co-applicant (d)	0.505	0.500	0.278	0.448	0.357	0.479	0.531	0.499
# Loans	800,806		32,753		43,269		676,986	
Panel B: Time-Varying Characteristics								
	All		Black		Hispanic White		Non-Hispanic White	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Qtrs since Orig	12.5	9.9	14.2	11.1	13.5	10.6	12.3	9.8
Interest Rate (ppts)	5.15	1.01	5.61	1.06	5.45	1.04	5.12	1.00
Call Option (ppts)	4.77	6.40	7.16	6.87	6.38	6.71	4.57	6.31
SATO (ppts)	0.150	0.411	0.281	0.478	0.235	0.445	0.139	0.403
LTV Change	-4.60	14.70	-1.92	16.96	-1.74	20.38	-4.84	13.96
Negative Equity (d)	0.045	0.207	0.090	0.287	0.104	0.306	0.038	0.192
Risk Score Change (100s points)	0.070	0.530	-0.030	0.686	0.005	0.659	0.077	0.510
Prepay Refinance (ppts)	1.71	12.95	1.21	10.95	1.21	10.91	1.74	13.09
Prepay Sale (ppts)	0.96	9.76	0.54	7.35	0.63	7.93	1.02	10.03
Default (ppts)	0.35	5.90	0.87	9.28	0.80	8.92	0.30	5.44
# Loan-quarters	15,460,588		730,648		924,765		12,970,785	

Notes: This table reports summary statistics from a 7.5% random sample of loans originated from 2005–2015 (inclusive) and held by the GSEs (Fannie Mae and Freddie Mac) from a matched HMDA-McDash-CRISM data set. The unit of observation in Panel A is a loan, while the unit of observation in Panel B is a loan-quarter. The label (d) denotes dummy variables. “SATO” is the spread between the mortgage rate and the average rate associated with newly originated 30-year FRMs according to the FHLMC survey. “Call Option” is a measure of the incentive to refinance taken from Deng et al. (2000).

Table 2: Summary Statistics: FHA Sample

Panel A: Fixed Characteristics								
	All		Black		Hispanic White		Non-Hispanic White	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Equifax Risk Score (100s points)	6.84	0.67	6.53	0.71	6.77	0.64	6.89	0.66
LTV (%)	93.6	7.5	93.1	8.2	94.1	7.2	93.5	7.4
Loan Amount (\$100k)	1.73	0.91	1.68	0.90	1.67	0.88	1.72	0.89
Interest Rate (ppts)	4.93	1.00	5.10	1.05	4.87	0.98	4.92	0.99
Income (\$1k)	65.8	37.5	61.0	33.3	56.2	30.3	67.6	38.5
Borrower Age (years)	38.5	11.9	41.9	12.1	37.8	11.2	38.2	11.9
Refinance (d)	0.294	0.456	0.310	0.462	0.181	0.385	0.312	0.463
Condo (d)	0.115	0.318	0.155	0.362	0.110	0.312	0.106	0.308
2-4 Family (d)	0.014	0.119	0.024	0.154	0.031	0.174	0.010	0.101
Low Documentation (d)	0.190	0.393	0.207	0.405	0.164	0.370	0.192	0.394
Non-Occupant Owner (d)	0.033	0.178	0.033	0.179	0.026	0.158	0.034	0.181
Female (d)	0.353	0.478	0.530	0.499	0.318	0.466	0.333	0.471
Co-applicant (d)	0.414	0.493	0.248	0.432	0.367	0.482	0.445	0.497
# Loans	295,487		31,764		33,717		222,236	
Panel B: Time-Varying Characteristics								
	All		Black		Hispanic White		Non-Hispanic White	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Qtrs since Orig	13.3	10.2	15.0	11.1	13.7	10.3	12.9	10.0
Interest Rate (ppts)	4.93	0.98	5.11	1.01	4.90	0.96	4.92	0.97
Call Option (ppts)	4.77	6.53	5.68	6.68	4.88	6.50	4.64	6.50
SATO (ppts)	0.116	0.346	0.165	0.376	0.158	0.356	0.104	0.338
Equity (%)	-9.25	14.85	-8.84	16.42	-12.38	16.86	-8.70	14.10
Negative Equity (d)	0.117	0.322	0.146	0.353	0.105	0.307	0.115	0.319
Risk Score Change (100s points)	0.016	0.697	-0.109	0.778	0.002	0.727	0.036	0.676
Prepay Refinance (ppts)	1.33	11.47	0.89	9.40	1.03	10.10	1.44	11.93
Prepay Sale (ppts)	0.94	9.67	0.47	6.87	0.62	7.86	1.08	10.33
Default (ppts)	0.89	9.41	1.58	12.47	0.90	9.42	0.81	8.94
# Loan-quarters	6,184,502		765,502		749,691		4,518,876	

Notes: This table reports summary statistics from a 7.5% random sample of FHA loans originated from 2005–2015 (inclusive) from a matched HMDA-McDash-CRISM data set. The unit of observation in Panel A is a loan, while the unit of observation in Panel B is a loan-quarter. The label (d) denotes dummy variables. “SATO” is the spread between the mortgage rate and the average rate associated with newly originated 30-year FRMs according to the FHLMC survey. “Call Option” is a measure of the incentive to refinance taken from Deng et al. (2000).

Table 3: Baseline Prepayment due to Refinance Results

Dependent Variable: Prepay Refinance (d)										
	GSE Loans						FHA Loans			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Black (d)	-0.746*** (0.086)	-0.380*** (0.049)	-0.330*** (0.038)	-0.255*** (0.032)	-0.148*** (0.025)	-0.149*** (0.025)	-0.600*** (0.053)	-0.364*** (0.031)	-0.235*** (0.028)	-0.163*** (0.028)
Hispanic White (d)	-0.687*** (0.118)	-0.454*** (0.066)	-0.462*** (0.064)	-0.421*** (0.057)	-0.278*** (0.038)	-0.289*** (0.036)	-0.401*** (0.076)	-0.384*** (0.042)	-0.400*** (0.047)	-0.315*** (0.041)
Asian (d)	0.436*** (0.143)	0.258*** (0.093)	0.268*** (0.097)	0.256** (0.098)	0.180** (0.068)	0.176** (0.068)	0.417*** (0.088)	-0.020 (0.059)	0.071 (0.073)	0.034 (0.073)
Equifax Risk Score		0.381*** (0.068)						0.449*** (0.060)		
LTV Origination		-0.011*** (0.002)						-0.010*** (0.002)		
Loan Amount		0.567*** (0.059)						0.789*** (0.053)		
LTV Change		-0.001 (0.004)	-0.050*** (0.004)	-0.046*** (0.004)	-0.045*** (0.004)	-0.038*** (0.004)		-0.015*** (0.003)	-0.038*** (0.003)	-0.014*** (0.002)
Refinance (d)		-0.208*** (0.061)	-0.239*** (0.052)	-0.208*** (0.050)	-0.217*** (0.051)	-0.256*** (0.048)		-0.123*** (0.034)	-0.147*** (0.041)	-0.214*** (0.044)
Female (d)			-0.061*** (0.012)	-0.060*** (0.012)	-0.078*** (0.013)	-0.079*** (0.014)			-0.082*** (0.016)	-0.101*** (0.017)
Call Option			0.308*** (0.021)	0.315*** (0.021)	0.316*** (0.021)	0.721*** (0.061)			0.194*** (0.020)	0.595*** (0.081)
SATO			-1.597*** (0.128)	-1.549*** (0.122)	-1.518*** (0.123)	-4.560*** (0.427)			-0.293** (0.122)	-3.302*** (0.509)
Risk Score Change				0.775*** (0.085)	0.764*** (0.084)	0.778*** (0.077)			0.843*** (0.084)	0.836*** (0.083)
Loan Age	X	X	X	X	X	X	X	X	X	X
Underwriting Vars		X	X	X	X	X		X	X	X
HMDA Vars			X	X	X	X			X	X
Vintage Year-Qtr FE	X	X	X	X	X	X	X	X	X	X
State FE		X	X	X				X		
Zip Code FE					X				X	
Zip Code-by-Year-Qtr FE						X				X
# Observations	15,460,588	11,983,398	11,547,035	11,469,141	11,469,141	11,318,445	6,184,502	4,316,733	3,732,349	3,559,947
# Loans	792,823	622,936	601,094	601,028	601,028	590,643	291,587	209,827	182,517	170,234
R ²	0.008	0.012	0.019	0.020	0.022	0.079	0.004	0.012	0.017	0.145

Notes: This table reports LPM estimates of equation (1)—the likelihood of voluntary mortgage prepayment due to refinancing on a set of race/ethnicity indicator variables. The estimation is performed at the quarterly frequency on a 7.5% random sample of loans from a matched HMDA-McDash-CRISM data set. The unit of observation is a loan-quarter. Underwriting variables include the borrower's Equifax Risk Score at origination, LTV at origination, loan amount, change in LTV since origination, indicators for condos and 2–4 multi-family properties, low documentation loans, non-owner occupant properties, and refinance loans. HMDA variables include borrower age (2nd order polynomial), borrower income and indicators for gender and co-applicants. All columns include a 3rd order polynomial for the number of quarters since origination (duration). “SATO” is the spread between the mortgage rate and the average rate associated with newly originated 30-year mortgages according to the FHLMC survey. “Call Option” is a measure of the incentive to refinance taken from Deng et al. (2000). Standard errors are double clustered by county and vintage year-quarter. (***) $p < 0.01$, (**) $p < 0.05$, (*) $p < 0.1$

Table 4: Baseline Prepayment due to Sale Results

Dependent Variable: Prepay Sale (d)										
	GSE Loans						FHA Loans			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Black (d)	-0.524*** (0.019)	-0.424*** (0.018)	-0.385*** (0.017)	-0.383*** (0.019)	-0.346*** (0.022)	-0.346*** (0.023)	-0.644*** (0.030)	-0.554*** (0.036)	-0.449*** (0.035)	-0.425*** (0.033)
Hispanic White (d)	-0.430*** (0.028)	-0.338*** (0.020)	-0.329*** (0.021)	-0.328*** (0.021)	-0.273*** (0.018)	-0.273*** (0.018)	-0.515*** (0.029)	-0.559*** (0.035)	-0.522*** (0.035)	-0.472*** (0.035)
Asian (d)	-0.185*** (0.031)	-0.196*** (0.027)	-0.219*** (0.029)	-0.220*** (0.029)	-0.215*** (0.030)	-0.211*** (0.031)	-0.233*** (0.041)	-0.357*** (0.033)	-0.336*** (0.038)	-0.343*** (0.036)
Equifax Risk Score		0.030* (0.016)						0.121*** (0.011)		
LTV Origination		-0.001 (0.001)						-0.003*** (0.001)		
Loan Amount		0.136*** (0.015)						0.189*** (0.014)		
LTV Change		-0.016*** (0.001)	-0.023*** (0.001)	-0.023*** (0.001)	-0.023*** (0.001)	-0.024*** (0.002)		-0.026*** (0.002)	-0.027*** (0.002)	-0.017*** (0.002)
Refinance (d)		-0.191*** (0.019)	-0.134*** (0.018)	-0.133*** (0.018)	-0.116*** (0.018)	-0.131*** (0.018)		-0.223*** (0.025)	-0.112*** (0.021)	-0.114*** (0.022)
Female (d)			0.026*** (0.008)	0.024*** (0.008)	0.019** (0.009)	0.016 (0.010)			0.023** (0.011)	0.011 (0.012)
Call Option			0.041*** (0.003)	0.042*** (0.003)	0.043*** (0.003)	0.271*** (0.014)			0.010** (0.004)	0.292*** (0.019)
SATO			-0.141*** (0.031)	-0.138*** (0.030)	-0.122*** (0.030)	-1.830*** (0.100)			0.101** (0.047)	-1.996*** (0.128)
Risk Score Change				0.030 (0.033)	0.020 (0.033)	0.031 (0.030)			0.277*** (0.015)	0.260*** (0.015)
Loan Age	X	X	X	X	X	X	X	X	X	X
Underwriting Vars		X	X	X	X	X		X	X	X
HMDA Vars			X	X	X	X			X	X
Vintage Year-Qtr FE	X	X	X	X	X	X	X	X	X	X
State FE		X	X	X				X		
Zip Code FE					X				X	
Zip Code-by-Year-Qtr FE						X				X
# Observations	15,460,588	11,983,398	11,547,035	11,469,141	11,469,141	11,318,445	6,184,502	4,316,733	3,732,349	3,559,947
# Loans	792,823	622,936	601,094	601,028	601,028	590,643	291,587	209,827	182,517	170,234
R ²	0.002	0.003	0.004	0.004	0.006	0.062	0.003	0.005	0.006	0.131

Notes: This table reports LPM estimates of equation (1)—the likelihood of voluntary mortgage prepayment due to home sale on a set of race/ethnicity indicator variables. The estimation is performed at the quarterly frequency on a 7.5% random sample of loans from a matched HMDA-McDash-CRISM data set. The unit of observation is a loan-quarter. Underwriting variables include the borrower's Equifax Risk Score at origination, LTV at origination, loan amount, change in LTV since origination, indicators for condos and 2–4 multi-family properties, low documentation loans, non-owner occupant properties, and refinance loans. HMDA variables include borrower age (2nd order polynomial), borrower income and indicators for gender and co-applicants. All columns include a 3rd order polynomial for the number of quarters since origination (duration). “SATO” is the spread between the mortgage rate and the average rate associated with newly originated 30-year mortgages according to the FHLMC survey. “Call Option” is a measure of the incentive to refinance taken from Deng et al. (2000). Standard errors are double clustered by county and vintage year-quarter. (***) $p < 0.01$, (**) $p < 0.05$, (*) $p < 0.1$

Table 5: Baseline Default Results

Dependent Variable: Default (d)									
	GSE Loans					FHA Loans			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Black (d)	0.443*** (0.076)	0.285*** (0.053)	0.223*** (0.043)	0.146*** (0.030)	0.135*** (0.029)	0.733*** (0.057)	0.466*** (0.038)	0.421*** (0.033)	0.318*** (0.033)
Hispanic White (d)	0.422*** (0.097)	0.274*** (0.066)	0.235*** (0.061)	0.194*** (0.049)	0.188*** (0.049)	0.165*** (0.044)	0.155*** (0.047)	0.085* (0.046)	0.071 (0.044)
Asian (d)	0.026 (0.018)	0.028** (0.012)	0.048*** (0.014)	0.027** (0.012)	0.021* (0.012)	-0.125*** (0.023)	-0.052** (0.024)	-0.041 (0.026)	-0.052 (0.039)
Equifax Risk Score		-0.449*** (0.062)					-0.894*** (0.075)		
LTV Origination		0.010*** (0.001)					0.014*** (0.001)		
Loan Amount		0.046*** (0.012)					0.125*** (0.026)		
LTV Change		0.034*** (0.003)	0.036*** (0.004)	0.037*** (0.004)	0.039*** (0.004)		0.036*** (0.004)	0.037*** (0.004)	0.051*** (0.006)
Refinance (d)		0.129*** (0.018)	0.069*** (0.016)	0.065*** (0.015)	0.060*** (0.015)		0.252*** (0.043)	0.140*** (0.034)	0.135*** (0.034)
Female (d)			-0.017*** (0.005)	-0.016*** (0.005)	-0.015*** (0.005)			-0.027* (0.014)	-0.025 (0.015)
Call Option			-0.014*** (0.003)	-0.013*** (0.003)	0.038*** (0.013)			-0.001 (0.004)	0.323*** (0.026)
SATO			0.485*** (0.091)	0.477*** (0.087)	0.073 (0.120)			0.447*** (0.062)	-2.104*** (0.166)
Loan Age	X	X	X	X	X	X	X	X	X
Underwriting Vars		X	X	X	X		X	X	X
HMDA Vars			X	X	X			X	X
Vintage Year-Qtr FE	X	X	X	X	X	X	X	X	X
State FE		X	X				X		
Zip Code FE				X				X	
Zip Code-by-Year-Qtr FE					X				X
# Observations	14,883,532	11,555,401	11,135,402	11,135,402	10,983,861	5,484,924	3,840,247	3,328,566	3,154,707
# Loans	792,823	622,936	601,094	601,094	590,534	291,587	209,827	182,527	169,608
R ²	0.006	0.012	0.013	0.016	0.084	0.006	0.011	0.012	0.146

Notes: This table reports LPM estimates of equation (1)—the likelihood of mortgage default (defined as 90-day delinquency) on a set of race/ethnicity indicator variables. The estimation is performed at the quarterly frequency on a 7.5% random sample of loans from a matched HMDA-McDash-CRISM data set. The unit of observation is a loan-quarter. Underwriting variables include the borrower's Equifax Risk Score at origination, LTV at origination, loan amount, change in LTV since origination, indicators for condos and 2–4 multi-family properties, low documentation loans, non-owner occupant properties, and refinance loans. HMDA variables include borrower age (2nd order polynomial), borrower income and indicators for gender and co-applicants. All columns include a 3rd order polynomial for the number of quarters since origination (duration). “SATO” is the spread between the mortgage rate and the average rate associated with newly originated 30-year mortgages according to the FHLMC survey. “Call Option” is a measure of the incentive to refinance taken from Deng et al. (2000). Standard errors are double clustered by county and vintage year-quarter. (***) $p < 0.01$, (**) $p < 0.05$, (*) $p < 0.1$

Table 6: Prepayment due to Refinance with Interaction Effects

Dependent Variable: Prepay Refinance (d)								
	GSE Loans				FHA Loans			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Black (d)	-0.178*** (0.027)	0.465*** (0.073)	0.470*** (0.075)	2.665*** (0.468)	-0.158*** (0.025)	0.233** (0.100)	0.227** (0.100)	1.861*** (0.339)
Hispanic White (d)	-0.293*** (0.038)	0.284*** (0.065)	0.295*** (0.070)	2.596*** (0.445)	-0.297*** (0.040)	0.062 (0.081)	0.072 (0.082)	1.714*** (0.352)
Call Option	0.310*** (0.021)	0.320*** (0.022)	0.321*** (0.022)	0.321*** (0.022)	0.193*** (0.020)	0.209*** (0.021)	0.209*** (0.021)	0.211*** (0.022)
Risk Score Change	0.754*** (0.084)	0.742*** (0.083)	0.756*** (0.093)	2.428*** (0.203)	0.836*** (0.083)	0.832*** (0.083)	0.889*** (0.087)	0.456*** (0.108)
Black * Call Option		-0.100*** (0.007)	-0.101*** (0.007)	-0.112*** (0.008)		-0.068*** (0.009)	-0.070*** (0.009)	-0.076*** (0.010)
Hispanic White * Call Option		-0.097*** (0.007)	-0.099*** (0.007)	-0.109*** (0.008)		-0.068*** (0.008)	-0.070*** (0.009)	-0.075*** (0.009)
Black * Risk Score Change			-0.052 (0.055)	-1.312*** (0.197)			-0.207*** (0.026)	-0.452*** (0.131)
Hispanic White * Risk Score Change			-0.097 (0.065)	-1.327*** (0.245)			-0.177*** (0.038)	-0.388** (0.164)
Equifax Risk Score				0.614*** (0.090)				0.758*** (0.087)
Equifax Risk Score * Risk Score Change				-0.352*** (0.031)				-0.052*** (0.018)
Black * Equifax Risk Score * Risk Score Change				0.226*** (0.030)				0.082*** (0.026)
Hispanic White * Equifax Risk Score * Risk Score Change				0.224*** (0.038)				0.072*** (0.025)
Loan Age	X	X	X	X	X	X	X	X
Underwriting Vars	X	X	X	X	X	X	X	X
HMDA Vars	X	X	X	X	X	X	X	X
Vintage Year-Qtr FE	X	X	X	X	X	X	X	X
Zip Code FE	X	X	X	X	X	X	X	X
# Observations	10,816,263	10,816,263	10,816,263	10,816,263	3,636,573	3,636,573	3,636,573	3,636,573
# Loans	563,001	563,001	563,001	563,001	177,437	177,437	177,437	177,437
R ²	0.022	0.022	0.022	0.022	0.023	0.023	0.023	0.023

Notes: This table reports LPM estimates of equation (2). The estimation is performed at the quarterly frequency on a 7.5% random sample of loans from a matched HMDA-McDash-CRISM data set. The unit of observation is a loan-quarter. Underwriting variables include the borrower's Equifax Risk Score at origination, LTV at origination, loan amount, change in LTV since origination, indicators for condos and 2–4 multi-family properties, low documentation loans, non-owner occupant properties, and refinance loans. HMDA variables include borrower age (2nd order polynomial), borrower income and indicators for gender and co-applicants. All columns include a 3rd order polynomial for the number of quarters since origination (duration). “SATO” is the spread between the mortgage rate and the average rate associated with newly originated 30-year mortgages according to the FHLMC survey. “Call Option” is a measure of the incentive to refinance taken from Deng et al. (2000). Standard errors are double clustered by county and vintage year-quarter. (***) p<0.01, ** p<0.05, * p< 0.1)

Table 7: Effect of QE1 on Differences in Refinance Propensities

Dependent Variable: Prepay Refinance (d) Window:	1-Year			2-Year		
	(1)	(2)	(3)	(4)	(5)	(6)
postQE1 (d)	3.221*** (0.496)	3.004*** (0.219)	0.771* (0.414)	1.762*** (0.438)	2.361*** (0.222)	0.970*** (0.283)
Black * postQE1	-2.236*** (0.417)	-2.671*** (0.271)	-1.798*** (0.205)	-1.241*** (0.338)	-1.593*** (0.187)	-1.037*** (0.142)
Hispanic White * postQE1	-2.288*** (0.413)	-2.624*** (0.275)	-2.119*** (0.239)	-1.263*** (0.346)	-1.515*** (0.216)	-1.191*** (0.191)
Black (d)	-0.112** (0.040)	1.056*** (0.113)	0.595*** (0.090)	-0.268*** (0.092)	0.608*** (0.105)	0.305*** (0.094)
Hispanic White (d)	-0.311*** (0.051)	0.914*** (0.093)	0.641*** (0.082)	-0.480*** (0.103)	0.423*** (0.094)	0.241** (0.089)
600 ≤ Equifax Risk Score < 740 (d)		0.511*** (0.062)	-0.185** (0.072)		0.546*** (0.060)	0.074 (0.059)
Equifax Risk Score ≥ 740 (d)		1.643*** (0.122)	-0.243** (0.094)		1.386*** (0.128)	0.186 (0.109)
postQE1 * (600 ≤ Equifax Risk Score < 740)			1.484*** (0.162)			1.033*** (0.119)
postQE1 * (Equifax Risk Score ≥ 740)			3.754*** (0.331)			2.377*** (0.227)
Constant	0.615*** (0.066)	-3.886*** (1.275)	-3.312** (1.271)	1.095*** (0.177)	0.256 (0.261)	0.912** (0.343)
Loan Age		X	X		X	X
Underwriting Vars		X	X		X	X
HMDA Vars		X	X		X	X
Vintage Year-Qtr FE		X	X		X	X
Zip Code FE		X	X		X	X
# Observations	1,066,525	782,523	782,523	2,129,912	1,563,213	1,563,213
R ²	0.012	0.055	0.058	0.004	0.038	0.039

Notes: This table reports LPM estimates of equation (3). The estimation is performed at the quarterly frequency on a 7.5% random sample of GSE 30-year FRMs from a matched HMDA-McDash-CRISM data set. The unit of observation is a loan-quarter. Underwriting variables include the borrower's Equifax Risk Score at origination, LTV at origination, loan amount, change in LTV since origination, indicators for condos and 2–4 multi-family properties, low documentation loans, non-owner occupant properties, and refinance loans. HMDA variables include borrower age (2nd order polynomial), borrower income and indicators for gender and co-applicants. All columns include a 3rd order polynomial for the number of quarters since origination (duration). “QE1” is an indicator variable that takes a value of 1 for year-quarters after 2008:Q4. Standard errors are double clustered by county and vintage year-quarter. (***) p<0.01, ** p<0.05, * p<0.1)

Table 8: Effect of QE1 on Differences in the Stock of Outstanding Mortgage Rates

Dependent Variable: Mortgage Rate						
Window:	1-Year		2-Years		4-Years	
	(1)	(2)	(3)	(4)	(5)	(6)
Black (d)	0.224*** (0.014)	0.189*** (0.012)	0.222*** (0.014)	0.187*** (0.012)	0.210*** (0.014)	0.176*** (0.012)
Hispanic White (d)	0.135*** (0.019)	0.112*** (0.016)	0.132*** (0.019)	0.109*** (0.016)	0.123*** (0.019)	0.101*** (0.016)
postQE1 (d)	-0.209*** (0.004)	-0.006*** (0.000)	-0.313*** (0.005)	-0.004*** (0.000)	-0.463*** (0.007)	-0.008*** (0.001)
Black * postQE1	0.115*** (0.004)	-0.005*** (0.002)	0.162*** (0.005)	-0.009*** (0.002)	0.226*** (0.007)	-0.007* (0.003)
Hispanic White * postQE1	0.111*** (0.004)	0.004** (0.002)	0.157*** (0.006)	0.003 (0.002)	0.214*** (0.010)	0.006* (0.004)
Constant	6.239*** (0.004)	6.142*** (0.004)	6.245*** (0.004)	6.090*** (0.004)	6.252*** (0.004)	6.000*** (0.004)
Vintage Year-Qtr FE	X		X		X	
# Observations	1,066,525	1,066,525	2,129,912	2,129,912	4,085,825	4,085,825
R ²	0.044	0.528	0.070	0.588	0.114	0.660

Notes: This table reports LPM estimates of equation (4). The estimation is performed at the quarterly frequency on a 7.5% random sample of GSE 30-year FRMs from a matched HMDA-McDash-CRISM data set. The unit of observation is a loan-quarter. “QE1” is an indicator variable that takes a value of 1 for year-quarters after 2008:Q4. Standard errors are double clustered by county and vintage year-quarter. (***) $p < 0.01$, (**) $p < 0.05$, * $p < 0.1$)

Mortgage Prepayment, Race, and Monetary Policy

Appendix

This appendix supplements the empirical analysis in Gerardi, Willen, and Zhang (2020). Below is a list of the sections contained in this appendix.

Table of Contents

A.1	HMDA-McDash and HMDA-McDash-CRISM Match Rates	2
A.1.1	HMDA-McDash Database	2
A.1.2	CRISM Database	6
A.2	Sample Restrictions	9
A.3	LPM Estimates for All Covariates	10
A.4	Logit Models	18
A.5	Involuntary Prepayments	22
A.6	Agarwal, Driscoll, and Laibson Closed-Form Refinance Rule	24
A.7	Evidence from Survey of Consumer Finances	27
A.8	Secondary Market Pricing Estimates for All Covariates	29
A.9	Additional Tables and Figures	31

A.1 HMDA-McDash and HMDA-McDash-CRISM Match Rates

As we discussed in section 2, our analysis employs a novel data set that combines three sources of administrative data: Home Mortgage Disclosure Act (HMDA) data, Black Knight McDash mortgage servicing data, and credit bureau data from Equifax. The three data sources are linked together through two separate loan-level matches: a match between the HMDA and McDash databases, which we will refer to as the HMDA-McDash dataset, and a match between the McDash and Equifax databases, which is referred to as CRISM (Equifax Credit Risk Insight Servicing McDash Database). We are then able to merge the two matched data sets, creating a final data set with information from all three sources, which we will refer to as the HMDA-McDash-CRISM data set. Below we will discuss some of the details of both matches and show match rates by loan vintage (year) to provide information on the quality and scope of the final data set used in the analysis.

A.1.1 HMDA-McDash Database

The HMDA-McDash matched data set is available to users within the Federal Reserve System and includes more than 93 million loans originated from 1992 through 2015 (inclusive). The matching algorithm was written by the Risk Assessment, Data Analysis and Research (RADAR) group at the Federal Reserve Bank of Philadelphia and matches HMDA and McDash loans by the origination date, origination amount, property Zip code, lien type, loan purpose (that is, purchase or refinance), loan type (for example, conventional or FHA), and occupancy type. Tables A.1 and A.1 display match rates by origination year; the former table calculates rates by dividing by the number of McDash loans, while the latter table divides by the total number of HMDA loans. Overall, approximately two-thirds of McDash loans are successfully matched to HMDA, while almost 40 percent of HMDA loans are successfully matched to loans in McDash. Since the HMDA database covers a greater fraction of the mortgage market, the match rates normalized by HMDA loans are significantly lower than the rates normalized by McDash loans.

Our sample includes only loans originated in 2005 and later due to lower coverage in the pre-2005 McDash database. In 2005 McDash added a large servicer to its database, which substantially increased the overall coverage of the database. The last column in Table A.1 shows that the coverage (relative to the total number of HMDA loan originations) goes from 65 percent in 2004 to 81 percent in 2006. When servicers are added to the McDash database, they typically provide information on only their active loans. This raises concerns of attrition bias, and thus we focus only on loans originated in 2005 and later.

The matching algorithm is based on the following logic:

- Origination date (McDash) and action date (HMDA) must be within five days of each other.
- Origination amounts must be within \$500.

- Property Zip codes must match.
- Lien types must match.
- Loan purposes (purchase, refinance) must match.
- Loan types (conventional, jumbo, etc.) must match.
- Occupancy types must match.

In our analysis, we use only loans that were uniquely matched. The last column in Table A.2 shows that during our sample period (2005 through 2015) our sample covers from 34 percent to 47 percent of all loan originations in HMDA.

Table A.1: Match Rate by Origination Year (Matched McDash Mortgages/All McDash Mortgages)

Origination Year	McDash Loans Matched	Only 1 HMDA Candidate	McDash Loans Uniquely Matched	McDash Coverage
1992	51%	48%	20%	58%
1993	55%	50%	19%	70%
1994	58%	53%	24%	52%
1995	61%	57%	29%	46%
1996	63%	58%	33%	42%
1997	62%	58%	35%	39%
1998	65%	60%	36%	52%
1999	65%	60%	35%	46%
2000	64%	61%	50%	31%
2001	64%	60%	49%	44%
2002	65%	59%	50%	50%
2003	71%	64%	53%	67%
2004	69%	64%	55%	65%
2005	67%	61%	51%	73%
2006	63%	59%	49%	81%
2007	63%	59%	50%	87%
2008	65%	62%	54%	79%
2009	67%	64%	59%	79%
2010	69%	67%	61%	77%
2011	69%	67%	61%	73%
2012	73%	71%	64%	67%
2013	75%	74%	67%	62%
2014	77%	76%	71%	48%
2015	79%	78%	75%	45%
Total	66%	62%	49%	61%

Notes: Match rates are calculated by the Risk Assessment, Data Analysis and Research (RADAR) group. McDash coverage is estimated by dividing the number of originations in the McDash database by the number of originations in HMDA.

Table A.2: Match Rate by Origination Year (Matched HMDA Mortgages/All HMDA Mortgages)

Origination Year	HMDA Loans Matched	Only 1 McDash Candidate	HMDA Loans Uniquely Matched
1992	21%	14%	12%
1993	27%	16%	13%
1994	22%	15%	12%
1995	22%	15%	13%
1996	21%	16%	14%
1997	21%	16%	14%
1998	30%	23%	19%
1999	25%	19%	16%
2000	19%	17%	16%
2001	27%	24%	22%
2002	33%	30%	25%
2003	48%	43%	36%
2004	45%	41%	36%
2005	48%	43%	37%
2006	50%	45%	40%
2007	53%	48%	43%
2008	49%	46%	43%
2009	53%	50%	47%
2010	53%	50%	47%
2011	49%	47%	45%
2012	47%	45%	42%
2013	46%	44%	42%
2014	37%	35%	35%
2015	36%	35%	34%
Total	38%	34%	30%

Notes: Match rates are calculated by the Risk Assessment, Data Analysis and Research (RADAR) group.

A.1.2 CRISM Database

CRISM is a data set that consists of McDash mortgages matched to credit bureau data from Equifax at the borrower level. The Equifax credit bureau data are updated at a monthly frequency and include information on outstanding consumer loans and credit lines for the primary borrower as well as all co-borrowers associated with the McDash mortgage. The matching process was conducted by Equifax using confidential and proprietary data. The exact matching algorithm is proprietary, but according to Equifax, anonymous fields such as the original and current mortgage balance, date of origination, ZIP code, and monthly payment history are all used in the algorithm.

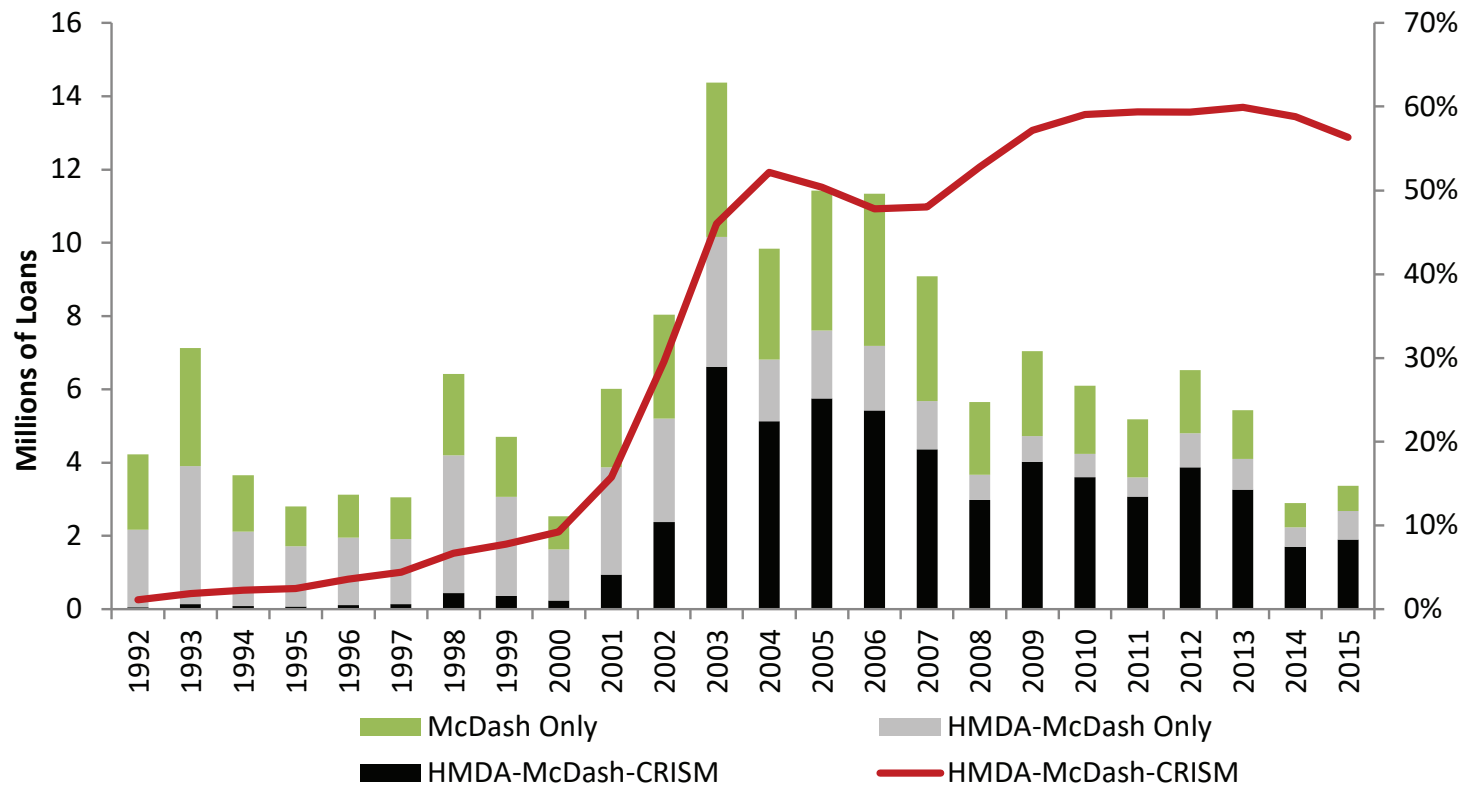
CRISM coverage begins in June 2005, and according to Equifax, approximately 90 percent of McDash mortgages were matched to a credit bureau account with high confidence.¹ We keep only observations that pertain to the primary mortgage borrower to avoid double counting. Borrower credit information is included in the data set for the life of each loan as well as for the six months preceding origination and the six months following termination.

Figure A.1 displays the match rate by vintage for the HMDA-McDash-CRISM matched data set as a ratio of the total number of McDash originations (solid red line). For 2005–2015 originations, the match rate is between 50 percent and 60 percent. The figure also shows the total number of mortgage originations for the McDash data set, the HMDA-McDash matched data set, and the HMDA-McDash-CRISM matched data set. The largest decline in the sample occurs when the McDash database is matched to HMDA. The addition of CRISM data results in only a small decline in loan originations during our sample period.

Finally, in Table A.3 we compare summary statistics for the HMDA-McDash and HMDA-McDash-CRISM GSE (Panel A) and FHA (Panel B) samples, respectively. The tables show that the summary statistics are almost identical across the two samples, which suggests that the addition of the Equifax credit bureau data does not significantly alter the composition of mortgages.

¹Equifax provides a “Match Confidence Score” that is based on a scale of 0 to 0.9, where a higher score indicates that the McDash and Equifax data align better on the matching fields. Approximately 90 percent of McDash loans have a match confidence score of 0.8 or higher. Equifax recommends using 0.8 as a threshold for modeling purposes, and we follow this advice, keeping only matches with scores above 0.8.

Figure A.1: Loans in the HMDA-McDash-CRISM Match, HMDA-CRISM Match, and McDash Data Sets by Vintage



Notes: This figure shows the number of loans in the McDash data set, the HMDA-McDash data set, and the HMDA-McDash-CRISM data set by vintage (bars). In addition, the solid red line shows the match rate for the HMDA-McDash-CRISM data set, calculated as a percentage of the number of loans in the McDash database by vintage. This figure was created by the Risk Assessment, Data Analysis and Research (RADAR) group, which conducted the matching exercise.

Table A.3: Comparison of Summary Statistics: HMDA-McDash vs. HMDA-McDash-CRISM Databases

Panel A: GSE Loans				
	HMDA-McDash-Equifax		HMDA-McDash	
	Mean	Std. Dev.	Mean	Std. Dev.
FICO Origination (100s points)	7.44	0.54	7.45	0.53
LTV (%)	72.6	15.9	72.7	15.9
Loan Amount (\$100k)	2.12	1.13	2.12	1.13
Interest Rate (ppts)	5.20	1.02	5.20	1.02
Income (\$1k)	97.6	64.0	97.5	63.9
Refinance (d)	0.538	0.499	0.539	0.498
Condo (d)	0.140	0.347	0.139	0.346
2-4 Family (d)	0.018	0.133	0.018	0.133
Low Documentation (d)	0.308	0.462	0.309	0.462
Non-Occupant Owner (d)	0.140	0.347	0.140	0.347
Female (d)	0.294	0.455	0.294	0.456
Co-applicant (d)	0.505	0.500	0.503	0.500
# Loans	800,806		1,076,117	

Panel B: FHA Loans				
	HMDA-McDash-Equifax		HMDA-McDash	
	Mean	Std. Dev.	Mean	Std. Dev.
FICO Origination (100s points)	6.85	0.60	6.88	0.59
LTV (%)	93.6	7.5	93.6	7.4
Loan Amount (\$100k)	1.73	0.91	1.73	0.91
Interest Rate (ppts)	4.93	1.00	4.93	1.00
Income (\$1k)	65.8	37.5	65.8	37.3
Refinance (d)	0.294	0.456	0.295	0.456
Condo (d)	0.115	0.318	0.114	0.317
2-4 Family (d)	0.014	0.119	0.015	0.120
Low Documentation (d)	0.190	0.393	0.191	0.393
Non-Occupant Owner (d)	0.033	0.178	0.033	0.178
Female (d)	0.353	0.478	0.352	0.478
Co-applicant (d)	0.414	0.493	0.415	0.493
# Loans	295,487		397,686	

Notes: This table reports summary statistics from a 7.5% random sample of GSE loans originated between 2005 and 2015 (inclusive) from a matched HMDA-McDash-CRISM data set and a 10% random sample of GSE and FHA loans originated between 2005 and 2015 (inclusive) from a matched HMDA-McDash data set. The label (d) denotes dummy variables.

A.2 Sample Restrictions

Table A.4 below displays all of the restrictions that we impose in constructing our 7.5 percent random sample of the HMDA-McDash-CRISM data set. We adopt most of the restrictions implemented in Fuster et al. (2018). We implement most of our restrictions while querying the database (and thus, we do not know how many loans are lost as a result of those restrictions).² For the restrictions that we implement while applying code to clean and create our variables, we display the number of loans that are dropped.

Table A.4: Sample Restrictions

Sample Restriction:	# Loans Lost	# Loans Remaining
Originations between 01/2005 and 12/2015		
Loans with “conf” ≥ 0.80		
Fixed Rate Loans		
First Liens		
Fully Amortizing Loans No Prepayment Penalty		
$20 \leq \text{LTV} \leq 100$		
Occupancy Non-missing		
Loan Amount $\leq \$1\text{m}$		
Income $\leq \$500\text{k}$		
Term = 30 years		
No Home Improvement Loans		1,681,252
Seasoning ≤ 6 Months	193,898	1,487,354
Black, Hispanic White, Asian, and White Borrowers	208,817	1,278,537
GSE and FHA Loans	179,810	1,098,727
$3\% \leq \text{Mortgage Rate} \leq 8\%$	2,434	1,096,293

²Because the HMDA-McDash-CRISM database is a monthly panel and extremely large, we were unable to download more than a 7.5 percent sample given computing constraints.

A.3 LPM Estimates for All Covariates

In Table A.5 below we display the full set of regression estimates from the specifications estimated in Table 3. The column numbers correspond to identical specifications across the two tables.

Table A.5: Baseline Refinance Results with All Covariates

Dependent Variable: Prepay Refinance (d)							
	GSE Loans					FHA Loans	
	(2)	(3)	(4)	(5)	(6)	(9)	(10)
Black (d)	-0.380*** (0.049)	-0.330*** (0.038)	-0.255*** (0.032)	-0.148*** (0.025)	-0.149*** (0.025)	-0.235*** (0.028)	-0.163*** (0.028)
White Hispanic (d)	-0.454*** (0.066)	-0.462*** (0.064)	-0.421*** (0.057)	-0.278*** (0.038)	-0.289*** (0.036)	-0.400*** (0.047)	-0.315*** (0.041)
Asian (d)	0.258*** (0.093)	0.268*** (0.097)	0.256** (0.098)	0.180** (0.068)	0.176** (0.068)	0.071 (0.073)	0.034 (0.073)
Qtrs since Orig	0.340*** (0.053)	0.130*** (0.021)	0.119*** (0.021)	0.123*** (0.021)		0.114*** (0.018)	
Qtrs since Orig $\hat{2}$	-0.014*** (0.003)	-0.008*** (0.001)	-0.007*** (0.001)	-0.007*** (0.001)		-0.007*** (0.001)	
Qtrs since Orig $\hat{3}$	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)		0.000*** (0.000)	
Call Option		0.308*** (0.021)	0.315*** (0.021)	0.316*** (0.021)	0.721*** (0.061)	0.194*** (0.020)	0.595*** (0.081)
SATO		-1.597*** (0.128)	-1.549*** (0.122)	-1.518*** (0.123)	-4.560*** (0.427)	-0.293** (0.122)	-3.302*** (0.509)
Refinance (d)	-0.208*** (0.061)	-0.239*** (0.052)	-0.208*** (0.050)	-0.217*** (0.051)	-0.256*** (0.048)	-0.147*** (0.041)	-0.214*** (0.044)
Condo (d)	-0.438*** (0.050)	-0.462*** (0.050)	-0.489*** (0.053)	-0.571*** (0.061)	-0.585*** (0.062)	-0.373*** (0.040)	-0.528*** (0.053)
2-4 Family (d)	-0.919*** (0.090)	-0.992*** (0.083)	-1.009*** (0.083)	-0.959*** (0.076)	-0.951*** (0.077)	-0.518*** (0.087)	-0.445*** (0.086)
Prop Type Missing (d)	0.247*** (0.060)	0.219*** (0.060)	0.216*** (0.060)	0.209*** (0.058)	0.238*** (0.061)	0.258*** (0.058)	0.332*** (0.068)

Continued on next page

Table A.5 – continued from previous page

Dependent Variable: Prepay Refinance (d)							
	GSE Loans					FHA Loans	
	(2)	(3)	(4)	(5)	(6)	(9)	(10)
Low Documentation (d)	0.162*** (0.056)	0.128** (0.053)	0.136** (0.053)	0.142*** (0.052)	0.133** (0.054)	-0.099* (0.058)	-0.121** (0.054)
Documentation Missing (d)	0.853*** (0.102)	0.830*** (0.098)	0.819*** (0.098)	0.822*** (0.095)	0.762*** (0.092)	0.849*** (0.129)	0.787*** (0.132)
Non-Occupant Owner (d)	-0.469*** (0.052)	-0.742*** (0.064)	-0.765*** (0.065)	-0.682*** (0.058)	-0.676*** (0.056)	4.031*** (0.729)	4.450*** (0.704)
Risk Score	0.381*** (0.068)						
LTV	-0.011*** (0.002)						
LTV = 80 (d)	0.567*** (0.059)						
Loan Amount	-0.019 (0.022)						
LTV Change	-0.001 (0.004)	-0.050*** (0.004)	-0.046*** (0.004)	-0.045*** (0.004)	-0.038*** (0.004)	-0.038*** (0.003)	-0.014*** (0.002)
Risk Score Change			0.775*** (0.085)	0.764*** (0.084)	0.778*** (0.077)	0.843*** (0.084)	0.836*** (0.083)
600 < Risk Score ≤ 620 (d)		0.256*** (0.052)	0.412*** (0.057)	0.430*** (0.060)	0.384*** (0.053)	0.519*** (0.054)	0.497*** (0.050)
620 < Risk Score ≤ 640 (d)		0.347*** (0.060)	0.526*** (0.065)	0.535*** (0.062)	0.481*** (0.053)	0.687*** (0.083)	0.666*** (0.071)
640 < Risk Score ≤ 660 (d)		0.483*** (0.066)	0.681*** (0.073)	0.698*** (0.070)	0.639*** (0.055)	0.856*** (0.109)	0.829*** (0.090)
660 < Risk Score ≤ 680 (d)		0.618*** (0.085)	0.832*** (0.093)	0.855*** (0.094)	0.797*** (0.077)	0.941*** (0.109)	0.914*** (0.096)

Continued on next page

Table A.5 – continued from previous page

Dependent Variable: Prepay Refinance (d)							
	GSE Loans					FHA Loans	
	(2)	(3)	(4)	(5)	(6)	(9)	(10)
680 < Risk Score \leq 700 (d)		0.703*** (0.091)	0.934*** (0.102)	0.949*** (0.102)	0.897*** (0.082)	1.036*** (0.131)	1.004*** (0.116)
700 < Risk Score \leq 720 (d)		0.856*** (0.100)	1.100*** (0.112)	1.115*** (0.113)	1.066*** (0.090)	1.186*** (0.144)	1.155*** (0.134)
720 < Risk Score \leq 740 (d)		0.983*** (0.111)	1.254*** (0.125)	1.268*** (0.125)	1.225*** (0.103)	1.310*** (0.154)	1.275*** (0.139)
740 < Risk Score \leq 760 (d)		1.128*** (0.122)	1.421*** (0.139)	1.431*** (0.139)	1.391*** (0.116)	1.465*** (0.152)	1.421*** (0.139)
760 < Risk Score \leq 780 (d)		1.256*** (0.134)	1.582*** (0.155)	1.584*** (0.154)	1.551*** (0.134)	1.554*** (0.161)	1.526*** (0.149)
780 < Risk Score \leq 800 (d)		1.304*** (0.146)	1.677*** (0.171)	1.674*** (0.170)	1.643*** (0.152)	1.633*** (0.172)	1.577*** (0.160)
800 < Risk Score \leq 820 (d)		1.297*** (0.146)	1.740*** (0.176)	1.730*** (0.176)	1.705*** (0.159)	1.657*** (0.176)	1.610*** (0.168)
Risk Score > 820 (d)		1.294*** (0.152)	1.783*** (0.186)	1.732*** (0.181)	1.703*** (0.167)	1.760*** (0.195)	1.620*** (0.192)
25 < LTV \leq 30 (d)		-0.039 (0.065)	-0.019 (0.069)	0.049 (0.069)	0.029 (0.076)	0.875 (0.801)	0.429 (0.860)
30 < LTV \leq 35 (d)		0.006 (0.066)	0.027 (0.067)	0.114 (0.072)	0.094 (0.077)	0.765 (0.623)	0.098 (0.638)
35 < LTV \leq 40 (d)		0.011 (0.050)	0.031 (0.053)	0.166*** (0.061)	0.142** (0.063)	0.476 (0.526)	0.329 (0.654)
40 < LTV \leq 45 (d)		-0.053 (0.057)	-0.020 (0.059)	0.154** (0.071)	0.120 (0.077)	0.280 (0.558)	0.108 (0.624)
45 < LTV \leq 50 (d)		-0.075 (0.056)	-0.037 (0.058)	0.173** (0.075)	0.139* (0.072)	0.840 (0.562)	0.655 (0.651)

Continued on next page

Table A.5 – continued from previous page

Dependent Variable: Prepay Refinance (d)		GSE Loans				FHA Loans		
		(2)	(3)	(4)	(5)	(6)	(9)	(10)
50 < LTV ≤ 55 (d)			-0.159*** (0.050)	-0.115** (0.052)	0.106 (0.070)	0.082 (0.066)	0.701 (0.573)	0.471 (0.709)
55 < LTV ≤ 60 (d)			-0.213*** (0.056)	-0.160*** (0.056)	0.081 (0.073)	0.056 (0.072)	0.554 (0.555)	0.316 (0.648)
60 < LTV ≤ 65 (d)			-0.314*** (0.061)	-0.256*** (0.061)	0.009 (0.079)	-0.023 (0.077)	0.506 (0.536)	0.292 (0.619)
65 < LTV ≤ 70 (d)			-0.411*** (0.063)	-0.347*** (0.063)	-0.071 (0.080)	-0.099 (0.078)	0.391 (0.517)	0.305 (0.626)
70 < LTV ≤ 75 (d)			-0.574*** (0.067)	-0.503*** (0.065)	-0.203** (0.086)	-0.229*** (0.081)	0.388 (0.509)	0.338 (0.605)
75 < LTV ≤ 80 (d)			-0.669*** (0.069)	-0.592*** (0.066)	-0.292*** (0.084)	-0.319*** (0.078)	0.319 (0.510)	0.279 (0.603)
80 < LTV ≤ 85 (d)			-0.874*** (0.082)	-0.790*** (0.078)	-0.480*** (0.091)	-0.501*** (0.087)	0.176 (0.525)	0.142 (0.622)
85 < LTV ≤ 90 (d)			-1.116*** (0.112)	-1.022*** (0.106)	-0.693*** (0.111)	-0.717*** (0.106)	0.104 (0.518)	0.088 (0.613)
90 < LTV ≤ 95 (d)			-1.253*** (0.124)	-1.143*** (0.115)	-0.813*** (0.117)	-0.833*** (0.111)	0.075 (0.525)	0.080 (0.617)
95 < LTV ≤ 100 (d)			-1.383*** (0.124)	-1.263*** (0.114)	-0.911*** (0.118)	-0.933*** (0.110)	-0.016 (0.523)	0.002 (0.622)
85k < Orig Amount ≤ 110k (d)			0.418*** (0.056)	0.417*** (0.056)	0.378*** (0.057)	0.373*** (0.052)	0.350*** (0.049)	0.252*** (0.042)
110k < Orig Amount ≤ 125k (d)			0.593*** (0.075)	0.595*** (0.076)	0.537*** (0.077)	0.529*** (0.072)	0.573*** (0.069)	0.456*** (0.061)
125k < Orig Amount ≤ 150k (d)			0.793*** (0.093)	0.794*** (0.093)	0.734*** (0.095)	0.731*** (0.087)	0.763*** (0.091)	0.621*** (0.077)

Continued on next page

Table A.5 – continued from previous page

Dependent Variable: Prepay Refinance (d)							
	GSE Loans					FHA Loans	
	(2)	(3)	(4)	(5)	(6)	(9)	(10)
150k < Orig Amount ≤ 175k (d)		0.953*** (0.112)	0.954*** (0.114)	0.878*** (0.116)	0.877*** (0.107)	1.054*** (0.110)	0.867*** (0.093)
Orig Amount > 175k (d)		1.493*** (0.148)	1.496*** (0.151)	1.350*** (0.154)	1.356*** (0.142)	1.710*** (0.152)	1.382*** (0.124)
25k < Income ≤ 50k (d)		0.057* (0.029)	0.051 (0.031)	0.032 (0.030)	0.049* (0.028)	0.130*** (0.039)	0.103*** (0.035)
50k < Income ≤ 75k (d)		0.084** (0.036)	0.064* (0.037)	0.037 (0.035)	0.060* (0.035)	0.130*** (0.047)	0.120*** (0.036)
75k < Income ≤ 100k (d)		0.164*** (0.049)	0.137** (0.052)	0.086* (0.047)	0.112** (0.048)	0.262*** (0.051)	0.233*** (0.044)
100k < Income ≤ 125k (d)		0.262*** (0.060)	0.227*** (0.062)	0.142** (0.054)	0.171*** (0.053)	0.374*** (0.064)	0.321*** (0.051)
125k < Income ≤ 150k (d)		0.320*** (0.077)	0.280*** (0.079)	0.168** (0.067)	0.200*** (0.066)	0.433*** (0.082)	0.287*** (0.065)
150k < Income ≤ 175k (d)		0.371*** (0.094)	0.331*** (0.096)	0.207** (0.080)	0.220*** (0.077)	0.673*** (0.101)	0.440*** (0.085)
175k < Income ≤ 200k (d)		0.336*** (0.101)	0.301*** (0.103)	0.160* (0.085)	0.182** (0.082)	0.616*** (0.174)	0.268 (0.179)
200k < Income ≤ 225k (d)		0.354*** (0.113)	0.310** (0.115)	0.158 (0.095)	0.169* (0.091)	0.851*** (0.136)	0.526*** (0.168)
225k < Income ≤ 250k (d)		0.246** (0.122)	0.206 (0.124)	0.051 (0.103)	0.072 (0.099)	0.694** (0.283)	0.376 (0.290)
250k < Income ≤ 275k (d)		0.281* (0.142)	0.246* (0.144)	0.089 (0.117)	0.094 (0.116)	0.556** (0.264)	0.249 (0.256)
275k < Income ≤ 300k (d)		0.228 (0.137)	0.192 (0.137)	0.043 (0.114)	0.049 (0.110)	0.205 (0.271)	-0.349 (0.331)

Continued on next page

Table A.5 – continued from previous page

Dependent Variable: Prepay Refinance (d)		GSE Loans				FHA Loans	
	(2)	(3)	(4)	(5)	(6)	(9)	(10)
300k < Income ≤ 325k (d)		0.140 (0.148)	0.125 (0.145)	-0.029 (0.127)	-0.033 (0.123)	0.323 (0.487)	-0.461 (0.545)
325k < Income ≤ 350k (d)		0.141 (0.146)	0.119 (0.145)	0.000 (0.125)	-0.005 (0.123)	0.960* (0.483)	0.458 (0.654)
350k < Income ≤ 375k (d)		0.086 (0.163)	0.053 (0.166)	-0.069 (0.154)	-0.063 (0.160)	0.830 (0.584)	0.421 (0.764)
375k < Income ≤ 400k (d)		0.106 (0.182)	0.067 (0.187)	-0.023 (0.178)	-0.045 (0.173)	-0.048 (0.752)	-1.031 (0.954)
400k < Income ≤ 425k (d)		-0.074 (0.172)	-0.111 (0.171)	-0.189 (0.163)	-0.220 (0.164)	-0.145 (0.608)	0.014 (0.663)
425k < Income ≤ 450k (d)		0.079 (0.196)	0.034 (0.196)	-0.129 (0.192)	-0.067 (0.192)	0.762 (1.015)	0.575 (1.050)
450k < Income ≤ 475k (d)		-0.155 (0.192)	-0.193 (0.194)	-0.312* (0.186)	-0.298 (0.184)	0.578 (0.741)	1.094 (0.918)
475k < Income ≤ 500k (d)		-0.055 (0.167)	-0.089 (0.167)	-0.203 (0.182)	-0.174 (0.185)	0.115 (1.210)	-0.515 (1.333)
Female (d)		-0.061*** (0.012)	-0.060*** (0.012)	-0.078*** (0.013)	-0.079*** (0.014)	-0.082*** (0.016)	-0.101*** (0.017)
Co-applicant (d)		0.171*** (0.033)	0.150*** (0.028)	0.161*** (0.026)	0.161*** (0.027)	-0.103*** (0.018)	-0.080*** (0.019)
Borrower Age		-0.007 (0.004)	-0.007* (0.004)	-0.004 (0.004)	0.000 (0.003)	0.021*** (0.004)	0.021*** (0.004)
Borrower Age $\hat{2}$		-0.000 (0.000)	-0.000 (0.000)	-0.000* (0.000)	-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)
Constant	-2.967***	-0.846***	-1.265***	-1.556***	-2.865***	-2.487***	-3.746***

Continued on next page

Table A.5 – continued from previous page

Dependent Variable: Prepay Refinance (d)							
	GSE Loans					FHA Loans	
	(2)	(3)	(4)	(5)	(6)	(9)	(10)
	(0.592)	(0.221)	(0.247)	(0.248)	(0.379)	(0.608)	(0.949)
Underwriting Vars	X	X	X	X	X	X	X
HMDA Vars		X	X	X		X	
Vintage Year-Qtr FE	X	X	X	X	X	X	X
State FE	X	X	X				
Zip Code FE				X		X	
Zip Code-by-Year-Qtr FE					X		X
# Observations	11,983,398	11,547,035	11,469,141	11,469,141	11,318,445	3,732,349	3,559,947
# Loans							
R ²	0.012	0.018	0.019	0.022	0.077	0.017	0.145

Notes: This table reports LPM estimates of equation (1)—the likelihood of voluntary mortgage prepayment due to refinancing on a set of race/ethnicity indicator variables. The estimation is performed at the quarterly frequency on a 7.5% random sample of loans from a matched HMDA-McDash-Equifax data set. The unit of observation is a loan-quarter. Underwriting variables include the borrower's risk score at origination, LTV at origination, loan amount, change in LTV since origination, indicators for condos and 2–4 multi-family properties, low-documentation loans, non-owner occupant properties, and refinance loans. HMDA variables include borrower age (2nd order polynomial), borrower income, and indicators for gender and co-applicants. All columns include a 3rd order polynomial for the number of quarters since origination (duration). "SATO" is the spread between the mortgage rate and the average rate associated with newly originated 30-year mortgages according to the FHLMC survey. "Refi Money" is a measure of the incentive to refinance taken from Deng et al. (2000). Standard errors are double clustered by county and vintage year-quarter. (***) $p < 0.01$, (**) $p < 0.05$, (*) $p < 0.1$

A.4 Logit Models

In this section we present prepayment due to refinance and home sale results as well as default results from logit models. These models are estimated on a 7.5 percent random sample of our HMDA-McDash-Equifax matched data set. Table A.6 contains the refinance results, Table A.7 contains the home sale results, and Table A.8 displays the default results. Both tables show the estimated average marginal effects associated with the racial/ethnic indicator variables. The covariates and fixed effects in each column correspond exactly to their counterparts in Tables 3, 4, and 5 in the main text. The omitted specifications are those with Zip code and Zip- code-by-year-quarter fixed effects. It was not possible to estimate those specifications using the logit framework.

Table A.6: Logit Prepayment due to Refinance Hazard Estimates

Dependent Variable: Prepay Refinance (d)						
	GSE Loans				FHA Loans	
	(1)	(2)	(3)	(4)	(7)	(8)
Black (d)	-0.686*** (0.033)	-0.421*** (0.030)	-0.353*** (0.029)	-0.282*** (0.030)	-0.585*** (0.037)	-0.419*** (0.024)
Hispanic (d)	-0.654*** (0.057)	-0.475*** (0.024)	-0.489*** (0.030)	-0.449*** (0.031)	-0.405*** (0.059)	-0.389*** (0.028)
Asian (d)	0.466*** (0.132)	0.275*** (0.070)	0.259*** (0.070)	0.247*** (0.070)	0.455*** (0.088)	-0.030 (0.043)
Loan Age	X	X	X	X	X	X
Underwriting Vars		X	X	X		X
HMDA Vars			X	X		X
Vintage Year-Qtr FE	X	X	X	X	X	X
State FE		X	X	X		X
# Observations	15,460,588	11,983,398	11,547,035	11,469,141	6,184,502	4,316,733

Notes: This table reports estimated marginal effects estimates from a logit model of equation (1)—the likelihood of voluntary mortgage prepayment due to refinancing on a set of race/ethnicity indicator variables. The estimation is performed at the quarterly frequency on a 5% random sample of loans from a matched HMDA-McDash-Equifax data set. The unit of observation is a loan-quarter. Underwriting variables include the borrower's risk score at origination, LTV at origination, loan amount, change in LTV since origination, indicators for condos and 2–4 multi-family properties, low-documentation loans, non-owner occupant properties, and refinance loans. HMDA variables include borrower age (2nd order polynomial), borrower income, and indicators for gender and co-applicants. All columns include a 3rd order polynomial for the number of quarters since origination (duration). Standard errors are clustered by county. (***) $p < 0.01$, (**) $p < 0.05$, (*) $p < 0.1$

Table A.7: Logit Prepayment due to Sale Hazard Estimates

Dependent Variable: Prepay Sale (d)						
	GSE Loans				FHA Loans	
	(1)	(2)	(3)	(4)	(7)	(8)
Black (d)	-0.505*** (0.014)	-0.440*** (0.013)	-0.415*** (0.012)	-0.414*** (0.012)	-0.633*** (0.017)	-0.587*** (0.016)
Hispanic (d)	-0.417*** (0.017)	-0.347*** (0.015)	-0.340*** (0.018)	-0.340*** (0.018)	-0.501*** (0.022)	-0.524*** (0.020)
Asian (d)	-0.189*** (0.020)	-0.190*** (0.019)	-0.208*** (0.020)	-0.209*** (0.020)	-0.234*** (0.031)	-0.339*** (0.024)
Loan Age	X	X	X	X	X	X
Underwriting Vars		X	X	X		X
HMDA Vars			X	X		X
Vintage Year-Qtr FE	X	X	X	X	X	X
State FE		X	X	X		X
# Observations	15,460,588	11,983,398	11,547,035	11,469,141	6,184,502	4,316,733

Notes: This table reports estimated marginal effects estimates from a logit model of equation (1)—the likelihood of voluntary mortgage prepayment due to sale on a set of race/ethnicity indicator variables. The estimation is performed at the quarterly frequency on a 5% random sample of loans from a matched HMDA-McDash-Equifax data set. The unit of observation is a loan-quarter. Underwriting variables include the borrower's risk score at origination, LTV at origination, loan amount, change in LTV since origination, indicators for condos and 2–4 multi-family properties, low-documentation loans, non-owner occupant properties, and refinance loans. HMDA variables include borrower age (2nd order polynomial), borrower income, and indicators for gender and co-applicants. All columns include a 3rd order polynomial for the number of quarters since origination (duration). Standard errors are clustered by county. (***) $p < 0.01$, (**) $p < 0.05$, (*) $p < 0.1$

Table A.8: Logit Default Hazard Estimates

Dependent Variable: Default (d)					
	GSE Loans			FHA Loans	
	(1)	(2)	(3)	(6)	(7)
Black (d)	0.350*** (0.023)	0.149*** (0.013)	0.101*** (0.012)	0.719*** (0.031)	0.340*** (0.025)
Hispanic White (d)	0.362*** (0.038)	0.185*** (0.012)	0.132*** (0.012)	0.162*** (0.028)	0.158*** (0.025)
Asian (d)	0.015 (0.015)	0.011 (0.012)	0.010 (0.012)	-0.163*** (0.030)	-0.098*** (0.037)
Loan Age	X	X	X	X	X
Underwriting Vars		X	X		X
HMDA Vars			X		X
Vintage Year-Qtr FE	X	X	X	X	X
State FE		X	X		X
# Observations	9,929,254	7,705,281	7,424,419	3,653,447	2,558,071

Notes: This table reports estimated marginal effects estimates from a logit model of the likelihood of mortgage default on a set of race/ethnicity indicator variables. The estimation is performed at the quarterly frequency on a 5% random sample of loans from a matched HMDA-McDash-Equifax data set. The unit of observation is a loan-quarter. Underwriting variables include the borrower's risk score at origination, LTV at origination, loan amount, change in LTV since origination, indicators for condos and 2–4 multi-family properties, low-documentation loans, non-owner occupant properties, and refinance loans. HMDA variables include borrower age (2nd order polynomial), borrower income, and indicators for gender and co-applicants. All columns include a 3rd order polynomial for the number of quarters since origination (duration). Standard errors are clustered by county. (***) $p < 0.01$, (**) $p < 0.05$, (*) $p < 0.1$)

A.5 Involuntary Prepayments

In Section 3.4 we showed that minority borrowers are more likely to default on their loans. The default definition that we use in that section is based on borrowers becoming seriously delinquent on their loans by missing at least three payments (that is, 90-plus days past due). We now consider an alternative definition of default that focuses on involuntary mortgage prepayment due to foreclosure and/or distressed sale (that is, short sales). Like our voluntary prepayment variables (refinance and home sale), this default definition identifies a terminal state, and is likely more correlated with the actual losses that lenders experience on distressed loans. As such, it is likely more relevant to mortgage pricing than a serious delinquency definition of default.

Table A.9 displays the estimation results. The table is identical in structure to Table 5, with the only difference being the dependent variable. The results are very different, however. In column (1) we see that minority borrowers are significantly more likely to lose their homes due to foreclosure, and the magnitudes are large.³ However, as we add more controls and fixed effects, the differences disappear. In our most saturated model with Zip-code-by-year-quarter fixed effects, minority GSE borrowers are significantly *less* likely to lose their homes to foreclosure. We see a similar pattern in the FHA sample, as Black borrowers are more than 8 percentage points less likely to lose their homes to foreclosure compared with non-Hispanic white borrowers (column (8)).

³The sample average for involuntary prepayment is approximately 0.11 percentage point.

Table A.9: Involuntary Prepayment/Foreclosure Results

Dependent Variable: Involuntary Prepayment/Foreclosure (d)									
	GSE Loans					FHA Loans			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Black (d)	0.095*** (0.024)	0.032* (0.017)	0.001 (0.016)	-0.028** (0.012)	-0.034*** (0.013)	-0.007 (0.014)	-0.059*** (0.016)	-0.084*** (0.017)	-0.089*** (0.017)
Hispanic White (d)	0.142*** (0.036)	0.048*** (0.016)	0.031** (0.015)	0.035** (0.014)	0.030** (0.014)	-0.027* (0.014)	-0.018 (0.012)	-0.037*** (0.013)	-0.022 (0.016)
Asian (d)	0.020* (0.011)	0.026*** (0.008)	0.033*** (0.008)	0.020** (0.008)	0.018** (0.008)	-0.066*** (0.015)	-0.021 (0.014)	-0.009 (0.016)	-0.009 (0.023)
Risk Score Origination		-0.161*** (0.024)					-0.207*** (0.017)		
LTV Origination		0.006*** (0.001)					0.010*** (0.001)		
Loan Amount		0.016*** (0.004)					0.031*** (0.008)		
LTV Change		0.023*** (0.002)	0.023*** (0.002)	0.023*** (0.002)	0.029*** (0.003)		0.023*** (0.002)	0.022*** (0.002)	0.028*** (0.003)
Refinance (d)		0.023*** (0.007)	0.004 (0.008)	0.003 (0.008)	0.005 (0.008)		0.109*** (0.020)	0.078*** (0.018)	0.081*** (0.019)
Female (d)			-0.013*** (0.004)	-0.012*** (0.003)	-0.010*** (0.004)			-0.029*** (0.008)	-0.027*** (0.009)
Refi Money			0.004*** (0.001)	0.004*** (0.001)	-0.059*** (0.008)			0.015*** (0.003)	0.025* (0.014)
SATO			0.162*** (0.036)	0.157*** (0.034)	0.615*** (0.077)			0.042** (0.018)	-0.052 (0.106)
Loan Age	X	X	X	X	X	X	X	X	X
Underwriting Vars		X	X	X	X		X	X	X
HMDA Vars			X	X	X			X	X
Vintage Year-Qtr FE	X	X	X	X	X	X	X	X	X
State FE		X	X				X		
Zip Code FE				X				X	
Zip Code-by-Year-Qtr FE					X				X
# Observations	15,460,588	11,983,398	11,547,035	11,547,035	11,396,543	6,184,502	4,316,733	3,748,150	3,575,715
# Loans	792,823	622,936	601,094	601,028	590,643	291,587	209,827	182,517	170,234
R ²	0.004	0.008	0.009	0.011	0.076	0.004	0.006	0.007	0.126

Notes: This table reports LPM estimates of the likelihood of involuntary mortgage prepayment due to foreclosure and/or distressed sale default on a set of race/ethnicity indicator variables. The estimation is performed at the quarterly frequency on a 7.5% random sample of loans from a matched HMDA-McDash-Equifax data set. The unit of observation is a loan-quarter. Underwriting variables include the borrower's risk score at origination, LTV at origination, loan amount, change in LTV since origination, indicators for condos and 2-4 multi-family properties, low-documentation loans, non-owner occupant properties, and refinance loans. HMDA variables include borrower age (2nd order polynomial), borrower income, and indicators for gender and co-applicants. All columns include a 3rd order polynomial for the number of quarters since origination (duration). "SATO" is the spread between the mortgage rate and the average rate associated with newly originated 30-year mortgages according to the FHLMC survey. "Refi Money" is a measure of the incentive to refinance taken from Deng et al. (2000). Standard errors are double clustered by county and vintage year-quarter. (***) p<0.01, (**) p<0.05, (*) p<0.1)

A.6 Agarwal, Driscoll, and Laibson Closed-Form Refinance Rule

In this section we proxy for the moneyness of the prepayment option using an alternative measure developed by Agarwal et al. (2013) (hereafter ADL). ADL derived a closed-form solution for the optimal time to refinance from a borrower’s perspective. Specifically, the rule states that a borrower should refinance when the current mortgage interest rate falls below the original rate by at least:

$$\frac{1}{\psi} [\phi + W(-exp(-\phi))]$$

where W is the Lambert W-function and

$$\psi = \frac{\sqrt{2\rho+\lambda}}{\sigma}$$

$$\phi = 1 + \psi(\rho + \lambda) \frac{\kappa/M}{(1-\tau)}$$

$$\lambda = \mu + \frac{i_0}{exp[i_0\Gamma]-1} + \pi$$

In these expressions ρ is the discount rate, μ is the expected probability of moving, σ is the standard deviation of the mortgage rate, $\frac{\kappa/M}{1-\tau}$ is the ratio of the tax-adjusted refinancing cost and the remaining mortgage value, Γ is the remaining maturity of the mortgage, i_0 is the original mortgage rate, π is the expected inflation rate, and τ is the marginal tax rate. We assume the following parameter values, where σ is estimated by taking the standard deviation of changes in the Freddie Mac Primary Market Mortgage Survey rate from April 1971 to August 2020:

$\rho =$	0.02
$\sigma =$	0.95
$\pi =$	0.02
$\mu_b =$	0.02
$\mu_w =$	0.04
$\frac{\kappa/M}{1-\tau} =$	$\frac{2000}{M} + 0.01$

We assume different mobility rates, μ_b, μ_w , for Black and non-Hispanic white borrowers, respectively, which we annualize based on the quarterly hazards from Table 1.⁴ We specify two variables based on the above threshold. First, we create an indicator variable, *ADL Dummy*, which takes a value of 1 if the difference between the borrower’s current interest rate and the market rate (PMMS survey) is greater than the ADL threshold. Second, we create a continuous variable, *ADL*, which measures how much higher/lower the difference between the current rate and market rate is from the ADL threshold. Positive values of *ADL* imply that the refinance option is in the money given the borrower type specific moving hazards

⁴For simplicity we assume the same mobility rate for Black and Hispanic white households.

and refi costs, while negative values imply that it is not.

We then re-estimate equation (2) and substitute our ADL variables for *Call Option*, which is our proxy for the moneyness of the refinance option from Deng et al. (2000). We focus on the specifications in columns (1) and (2) of Table 6. Column (1) includes only a control for the moneyness of the option, while column (2) includes interactions between the moneyness of the option and the race dummies. Table A.10 displays the results. In columns (1) and (2) we show results for the *Call Option* variable applied to the sample of loans with non-missing ADL values. Columns (3) and (4) display results for the *ADL Dummy*, and columns (5) and (6) display results for the *ADL* continuous variable.

A few notable patterns emerge from Table A.10. First, both ADL variables are positive and statistically significant as expected, which suggests that borrowers are more likely to refinance when their option is in the money. However, columns (2), (4), and (6) show that the refinancing behavior of minority borrowers is much less sensitive to changes in the value of the option. In fact, these differences appear to be much larger when we use the ADL variables, as the interaction coefficients in columns (4) and (6) are of about the same magnitude, but with the opposite sign as the ADL coefficients by themselves. This implies that minority borrowers are actually insensitive to macroeconomic changes in rates that make their prepayment option more valuable. Finally, as we saw in Table 6, the inclusion of the interaction terms causes the sign associated with the race dummies to flip and become positive. This means that the racial differences in refinance propensities is driven entirely by differential sensitivities of minority borrowers to respond to declining rates.

Table A.10: Prepayment due to Refinance with Interaction Effects

Dependent Variable: Prepay Refinance (d)						
	GSE Loans					
	(1)	(2)	(3)	(4)	(5)	(6)
Black (d)	-0.118*** (0.021)	0.309*** (0.048)	-0.136*** (0.021)	0.065** (0.029)	-0.149*** (0.020)	0.044 (0.028)
Hispanic White (d)	-0.197*** (0.025)	0.172*** (0.045)	-0.224*** (0.026)	-0.019 (0.026)	-0.234*** (0.025)	-0.049* (0.025)
Call Option	0.225*** (0.016)	0.232*** (0.016)				
Black * Call Option		-0.068*** (0.005)				
Hispanic White * Call Option		-0.064*** (0.005)				
ADL Dummy			0.525*** (0.079)	0.605*** (0.086)		
Black * ADL Dummy				-0.572*** (0.061)		
Hispanic White * ADL Dummy				-0.625*** (0.067)		
ADL					0.530*** (0.049)	0.560*** (0.049)
Black * ADL						-0.550*** (0.054)
Hispanic White * ADL						-0.564*** (0.059)
Loan Age	X	X	X	X	X	X
Underwriting Vars	X	X	X	X	X	X
HMDA Vars	X	X	X	X	X	X
Vintage Year-Qtr FE	X	X	X	X	X	X
Zip Code FE	X	X	X	X	X	X
# Observations	10,544,968	10,544,968	10,544,968	10,544,968	10,544,968	10,544,968
# Loans	557,848	557,848	557,848	557,848	557,848	557,848
R ²	0.016	0.016	0.011	0.011	0.012	0.012

A.7 Evidence from Survey of Consumer Finances

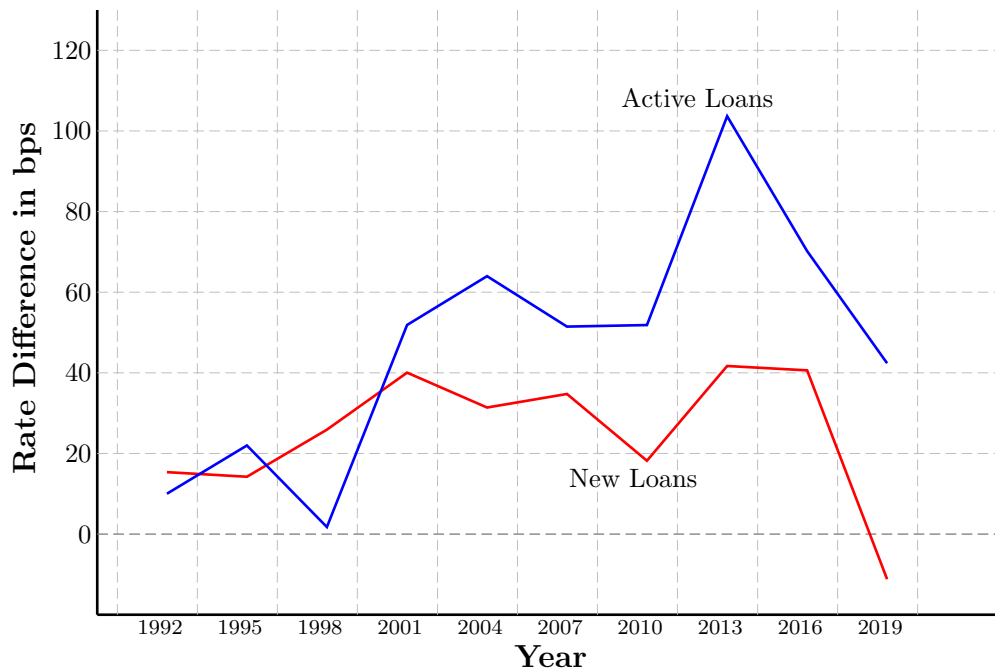
In this section we use data from the 1992–2019 Survey of Consumer Finances (SCF) to examine the rate gap between Black and non-Hispanic white borrowers for active loans as well as new loans originated that year.

Data construction is as follows. For comparison with the fixed-rate, conforming, and FHA mortgages used in our main analysis, the observations are from respondents who report having a non-adjustable-rate mortgage (X820=5) that is either non-federally guaranteed (X724=1) or a FHA loan (X726=1) with a loan amount at origination (X804) of less than \$450,000. The current interest rate is reported in X816, from which we remove outlier rates that are less than 2 percent or more than 4 percent over the average Freddie Mac PMMS rate during the year of origination, which is comparable to the restriction of rate to 3 percent to 8 percent in our main analysis.

The SCF definition of race underwent a slight revision in 1998 to include more categories. For the 1992–1995 SCF, we define respondent race based on Question X5909, “Are you Native American, Asian, Hispanic, black, white, or another race?”, with an answer of 4 (“black or African-American”) being our definition of a Black respondent and an answer of 5 (“white”) being our definition of a White respondent. In the 1998–2019 SCF, we define respondent race based on the revised Question X6809, which asks, “Which of these categories do you feel best describe you: (white, black or African-American, Hispanic or Latino, Asian, American Indian or Alaska Native, Hawaiian Native or other Pacific Islander, or another race?),” with an answer of 2 (“black or African-American”) being our definition of a Black respondent and answer of 1 (“white”) being our definition of a white respondent.

We compute the mean rate for all active loans by respondent race using the provided survey weights by race to compute the active loan-rate gap. For the new loans rate gap, we take means by respondent race and by the year of origination (X802) rounded to the nearest SCF survey year. The mean rate differences between Black and non-Hispanic white borrowers for active and new loans are shown in Figure A.2. While the estimates are much more noisy due to a smaller sample size (and potential survey error), we do find that the rate gap for active loans is higher than the rate gap for new loans, consistent with Figure 1 in the main text.

Figure A.2: Gap between interest rates for Black and non-Hispanic white borrowers based on data from the SCF



A.8 Secondary Market Pricing Estimates for All Covariates

Table A.11: Payups regression with all covariates

	(1) ≥\$85k trades	(2) ≥\$1mil trades
(mean) tract_black	0.998*** (0.303)	1.411*** (0.219)
Loan size under 85k	0 (.)	0 (.)
85-110k	-0.376*** (0.0296)	-0.369*** (0.0281)
110-125k	-0.609*** (0.0310)	-0.591*** (0.0291)
125-150k	-0.715*** (0.0309)	-0.687*** (0.0285)
150-175k	-0.927*** (0.0299)	-0.907*** (0.0280)
175-200k	-1.147*** (0.0338)	-1.132*** (0.0319)
over 200k	-1.244*** (0.0332)	-1.251*** (0.0311)
fico_cat=680	0 (.)	0 (.)
fico_cat=720	0.302*** (0.0431)	0.325*** (0.0427)
fico_cat=750	0.399*** (0.0439)	0.399*** (0.0438)
ltv_cat=80	0 (.)	0 (.)
ltv_cat=90	-0.0767*** (0.0168)	-0.0583*** (0.0153)
ltv_cat=95	-0.397*** (0.0522)	-0.319*** (0.0394)
ltv_cat=100	-0.384*** (0.0771)	-0.261*** (0.0733)
refi_diff	0.326*** (0.101)	0.242** (0.0948)
refi_diff_sq	-1.699*** (0.165)	-1.595*** (0.158)
refi_diff_cube	0.408*** (0.0724)	0.362*** (0.0714)

log_vol	6.227*** (1.007)	3.534*** (1.329)
log_vol_sq	-0.389*** (0.0632)	-0.224*** (0.0820)
log_vol_cube	0.00800*** (0.00132)	0.00464*** (0.00168)
group(dt_week coupon)=0	0 (.)	0 (.)
group(seller)=0	0 (.)	0 (.)
Observations	14374	13570
R^2	0.731	0.754

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

A.9 Additional Tables and Figures

Table A.12: Prepayment, Race, and Loan Amount

Dependent Variable: Prepay (d)	GSE Loans	
	(1)	(2)
Black (d)	-1.628*** (0.117)	
Hispanic White (d)	-1.342*** (0.158)	
Asian (d)	0.440** (0.188)	
Orig Amount \leq 85k (d)		-1.697*** (0.165)
85k < Orig Amount < 110k (d)		-1.225*** (0.137)
110k < Orig Amount < 125k (d)		-1.050*** (0.127)
125k < Orig Amount < 150k (d)		-0.854*** (0.109)
150k < Orig Amount < 175k (d)		-0.681*** (0.096)
Loan Age	X	X
Underwriting Vars		
HMDA Vars		
Vintage Year-Qtr FE	X	X
State FE		
Zip Code FE		
Zip Code-by-Year-Qtr FE		
# Observations	15,460,588	15,460,588
R ²	0.009	0.009

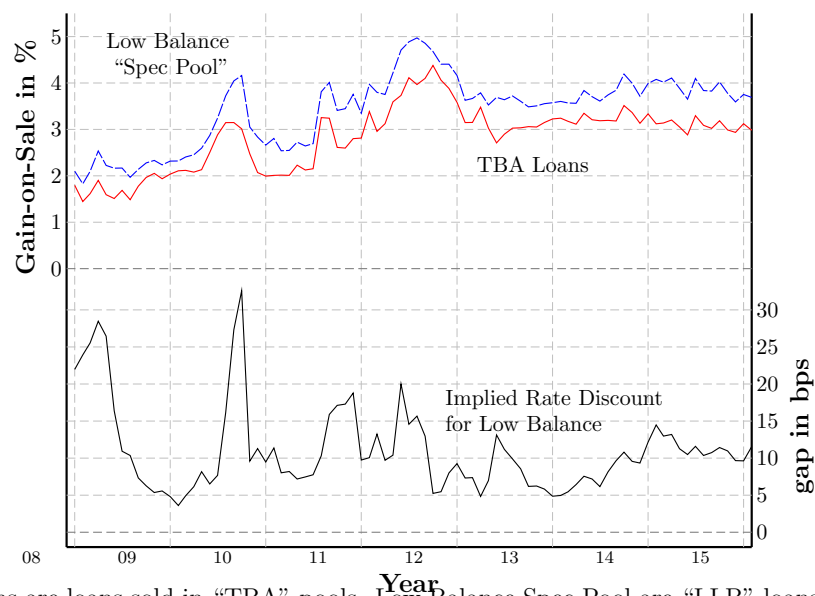
Notes: This table reports LPM estimates of the likelihood of voluntary prepayment due to either refinance or home sale on a set of race/ethnicity indicator variables (column (1)) and a set of indicator variables for loan amount bins at origination (column (2)). The estimation is performed at the quarterly frequency on a 7.5% random sample of loans from a matched HMDA-McDash-Equifax data set. The unit of observation is a loan-quarter. All columns include a 3rd order polynomial for the number of quarters since origination (duration). Standard errors are double clustered by county and vintage year-quarter. (***) $p < 0.01$, (**) $p < 0.05$, (*) $p < 0.1$

Table A.13: Prepayment LPM Results

Dependent Variable: Prepay (d)										
	GSE Loans						FHA Loans			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Black (d)	-1.628*** (0.117)	-1.040*** (0.060)	-0.967*** (0.050)	-0.867*** (0.047)	-0.690*** (0.040)	-0.694*** (0.042)	-1.447*** (0.066)	-1.074*** (0.050)	-0.818*** (0.056)	-0.723*** (0.053)
Hispanic White (d)	-1.342*** (0.158)	-0.977*** (0.089)	-0.948*** (0.087)	-0.887*** (0.078)	-0.667*** (0.049)	-0.682*** (0.048)	-0.932*** (0.099)	-1.017*** (0.048)	-0.972*** (0.054)	-0.852*** (0.049)
Asian (d)	0.440** (0.188)	0.176 (0.125)	0.202 (0.131)	0.191 (0.131)	0.106 (0.098)	0.107 (0.097)	0.290** (0.128)	-0.329*** (0.071)	-0.216** (0.085)	-0.269*** (0.073)
Risk Score Origination		0.482*** (0.089)						0.653*** (0.071)		
LTV Origination		-0.024*** (0.004)						-0.018*** (0.002)		
Loan Amount		0.765*** (0.084)						0.984*** (0.066)		
LTV Change		-0.025*** (0.006)	-0.089*** (0.006)	-0.084*** (0.005)	-0.084*** (0.005)	-0.075*** (0.006)		-0.051*** (0.004)	-0.077*** (0.005)	-0.038*** (0.004)
Refinance (d)		-0.602*** (0.090)	-0.593*** (0.079)	-0.561*** (0.075)	-0.553*** (0.075)	-0.619*** (0.073)		-0.438*** (0.048)	-0.350*** (0.042)	-0.415*** (0.047)
Female (d)			-0.079*** (0.017)	-0.072*** (0.017)	-0.102*** (0.019)	-0.107*** (0.020)			-0.075*** (0.021)	-0.110*** (0.021)
Refi Money			0.404*** (0.026)	0.410*** (0.026)	0.411*** (0.027)	1.128*** (0.083)			0.220*** (0.025)	0.960*** (0.102)
SATO			-1.965*** (0.171)	-1.907*** (0.162)	-1.846*** (0.163)	-7.215*** (0.586)			-0.216 (0.168)	-5.761*** (0.625)
Risk Score Change				0.899*** (0.130)	0.874*** (0.129)	0.908*** (0.116)			1.198*** (0.100)	1.172*** (0.098)
Loan Age	X	X	X	X	X	X	X	X	X	X
Underwriting Vars		X	X	X	X	X		X	X	X
HMDA Vars			X	X	X	X			X	X
Vintage Year-Qtr FE	X	X	X	X	X	X	X	X	X	X
State FE		X	X	X				X		
Zip Code FE					X				X	
Zip Code-by-Year-Qtr FE						X				X
# Observations	15,460,588	11,983,398	11,547,035	11,469,141	11,469,141	11,318,445	6,184,502	4,316,733	3,732,349	3,559,947
R ²	0.009	0.012	0.020	0.020	0.023	0.080	0.006	0.013	0.019	0.146

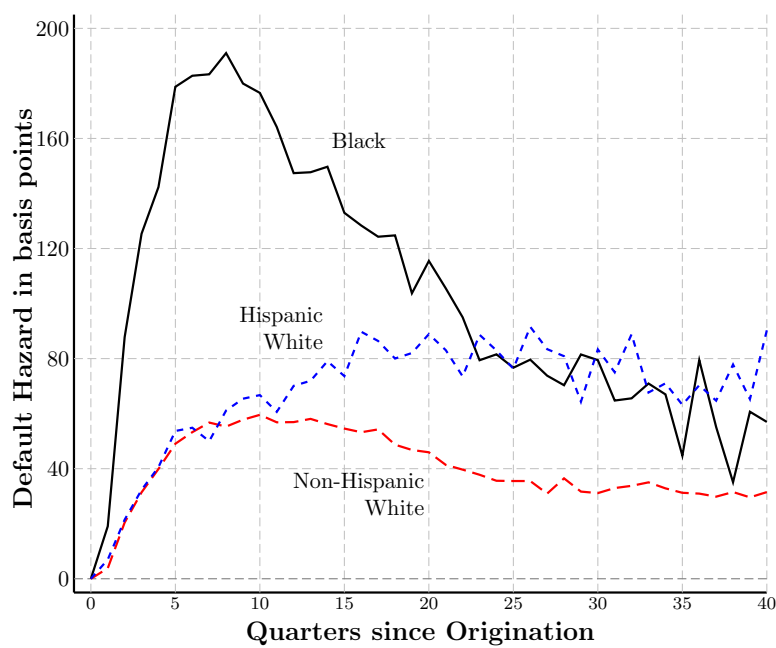
Notes: This table reports LPM estimates of the likelihood of voluntary prepayment due to either refinance or home sale on a set of race/ethnicity indicator variables. The estimation is performed at the quarterly frequency on a 7.5% random sample of loans from a matched HMDA-McDash-Equifax data set. The unit of observation is a loan-quarter. Underwriting variables include the borrower's risk score at origination, LTV at origination, loan amount, change in LTV since origination, indicators for condos and 2-4 multi-family properties, low-documentation loans, non-owner occupant properties, and refinance loans. HMDA variables include borrower age (2nd order polynomial), borrower income, and indicators for gender and co-applicants. All columns include a 3rd order polynomial for the number of quarters since origination (duration). "SATO" is the spread between the mortgage rate and the average rate associated with newly originated 30-year mortgages according to the FHLMC survey. "Refi Money" is a measure of the incentive to refinance taken from Deng et al. (2000). Standard errors are double clustered by county and vintage year-quarter. (***) $p < 0.01$, (**) $p < 0.05$, (*) $p < 0.1$

Figure A.3: Mortgage pricing for low prepayment loans.



Notes: TBA loans are loans sold in “TBA” pools. Low Balance Spec Pool are “LLB” loans defined as loans with balances of less than \$85K. Gain-on-sale is the gap between par and the interpolated price of an MBS paying a coupon equal to the FHLMC Primary Mortgage Market Survey 30-year FRM rate less the g-fee. Implied rate discount is the gap between the FHLMC PMMS 30-year FRM rate and the interest rate that yields the same gain-on-sale for an LLB mortgage.

Figure A.4: Kaplan Meier unconditional default hazard rates



Notes: This figure displays the Kaplan-Meier hazard estimates of default broken down by racial/ethnic groups. The Kaplan-Meier estimate of the hazard function is: $\lambda_p(t_j) = \frac{d_{pj}}{n_j}$, where the number of loans that have reached time t_j without being terminated or censored is given by n_j , and the number of terminations due to default at t_j is given by d_{pj} . The underlying data come from the Black Knight McDash database.