Mortgage Prepayment, Race, and Monetary Policy

Kristopher Gerardi, Paul Willen, and David Zhang *

September 2020

Abstract

This paper documents large differences in mortgage prepayment behavior across racial and ethnic groups in the US which have significant implications for monetary policy, inequality, and pricing. Using a novel dataset that combines administrative data on mortgage performance with information on race and ethnicity, we show that Black and Hispanic White borrowers have significantly lower prepayment rates than Non-Hispanic White borrowers, holding income, credit score and equity constant. This gap is on the order of 50 percent and largely reflects different sensitivities to movements in market interest rates, and was particularly pronounced during QE1. Differences in prepayment speeds result in large disparities between White and minority borrowers in the distribution of rates paid on outstanding mortgages, which widens during periods of low mortgage rates and high refinance volumes. Between 2010 and 2014, Black borrowers were paying 30 to 45 basis points more on average than Non-Hispanic Whites despite only a small gap of around 5 basis points between the groups at the time of mortgage origination. The large differences in prepayment behavior have important pricing implications as they suggest that minority borrowers are overpaying for their prepayment option. Our results show that inequality in mortgage markets is larger than previously realized and is exacerbated by expansionary monetary policy.

*We thank Larry Wall and Jon Willis for helpful comments. We especially thank Danny 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. The opinions expressed herein are those of the authors and do not represent the official positions of the Federal Reserve Bank of Atlanta, Federal Reserve Bank of Boston, or the Federal Reserve System. All remaining errors are our own.

1 Introduction

When traders started dealing in Mortgage Backed Securities (MBS) in the late 1970s, they needed estimates of the duration of the underlying loans. In the absence of any data, they settled on the assumption that all mortgages prepay seven years after origination. It quickly became apparent that such an assumption was wrong and costly and intensive research on prepayment commenced on Wall Street.[1] For the next 30 years, prepayment research on Wall Street and among a few academics burgeoned, largely focused on the question of securities valuation. In recent years, especially after the Global Financial Crisis in 2008, prepayment has emerged as a topic of much broader interest, particularly among macroeconomists. The reason is that prepayment has significant welfare effects on households.

Mortgage interest is a big item in most household budgets and households can reap a large windfall when interest rates fall. However, since most mortgages in the U.S. are fixed-rate, borrowers need to prepay their loans in order to exploit falling interest rates. Figure 1 shows the fall in mortgage rates that occurred from 2009 to 2012, which resulted both from broad Federal Reserve policies to lower rates and specific policies targeting mortgage rates. The bottom part of the figure shows that households reduced annual mortgage interest payments by $200 billion. Research using spending data shows that households that refinanced during this period spent a significant portion of the money and were less likely to default on their mortgages.[2]

Previous research has identified factors that influence prepayment behavior. The gap between the borrower's rate and the market rate, equity, and credit score are all associated with higher prepayment rates. But researchers remain puzzled by the large cross-sectional variation in prepayment behavior. In this paper we document a significant source of heterogeneity in prepayment speeds: race and ethnicity. Our findings have significant implications for monetary policy, inequality, and pricing in mortgage markets. In a large sample of mortgages going back to 2005, we find that, on average, 2.84 percent of Non-Hispanic White borrowers prepay their mortgages every quarter as compared to only 1.51 percent of Black borrowers. Our novel data allows us to control for a detailed list of underwriting variables and demographic differences including credit score, LTV ratio, loan amount, current equity, income, and gender, which shrinks the prepayment gap from 1.33 percent to 1.04 percent. However, even after controlling for observables, Non-Hispanic White borrowers remain more than 50 percent more likely to prepay than Black borrowers.

While prepayments are driven by both property sales and refinances we present evidence

[1] See article by Ranieri in Kendall and Fishman (2000).

[2] See Di Maggio et al. (2017) and Abel and Fuster (2019) for evidence on spending and Fuster and Willen (2017) for evidence on defaults.

that these racial gaps in prepayment behavior are most likely driven by differences in the propensity to refinance in response to rate declines. Specifically, we show that the racial gap in prepayment propensities is largely explained by different sensitivities to declines in market rates. Non-Hispanic White borrowers prepay at significantly higher rates when their refinance option is in the money compared to Black and Hispanic White borrowers. For mortgages insured by Fannie Mae and Freddie Mac (the GSEs), the entire racial gap in prepayment rates is explained by the differential sensitivities to market rates. For loans insured by the Federal Housing Administration (FHA), differential sensitivities to the option value of refinancing explains about one-third of the gap.

In recent years monetary policy has focused on lowering mortgage rates through large-scale MBS purchases. Using a simple difference-in-differences framework, we show that the first round of the Fed's MBS purchases, QE1, significantly exacerbated the prepayment gap between minority and White borrowers. In the two quarters before QE1, Black borrowers were about 0.5 percentage points (unconditionally) less likely to prepay compared to Non-Hispanic White borrowers. In the two quarters after QE1, the gap grew by more than a factor of 6 to approximately 3.3 percentage points. This pattern is robust to controlling for a detailed set of borrower and loan characteristics along with geographic and vintage fixed effects.

We show that these large differences in prepayment behavior resulted in huge disparities in the mortgage rates paid by minority borrowers compared to Non-Hispanic White borrowers over time. Figure 2 shows that differences in the *stock* of mortgage rates paid between Black and Non-Hispanic White borrowers (dotted blue line) dwarf the disparities in the *flow* of mortgage rates (solid red line) between the two groups. The racial differences in the flow of rates (i.e. average rates paid for newly originated loans) have been the focus of a recent literature on mortgage discrimination (Bartlett et al. (2019), Bhutta and Hizmo (2020), and Willen and Zhang (2020)), but as the figure shows, those differences have fallen considerably over time. In contrast, the difference in the stock of average rates (i.e. the difference in the average rates on outstanding loans at a given point in time) has grown significantly. We show that this divergence is directly due to the differential prepayment speeds between minority and Non-Hispanic White borrowers.

The paper also provides further evidence on differences in default rates between minority and Non-Hispanic White borrowers. Several previous studies have documented higher default and foreclosure rates for minority households, especially during the Global Financial Crisis.[3] While the differences in prepayment propensities that we document are quite robust,

[3]See for example Bayer et al. (2016), Bhutta and Canner (2013), Gerardi and Willen (2009) and Li and Mayock (2019).

the differences in default behavior are very sensitive to the set of conditioning variables. Unconditionally, Black and Hispanic White borrowers default at more than double the rate of Non-Hispanic White borrowers during our sample period. However, those differences completely disappear when we condition on basic underwriting variables like credit score and income, and even reverse themselves in certain specifications. For example, for our preferred regression specification estimated on a sample of FHA loans, Black and Hispanic White borrowers are about 6 basis points *less* likely to default per quarter than Non-Hispanic White borrowers.

The observation that minority borrowers have lower prepayment speeds 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 to 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.[4]

Our paper contributes to the literature on heterogeneity in monetary policy transmission in mortgage market. Factors such as the type of mortgage contract (Calza et al. (2013), Di Maggio et al. (2017)), house price growth (Beraja et al., 2018), renting vs owning a home (Cloyne et al., 2019), borrower age (Wong, 2019), and lender concentration (Scharfstein and Sunderam (2017), Agarwal et al. (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 et al. (2001), Deng and Gabriel (2006), Firestone et al. (2007), and Kau et al. (2019) have documented 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 are 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. Finally, previous studies focus exclusively on the pricing implications of prepayment differences and do not establish their implications for disparities in outstanding mortgage rates

[4] Higher interest rates for black borrowers at origination was found in Black and Schweitzer (1985), Boehm et al. (2006), Bocian et al. (2008), Ghent et al. (2014), Cheng et al. (2015), Bartlett et al. (2019), and Willen and Zhang (2020).

and the effect of monetary policy in exacerbating those differences.

Most of the literature on racial differences in mortgage performance has focused on default,[5] however as noted in Firestone et al. (2007) differential default rates tend to have small implications for pricing because defaults are rare in general.

2 Empirical Setup

We examine differences in prepayment and default behavior 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.[6] While linear probability models have some notable drawbacks,[7] 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 will also consider logit models below and show that the estimated marginal effects are very similar to the LPM coefficient estimates.

Our primary specifications take the following general form:

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

where i indexes the individual mortgage and t indexes the year-quarter. We focus on two mortgage outcomes: the likelihood of voluntary prepayment and the likelihood of default. Specifically, $Prepay_{it}$ is an indicator variable that takes a value of 1 if loan i prepays in year-quarter t, and $Default_{it}$ is an analogous indicator variable that identifies when a loan defaults. Our focus will be on testing for differences in mortgage 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 to Non-Hispanic White borrowers. X_{it} is a vector of control variables that includes numerous mortgage and borrower characteristics, which we describe in detail below. Most of the control variables are time-invariant, but there are a

[5] See e.g. Canner et al. (1991), Berkovec et al. (1994), Berkovec et al. (1998), Gerardi and Willen (2009), Bayer et al. (2016), Bhutta and Canner (2013) and Li and Mayock (2019)

[6] Our dataset only provides the year-quarter in which each mortgage was originated due to privacy concerns. We describe the data in detail below.

[7] For example, Horrace and Oaxaca (2006) proves that the LPM can lead to biased and inconsistent estimates of structural parameters, particularly 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 may be a good approximation to the conditional expectation function: http://www.mostlyharmlesseconometrics.com/2012/07/probit-better-than-lpm/.

few that vary at the quarterly frequency. In some specifications we will include geographic fixed effects, ν_g, typically at the state 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 where we are modeling 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 3 years at which point it prepays, it will contribute 12 observations with the $Prepay_{it}$ taking a value of 0 for the first 11 observations and a value of 1 for the last observation. Hazard models are commonly employed in the mortgage literature due to their ability to account for right-censored data (i.e. loans that neither prepay or default during the sample period and are either still active at the end of the sample or exit the dataset prior to the end of the sample period for other reasons).[8]

3 Data and Summary Statistics

To test for differences in prepayment and default propensities we use a novel dataset that combines two sources of administrative mortgage data: Home Mortgage Disclosure Act (HMDA) data and Black Knight McDash mortgage servicing data (hereafter referred to as the McDash data).

The HMDA database provides information on approximately 90% of U.S. mortgage originations (see National Mortgage Database, 2017). It has been frequently used in the literature to study issues around mortgage market discrimination.[9] 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 dataset is constructed using information from mortgage servicers, financial

[8]There are a non-trivial number of loans in our sample that 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.

[9]Examples include Carr and Megbolugbe (1993), Schill and Wachter (1993), Schill and Wachter (1994), Munnell et al. (1996), Tootell (1996), Avery et al. (1997), Black et al. (1997), Holloway (1998), Reibel (2000), Black et al. (2001), Cherian (2014), Haupert (2019), Bartlett et al. (2019), Bhutta and Hizmo (2020), Willen and Zhang (2020).

institutions that are responsible for collecting payments from borrowers. It covers between 60% and 80% 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 (FICOs), 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 questions around loan performance.[10]

We use a dataset constructed by merging the HMDA and McDash databases. This matched dataset is available to users within the Federal Reserve System and includes over 93 million loans originated between 1992 and 2015. 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 zipcode, lien type, loan purpose (i.e., purchase or refinance), loan type (e.g., conventional or FHA), and occupancy type. Tables A.1 and A.2 in the Internet Appendix display match rates by origination year. 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. The match rates are significantly higher in the 2005–2015 period, which is the focus of our analysis. In order to minimize measurement error created by poor matches, we only retain loans that can be uniquely matched between HMDA and McDash.

Our analysis focuses on loans originated in the 2005–2015 period (inclusive) due to poorer coverage of pre-2005 mortgage originations in the McDash dataset.[11] 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).[12] We impose some addi-

[10]Examples include Keys et al. (2012), Piskorski et al. (2010), Jiang et al. (2013), Jiang et al. (2014), Kaufman (2014), Ding (2017), Fuster et al. (2018), Adelino et al. (2019), Agarwal et al. (2020) and Berger et al. (2020).

[11]In 2005 McDash added a large servicer to its database, which substantially increased the overall coverage of the database. In addition, the large servicer only provided information on all of 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.

[12]GSE 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

tional sample restrictions to address outliers and missing information on key underwriting variables. Table A.3 in the Internet 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 also uses the McDash-HMDA matched database.[13] 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.[14]

After imposing all of the sample restrictions we are left with approximately 14.7 million unique mortgages. Since most of our analysis is conducted on panel dataset at the quarterly frequency where the unit of observation is a loan-quarter, we work with a 10% random sample to ease the computational burden. We also distinguish 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 downpayments compared to 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 (i.e. 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 (i.e. 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.[15] There are large differences across the racial/ethnic categories for many of the observable variables in both tables. Focusing on the GSE sample,

not technically backed by the federal government. However, most market participants believed those loans to be implicitly guaranteed by the government.

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

[14]The 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 only make the distinction for White borrowers. We combine Hispanic and non-Hispanic Black borrowers into the single "Black" category.

[15]Asian borrowers are included in the pooled sample, but due to space constraints we do not include separate statistics for them in the table. Asian borrowers look very similar to Non-Hispanic White borrowers across most observable variables.

for example, Non-Hispanic White borrowers have significantly higher average FICO scores and household incomes compared to Black and Hispanic White borrowers (747 vs. 711 and 726 and \$98.5k vs. \$81.4k and \$79.1k, respectively). Non-Hispanic White borrowers obtain significantly lower mortgage rates on average (5.17 vs. 5.64 and 5.45, respectively), which has been documented by several papers in the literature.[16] Interestingly, Black borrowers are much more likely to be female (48.1%) compared to both Hispanic White (31.0%) and Non-Hispanic White (28.4%) borrowers, while Non-Hispanic White borrowers are much more likely to have a co-applicant on the mortgage (53%) compared to Black (27.3%) and Hispanic White (35.4%) 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 FICO scores and household income levels are significantly lower for all groups in the FHA sample compared to the GSE sample. In addition, LTV ratios are much higher in the FHA sample (93.6% vs. 72.7%).

The bottom panel of Table 1 shows that the average overall prepayment rate is 3.37% per quarter in our GSE sample, while the average default rate is only 0.115%.[17] The overall prepayment rate is slightly lower in the FHA sample (2.58%) while the default rate is more than twice as high (0.238%). There are large differences in average prepayment rates across racial/ethnic groups in both loan samples. In the GSE sample, Non-Hispanic White borrowers prepay at an average rate of 3.5% per quarter compared to only 2.1% and 2.4% for Black and Hispanic White borrowers, respectively. The difference between Black and Non-Hispanic White prepayment rates in the FHA sample is even larger (2.8% vs. 1.5%).

The left panel in Figure 4 plots Kaplan-Meier estimates of the hazard rates of prepayment by racial/ethnic group. These are unconditional, average quarterly prepayment rates as a function of duration that account for right-censoring.[18] The figure shows that the unconditional hazard estimates of prepayment for Non-Hispanic White and Asian borrowers are approximately twice as high as 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 prepayment hazards compared to Non-Hispanic White and Asian borrowers, although the difference is not as large as for Black borrowers. While these are sim-

[16] See e.g. 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), Willen and Zhang (2020).

[17] The default definition used in this analysis includes loans that terminated due to foreclosure (both auction sales and bank/REO sales) or pre-foreclosure distressed sales (i.e. short-sales). The prepayment definition includes all voluntary payoffs, which include both refinances and property sales.

[18] Specifically, the Kaplan-Meier estimates are calculated as follows: Assuming that hazards occur at discrete times t_j where $t_j = t_{0+j}$, $j = 1, 2, ..., J$, 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 prepayment at t_j as d_{pj}, then the Kaplan-Meier estimate of the hazard function is: $\lambda_p(t_j) = \frac{d_{pj}}{n_j}$.

ply unconditional sample averages, and do not account for the large differences in observable characteristics between Black, Hispanic White, and Non-Hispanic White borrowers, as we will show below, the patterns are quite robust to conditioning on conventional underwriting variables.

There are also significant differences in quarterly default rates across the racial/ethnic groups. Table 1 shows that in the GSE sample, Black and Hispanic White borrowers are more than twice as likely to default as Non-Hispanic White borrowers (0.23% vs. 0.10%). These differences ares smaller in the FHA sample where Black borrowers have only slightly higher default rates compared to Non-Hispanic White borrowers (0.29% vs. 0.24%), while Hispanic White borrowers actually have slightly lower default rates compared to Non-Hispanic White borrowers. The right panel in Figure 4 shows the unconditional Kaplan-Meier default hazard estimates, which are consistent with the patterns in Tables 1 and 2. However, in contrast to the prepayment patterns, we will show below that the differences in default rates largely disappear when we condition on observable loan and borrower characteristics.

4 Results

In this section we present our main empirical results. We start by showing estimates of the gap in prepayment and default propensities between minority and Non-Hispanic White households. We then present evidence that differences in prepayment rates are likely driven by differences in refinancing behavior rather than differences in mobility. Finally, we provide evidence that monetary policy has exacerbated the gaps in prepayment propensities.

4.1 Prepayment

We begin by estimating the LPM model in equation (1) for prepayment. Table 3) contains the results. Columns (1)–(7) report estimates for the GSE sample while columns (8)–(9) show estimates for the FHA sample. In all columns we have multiplied the prepayment indicator by 100 so that the coefficients can be interpreted in terms of percentage points. Column (1) reports estimates from a specification without any controls or fixed effects. The differences in prepayment rates between Black, Hispanic White and Non-Hispanic White borrowers are virtually identical to the differences reported in Panel B of Table 1. Unconditionally, Black (Hispanic White) borrowers prepay at a rate that is 1.3 (1.1) percentage points lower than Non-Hispanic White borrowers on average, while Asian borrowers prepay at a slightly higher rate on average. In column (2) we control for duration using a third-order polynomial and include vintage year-quarter fixed effects to control for unobservable changes in underwriting standards over time.[19] The differences in prepayment actually increase slightly as the Black (Hispanic White) coefficient increases (in absolute value) to -1.7 (-1.3) percentage points.

In column (3) of Table 1 we include controls for some basic underwriting characteristics at origination such as the borrower's FICO 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 2-4 units).[20] In addition we include an estimate for the borrower's equity in the property, 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 literature. Borrowers with higher credit scores and larger loan sizes prepay at faster rates. Borrowers with more equity in their properties also prepay at higher rates. The prepayment differences between racial/ethnic groups slightly decrease with the addition of these controls, but the gaps remain economically large and statistically significant. Black and Hispanic White borrowers prepay at rates that are approximately 1 percentage point lower than Non-Hispanic White borrowers per quarter, which is approximately 30 percent of the average quarterly prepayment rate in our GSE sample (see Table 1).

Column (4) includes the same set of controls as column (3) but specifies FICO, LTV, and loan size in small, discrete bins in order to allow for any non-linearities that might exist in their relationship with prepayment. We do not display the estimates due to space constraints, but they can be found in Table A.4 in the Internet Appendix. This alternative specification has no discernible affect on the prepayment differences across groups. The specification does provide some additional insight regarding the economic magnitude of the prepayment differences, however. According to Table A.4 the difference in prepayment hazards between Black and Non-Hispanic White borrows (1.1 ppts) is approximately equivalent to the difference in prepayment rates between borrowers with FICO scores below 600 and between 720 and 740. The difference is also similar in magnitude to the difference in prepayment rates between borrowers with loans below \$85k and those with loans between \$150k and \$175k. We will come back to this comparison below when we discuss pricing implications.

In column (5) 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. Borrowers with higher income are more likely to prepay

[19]We experimented with higher order polynomials as well as 1-year bins for duration but the results did not materially change.

[20]We also include indicators for missing information about documentation and property type.

while female borrowers and borrowers with a co-applicant are slightly less likely to prepay. The differences across income categories (displayed in Table A.4 are economically large and comparable to the racial/ethnic group differences. For example the difference in prepay hazards between Black and Non-Hispanic White borrowers is about the same as the difference in rates between borrowers with income less than $25k and more than $175k. Note that the inclusion of these controls has only a small effect on the prepayment gaps between racial/ethnic groups.

In column (6) we add two additional controls. "Refi Money" is the difference between the borrower's mortgage rate and the value of the FHLMC 30-year FRM index in the current year-quarter. This is a measure of the "moneyness" of the refinance option, as the larger the difference, the more the borrower would benefit from refinancing into a new loan with a lower rate and payment. "SATO" (spread at origination) is the difference between the borrower's mortgage rate and the value of the FHLMC 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. To the extent that differences in prepayment are driven by differences in the propensity to refinance, we would expect these two variables to be important predictors of prepayment behavior. This is exactly what we find in column (6). A one standard deviation increase in "Refi Money" (0.90 ppts) is associated with a 2.4 percentage point increase in the hazard of prepayment while a one standard deviation increase in SATO (0.15 ppts) is associated with a -0.25 percentage point decrease in prepayment hazards. The inclusion of these variables does not materially affect the racial/ethnic gaps in prepayment behavior however.

The specification reported in column (7) of Table 3 includes Zip Code fixed effects, so that differences in prepayment hazards between groups in column (7) are estimated using only variation 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. However, the inclusion of tens of thousands of additional variables in the LPM substantially increases the computational burden. While the gap between the racial/ethnic groups does narrow, the magnitudes remain economically large and statistically significant. Finally, in column (8) 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. Thus, Black borrowers prepay by approximately 0.88 percentage points less per quarter compared to Non-Hispanic White borrowers in the same year-quarter in the same Zip Code controlling for credit score, LTV, income, gender, and our additional underwriting variables.[21] This specification controls for

[21]There are almost 800 thousand Zip Code-by-year quarter fixed effects. A few thousand are dropped

local economic shocks as well as house price dynamics at the ZIP Code level.

The final three columns in Table 3 display estimation results for three specifications from our sample of FHA loans. Column (9) is analogous to column (1) and does not include any covariates, while column (10) is the same specification displayed in column (6), which includes all of our controls along with state and vintage-year fixed effects. Column (11) substitutes Zip Code for state fixed effects and is analogous to column (7). The differences in prepay hazards across the racial/ethnic groups in both specifications essentially mirror the differences in the GSE sample.

In Table A.5 in the Internet Appendix we show that the results in Table 3 are not sensitive to our choice of the LPM, which assumes that the prepayment 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[22] The average marginal effects associated with the logits in all specifications are very close to the corresponding LPM coefficients.

4.2 Default

In this section we present results on the differences in default hazards across racial/ethnic groups. Table 4 presents estimation results for the same LPM specifications in Table 3. Again we have multiplied the default indicator by 100 so that the coefficients can be interpreted in terms of percentage points. The unconditional differences reported in column (1) are large relative to the average quarterly default rate in the sample (0.115 ppts). Both Black and Hispanic White borrowers are approximately 0.13 percentage points more likely to default than Non-Hispanic White borrowers. Unlike the prepayment results however, the estimates are very sensitive to the inclusion of fixed effects and control variables. Controlling for duration and including vintage year-quarter effects in column (2) cuts the differences in half for Black borrowers, and including additional controls for basic underwriting variables dramatically reduces the differences for both Black and Hispanic White borrowers in columns (3) and (4). Further controlling for borrower income and gender in column (5) completely eliminates differences in default propensities between Black and Non-Hispanic White borrowers. Hispanic White borrowers are still slightly more likely to default compared to Non-Hispanic White borrowers, but the effect is small in magnitude and marginally significant. Once we add Zip Code fixed effects to the specification in column (7), the Black coefficient flips signs and we find that Black borrowers are significantly *less* likely to default than non-Hispanic

due to there only being 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.

[22] The one exception is the specification with Zip Code fixed effects. We were unable to obtain convergence for the maximum likelihood estimator.

White borrowers. The inclusion of Zip Code-by-year quarter fixed effects does not affect the estimate.

The default patterns are similar in the FHA sample as again we find that the sign of the coefficient flips for Black borrowers. Unconditionally, Black FHA borrowers are more likely to default than Non-Hispanic White FHA borrowers (column (8)). However, once we control for underwriting variables and income and gender, Black FHA borrowers are actually less likely to default (column (9)). Adding Zip Code fixed effects in column (11) increases the difference (in absolute magnitude). Interestingly, unlike in the GSE sample where we found that Hispanic White borrowers are more likely to default in all specifications, in the FHA sample we find that Hispanic White borrowers are significantly less likely to default compared to Non-Hispanic White borrowers both conditionally and unconditionally.

In summary, while the differences in prepayments between borrower races are large, the differences in default are much smaller after conditioning on covariates. This suggests that prepayment is a much more relevant aspect of mortgage performance by race even though it has been under-studied relative to default risk.[23]

4.3 Evidence on the Refinance Channel

Thus far we have presented evidence of economically large, robust differences in prepayment behavior across racial/ethnic groups. Prepayments could be driven by two activities in our context: borrowers selling their properties and moving or borrowers refinancing their loans. Distinguishing between these two activities is important for monetary policy since the refinance channel is one of the most important ways that borrowers benefit from lower nominal interest rates.

Ideally our data would provide information about the exact reason for prepayment, but one of the drawbacks of virtually all loan-level datasets, including ours, is that it is impossible to distinguish between prepayments due to refinances and prepayments due to home sales. In this section we attempt to distinguish between these two factors. We design an indirect test that compares across Black/Hispanic White and Non-Hispanic White borrowers the sensitivity of prepayment to movements in market interest rates relative to the current rates that they pay on their loans. Specifically we estimate a version of equation (1) where we interact our race variables with a variable that measures a borrower's incentive to refinance

[23] Earlier studies have primarily focused on default, and found that black borrowers tend to have higher cumulative default probabilities (Canner et al. (1991), Berkovec et al. (1994), and Berkovec et al. (1998)). One explanation for the difference in our results is that the higher cumulative default probabilities may be due to lower prepayment probabilities on the part of minority borrowers combined with a similar default rate.

into a loan with a lower interest rate:

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

where $Refi\ Money$ is the difference between the borrower's current mortgage rate and the prevailing market rate and measures the extent to which the option to refinance is in the money. If differences in prepayment behavior between Black/Hispanic White and Non-Hispanic White borrowers are explained by differences in the propensity to refinance into lower rates then we should expect to find the inclusion of the interaction term in equation (2) lower the estimate of β.

Table 5 displays the estimation results separately for GSE and FHA loans. We focus on the specification from column (6) in Tables 3 and 4, which includes state and year-quarter fixed effects along with controls for underwriting variables and the additional variables that we obtain from HMDA (income, gender, co-applicant status). We restrict the sample to only Black, Hispanic White and Non-Hispanic White borrowers, since the prepayment differences between Asian and Non-Hispanic White borrowers is negligible. In columns (1) and (4) we do not include any interaction effects. In columns (2) and (5) we include interactions between the Black and Hispanic White dummies and the variable that measures the incentive to refinance, $Refi\ Money$. In columns (3) and (6) we add interactions between the Black and Hispanic dummies and the $SATO$ variable.

The results for the GSE sample are striking. The addition of the $Refi\ Money$ interaction essentially explains the entire discrepancy in prepayment behavior between minority and Non-Hispanic White borrowers. That is, differences in prepayment behavior between minority GSE borrowers and Non-Hispanic White GSE borrowers comes entirely from differences in the sensitivity of prepayment behavior to interest rate movements. Column (2) shows that Black and Hispanic White borrowers are significantly less likely to prepay their loans in response to market rates declining. A one standard deviation increase in $Refi\ Money$ (which corresponds to market rates declining relative to the borrower's current rate by 0.9 ppts.) increases the likelihood of prepayment by 2.5 percentage points for Non-Hispanic White borrowers but only 1.5 percentage points for Black borrowers and 1.6 percentage points for Hispanic White borrowers. This pattern is not nearly as dramatic in the sample of FHA loans, which may be expected since other factors (e.g. removal of private mortgage insurance payments) may also drive refinancing behavior in FHA mortgages. Column (5) shows that minority borrowers are less responsive to market rate movements compared to Non-Hispanic Whites, but the differences are not as big as we see in the GSE sample. In addition, the differential sensitivity to market rates explains only a small part of the difference in prepayment propensities between minority and Non-Hispanic White borrowers. The coefficient associated with the Black and Hispanic White dummies remains economically large and statistically significant.

One important caveat to mention is that our estimate of the sensitivity of prepayment to interest rates for the different racial/ethnic groups (δ in equation (2)) is not necessarily isolating variation in refinance behavior alone. We believe it largely reflects such variation, but it could also be the case that property sales are differentially correlated with movements in interest rates across groups. If that is true then the δ estimates would at least partially reflect differences across groups in mobility rather than refinancing.[24]

4.4 The Effect of Monetary Policy on Prepayment Gaps

Figure 5 displays unconditional, quarterly prepayment 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 figures shows that prepayment gaps are relatively small in the first few years of the sample period, but then increase dramatically beginning in early 2009 right around 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 falls in late-2009/early-2010, but then spikes again in the third quarter of 2010, which coincides with the first Federal Reserve discussions of the second LSAP, QE2.[25] Finally, the third spike in prepayment gaps in the figure occurs around the time of the announcement of the Fed's final LSAP, QE3, in the third quarter of 2012.[26]

While Figure 5 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.[27] For that reason, we implement a more direct test for monetary policy effects on the gaps in prepayment behavior between 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 prepayment behavior among FHA borrowers are not explained by differences in the sensitivity of prepayment to market rates. We also explicitly focus on QE1. Beraja et al. (2018) shows that mortgage rates fell significantly and refinancing activity expanded considerably when QE1 was announced.[28] 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 24, 2008 and initially called for purchases of up to \$500 billion in MBS guaranteed by the GSEs.[29] In March 2009, the Federal

[24] Berger et al. (2020) find that household mobility is correlated with interest rate changes during our sample period. However, there is no evidence on whether the sensitivity of moving to interest rate movements differs by race.

[25] On 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."

[26] QE3 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.

[27] For 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.

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

[29] It also announced purchases of up to \$100 billion in debt obligations of Fannie Mae, Freddie Mac, Ginnie

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.[30]

We test whether QE1 exacerbated the gap in prepayment rates between 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):[31]

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

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).[32] We consider two different sample windows around the QE1 announcement: six months and one year.

Table 6 displays the estimation results. In columns (1)–(3) we restrict the sample to a six month window around QE1 and in columns (3)–(5) we expand the sample to a one 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 state and origination year-quarter fixed effects (the specification in column (6) in Table 3). Finally we estimate a specification that adds interaction terms between our $postQE1$ dummy and FICO 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 is motivated by anecdotal evidence that has suggested that the refinancing boom that followed QE1 was mainly driven by borrowers with high credit scores and low LTVs.

The estimation results suggest that QE1 had a large effect on the racial gap in prepayment propensities. According to column (1), Black borrowers were about 0.5 percentage points less likely to prepay in the six months prior to QE1 compared to Non-Hispanic White borrowers and the gap increases substantially to approximately 3.3 percentage points after QE1. While prepayment propensities for Non-Hispanic White borrowers increased by more than 4 percentage points, an increase of approximately 250% of their rate prior to QE1 (1.6 ppts), Black and Hispanic White borrowers increased their prepayment rates by only 1.2 percentage points, an increase of approximately 110% of their pre-QE1 rate (1.1 ppts). Including our controls and fixed effects slightly changes the magnitudes, but the large effect of QE1 on prepayment gaps remains. In column (2) Black and Hispanic White conditional prepayment rates are actually significantly higher than Non-Hispanic White borrowers right before QE1, but afterwards their rates fall more than 3 percentage

Mae and the Federal Home Loan Banks.

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

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

[32]Since QE1 was announced at the end of November, prepayments driven by QE1 would not show up until the beginning of 2009:Q1.

points below the rates for Non-Hispanic White borrowers.

In column (3) the addition of the interactions between the $postQE1$ dummy and FICO scores and LTVs slightly attenuates the gaps in prepayments between minority and Non-Hispanic White borrowers that emerged after QE1, but the differences remain large and statistically significant. The interactions with FICO score, which are displayed in the table are striking.[33] High FICO borrowers (FICO $>$ 740) increased their prepayment rates by more than 7 percentage points after QE1 compared to an increase of about 4 percentage points for low FICO borrowers (FICO $\leq$ 660). Since the prepayment differences across FICO bins are small in the period before QE1, these findings are consistent with the claim that the refinancing boom driven by QE1 disproportionately affected borrowers with high credit scores.

Columns (4)–(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 to minority borrowers as well as high credit score borrowers compared to those with lower credit scores.

5 Implications for Mortgage Rate Disparities

The literature on statistical discrimination in mortgage market pricing has focused 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 3 displays the difference in the flow of average mortgage rates (solid red line) between Black and Non-Hispanic White borrowers during our sample period and the difference in the stock of average rates (dotted 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 A.1 with the only difference being that they are constructed using our estimation sample of loans originated between 2005–2015. Figure A.1 uses loans originated between 1996–2015. 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 almost 35 bps in 2008 down to 10 bps in 2010 and then fluctuates between 10 and

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

20 bps for the remainder of the period.[34]

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 40 basis points in 2013, while it climbs above 60 basis points in the GSE sample. To isolate the disparities in the stock of rates that is only due to prepayment behavior (as opposed to differences in pricing at origination) in the bottom panel of Figure 3 instead of using actual interest rates paid by borrowers, we assume that every mortgage origination receives that quarter's FHLMC rate index value. Thus, by construction, there are no disparities in the rate of mortgage flows between Black and Non-Hispanic White borrowers, so that the disparities in the stock of rates are only driven by the differences in prepayment propensities. The bottom panel of Figure 3 shows that beginning in 2009, the tendency of Black borrowers to pay higher than market rates for longer than Non-Hispanic White borrowers drives the rate gap up by more than 30 bps in the pooled sample and by approximately 50 bps in the GSE sample.

If we go back to Figure A.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 in mortgage-backed securities (MBS), drove mortgage rates down 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 A.1 by estimating a difference-in-differences specification that is similar to equation 3 above. Specifically we estimate the following regression:

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

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 7 displays the estimation results for three windows around the announcement of QE1: six months, one year, and two years. For each window we display two different specifications. In columns (1), (3), and (5) we estimate specifications with no additional controls, while 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 the upper panel of Figure A.1. Rates paid by non-Hispanic White borrowers drop significantly after QE1 — 22 bps in the 6-month

[34] These are slightly larger differences compared to the results in Bartlett et al. (2019), which finds differences in interest rates between minority and Non-Hispanic White borrowers of 7.9 and 3.6 basis points for purchase and refinance 30-year FRMs originated between 2009–2015 and insured by the GSEs. However, the gap in Figure 3 is unconditional while the differences documented in Bartlett et al. (2019) are conditional on credit scores and LTV ratios.

window and 47 bps in the 2-year window. At the same time, rates paid by minority borrowers also decline, but by much smaller magnitudes. For the 6-month window average rates paid by black borrowers drop by 10 bps after QE1 and by about 24 bps in the 2-year window. This causes the gap in outstanding rates to grow from 22 bps in the two years before QE1 to 46 bps 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 that drives the unconditional results, and is consistent with differential refinancing behavior driving the large divergence in mortgage rates between minority and non-Hispanic White borrowers in the period after QE1.

6 Pricing Implications

Differential prepayment behavior of Black and Hispanic behavior 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. Third, lower prepayment rates mean that the cost of default 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 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

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

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

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

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.1, 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 Primary Mortgage Market Survey 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 can validate our claim that lower prepayment speeds increase the value of mortgages and get some idea of the quantitative magnitudes by looking at low balance mortgages. It is well-known in the industry that borrowers with low balances are less likely to prepay. The reason is that some costs of refinancing are fixed but the benefits are proportional to the balance of the loan. Because of their different prepayment properties, low balance loans trade in their own specified or "spec" pools In our prepayment regressions, we found that loans in a 110 to 125 thousand dollar loan balanace spec pool had similar prepayment speed differentials (compared to loans with balances over \$175k) to the differential between Black and Non-Hispanic White borrowers (see Table A.4 in the Appendix.) The gain-on-sale premium for pools of loans in these spec pools is typically between 50 and 100 basis points.[35]

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. If we assume that a lender wants to maintain a constant gain-on-sale across all loans and ask what the rate reduction on loans to Black borrowers would need to be to ensure that outcome. If MBS price differences were fully passed through to Black borrowers, they would typically pay between 5 and 10 basis points less than they currently do.[36]

There is however, a potentially offsetting effect that would make minority borrowers less attractive to default insurers and, *ceteris paribus*, increase the rates that they might face relative to Non-Hispanic White borrowers. It is easiest to see this if we consider a mortgage insurer like Fannie Mae or Freddie Mac. 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

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

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.

Because of a quirk in the way Fannie Mae and Freddie Mac price insurance, higher prepayment speeds may make Non-Hispanic White borrowers more attractive to insure. The issue is that Fannie

[35] Compare the lines labelled "Low-balance spec pool" and TBA in Figure A.1 in the appendix.

[36] Figure A.1 shows that there are periods, such as early 2009 and late 2010, that they would pay substantially less ($\sim$ 30 bps).

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 to an otherwise identical Non-Hispanic White borrower who has a higher λ_p.

Note that this is only an issue for the GSEs and not the lender: the lender receives a more valuable stream of cashflows from the MBS of Black and Hispanic mortgages as shown in Equation (5). If instead there were no GSE insurance, then the value of the loan is:

$$V - S_0 = \frac{m - r - LDG\lambda_d}{r + \lambda_p + \lambda_d} \tag{6}$$

Where the similar default rates λ_d and lower prepayment probabilities λ_p of Black and Hispanic borrowers would again make the loan more valuable to lenders as long as the lenders continue to make a positive gain on sale such that $m - r - LDG\lambda_d > 0$.

7 Conclusion

In this paper we have shown that minority borrowers prepay their fixed-rate mortgages at a significantly lower rate than majority White borrowers, and that expansionary monetary policy appears to have exacerbated these differences. In turn, the large differences in prepayment propensities have resulted in significant disparities in the average interest rate that minority borrowers pay on the stock of outstanding mortgages compared to 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.

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 some of the difference but a large gap remains.

The second question is what policy makers can do to reduce racial differences? The prepayable, fixed-rate mortgage plays a central role in the story. Many commentators have argued 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 the willingness of borrowers to exercise the option and the data shows systematic variation across racial groups in that willingness and thus, in a sense, the value of the option.

How could a policy maker enable Black and Hispanic White borrowers to exploit rate reductions more effectively? One way would be to expand the use of ARMs. The US is almost unique in its reliance on FRMs. In many countries, the mortgage ecosystem is largely populated with ARMs and those countries enjoy high homeownership rates and have foreclosure problems that are no worse than the US. 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.

More broadly, our results show how a race-blind policy – lower mortgage rates – can have disparate effects on racial groups. It is also an example of how a total ban on statistical discrimination can hurt as well as help minority groups. Under the Equal Credit Opportunity Act (ECOA), a mortgage lender is not allowed to price discriminate on the basis of race, color, or national origin. We show that based on the differences in prepayment behavior, and the lack of differences in default propensities, minority groups should face lower mortgage rates, ceteris paribus. Therefore, allowing lenders the option of pricing loans to minority borrowers at lower rates than White borrowers, but not the other way around, may allow minority borrowers to receive more advantageous terms on their loans.

References

Abel, J. and A. Fuster (2019). How do mortgage refinances affect debt, default, and spending? Evidence from harp. *American Economic Journal: Macroeconomics*, Forthcoming.

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*.

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*.

Bayer, P., F. Ferreira, and S. L. Ross (2016). The vulnerability of minority homeowners in the housing boom and bust. *American Economic Journal: Economic Policy 8*(1), 1–27.

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 G. B. Canner (2013). Mortgage market conditions and borrower outcomes: Evidence from the 2012 hmda data and matched hmda-credit record data. *Federal Reserve Bulletin 99*(4).

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.

Calza, A., T. Monacelli, and L. Stracca (2013). Housing finance and monetary policy. *Journal of the European Economic Association 11*(s1), 101–122.

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.

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.

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).

Fuster, A. and P. Willen (2017). Payment size, negative equity, and mortgage default. *American Economic Journal: Economic Policy 9*(4), 167–191.

Gerardi, K. and P. Willen (2009). Subprime mortgages, foreclosures, and urban neighborhoods. *The BE Journal of Economic Analysis & Policy 9*(3).

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.

Haupert, 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. Vytlacil (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. Vytlacil (2014). Liar's loan? effects of origination channel and information falsification on mortgage delinquency. *The Review of Economics and Statistics 96*(1), 1–18.

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.

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.

Kendall, L. and M. Fishman (2000). *A Primer on Securitization.* The MIT Press. MIT Press.

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.

Li, P. and T. Mayock (2019). Mortgage characteristics and the racial incidence of default. *Journal of Housing Economics 46*, 101655.

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.

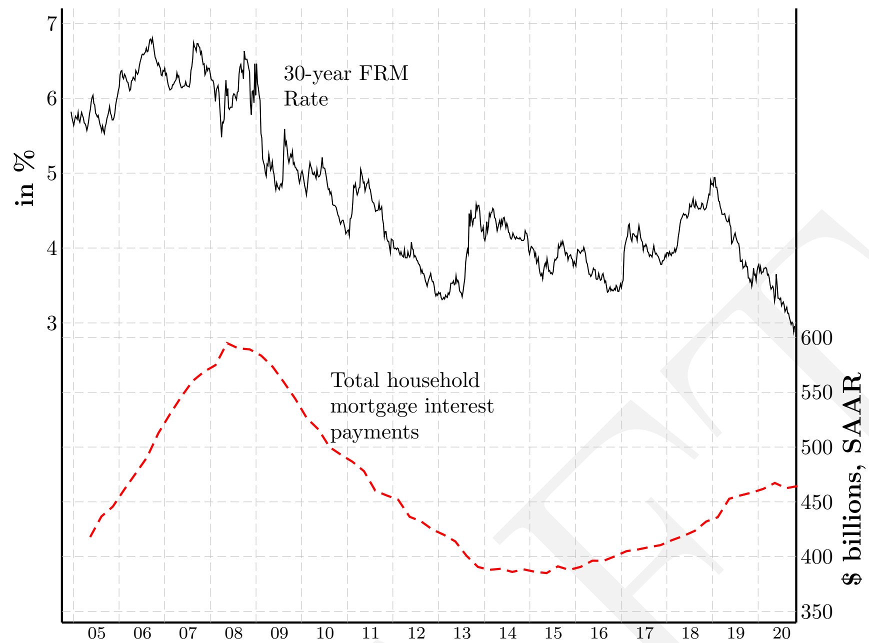
Schill, M. H. and S. M. Wachter (1994). Borrower and neighborhood racial and income characteristics and financial institution mortgage application screening. *The Journal of Real Estate Finance and Economics 9*, 223–239.

Tootell, G. M. B. (1996, 11). Redlining in Boston: Do Mortgage Lenders Discriminate Against Neighborhoods?*. *The Quarterly Journal of Economics 111*(4), 1049–1079.

Willen, P. S. and D. H. Zhang (2020). Do lenders still discriminate? a robust methodology for detecting differences in menus. *Working Paper*.

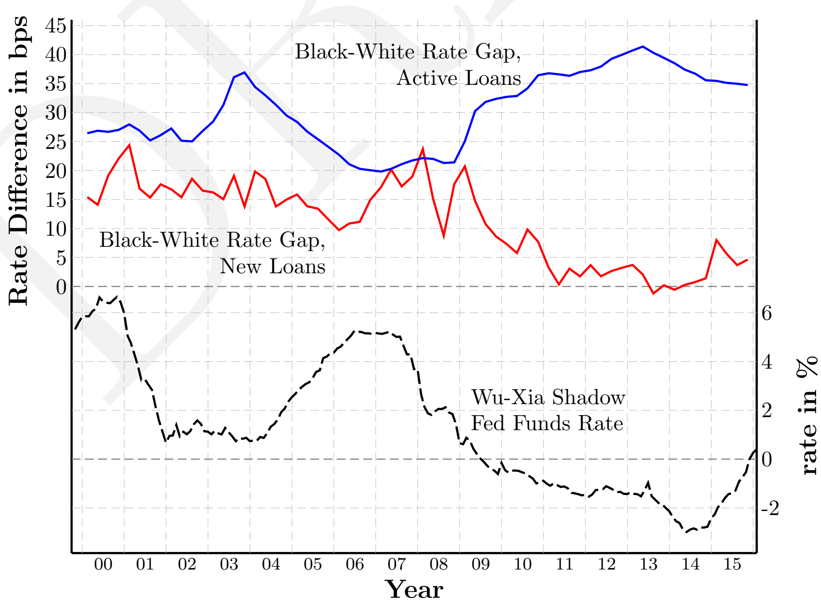
Wong, A. (2019). Refinancing and the transmission of monetary policy to consumption. *Working Paper*.

Figure 1: The evolution of interest rates and mortgage interest payments, 2005-2020.



Notes: The 30-year FRM rate is the Freddie Mac Primary Mortgage Market Survey rate (http://www.freddiemac.com/pmms/). Mortgage Interest paid comes from the Bureau of Economic Analysis (https://www.bea.gov/national/supplementary)

Figure 2: Rates on outstanding mortgages: Black versus Non-Hispanic White Borrowers for mortgages originated between 1996-2015

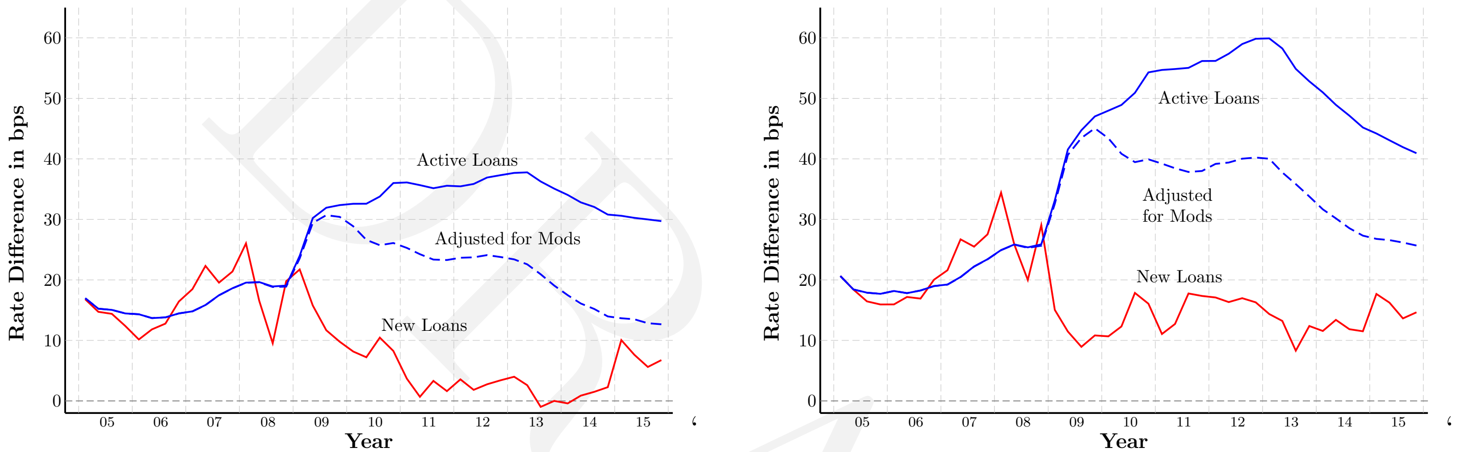


Notes: This figure displays the rate gap for Black and Non-Hispanic White borrowers with 30-year FRMs. New Loans are originated in the quarter and Active loans are all outstanding loans. Data to compute the rate gaps comes from the Black Knight McDash database. The Wu-Xia Shadow Fed Funds rate comes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

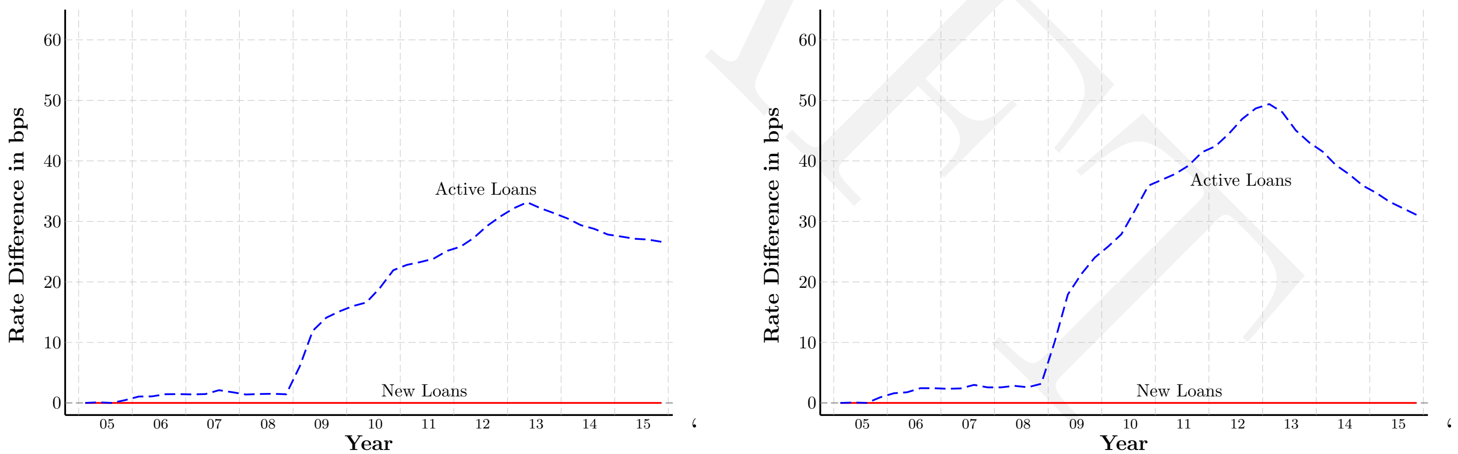
Figure 3: Gap between interest rates for Black and Non-Hispanic White Borrowers for mortgages originated between 2005-2015

1. GSE and FHA Loans **2. GSE Loans Only**

A. With Actual Rates

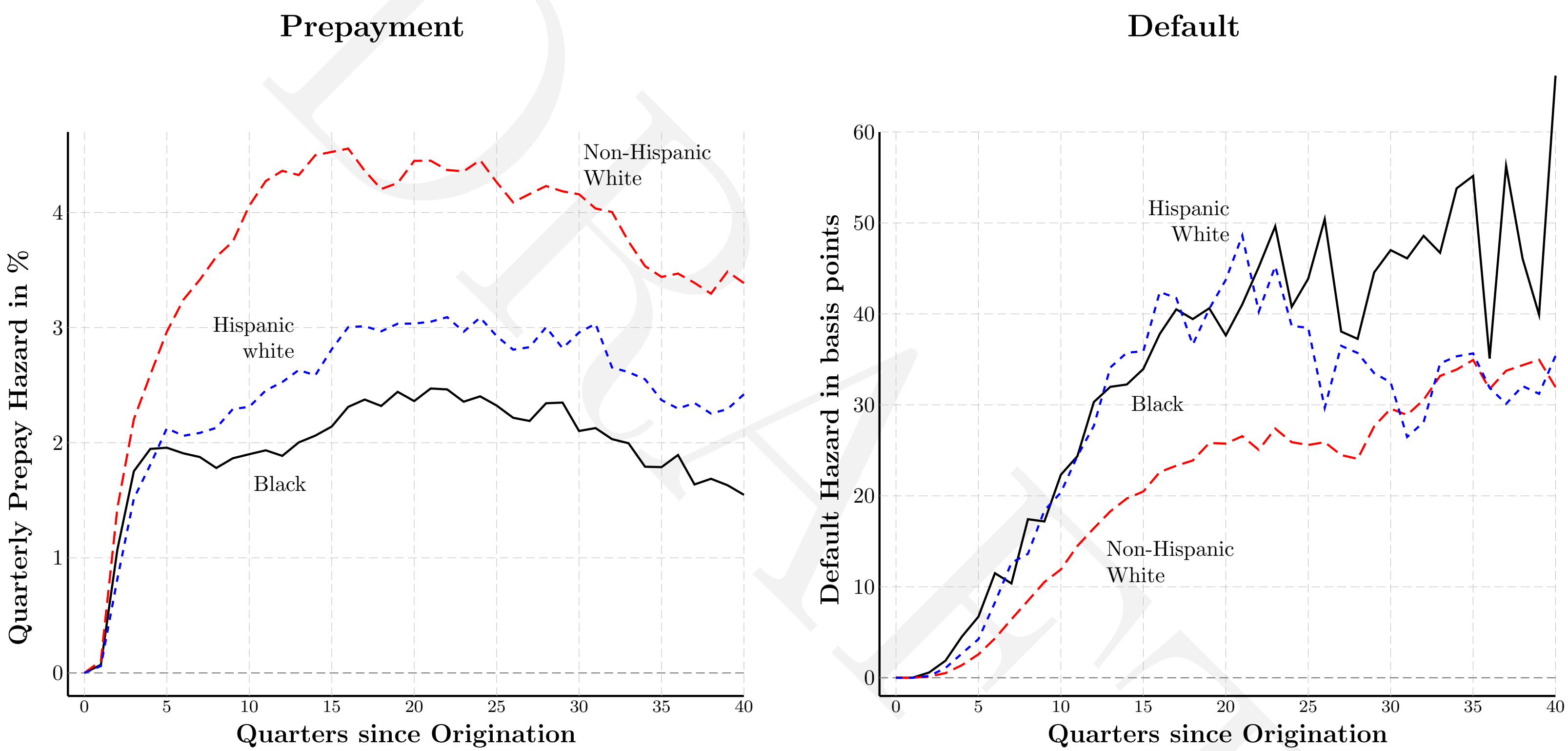


B. Assuming all borrowers receive average quarterly rate at origination



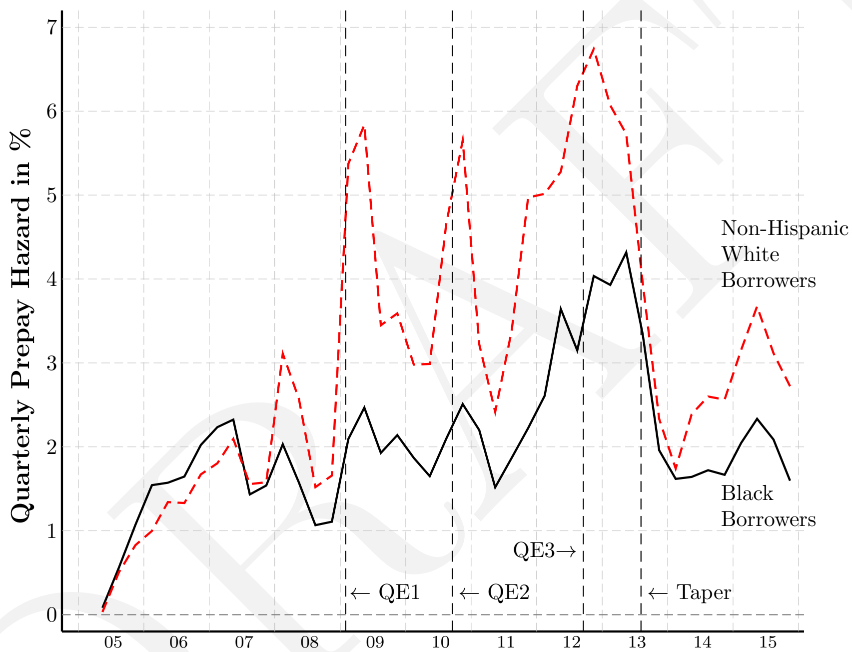
Notes: This figure displays the difference between the average interest rate paid by a Black versus a Non-Hispanic White borrower. "New Loans" are loans originated in the quarter. "Active Loans" are all loans outstanding in the quarter (including new loans.) The top two panels shows the actual rates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

Figure 4: Kaplan Meier Unconditional Prepayment and Default Hazard Rates



Notes: This figure displays the Kaplan-Meier hazard estimates of prepayment 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 Black Knight McDash database.

Figure 5: Unconditional quarterly prepayment hazards for Black and White borrowers.



Notes: Hazard is defined as the percentage of matched HMDA-McDash loans at the beginning of a quarter that prepayed by the end of the quarter. Events are QE1: Annoucement 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 2013 FOMC press conference suggesting that FOMC would wind down purchases of MBS.

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.
FICO (100s points)	7.45	0.53	7.11	0.63	7.26	0.56	7.47	0.52
LTV (%)	72.7	15.9	75.8	15.4	74.1	16.0	72.5	16.0
Loan Amount ($100k)	2.12	1.13	1.84	1.01	1.98	1.02	2.10	1.11
Interest Rate (ppts)	5.20	1.02	5.64	1.10	5.45	1.06	5.17	1.01
Income ($1k)	97.5	63.9	81.4	51.4	79.1	51.7	98.5	64.6
Refinance (d)	0.539	0.498	0.584	0.493	0.514	0.500	0.544	0.498
Condo (d)	0.139	0.346	0.149	0.356	0.138	0.345	0.133	0.340
2-4 Family (d)	0.018	0.133	0.039	0.193	0.040	0.196	0.015	0.120
Low Documentation (d)	0.309	0.462	0.326	0.469	0.313	0.464	0.309	0.462
Non-Occupant Owner (d)	0.140	0.347	0.163	0.369	0.142	0.349	0.136	0.343
Female (d)	0.294	0.456	0.481	0.500	0.310	0.462	0.284	0.451
Co-applicant (d)	0.503	0.500	0.273	0.445	0.354	0.478	0.530	0.499
# Loans	1,076,117		43,882		58,618		909,771	

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	12.8	10.3	14.6	11.5	13.8	11.0	12.6	10.2
Refi Money (ppts)	0.677	0.903	1.050	1.039	0.928	0.989	0.646	0.883
SATO (ppts)	0.151	0.407	0.288	0.477	0.235	0.438	0.139	0.398
Equity (%)	32.2	20.9	26.3	22.5	28.0	25.1	32.7	20.3
Negative Equity (d)	0.047	0.212	0.097	0.296	0.109	0.311	0.040	0.196
Prepay (ppts)	3.37	18.04	2.14	14.46	2.38	15.24	3.49	18.34
Default (ppts)	0.115	3.391	0.226	4.753	0.234	4.828	0.102	3.195
# Loan-quarters	21,546,863		1,023,635		1,302,065		18,057,583	

Notes: This table reports summary statistics from a 10% random sample of loans originated between 2005–2015 (inclusive) and held by the GSEs (Fannie Mae and Freddie Mac) from a matched HMDA-McDash dataset. 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. "Refi Money" is the difference between the mortgage rate and the FHLMC survey rate in the current quarter.

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.
FICO (100s points)	6.88	0.59	6.64	0.60	6.82	0.55	6.92	0.59
LTV (%)	93.6	7.4	93.0	8.2	94.1	7.3	93.6	7.3
Loan Amount ($100k)	1.73	0.91	1.67	0.90	1.68	0.88	1.72	0.89
Interest Rate (ppts)	4.93	1.00	5.09	1.04	4.88	0.98	4.92	0.99
Income ($1k)	65.8	37.3	60.8	33.5	56.5	30.6	67.6	38.1
Refinance (d)	0.295	0.456	0.310	0.463	0.181	0.385	0.314	0.464
Condo (d)	0.114	0.317	0.154	0.361	0.111	0.314	0.105	0.307
2-4 Family (d)	0.015	0.120	0.025	0.156	0.032	0.175	0.010	0.101
Low Documentation (d)	0.191	0.393	0.209	0.406	0.164	0.370	0.193	0.395
Non-Occupant Owner (d)	0.033	0.178	0.034	0.182	0.025	0.156	0.034	0.182
Female (d)	0.352	0.478	0.528	0.499	0.317	0.465	0.333	0.471
Co-applicant (d)	0.415	0.493	0.250	0.433	0.365	0.481	0.446	0.497
# Loans	397,686		42,741		45,336		299,354	

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	13.6	10.7	15.5	11.7	14.2	10.9	13.2	10.4
Refi Money (ppts)	0.684	0.913	0.843	0.974	0.711	0.916	0.659	0.901
SATO (ppts)	0.108	0.347	0.154	0.377	0.154	0.355	0.095	0.340
Equity (%)	16.1	16.8	16.2	18.7	18.8	18.4	15.6	16.1
Negative Equity (d)	0.118	0.322	0.148	0.355	0.107	0.309	0.115	0.319
Prepay (ppts)	2.58	15.86	1.51	12.18	2.00	14.01	2.84	16.62
Default (ppts)	0.238	4.875	0.287	5.346	0.208	4.561	0.238	4.873
# Loan-quarters	8,622,690		1,077,195		1,051,701		6,289,409	

Notes: This table reports summary statistics from a 10% random sample of FHA loans originated between 2005–2015 (inclusive) from a matched HMDA-McDash dataset. 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. "Refi Money" is the difference between the mortgage rate and the FHLMC survey rate in the current quarter.

Table 3: Baseline Prepayment Results

Dependent Variable: Prepay (d)	GSE Loans								FHA Loans		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Black (d)	-1.347***	-1.677***	-1.065***	-1.069***	-0.997***	-1.045***	-0.826***	-0.883***	-1.339***	-1.073***	-0.893***
	(0.112)	(0.125)	(0.059)	(0.053)	(0.050)	(0.051)	(0.037)	(0.040)	(0.048)	(0.044)	(0.036)
Hispanic White (d)	-1.105***	-1.353***	-0.967***	-1.043***	-0.949***	-0.935***	-0.726***	-0.756***	-0.841***	-0.990***	-0.846***
	(0.124)	(0.151)	(0.075)	(0.080)	(0.074)	(0.072)	(0.054)	(0.055)	(0.094)	(0.043)	(0.038)
Asian (d)	0.247	0.385**	0.051	0.114	0.157	0.111	0.015	-0.016	0.243**	-0.293***	-0.365***
	(0.152)	(0.170)	(0.105)	(0.118)	(0.118)	(0.112)	(0.081)	(0.077)	(0.115)	(0.078)	(0.069)
FICO Score			0.627***								
			(0.102)								
LTV Ratio			0.002								
			(0.003)								
Loan Amount			0.746***								
			(0.084)								
Equity			0.025***	0.025***	0.025***	0.083***	0.082***	0.071***		0.071***	0.071***
			(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)		(0.004)	(0.005)
Refinance (d)			-0.497***	-0.524***	-0.503***	-0.575***	-0.562***	-0.601***		-0.403***	-0.383***
			(0.075)	(0.080)	(0.079)	(0.072)	(0.072)	(0.070)		(0.041)	(0.040)
Female (d)					-0.079***	-0.078***	-0.097***	-0.088***		-0.093***	-0.129***
					(0.014)	(0.014)	(0.015)	(0.015)		(0.018)	(0.018)
Refi Money						2.700***	2.704***	1.099***		1.240***	1.229***
						(0.160)	(0.159)	(0.123)		(0.138)	(0.140)
SATO						-1.640***	-1.570***			-0.174	-0.030
						(0.147)	(0.144)			(0.124)	(0.126)
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
Vintage Year-Qtr FE		X	X	X	X	X	X	X		X	X
State FE			X	X	X	X				X	
Zip Code FE							X				X
Zip Code-by-Year-Qtr FE								X			
# Observations	21,546,863	21,546,863	19,754,187	19,777,147	19,122,272	19,122,272	19,122,272	18,978,349	8,622,690	7,039,152	7,039,152
# Loans	1,076,117	1,076,117	979,938	980,688	949,567	949,567	949,567	935,939	397,686	323,391	323,391
R^2	0.000	0.009	0.013	0.013	0.013	0.018	0.020	0.059	0.001	0.015	0.019

Notes: This table reports LPM estimates of equation (1)—the likelihood of voluntary mortgage prepayment on a set of race/ethnicity indicator variables. The estimation is performed at the quarterly frequency on a 10% random sample of loans from a matched HMDA-McDash dataset. The unit of observation is a loan-quarter. Underwriting variables include FICO, LTV, loan amount, mark-to-market equity, indicators for condos and 2-4 multi-family properties, low documentation loans, non-owner occupant properties, and refinance loans. HMDA variables include borrower income and indicators for gender and co-applicants. All columns except (1),(8), and (9) 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 the difference between the mortgage rate and the current FHLMC survey rate. Standard errors are double-clustered by county and vintage year-quarter. (*** p<0.01, ** p<0.05, * p< 0.1)

Table 4: Baseline Default Results

Dependent Variable: Default (d)	GSE Loans								FHA Loans		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Black (d)	0.124***	0.060***	0.010	0.004	-0.003	-0.008	-0.031***	-0.032***	0.049***	-0.046***	-0.058***
	(0.020)	(0.014)	(0.010)	(0.010)	(0.009)	(0.009)	(0.008)	(0.008)	(0.011)	(0.012)	(0.012)
Hispanic White (d)	0.131***	0.084***	0.023**	0.023**	0.016*	0.013*	0.017**	0.017**	-0.030***	-0.050***	-0.046***
	(0.028)	(0.021)	(0.009)	(0.009)	(0.008)	(0.008)	(0.007)	(0.008)	(0.010)	(0.008)	(0.008)
Asian (d)	-0.017**	0.002	0.009*	0.015***	0.008*	0.009**	0.003	0.003	-0.095***	-0.019**	-0.025**
	(0.007)	(0.007)	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.012)	(0.008)	(0.009)
FICO Score			-0.100***								
			(0.014)								
LTV Ratio			-0.010***								
			(0.001)								
Loan Amount			0.005**								
			(0.002)								
Equity			-0.014***	-0.014***	-0.014***	-0.013***	-0.013***	-0.016***		-0.013***	-0.014***
			(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)		(0.001)	(0.001)
Refinance (d)			-0.005	0.001	0.003	-0.006	-0.007	-0.005		0.036***	0.039***
			(0.004)	(0.003)	(0.003)	(0.005)	(0.005)	(0.004)		(0.011)	(0.011)
Female (d)					-0.013***	-0.013***	-0.012***	-0.011***		-0.024***	-0.024***
					(0.003)	(0.003)	(0.003)	(0.003)		(0.004)	(0.005)
Refi Money						0.051***	0.051***	0.114***		0.048***	0.043***
						(0.008)	(0.008)	(0.020)		(0.014)	(0.014)
SATO						0.070***	0.066***			0.066***	0.065***
						(0.018)	(0.017)			(0.020)	(0.019)
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
Vintage Year-Qtr FE		X	X	X	X	X	X	X		X	X
State FE			X	X	X	X				X	
Zip Code FE							X				X
Zip Code-by-Year-Qtr FE								X			
# Observations	21,546,863	21,546,863	19,754,187	19,777,147	19,122,272	19,122,272	19,122,272	18,978,349	8,622,690	7,039,152	7,039,152
# Loans	1,076,117	1,076,117	979,938	980,688	949,567	949,567	949,567	935,939	397,686	323,391	323,391
R^2	0.0001	0.0023	0.0049	0.0050	0.0050	0.0052	0.0066	0.051	0.0000	0.0047	0.0073

Notes: This table reports LPM estimates of equation (1)—the likelihood of mortgage default on a set of race/ethnicity indicator variables. The estimation is performed at the quarterly frequency on a 10% random sample of loans from a matched HMDA-McDash dataset. The unit of observation is a loan-quarter. Underwriting variables include FICO, LTV, loan amount, mark-to-market equity, indicators for condos and 2-4 multi-family properties, low documentation loans, non-owner occupant properties, and refinance loans. HMDA variables include borrower income and indicators for gender and co-applicants. All columns except (1),(8), and (9) include a 3rd order polynomial for the number of quarters since origination (loan age). "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 the difference between the mortgage rate and the current FHLMC survey rate. Standard errors are double-clustered by county and vintage year-quarter. (*** $p<0.01$, ** $p<0.05$, * $p< 0.1$)

Table 5: Prepayment with Interaction Effects

Dependent Variable: Prepay (d)	GSE Loans			FHA Loans		
	(1)	(2)	(3)	(4)	(5)	(6)
Black (d)	-1.058***	0.073	0.005	-1.071***	-0.726***	-0.725***
	(0.052)	(0.11)	(0.125)	(0.045)	(0.114)	(0.117)
Hispanic White (d)	-0.945***	-0.09	-0.162	-0.994***	-0.721***	-0.725***
	(0.072)	(0.098)	(0.112)	(0.043)	(0.088)	(0.090)
Refi Money	2.637***	2.778***	2.794***	1.219***	1.328***	1.348***
	(0.155)	(0.162)	(0.165)	(0.136)	(0.147)	(0.150)
SATO	-1.588***	-1.559***	-1.676***	-0.164***	-0.156***	-0.286***
	(0.143)	(0.139)	(0.155)	(0.122)	(0.121)	(0.135)
Black * Refi Money		-1.158***	-1.281***		-0.428***	-0.502***
		(0.065)	(0.085)		(0.069)	(0.084)
Hispanic White * Refi Money		-0.981***	-1.088***		-0.386***	-0.456***
		(0.059)	(0.076)		(0.059)	(0.070)
Black * SATO			0.730***			0.445***
			(0.108)			(0.109)
Hispanic White * SATO			0.757***			0.419***
			(0.103)			(0.098)
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
State FE	X	X	X	X	X	X
# Observations	18,027,494	18,027,494	18,027,494	6,861,142	6,861,142	6,861,142
R^2	0.018	0.019	0.019	0.015	0.015	0.015

Notes: This table reports LPM estimates of equation (2). The estimation is performed at the quarterly frequency on a 10% random sample of loans from a matched HMDA-McDash dataset. The unit of observation is a loan-quarter. Underwriting variables include FICO, LTV, loan amount, mark-to-market equity, indicators for condos and 2-4 multi-family properties, low documentation loans, non-owner occupant properties, and refinance loans. HMDA variables include borrower income and indicators for gender and co-applicants. All columns except (1) and (8) include a 3rd order polynomial for the number of quarters since origination (loan age). "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 the difference between the mortgage rate and the current FHLMC survey rate. Standard errors are double-clustered by county and vintage year-quarter. (*** p<0.01, ** p<0.05, * p< 0.1)

Table 6: Effect of QE1 on Prepayment Differences

Dependent Variable: Prepay (d)						
Window around QE:		6 Months			1 Year	
	(1)	(2)	(3)	(4)	(5)	(6)
Black (d)	-0.503***	0.843***	0.350***	-0.755***	0.310**	-0.048
	(0.096)	(0.154)	(0.115)	(0.144)	(0.132)	(0.104)
Hispanic White (d)	-0.739***	0.777***	0.488***	-0.882***	0.265**	0.057
	(0.100)	(0.158)	(0.138)	(0.128)	(0.122)	(0.108)
postQE1 (d)	4.023***	5.093***	4.057***	2.350***	3.829***	3.149***
	(0.793)	(0.442)	(0.523)	(0.727)	(0.410)	(0.529)
Black * postQE1	-2.827***	-3.411***	-2.464***	-1.625***	-2.268***	-1.595***
	(0.667)	(0.387)	(0.289)	(0.564)	(0.297)	(0.222)
Hispanic White * postQE1	-2.794***	-3.253***	-2.705***	-1.739***	-2.165***	-1.779***
	(0.643)	(0.346)	(0.303)	(0.552)	(0.275)	(0.244)
$660 \leq$ FICO < 740 (d)		0.482***	0.296***		0.583***	0.396***
		(0.088)	(0.086)		(0.064)	(0.056)
FICO ≥ 740 (d)		1.720***	0.160		1.518***	0.389***
		(0.138)	(0.109)		(0.128)	(0.105)
postQE1 * ($660 \leq$ FICO < 740)			0.481**			0.507***
			(0.177)			(0.134)
postQE1 * (FICO ≥ 740)			3.083***			2.242***
			(0.334)			(0.242)
Constant	1.590***	-17.915***	-17.401***	2.186***	-6.755***	-6.513***
	(0.181)	(2.211)	(2.202)	(0.315)	(0.601)	(0.615)
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
State FE		X	X		X	X
# Observations	1,462,695	1,342,600	1,342,600	2,916,507	2,673,925	2,673,925
R^2	0.012	0.040	0.042	0.005	0.030	0.032

Notes: This table reports LPM estimates of equation (3). The estimation is performed at the quarterly frequency on a 10% random sample of GSE 30-year FRMs from a matched HMDA-McDash dataset. The unit of observation is a loan-quarter. Underwriting variables include FICO, LTV, loan amount, mark-to-market equity, indicators for condos and 2-4 multi-family properties, low documentation loans, non-owner occupant properties, and refinance loans. HMDA variables include borrower income and indicators for gender and co-applicants. “QE_1” 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 7: Effect of QE1 on Differences in the Stock of Outstanding Mortgage Rates

Dependent Variable: Mortgage Rate						
Window around QE:	6-Month		1-Year		2-Year	
	(1)	(2)	(3)	(4)	(5)	(6)
Black (d)	0.232***	0.202***	0.229***	0.199***	0.216***	0.188***
	(0.016)	(0.015)	(0.016)	(0.015)	(0.016)	(0.014)
Hispanic White (d)	0.134***	0.110***	0.132***	0.107***	0.124***	0.100***
	(0.019)	(0.016)	(0.019)	(0.016)	(0.019)	(0.016)
postQE1 (d)	-0.220***	-0.002***	-0.320***	-0.002***	-0.472***	-0.006***
	(0.004)	(0.000)	(0.005)	(0.000)	(0.007)	(0.001)
Black * postQE1	0.120***	-0.007***	0.166***	-0.011***	0.234***	-0.007**
	(0.004)	(0.001)	(0.005)	(0.002)	(0.007)	(0.003)
Hispanic White * postQE1	0.118***	0.003**	0.162***	0.003	0.219***	0.006*
	(0.005)	(0.002)	(0.007)	(0.002)	(0.010)	(0.003)
Constant	6.239***	6.135***	6.246***	6.086***	6.255***	5.998***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Vintage Year-Qtr FE		X		X		X
# Observations	1,462,695	1,462,695	2,916,507	2,916,507	5,610,628	5,610,628
R^2	0.046	0.534	0.073	0.589	0.118	0.661

Notes: This table reports LPM estimates of equation (4). The estimation is performed at the quarterly frequency on a 10% random sample of GSE 30-year FRMs from a matched HMDA-McDash dataset. The unit of observation is a loan-quarter. "QE_1" 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

Internet 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 Match Rates

In this section we present some details regarding the merge between the HMDA and McDash databases. The merge was performed by the RADAR group at the Federal Reserve Bank of Philadelphia. It includes loans originated between 1992 and 2015 (inclusive). Tables A.1 and A.2 display match rates over time, where 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. 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% in 2004 to 81% in 2006. When servicers are added to the McDash database, they typically only provide information on 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 5 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 (2005–2015) our sample covers between 34% and 47% 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.2 Sample Restrictions

Table A.3 below displays all of the restrictions that we impose in constructing our sample. The regressions in Tables 3–5 are estimated using a 10% of our final sample. We adopt most of the restrictions implemented in Fuster et al. (2018).

Table A.3: Sample Restrictions

# Initial Loans = 42,379,615 Sample Restriction:	# Loans Lost	# Loans Remaining
Fixed Rate Loans	6,610,527	35,769,088
First Liens	1,129,607	34,639,481
No Prepayment Penalty	645,796	33,993,685
Fully Amortizing Loans	6,816	33,986,869
$20 \leq$ LTV ≤ 100	3,343,779	30,643,090
No Home Improvement Loans	67,862	30,575,228
Seasoning ≤ 6 Months	4,038,490	26,536,738
Loan Amount $\leq$ \$1m	75,078	26,461,660
Occupancy Non-missing	61,844	26,399,816
Income $\leq$ \$500k	194,705	26,205,111
Term = 30 years	5,918,892	20,286,219
Black, Hispanic White, Asian, and White Borrowers	2,858,308	17,427,911
GSE and FHA Loans	2,656,763	14,771,148
$3\% \leq$ Mortgage Rate $\leq 8\%$	33,115	14,738,033

A.3 LPM Estimates for All Covariates

In Table A.4 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.4: Baseline Prepayment Results with All Covariates

Dependent Variable: Prepay (d)	GSE Loans					FHA Loans
	(3)	(4)	(5)	(6)	(7)	(9)
Black (d)	-1.069***	-1.066***	-0.985***	-1.039***	-0.866***	-1.073***
	(0.062)	(0.056)	(0.051)	(0.051)	(0.040)	(0.050)
Hispanic White (d)	-0.963***	-1.033***	-0.936***	-0.918***	-0.763***	-0.941***
	(0.071)	(0.076)	(0.068)	(0.069)	(0.047)	(0.051)
Asian (d)	0.088	0.148	0.188	0.145	0.022	-0.210***
	(0.100)	(0.113)	(0.114)	(0.107)	(0.074)	(0.075)
Qtrs since Orig	0.598***	0.594***	0.596***	0.383***	0.387***	0.333***
	(0.066)	(0.066)	(0.066)	(0.027)	(0.027)	(0.019)
Qtrs since Orig^2	-0.023***	-0.023***	-0.023***	-0.018***	-0.018***	-0.015***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.001)
Qtrs since Orig^3	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Refi Money				2.705***	2.699***	1.251***
				(0.163)	(0.160)	(0.145)
SATO				-1.668***	-1.619***	-0.152
				(0.151)	(0.146)	(0.138)
Refinance (d)	-0.503***	-0.530***	-0.511***	-0.583***	-0.573***	-0.388***
	(0.082)	(0.087)	(0.087)	(0.080)	(0.078)	(0.041)
Condo (d)	-0.401***	-0.434***	-0.421***	-0.451***	-0.652***	-0.242***
	(0.079)	(0.082)	(0.080)	(0.077)	(0.094)	(0.064)
2-4 Family (d)	-1.386***	-1.214***	-1.156***	-1.500***	-1.480***	-0.287***
	(0.135)	(0.118)	(0.116)	(0.118)	(0.113)	(0.078)
Prop Type Missing (d)	0.915***	0.894***	0.895***	0.871***	0.857***	0.457***
	(0.120)	(0.122)	(0.122)	(0.119)	(0.113)	(0.106)

Continued on next page

Table A.4 – continued from previous page

Dependent Variable: Prepay (d)	GSE Loans					FHA Loans
	(3)	(4)	(5)	(6)	(7)	(9)
Low Documentation (d)	0.229***	0.210***	0.207***	0.156**	0.162**	-0.036
	(0.074)	(0.074)	(0.073)	(0.072)	(0.073)	(0.091)
Documentation Missing (d)	1.708***	1.707***	1.705***	1.704***	1.722***	0.867***
	(0.279)	(0.280)	(0.281)	(0.276)	(0.280)	(0.317)
Non-Occupant Owner (d)	0.254***	0.192***	-0.060	-0.291***	0.004	4.440***
	(0.043)	(0.049)	(0.063)	(0.073)	(0.056)	(0.800)
FICO	0.627***					
	(0.102)					
LTV	0.002					
	(0.003)					
LTV = 80 (d)	-0.018					
	(0.027)					
Equity	0.746***					
	(0.084)					
Negative Equity (d)	-2.385***					
	(0.146)					
Loan Amount	0.007***					
	(0.001)					
$600 < \text{FICO} \leq 620$ (d)		0.021	-0.006	0.083	0.116*	0.394***
		(0.057)	(0.060)	(0.066)	(0.059)	(0.059)
$620 < \text{FICO} \leq 640$ (d)		0.162**	0.145**	0.313***	0.345***	0.727***
		(0.061)	(0.064)	(0.071)	(0.067)	(0.087)
$640 < \text{FICO} \leq 660$ (d)		0.442***	0.413***	0.637***	0.668***	1.112***
		(0.072)	(0.078)	(0.082)	(0.076)	(0.118)
$660 < \text{FICO} \leq 680$ (d)		0.629***	0.592***	0.870***	0.912***	1.338***
		(0.085)	(0.090)	(0.097)	(0.093)	(0.126)

Continued on next page

Table A.4 – continued from previous page

Dependent Variable: Prepay (d)						
	GSE Loans					FHA Loans
	(3)	(4)	(5)	(6)	(7)	(9)
680 < FICO ≤ 700 (d)		0.808***	0.764***	1.111***	1.140***	1.602***
		(0.093)	(0.099)	(0.107)	(0.101)	(0.127)
700 < FICO ≤ 720 (d)		0.958***	0.902***	1.306***	1.346***	1.747***
		(0.113)	(0.120)	(0.128)	(0.123)	(0.138)
720 < FICO ≤ 740 (d)		1.086***	1.028***	1.496***	1.527***	1.919***
		(0.125)	(0.130)	(0.139)	(0.135)	(0.151)
740 < FICO ≤ 760 (d)		1.317***	1.255***	1.759***	1.781***	1.997***
		(0.140)	(0.144)	(0.154)	(0.150)	(0.154)
760 < FICO ≤ 780 (d)		1.475***	1.405***	1.927***	1.946***	2.142***
		(0.159)	(0.162)	(0.176)	(0.172)	(0.163)
780 < FICO ≤ 800 (d)		1.503***	1.440***	1.974***	1.981***	2.121***
		(0.174)	(0.177)	(0.195)	(0.190)	(0.172)
800 < FICO ≤ 820 (d)		1.279***	1.258***	1.799***	1.770***	1.930***
		(0.170)	(0.176)	(0.194)	(0.186)	(0.166)
FICO > 820 (d)		0.782***	0.873***	1.434***	1.294***	0.858**
		(0.156)	(0.172)	(0.179)	(0.168)	(0.398)
25 < LTV ≤ 30 (d)		0.108	0.095	0.348***	0.440***	0.173
		(0.089)	(0.095)	(0.095)	(0.088)	(0.696)
30 < LTV ≤ 35 (d)		0.277***	0.280***	0.795***	0.929***	-0.164
		(0.098)	(0.102)	(0.101)	(0.107)	(0.804)
35 < LTV ≤ 40 (d)		0.370***	0.370***	1.145***	1.330***	-0.215
		(0.112)	(0.115)	(0.115)	(0.128)	(0.629)
40 < LTV ≤ 45 (d)		0.389***	0.399***	1.432***	1.639***	-0.586
		(0.112)	(0.116)	(0.125)	(0.142)	(0.669)
45 < LTV ≤ 50 (d)		0.452***	0.477***	1.770***	1.999***	0.362
		(0.132)	(0.135)	(0.142)	(0.163)	(0.724)

Continued on next page

Table A.4 – continued from previous page

Dependent Variable: Prepay (d)

	GSE Loans					FHA Loans
	(3)	(4)	(5)	(6)	(7)	(9)
50 < LTV ≤ 55 (d)		0.495***	0.526***	2.077***	2.317***	-0.016
		(0.147)	(0.150)	(0.158)	(0.182)	(0.690)
55 < LTV ≤ 60 (d)		0.512***	0.542***	2.345***	2.614***	0.685
		(0.161)	(0.164)	(0.176)	(0.201)	(0.710)
60 < LTV ≤ 65 (d)		0.580***	0.603***	2.626***	2.922***	0.936
		(0.167)	(0.171)	(0.184)	(0.212)	(0.711)
65 < LTV ≤ 70 (d)		0.520***	0.559***	2.826***	3.125***	0.924
		(0.185)	(0.190)	(0.203)	(0.232)	(0.726)
70 < LTV ≤ 75 (d)		0.511**	0.543***	3.017***	3.329***	1.122
		(0.192)	(0.197)	(0.207)	(0.236)	(0.692)
75 < LTV ≤ 80 (d)		0.529**	0.565***	3.278***	3.585***	1.389*
		(0.204)	(0.207)	(0.215)	(0.242)	(0.717)
80 < LTV ≤ 85 (d)		0.416*	0.459**	3.276***	3.601***	1.538**
		(0.211)	(0.215)	(0.211)	(0.238)	(0.725)
85 < LTV ≤ 90 (d)		0.331	0.390	3.500***	3.815***	1.755**
		(0.232)	(0.235)	(0.216)	(0.240)	(0.732)
90 < LTV ≤ 95 (d)		0.409*	0.478*	3.773***	4.096***	2.044***
		(0.242)	(0.245)	(0.224)	(0.246)	(0.737)
95 < LTV ≤ 100 (d)		0.417	0.498*	3.922***	4.245***	2.239***
		(0.249)	(0.252)	(0.237)	(0.259)	(0.747)
85k < Orig Amount ≤ 110k (d)		0.455***	0.401***	0.624***	0.565***	0.483***
		(0.064)	(0.063)	(0.067)	(0.067)	(0.041)
110k < Orig Amount ≤ 125k (d)		0.695***	0.595***	0.906***	0.847***	0.766***
		(0.092)	(0.092)	(0.098)	(0.097)	(0.061)
125k < Orig Amount ≤ 150k (d)		0.885***	0.741***	1.121***	1.069***	1.009***
		(0.108)	(0.106)	(0.114)	(0.112)	(0.082)

Continued on next page

Table A.4 – continued from previous page

Dependent Variable: Prepay (d)						
	GSE Loans					FHA Loans
	(3)	(4)	(5)	(6)	(7)	(9)
150k < Orig Amount ≤ 175k (d)		1.080***	0.878***	1.323***	1.299***	1.258***
		(0.126)	(0.123)	(0.134)	(0.134)	(0.105)
Orig Amount > 175k (d)		1.815***	1.391***	1.945***	1.973***	1.905***
		(0.194)	(0.175)	(0.192)	(0.193)	(0.138)
25k < Income ≤ 50k (d)			0.089**	0.109***		0.303***
			(0.038)	(0.038)		(0.038)
50k < Income ≤ 75k (d)			0.206***	0.204***		0.524***
			(0.044)	(0.045)		(0.040)
75k < Income ≤ 100k (d)			0.377***	0.368***		0.791***
			(0.057)	(0.059)		(0.047)
100k < Income ≤ 125k (d)			0.575***	0.553***		1.040***
			(0.077)	(0.077)		(0.050)
125k < Income ≤ 150k (d)			0.777***	0.742***		1.325***
			(0.097)	(0.095)		(0.091)
150k < Income ≤ 175k (d)			0.854***	0.802***		1.615***
			(0.118)	(0.115)		(0.098)
175k < Income ≤ 200k (d)			0.985***	0.924***		1.675***
			(0.132)	(0.129)		(0.173)
200k < Income ≤ 225k (d)			1.033***	0.958***		1.857***
			(0.146)	(0.140)		(0.160)
225k < Income ≤ 250k (d)			1.086***	1.003***		1.878***
			(0.157)	(0.154)		(0.252)
250k < Income ≤ 275k (d)			1.067***	0.978***		2.012***
			(0.172)	(0.170)		(0.288)
275k < Income ≤ 300k (d)			1.132***	1.049***		1.765***
			(0.176)	(0.174)		(0.372)

Continued on next page

Table A.4 – continued from previous page

Dependent Variable: Prepay (d)						
	GSE Loans					FHA Loans
	(3)	(4)	(5)	(6)	(7)	(9)
300k < Income ≤ 325k (d)			0.973***	0.891***		1.050**
			(0.162)	(0.165)		(0.426)
325k < Income ≤ 350k (d)			0.982***	0.896***		1.290***
			(0.199)	(0.196)		(0.370)
350k < Income ≤ 375k (d)			1.113***	1.014***		1.542***
			(0.181)	(0.177)		(0.511)
375k < Income ≤ 400k (d)			1.144***	1.041***		2.144***
			(0.212)	(0.207)		(0.621)
400k < Income ≤ 425k (d)			0.968***	0.883***		1.431***
			(0.189)	(0.185)		(0.460)
425k < Income ≤ 450k (d)			1.284***	1.155***		1.924**
			(0.223)	(0.218)		(0.756)
450k < Income ≤ 475k (d)			1.102***	1.053***		2.082***
			(0.190)	(0.188)		(0.568)
475k < Income ≤ 500k (d)			1.107***	1.011***		1.180
			(0.208)	(0.200)		(0.936)
Female (d)			-0.067***	-0.069***		-0.092***
			(0.014)	(0.014)		(0.015)
Co-applicant (d)			-0.086***	-0.089***		-0.099***
			(0.014)	(0.013)		(0.017)
Constant	-6.736***	-3.249***	-3.392***	-8.088***	-8.033***	-5.452***
	(1.199)	(0.687)	(0.707)	(0.642)	(0.637)	(0.799)
Underwriting Vars	X	X	X	X	X	X
HMDA Vars			X	X		X
Vintage Year-Qtr FE	X	X	X	X	X	X

Continued on next page

Table A.4 – continued from previous page

Dependent Variable: Prepay (d)						
	GSE Loans					FHA Loans
	(3)	(4)	(5)	(6)	(7)	(9)
State FE	X	X	X	X		X
Zip Code FE					X	
# Observations	19,729,750	19,752,426	19,107,663	19,107,663	19,752,426	7,053,144
# Loans	979,919	980,659	949,874	949,874	949,874	323,519
R^2	0.013	0.013	0.013	0.019	0.020	0.015

Notes: This table reports LPM estimates of equation (1)—the likelihood of voluntary mortgage prepayment on a set of race/ethnicity indicator variables. The estimation is performed at the quarterly frequency on a 10% random sample of loans from a matched HMDA-McDash dataset. The unit of observation is a loan-quarter. Underwriting variables include FICO, LTV, loan amount, mark-to-market equity, indicators for condos and 2-4 multi-family properties, low documentation loans, non-owner occupant properties, and refinance loans. HMDA variables include borrower income and indicators for gender and co-applicants. All columns except (1) and (8) 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 the difference between the mortgage rate and the current FHLMC survey rate. 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 and default results from logit models. These models are estimated on a 1% random sample of our HMDA-McDash matched dataset. Table A.5 contains the prepayment results while Table A.6 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 and 4 in the main text. The only specification that is omitted is column (7), which included Zip Code fixed effects instead of state fixed effects. It was not possible to estimate that specification using the logit framework.

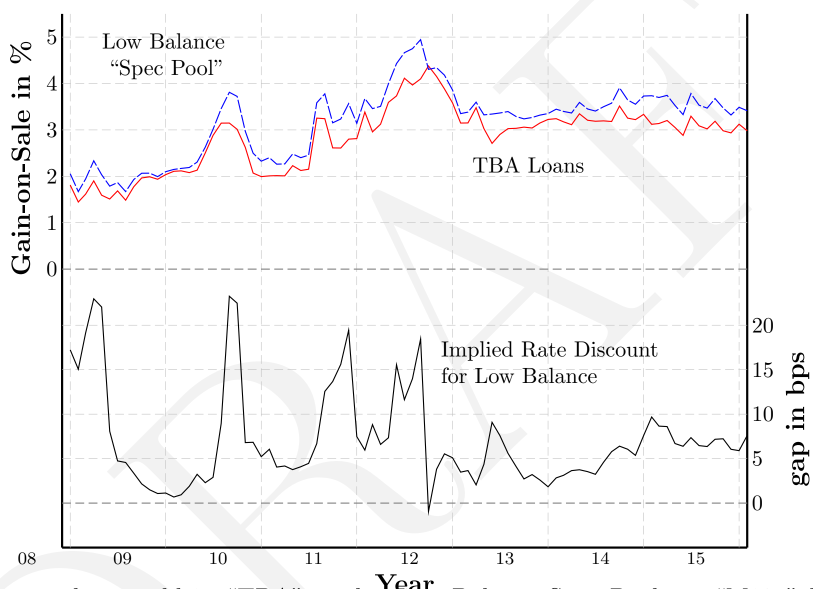
Table A.5: Logit Prepayment Hazard Estimates

Dependent Variable: Prepay (d)	GSE Loans						FHA Loans	
	(1)	(2)	(3)	(4)	(5)	(6)	(8)	(9)
Black (d)	-1.299***	-1.532***	-1.089***	-1.089***	-1.048***	-1.071***	-1.303***	-1.192***
	(0.057)	(0.059)	(0.058)	(0.056)	(0.058)	(0.059)	(0.063)	(0.062)
Hispanic White (d)	-1.053***	-1.232***	-0.904***	-0.962***	-0.887***	-0.871***	-0.857***	-0.936***
	(0.066)	(0.074)	(0.053)	(0.055)	(0.057)	(0.061)	(0.096)	(0.063)
Asian (d)	0.367***	0.519***	0.183*	0.253**	0.310***	0.264***	0.207	-0.421***
	(0.121)	(0.161)	(0.096)	(0.106)	(0.106)	(0.099)	(0.141)	(0.102)
Underwriting Vars			X	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		X
# Observations	2,160,134	2,160,134	1,983,018	1,985,518	1,919,856	1,919,856	864,498	703,847

Table A.6: Logit Default Hazard Estimates

Dependent Variable: Default (d)								
	GSE Loans						FHA Loans	
	(1)	(2)	(3)	(4)	(5)	(6)	(8)	(9)
Black (d)	0.114***	0.024**	-0.002	-0.003	-0.012	-0.014*	0.068***	-0.051***
	(0.019)	(0.011)	(0.009)	(0.009)	(0.008)	(0.008)	(0.020)	(0.015)
Hispanic White (d)	0.156***	0.069***	0.027***	0.027***	0.017*	0.015*	-0.036**	-0.079***
	(0.020)	(0.012)	(0.009)	(0.009)	(0.009)	(0.009)	(0.016)	(0.017)
Asian (d)	-0.012	0.003	0.012	0.012	0.005	0.006	-0.128***	-0.102***
	(0.012)	(0.013)	(0.014)	(0.014)	(0.015)	(0.015)	(0.028)	(0.036)
Underwriting Vars			X	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		X
# Observations	2,163,596	2,163,596	1,976,876	1,979,594	1,900,150	1,900,150	859,011	694,228

Figure A.1: Mortgage pricing for low prepayment loans.



Notes: TBA loans are loans sold in "TBA" pools. Low Balance Spec Pool are "M125" loans defined as loans with balances between \$100k and \$125k. 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 M125 mortgage.