The role of neighborhood characteristics in mortgage default risk: evidence from New York City

Chan, Sewin and Gedal, Michael and Been, Vicki and Haughwout, Andrew

New York University, Federal Reserve Bank of New York

2011

The Role of Neighborhood Characteristics in Mortgage Default Risk:

Evidence from New York City

Sewin Chan[a]

Michael Gedal[b]

Vicki Been[c]

Andrew Haughwout[d]

August 2011

Abstract: Using a rich database of non-prime mortgages from New York City, we find that census tract level neighborhood characteristics are important predictors of default behavior, even after controlling for an extensive set of controls for loan and borrower characteristics. First, default rates increase with the rate of foreclosure notices and the number of lender-owned properties (REOs) in the tract. Second, default rates on home purchase mortgages are higher in census tracts with larger shares of black residents, regardless of the borrower's own race. We explore possible explanations for this second finding and conclude that it likely reflects differential treatment of black neighborhoods by the mortgage industry in ways that are unobserved in our data.

Key words: mortgage, default, neighborhoods, race

[a] Corresponding author. Robert F. Wagner School of Public Service, New York University, 295 Lafayette Street, New York, NY 10012, USA. Tel: +1.212.998.7495. Fax: +1.212.998.3890. sewin.chan@nyu.edu

[b] Robert F. Wagner School of Public Service, New York University. michael.gedal@nyu.edu

[c] New York University School of Law. vicki.been@nyu.edu

[d] Federal Reserve Bank of New York. andrew.haughwout@ny.frb.org

The views represented here are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of New York or the Federal Reserve System.

1. Introduction

The wave of delinquencies and foreclosures that began in 2007 and the financial crisis that it engendered have drawn new attention to the different reasons why households may end up in foreclosure. In this paper, we use a newly assembled dataset from New York City to examine the role of borrower, loan and neighborhood characteristics on default rates of non-prime mortgages. The depth of the dataset allows us to provide a more complete set of controls than previous research. In particular we are able to examine the effects of census tract level neighborhood characteristics while controlling for detailed individual borrower and loan characteristics. The analysis of neighborhood effects is critical as policymakers consider whether and how public policies to address the foreclosure crisis and to regulate lending practices should be tailored to the needs and experiences of different types of neighborhoods.

Many researchers have used LoanPerformance from FirstAmerican CoreLogic, a commercial database that is the major source of non-prime mortgage performance information for the mortgage industry. A major limitation of this database is that its most detailed geographic identifier is the zip code of the mortgaged property. A zip code is a good deal bigger than what is generally thought of as a "neighborhood." Researchers studying neighborhoods typically examine census tracts: in New York City there are about 2,200 census tracts, compared to 180 zip codes, and on average, a tract is 13 city blocks, while a zip code is 160 city blocks.

We have matched LoanPerformance records to actual parcels of land with a high level of precision using deeds from New York City's Department of Finance. This allows us to merge information about borrowers, their payment histories, and the terms of their loans, with information about a borrower's race and ethnicity from Home Mortgage Disclosure Act (HMDA) records, as well as census tract level neighborhood characteristics from a variety of sources. In particular, we have precise information at the census tract level on the number of foreclosures and the share of properties that fail to sell for the lender's reservation price and thus become "real estate owned" (REO). While repeat sales house price indices are not available at the census tract level due to insufficient sales transactions, we use indices for 56 different community districts (political jurisdictions within New York City that average just over four square miles each). We are not aware of any other mortgage research that has examined such detailed information at so fine a level of geography, and that contains a full set of critical loan and borrower characteristics, especially the borrower's credit score and race.

We report results on the effect of loan and borrower risk characteristics on mortgage default. Not only do these estimates broadly confirm earlier results from the literature while using the additional controls provided by our much richer dataset, but the magnitude of these estimates do not change when we add the additional controls. This is an important finding that confirms the validity of existing research on the effect of loan and borrower risk characteristics on non-prime mortgage defaults that does not have the wealth of local level information that we are able to use here.

The primary contribution of the paper, however, is our finding that neighborhood characteristics have a powerful impact on the likelihood of default, even after an extensive set of controls. Moreover, when we add a set of analogous zip code level characteristics, they are generally insignificant and do not eliminate the significance of the census tract level variables, showing that the finer geographic variation that we use adds substantial explanatory information.

Two important neighborhood effects emerge. First, we find that mortgages in census tracts with high foreclosure and REO activity have a substantially higher chance of defaulting. In the body of the paper, we argue that this likely reflects a local contagion effect whereby more information about the default process, or reduced stigma surrounding it, leads to more defaults. In addition, the foreclosure and REO rates may serve as a proxy for very local housing market conditions that are not captured in our community district level house price indices.

Second, we find that home purchase mortgages in neighborhoods with large shares of black residents have a substantially higher chance of defaulting, regardless of the borrower's own race. We explore possible explanations for this finding and conclude that it likely reflects unobserved loan and borrower characteristics that are both correlated with higher default risk and are more prevalent in black neighborhoods. Moreover, we argue that the most plausible candidate for these unobservables is the differential treatment of black neighborhoods by the mortgage industry via marketing, underwriting or loan terms.

2. Background

There is an extensive and growing body of empirical literature on the determinants of mortgage default.[1] Rather than exhaustively review that literature, our goal in this section is to provide context for our analysis of the role that borrower, loan and neighborhood characteristics play in the default of non-prime mortgages.

Non-prime mortgages have traditionally extended credit to borrowers who could not qualify for prime mortgages, and so by their very nature tend to have higher default risk than do prime mortgages. Okah and Orr (2010) compare non-prime and prime mortgages in New York City and find that on average, non-prime borrowers have, at origination, much lower credit scores, higher debt-to-income ratios (DTIs) and loan-to-value ratios (LTVs), as well as substantially higher rates of default. Data from other locations show similar patterns, and several studies find that non-prime

[1] See Mayer et al. (2009) for a recent review.

status is one of the strongest predictors of default and foreclosure (Coulton et al., 2008; Ding et al., 2008; Gerardi et al., 2007).

2.1. The Relationship Between Loan and Borrower Characteristics and Default Risk

The research on both prime and non-prime mortgages has stressed the importance of credit scores, DTIs and LTVs on default behavior. Studies that have information on borrower credit scores find that they play a large role in predicting default (e.g., Amronim and Paulson, 2010; Haughwout et al., 2008). Several studies have found that higher initial DTIs contribute to a higher probability of default, although the effects seem to be less strong than that of LTV, and are somewhat inconsistent over time (e.g., Ding et al., 2008; Foote et al., 2009). Recent research on non-prime mortgages has emphasized that it is the current, and not the initial LTV that increases the probability of delinquency and default (e.g., Demyanyk, 2009). Foreclosures happen less frequently in appreciating markets, most likely because financially-distressed borrowers can more easily sell their properties or refinance and prepay the remaining balance on their loans (e.g., Danis and Pennington-Cross, 2005; Schloemer et al., 2006).

Few models of mortgage default using loan-level data are able to identify and control for borrower race as well as credit history and loan characteristics. In studies that do, there is some evidence that non-white borrowers are more likely to default than white borrowers. For example, Firestone et al. (2007), using data on conforming fixed rate mortgages, find that black and Hispanic borrowers have higher default rates. Jiang et al. (2009), using data from a national mortgage bank, find that delinquency rates are higher for black and Hispanic borrowers, and that the difference is even greater among broker-originated loans.

2.2. The Relationship Between Neighborhood Characteristics and Default Risk

Considerable research documents the negative externalities neighborhoods experience from foreclosures, particularly in terms of the values of nearby housing (e.g., Campbell et al., 2009; Rogers and Winter, 2009; Schuetz et al., 2008). The literature also shows that nearby foreclosures can lead to a heightened probability of default or foreclosure, for a number of reasons (see Campbell et al., 2009; Harding et al., 2009; Hartley, 2010; Immergluck and Smith, 2006; Ioannides, 2003; Lee, 2008; Leonard and Murdoch, 2009; Lin et al., 2009). Negative physical externalities caused by foreclosures, including visible deterioration, maintenance deferral or vandalism, may cause declines in neighboring home values causing potential homebuyers to view the neighborhood as less attractive. Foreclosure-induced mobility may increase the number of homes on the market, and thereby drive prices down. Further, the sale of foreclosed properties at discounted prices may lower property appraisals and sellers' reservation prices for nearby properties.

Research on whether neighborhood racial composition influences default is rather thin. While the limited number of studies using data on individual loans sometimes find a significant relationship between the share of minority residents or homeowners in a census tract and default or foreclosure rates, they are not able to rule out the possibility that minority share is merely a proxy for credit history (e.g., Berkovec et al., 1994; Van Order and Zorn, 2000).

Several recent studies document a relationship between default rates and the share of subprime loans in a zip code. Mian and Sufi (2009) find a positive relationship using national zip code level data. By contrast, using a set of non-prime individual mortgages from Phoenix, Agarwal et al. (2011) find that the zip code share of subprime mortgages *reduces* borrower defaults, which they argue is due to a positive relationship between subprime activity and house price appreciation in Phoenix. They also find that the zip code level concentration of riskier mortgages (hybrid ARMs and no- or low-documentation loans) increases the probability of borrower default.

3. Data Description

To investigate the determinants of default, we begin with all first lien hybrid 2/28 and 3/27 adjustable rate mortgages (ARMs)[2] and 30-year fixed rate mortgages (FRMs) originated in New York City from 2004 to 2007 in LoanPerformance, a database that covers over 90 percent of all non-prime securitized mortgages in the United States over this time period.[3] We observe monthly updates on these loans until December 2009. Although LoanPerformance provides detailed information on borrower characteristics, loan terms, and payment history, it contains no information on borrower race and provides little in terms of neighborhood characteristics. We therefore matched the LoanPerformance database to New York City mortgage deeds. This told us about additional liens on the mortgaged property and its exact location, which in turn allowed us to merge on a variety of additional variables at the census tract level. We also merged on Home Mortgage Disclosure Act (HMDA) data which gave us additional borrower information including race and ethnicity. Details of these merges and the additional data sources are provided in the Data Appendix.

In the analysis below, we use the 78 percent of LoanPerformance hybrid ARMs and FRMs that matched both the deeds records and a unique loan in the HMDA database. This sample is not significantly different from the full universe in terms of the loan, borrower, and neighborhood characteristics that we use in the analyses below. Figure 1 plots the number of originations by quarter.

---

[2] 2/28s are 30-year loans with an initial rate that remains in effect for the first 2 years, and then is reset every six months for the remaining 28 years, while on a 3/27, the initial rate is in effect for 3 years and floats for 27 years.
[3] These hybrid ARMs and FRMs represent almost two-thirds of all first lien LoanPerformance mortgages in New York City. Because LoanPerformance includes only securitized loans, any inferences should be limited to such loans.

3.1. Loan and Borrower Characteristics

Table 1 displays summary statistics for loan and borrower characteristics. Our sample is split roughly equally between FRMs and hybrid ARMs. In terms of borrower characteristics, ARM borrowers tend to have lower FICO scores[4], higher debt-to-income ratios (DTIs) and higher combined loan-to-value ratios (LTVs) at origination than FRM borrowers. The combined LTV measure is reported in LoanPerformance and includes any other liens in existence at the time of origination. Any new liens taken out afterwards will not be reflected in this measure; however, we have information on these liens from the mortgage deeds: 5 percent of ARM borrowers and 11 percent of FRM borrowers took on additional debt against the same property (in any form, including second mortgages, home equity loans and lines of credit) that totaled at least 5 percent of the first lien's original loan amount. Only about 40 percent of borrowers provided full loan documentation.

We rely on HMDA for some additional borrower characteristics: a majority of primary borrowers in our sample are male, and 21 percent of ARMs and 28 percent of the FRMs included a coborrower. Almost half of ARM borrowers and over one third of FRM borrowers are black.[5] By contrast, blacks made up just 20 percent of borrowers who originated a loan in New York City during 2004-2007, according to HMDA.

3.2. Neighborhood Characteristics

Figure 2 provides a snapshot of the distribution of foreclosures across New York City in 2009. Each dot represents one notice of foreclosure for a 1-4 family building. The map shows that foreclosures are concentrated, especially in neighborhoods where the majority of residents are black,

[4] The Fair Isaac Corporation (FICO) credit score is the most widely used credit score model in the U.S. and takes into account payment history, credit utilization, length of credit history, types of credit and recent searches for credit.

[5] We use these non-missing, mutually exclusive race/ethnicity categories: "white" (white non-Hispanic), "black" (black non-Hispanic), "Hispanic" (Hispanic white), "Asian" (Asian non-Hispanic), and "other" (all others).

according to the 2000 Census.[6] To investigate the link between neighborhood racial composition and foreclosures, we constructed a neighborhood foreclosure rate measure defined as the number of foreclosure notices (*lis pendens*) issued on 1-4 family buildings in a census tract during a six-month period, divided by the stock of 1-4 family buildings in that tract. Figure 3 shows average neighborhood foreclosure rates on a semiannual basis, broken down by neighborhood racial composition. Overall, neighborhood foreclosure rates in New York City have trended upwards since 2004. In tracts where more than 60 percent of residents are black, the average foreclosure rate in the first half of 2004 was 0.4 percent, almost double the rate in tracts that are less than 40 percent black. By the last half of 2009, foreclosure rates in black neighborhoods had doubled to about 0.8 percent, compared to 0.4 percent for non-black neighborhoods.

Table 2 summarizes the distribution of loans in our sample across neighborhoods in terms of demographics in 2000. A considerable number of loans in our sample are in tracts that were at least 60 percent black in 2000: almost half of ARMs and 41 percent of FRMs. This may be surprising considering that blacks made up just one quarter of New York City residents; but it likely reflects both the higher proportion of blacks in our non-prime sample (as noted earlier), and the fact that there are relatively high levels of residential racial segregation in New York City, so that black borrowers are more likely than non-black borrowers to live in neighborhoods that have a high concentration of black residents. In tracts that are predominantly non-black, FRMs are more prevalent than ARMs, while in predominantly black tracts, the majority of loans are ARMs.

Table 2 also considers the neighborhood share of non-prime loans at origination, calculated as the fraction of non-prime loans originated in LoanPerformance divided by total loans originated in HMDA, during the 2 years preceding the loan's origination month. The vast majority of loans in

[6] Throughout the paper, we classify neighborhoods based on the distribution of residents by race/ethnicity as follows: "black" residents include non-Hispanic blacks, "white" residents include non-Hispanic whites, "Asian" residents include non-Hispanic Asians, and "Hispanic" residents include all individuals reporting Hispanic origin regardless of race.

our sample are in tracts where fewer than 30 percent of mortgages were non-prime in the 2 years preceding origination.

The final two indicators in table 2 are measured across all loan-months in our sample. While the majority of loan-month observations are in census tracts that experienced a foreclosure rate of less than one percent in the six months prior, about one in five loan-months are in neighborhoods where the foreclosure rate was 2 percent or more. The dynamic neighborhood REO rate is calculated as the number of properties listed as being in REO at any point in time during the six months preceding the month of analysis, divided by the total number of 1-4 family buildings in the census tract. While the REO rate measures the *stock* of properties that were in REO at any point during the preceding six months, the foreclosure rate measures the *flow* of new properties into foreclosure during the preceding six months. About two-thirds of loan-months were in census tracts where the REO rate was less than 1 percent, while 14 percent of loan-months experienced neighborhood REO rates above 3 percent.[7]

3.3. Default Rates

We examine the hazard of default for each month since origination, with default defined as 90 days of delinquency. This definition is used in much of the literature on mortgage default risk as it is entirely in the borrower's control and excludes the behavior of the lender or servicer. The hazard is simply the probability that a loan enters default, conditional on not having defaulted earlier. Figure 4 shows hazard rates up to December 2009, by the year the loan was originated. Two clear patterns emerge. First, ARMs tend to experience much higher default rates on average than FRMs that were originated in the same year. For example, among 2004 originations, the default

[7] The higher stock of REOs compared with the flow of foreclosures reflects a large fraction of REO properties remaining in REO for long periods of time. For example, of properties that entered REO in 2007, fewer than half had exited within 12 months (Furman Center 2010).

hazard in month 18 after origination was 0.4 percent for FRMs, compared to 1.0 percent for ARMs. Second, for both loan types, later originations are more likely to experience default. The underlying risk characteristics of loans may have been changing over time. In addition, adverse changes in the macroeconomic environment - in particular, rising unemployment and weak house price appreciation - may have contributed to the striking increase in default risk for loans originated in more recent years.

Figure 5 plots default hazards by borrower race for census tracts that are predominantly non-black (less than 40 percent black) and neighborhoods that are predominantly black (more than 60 percent black). These displayed hazards do not control for any variables. In the non-black neighborhoods, black borrowers experience default rates that are higher than non-black borrowers, both for ARMs and FRMs. However, in black neighborhoods, black and non-black borrowers display similar default hazards.

4. Empirical Specification and Results

4.1. Empirical Specification

To examine the role of borrower, loan and neighborhood characteristics on default, we follow much of the literature on mortgage terminations and estimate semi-parametric Cox proportional hazard models of the form:

$$h_i(t) = h_0(t) \exp(\ \beta \text{ loan characteristics}_i + \gamma \text{ neighborhood characteristics}_i$$
$$+ \delta \text{ borrower characteristics}_i + \alpha \text{ calendar time and origination year fixed effects})$$

where $h_i(t)$ is the default hazard of mortgage i at time t, that is, the probability that mortgage i will experience a default at time t, conditional on not having previously defaulted. The proportional hazard model assumes that there is an underlying baseline hazard function $h_0(t)$ that is shared by all mortgages in the analysis sample. The model then allows time-varying explanatory variables to shift this baseline up or down proportionally, with α, β, γ and δ representing vectors of coefficient estimates. The Cox model provides no direct estimate of, and makes no assumptions about, the functional form of the baseline hazard, and is able to account for both right and left censoring of the longitudinal data. In our data, mortgage prepayments are treated as right-censored observations and there is also minor left-censoring as a few months typically elapse between the time of origination and the entry of the mortgage into the LoanPerformance database upon securitization.

The calendar time fixed effects in our models are in terms of quarter dummies. These will control for any city-, state- or nation- wide macroeconomic factors, including unemployment rates and city-wide house price movements. The origination year fixed effects are intended to pick up any city-wide systematic changes in mortgage characteristics over time, including changes in average borrower risk and underwriting standards.

By estimating separate hazard models for hybrid ARMs and FRMs, we remove any endogeneity effects due to borrowers selecting into different product types.[8] Our results are displayed in tables 3, 4 and 5. We report hazard ratios (the exponential of the estimated coefficients) that can be interpreted as the proportional shift in the baseline hazard as a result of a unit change in the variable of interest. Hazard ratios greater than one indicate a positive effect, while those less than one indicate a negative effect. Our robust standard errors are clustered at the census tract level to account for any neighborhood level spatial correlation of residuals.

[8] In the hybrid ARM models discussed below, we also included an indicator variable for 3/27 ARMs. The coefficient on this variable implies that 3/27s have a default hazard that is approximately 18 percent lower than the 2/28s. When we limit the sample to only 2/28s (almost 3 in 4 of the hybrids), all the patterns that we describe below remain.

4.2. Adjustable Rate Mortgage Results

Loan and property characteristics. The first rows in table 3 show the hazard ratios associated with loan pricing terms. All of these terms are strongly significant and in the expected direction, with the default hazard monotonically increasing with higher relative interest rates at origination and with higher margins. These loan pricing terms may reflect both the direct effect of higher mortgage payments on the likelihood of default and *ex ante* risk pricing by lenders, to the extent that our controls for borrower risk characteristics (discussed below) do not fully capture the lender's overall assessment of borrower risk. The next set of variables in table 3 measure the size of the payment shock upon initial adjustment of the interest rate. The direction and time path described by these coefficients are also as expected, with larger payment shocks associated with more defaults.

Consistent with virtually all prior research, lower credit scores and high DTIs at origination significantly increase the default hazard. The current combined LTV also has significant, large and monotonically increasing effects on default.[9] We might expect those who take out additional debt on the property to be more likely to default because of a higher debt burden. There is a countervailing screening argument, however, that only borrowers who are good risks will be able to secure additional financing, so new debt may be negatively associated with default. We find a negative effect for new debt, suggesting that screening might overwhelm the direct impact of having to shoulder a greater debt burden, but the effect is statistically insignificant.

Having a coborrower on the loan lowers the rate of default, probably because a coborrower diversifies the effect of income shocks. Higher original loan balances, entered in the model as a third order polynomial, are positively associated with significantly higher default rates (coefficients not displayed). Home purchase loans have higher default rates than do refinances, possibly

[9] To estimate current LTV, we use the monthly dynamic loan balance from LoanPerformance (the numerator) and adjust the property value (the denominator) using the appropriate community district level house price index.

reflecting the fact that refinancers have longer housing tenure, and also cannot be first time mortgage borrowers. Surprisingly, we find that owner-occupiers have elevated default rates compared with investors, though owner-occupancy is self-reported and may be unreliable. Further decomposition of this owner-occupier indicator by documentation level reveals that most of the positive effect that owner-occupation has on the default hazard is for no- or low-documentation loans. However, as shown in the table, full documentation loans as a group are not significantly less likely to default than loans without full documentation.

Non-racial neighborhood characteristics. In the subsequent models of table 3, we add a variety of neighborhood characteristics. It is noteworthy that current LTV aside, the loan and property coefficients are virtually unchanged when these neighborhood characteristics are added. This is an important finding that confirms the validity of most existing research on mortgage defaults that does not have the wealth of very local neighborhood level information that we are able to use here. The LTV coefficients are smaller in magnitude when we add the neighborhood characteristics, as local house price appreciation is likely correlated with these neighborhood characteristics.

The estimates on the neighborhood characteristics themselves highlight important spatial patterns in default. First, we include several census tract level demographic measures from the 2000 Census. As expected, we find that lower median income neighborhoods have higher rates of default, and loans against properties in neighborhoods with more high school graduates generally have lower rates of default. Using poverty rates instead of median income in these models gave similar results. Neighborhoods with more non-native born residents are also associated with lower rates of default, but differentiating these immigrant neighborhoods by their dominant race or ethnic group (black, Hispanic or Asian) did not reveal any clear differences among them.

The next set of rows in table 3 show the effects of neighborhood foreclosure notices and the fraction of properties held by lenders (REOs). We find a positive and generally monotonically increasing effect of both. Because foreclosure and REO rates are positively correlated (an REO must have originally generated a foreclosure notice), these two sets of coefficients should be considered together. A foreclosure rate of over 3 percent in the surrounding census tract increases the default hazard by over one quarter, compared with mortgages in neighborhoods where the foreclosure rate is less than 1 percent. In addition, REO rates higher than 3 percent increase the default hazard by at least 12 percent, compared to neighborhoods with REO rates of less than 1 percent.

We also included the non-prime share of mortgages originated in the census tract during the two years prior to the loan's origination. In model 2, this is positively associated with default: ARMs in tracts where this non-prime share is more than 30 percent have a default hazard that is over 25 percent higher than loans where the non-prime share was less than 10 percent. However, in models 3 and 4, when we add explanatory variables for neighborhood racial composition, the non-prime share estimates are reduced in magnitude and no longer significant. We found a similar pattern for analogous variables that measure the share of riskier mortgage types in a census tract (the fraction of hybrid ARMs, and no- and low-documentation loans). The coefficients were generally positive and significant when added to model 2, but became much reduced in magnitude and insignificant when added to models 3 and 4 (results not displayed).

To capture potential differences in the underwriting standards that prevailed at the neighborhood level at the time of the loan's origination, we also ran models that included mortgage application denial rates in the census tract in the six months prior to the loan's origination.[10] Higher mortgage denial rates at origination may indicate more stringent underwriting standards at the time

[10] The HMDA mortgage application denial rate is: the number of denied applications, divided by the sum of loans originated, denied applications, and approved applications that were not accepted by the applicant.

of a loan's origination and thus may be associated with lower default rates. We found that while these coefficients were negative, they were small and insignificant once all the other controls were included. To capture the ability of borrowers to avoid default by selling their property, we also tried including mortgage application denial rates in the census tract within the six months prior to the month of observation (as opposed to at origination). Higher current denial rates may indicate more stringent underwriting standards for potential buyers, which would hinder the ability of the borrower to sell the house, and thus result in higher default rates. These coefficients were also insignificant. The inclusion of these additional denial rate variables did not change the magnitudes or significance of our other results reported in table 3 and these alternate models are not shown in the table.[11]

Neighborhood racial composition. Model 3 of table 3 shows that residing in a census tract with a higher proportion of black residents is associated with higher default rates, even after controlling for an extensive set of loan and borrower characteristics. Almost one third of hybrid ARMs in our sample were made to borrowers living in tracts that are over 80 percent black. For borrowers in these tracts, the default hazard is over 25 percent higher than that of borrowers in tracts with fewer than 20 percent black residents (the reference category). We do not see significant or consistent patterns when we look at the effect of Hispanic or Asian population shares.

Comparing models 2 and 3, we see that the loan and borrower coefficients do not change as we add more neighborhood characteristics. However, the addition of the racial composition variables reduces the magnitude of the foreclosure rate coefficients, reflecting the positive correlation between foreclosure rates and predominantly black neighborhoods.[12] And as noted

[11] Other census tract level variables that we tried in alternate models included crime rates (property crime, violent crime and all crimes) from the city's police records (the number of crimes reported in the census tract in the preceding 6 months divided by the tract's population in the 2000 Census), and the fraction of residents who had moved into the tract within the last two years from the 2000 Census. None of these variables were statistically significant.
[12] We also interacted these racial composition measures with the foreclosure measures but found no effect.

above, the inclusion of the neighborhood race variables reduces the magnitude and significance of the share of non-prime loans that we found in model 2.

Borrower race. In model 4, we add the race of the primary borrower.[13] Specifically, we interact whether the borrower is black with the neighborhood share of black residents. The reference category is non-black borrowers in neighborhoods with fewer than 20 percent black residents. The estimates display a somewhat unexpected pattern. In neighborhoods that are more than 40 percent black, the estimates on the share of black residents are virtually identical across borrower race; i.e., the borrower's own race does not matter much. Only in neighborhoods with fewer than 40 percent black residents do black borrowers have significantly higher default hazards. In model 5, we repeat model 4, but remove the variables on neighborhood racial composition, leaving in only the primary borrower's race. These results reinforce the interpretation that being black, by itself, has a limited effect on default: the hazard is only 10 percent higher for black versus white borrowers (the reference category).[14] Overall, the pattern of coefficients in models 4 and 5 suggest that the neighborhood share of black residents is as important in explaining default as the borrower's own race. Results for other borrower and neighborhood races and ethnicity did not yield clear patterns, although we find that Hispanic white borrowers tend to have lower default hazards. We also included the gender of the primary borrower in models 4 and 5 (coefficients not displayed) but the coefficients were small and insignificant.

Time-location fixed effects. The models in table 3 all include calendar quarter fixed effects to account for any city-wide macroeconomic factors. Still, if neighbors have similar socioeconomic backgrounds or human capital, job losses in one industry or occupation may be

[13] Race and/or ethnicity were missing for 15 percent of ARMs. The models in table 3 included indicator variables to capture this category, but our results do not change if we simply omit these observations from the sample. Our results also do not change if we define "black" to also include Hispanic black, or if we allow "black" to include instances where the coborrower is black (almost 90 percent of coborrowers are of the same race and ethnicity as the primary borrower).
[14] Note that this is a different reference category than that used in model 4.

concentrated among residents in some neighborhoods. To control for these more localized time varying economic factors and further isolate the census tract level findings, we reestimated model 4 with a variety of fixed effects that interact calendar time with location. These include: (i) calendar year interacted with community district fixed effects, which removes any annual variation in community district level economic factors (the 56 community district boundaries included are shown in figure 2), and alternatively, (ii) calendar quarter interacted with 11 geographic areas within the City, which removes any quarterly variation in economic factors at this 11 area level of geography.[15]

When we include these fixed effects, the effect of foreclosure rates is slightly smaller than in the models displayed in table 3, but the effects are still significant at 1 percent, and substantial in magnitude. For example, in the second version of these time-location fixed effects, a neighborhood foreclosure rate of over 3 percent increases the default hazard by 24 percent relative to mortgages in neighborhoods where the foreclosure rate is less than 1 percent. All the other coefficients, including those on the neighborhood racial compositions, were virtually unchanged. We also ran the models in table 3 with annual unemployment rates at the community district level. These unemployment rates were never significant and did not affect the magnitude of the other coefficients. Taken together, these results strongly suggest that most of the estimated neighborhood effects in table 3 are due to variation at the very local neighborhood level.

Zip code level neighborhood characteristics. To assess whether tract level measures are more useful for understanding the determinants of default than measures at higher levels of geographic aggregration, we reestimated the models in table 3 including the tract level neighborhood variables as shown, as well as their zip code level analogs. These zip code level variables generally

[15] The boundaries for these 11 areas were designated by aggregating the community districts in Manhattan into two areas, and those in the Bronx, Brooklyn and Queens into three areas each. The 11 areas were roughly equal in size. Models that included calendar quarter interacted with community district fixed effects failed to converge.

were not statistically significant, and they had little impact on the size or significance of the other coefficients, with one exception: the non-prime share of mortgages originated in the zip code during the two years prior to the loan's origination. These coefficients were positive, monotonically increasing and significant, while their census tract level analogs remained insignificant in all of the table 3 models.

Purchases vs. refinances. Finally, we have estimated all our models separately for home purchases and refinance loans. In table 4, we present the neighborhood foreclosure, REO, racial composition and borrower race coefficients associated with model 4 from table 3. The estimated impact of foreclosure rates are similar across the two samples, except that high REO rates have a larger and more significant effect on the default hazard for home purchases than on the hazard for refinances. This is consistent with refinance borrowers being more accustomed to both booms and busts in the neighborhood and less sensitive to changes caused by neighboring REOs because of their presumably longer tenure.

We find that the neighborhood racial composition results are stronger in magnitude for home purchases than for refinances. Both black and non-black borrowers in neighborhoods that are over 80 percent black have default hazards that are about two thirds higher than the reference group (non-black borrowers in neighborhoods with fewer than 20 percent black residents). By contrast, the models for refinance loans do not display this pattern, and in fact, the neighborhood race effects are all much smaller in magnitude and statistically insignificant.

4.3. Fixed Rate Mortgage Results

Table 5 displays our results for fixed rate mortgages. In general, the patterns are similar to those for the hybrid ARMs. There are, however, a few notable differences. The magnitude of the coefficients on FICO score and current LTV are larger, which is perhaps unsurprising because there is no equivalent to the payment shocks that are very important in explaining the defaults for ARMs. FRM borrowers who take out additional new debt have a 12 percent higher default hazard than those who do not, whereas this variable had an insignificantly negative effect on ARM defaults. This suggests that being able to secure additional financing does not provide much more information on FRM borrower risk, possibly because they are better selected or because underwriting standards are more stringent for FRMs. Instead, the pure effect of a larger debt burden dominates, resulting in a positive coefficient on taking on new debt.

The effect of foreclosures and REOs remain similar in magnitude and significance as for the ARMs. In terms of neighborhood racial composition, the coefficients on predominantly black neighborhoods are larger for FRMs than for ARMs. In model 3, tracts with more than 40 percent black residents have a default hazard that is at least 30 percent higher than tracts with fewer than 20 percent black residents (the reference category). Model 4 shows that while borrowers of all races have higher default hazards in neighborhoods with higher shares of blacks, non-black borrowers tend to have even higher default hazards than do black borrowers in those neighborhoods. Relative to the reference category (non-black borrowers in 0-20 percent black neighborhoods), non-black borrowers in 80-100 percent black neighborhoods have default hazards that are 52 percent higher, while black borrowers in these mostly black neighborhoods have default hazards that are only 28 percent higher. In model 5, where we remove neighborhood racial composition, the coefficient on being a black borrower is itself small and insignificant.

5. Possible Interpretations of the Neighborhood Results

Our results indicate that neighborhood characteristics play a significant role in default outcomes, beyond the effects of loan and individual borrower characteristics. We find that as the rate of foreclosure notices filed and the number of REOs in the neighborhood increases, the hazard of default increases, even after controlling for a host of other variables. We also find that home purchase borrowers living in predominantly black neighborhoods are more likely to enter default, regardless of the race of the individual borrower living in such neighborhoods. In this section, we consider possible interpretations of these results. While this will be necessarily speculative, it is important for us to consider because the plausible explanations for our findings have important policy implications.

Neighborhood foreclosure rates and REO activity. The most direct interpretation of our findings on foreclosures and REOs is that there are contagion effects at play. Neighbors may share information about the efficacy of default or the foreclosure process, leading other neighbors struggling with mortgage payments and negative equity to enter into default. High neighborhood foreclosure rates may also reduce the stigma associated with defaulting on a mortgage, making it more likely that others default as well. Survey results indicate that underwater homeowners who know people who have defaulted are more willing to default than those who have not been exposed, although a majority of survey respondents still say it is "immoral" or "unacceptable" for homeowners to stop making their mortgage payments (Guiso et al., 2009; Morin, 2010).

Foreclosure and REO concentrations could also be serving as proxies for omitted variables. They could, for example, be picking up local economic conditions, because income and employment shocks are correlated with defaults and foreclosures. However, as discussed above, adding community district level unemployment rates had no impact on the foreclosure and REO estimates.

Further, because tracts with high foreclosure and REO rates are geographically concentrated in particular community districts[16], the fact that controlling for calendar year interacted with community district fixed effects in our hazard models only slightly reduces the estimated impact of foreclosures and REOs suggests that these omitted variable interpretations are quite limited in importance.

Measurement error in house price appreciation may also cause some of the apparent effect of foreclosures and REO. Because the foreclosure and REO variables are measured at the census tract level, while our house price indices are at the larger community district level, less house price appreciation in higher foreclosure tracts would lead to systematically overestimated housing values and underestimated current LTVs for loans in high foreclosure neighborhoods.[17] As noted earlier, other research has shown that foreclosures and REOs can lead to diminished house price appreciation. Further, foreclosure rates may reflect expectations about future house price depreciation that are not already captured in price indices based on recent sales transactions.

In sum, we interpret our finding that default hazards increase with higher recent foreclosure and REO activity in the neighborhood to be reflecting some combination of a contagion effect and a proxy effect for very local house price fluctuations. To a much lesser extent, the foreclosure and REO rates may also serve as a proxy for neighborhood economic conditions or unobserved borrower characteristics.

Neighborhood racial composition. We find that borrowers in predominantly black neighborhoods also have higher default rates, regardless of race, even after controlling for a host of other borrower, loan, and neighborhood characteristics. The most obvious interpretation of this

[16] For example, during the first six months of 2008, 55 percent of tracts experiencing a foreclosure rate of at least 3 percent were located in just five community districts.
[17] Indeed, because borrowers with positive equity can avoid default by selling their property, the fact that current LTVs ranging from 60 to 90 are positive and significant in our default models suggests that our LTV measure is underestimated. If it were possible to use good repeat sales indices at the census tract level, our current LTV measure would be more accurate and the foreclosure effects would likely be reduced. In a study of Los Angeles County, Aragon et al. (2010) argue that even repeat sales indices at the zip code level are poor predictors of individual property values.

finding is that the share of black residents is proxying for unobserved borrower and loan characteristics common in those neighborhoods that are correlated with higher default rates. But these unobservables must be ones that apply only to home purchase and not to refinance borrowers, because we find no neighborhood racial composition effect for the latter. In considering plausible candidates for such unobservables, it is useful to divide them into (i) factors that lenders and others involved in the loan's origination know (or can learn) and that may be used to treat borrowers in black neighborhoods differently from borrowers in other neighborhoods, and (ii) factors that are not known to lenders and others.

In the first category, variations in lending and underwriting practices across census tracts that are unobserved in our data could lead to systematically different borrower and loan characteristics in certain tracts. For example, if underwriters used less stringent or improper standards for mortgage applicants in black neighborhoods, then borrowers in those neighborhoods likely would have relatively higher default risk, even after controlling for FICO scores and other observable risk factors. Similarly, if borrowers in black neighborhoods paid higher upfront fees, they likely would be less able to overcome income shocks and thus more likely to default, all else equal. Despite our rich set of loan specific variables, we are not able to account for all underwriting standards or for any upfront fees that were paid at origination.[18] Another possibility is that borrowers in black neighborhoods were more likely to use a mortgage broker, perhaps due to relatively limited financial sophistication or just limited access to local bank branches that makes it more costly to shop around for a mortgage. While we cannot discern the use of brokers in our data, others have found that loans involving brokers are more likely to enter default (Coulton et al., 2008; Laderman and Reid, 2008). The higher default rate might be because brokers' earnings depend upon both the success of the application and the loan amount, and brokers therefore have more incentive than borrowers

---

[18] Haughwout et al. (2009) find that borrowers in zip codes with higher proportions of black or Hispanic residents pay slightly *lower* adjustable mortgage interest rates. However, they were also not able to control for upfront fees.

applying without a broker to provide inaccurate information in the application (for example, an inflated appraisal). Or, the higher default rate may reflect aggressive marketing tactics some brokers use, which may make it more difficult for borrowers to evaluate the broker's offerings and more likely the borrower will accept an inappropriate loan.

Each of these possible unobservables should be less relevant for refinance loans. Refinance borrowers have undergone earlier screening by another party (the underwriter of their previous mortgage), and have demonstrated that they can carry a mortgage, at least up to the time of their refinance, so unobserved underwriting standards should matter less for refinancings than for home purchase loans. Refinance borrowers should also be less susceptible to aggressive marketing tactics because they are not first time homebuyers and may be more financially literate as a result of their prior homeowning and borrowing experience.[19] While we cannot rule out the possibility that other unobservable factors explain the effects we find, the factors we have mentioned are likely to be the best candidates because they are consistent with the different effect neighborhood racial composition has on defaults for purchase and refinance loans.

The hypothesis that differential lending practices in black neighborhoods lead to higher default rates also is supported by our results on non-prime loan concentrations. Model 2 of table 3 shows that loans in neighborhoods with higher non-prime shares at origination have higher default hazards, even after controlling for a large set of loan and borrower characteristics. The non-prime share of mortgages may be proxying for an omitted variable, "differential lending practices," that results in both more non-prime loans and higher default rates. However, the coefficients on the share of non-prime loans become smaller and insignificant once we include the neighborhood racial

[19] Some have argued that refinancings were targeted for "equity-stripping" by unscrupulous originators seeking high fees or other unfavorable terms that would serve to transfer the homeowner's equity to the originator or lender. Refinances make an attractive target because there is often more equity to strip. Whatever the extent of such practices may have been, it does not undercut the present argument. Refinance borrowers have gone through the mortgage application process and have been approved at least once before. As such, they are likely to be more financially knowledgeable and more financially sound than home purchase borrowers. While originators may have targeted refinance borrowers for the very worst loans, it does not follow that refinance borrowers would be more likely to accept loans on poor terms.

composition variables in model 3 of table 3. This suggests that the share of black residents is an even better proxy for this missing indicator. We found a similar result when we added variables capturing the share of riskier mortgage types (hybrid ARMs and no- and low-documentation loans) in the loan's census tract at origination. Consistent with Agarwal et al. (2011), these variables are positive and significant in explaining default when we do not also control for neighborhood racial composition. But, the addition of the share of black residents in the tract renders the coefficients on the shares of riskier loans small and insignificant. Thus, the share of riskier loans at origination is in some sense picking up the *outcome* of differential lending practices, while the differential practices themselves are better captured by the share of black residents, as measured in the 2000 Census.

We now turn to the second category of factors, those that are not known to lenders and others. There may be a systematic difference, unobserved in our data and unknown to lenders, between individuals living in predominantly black neighborhoods and those living elsewhere. It is difficult, however, to think of plausible candidates for such a difference that applies to home purchase borrowers but not to refinance borrowers. The unobserved factor would have to be associated with home purchase borrowers of all races living in black neighborhoods being less likely to repay (compared to home purchase borrowers in non-black neighborhoods), that is distinct from factors that are observable to the underwriter like income stability or asset reserves, and that does not affect refinance borrowers differentially by neighborhood racial composition.[20] Moreover, given the magnitude of our estimates, the importance of these unobservables would need to be considerable if they were solely responsible for our findings. In the absence of reasonable candidates for such unobserved variables, we interpret our results as being less likely to be driven by

[20] Possible candidates might involve innate borrower characteristics such as the ethic to repay loans. The argument would be that borrowers living in black neighborhoods lack this unobservable ethic. However, it is hard to come up with a story as to why home purchase borrowers in black neighborhoods would lack this ethic compared to those living in other neighborhoods, whereas refinance borrowers do not vary across black and non-black neighborhoods.

such unknown unobservables, and more likely to be due to the unobserved factors associated with differential lending practices discussed above.

Another possible interpretation of our neighborhood race findings is that the share of black residents is proxying for local economic conditions and exposure to adverse shocks. However, as we noted in the discussion of foreclosure and REO rates above, predominantly black tracts are geographically concentrated[21] and adding calendar year interacted with community district fixed effects does not change the coefficients on neighborhood racial composition. This makes it unlikely that the share of black residents is just a proxy for local unemployment or other local concerns.

Finally, we consider how measurement error may influence our neighborhood race findings. Similar to the discussion above for foreclosures and REOs, it is possible that predominantly black census tracts experience less house price appreciation than the surrounding community district, leading to a systematic underestimate of the current LTV for loans in predominantly black neighborhoods. Of course, this raises the question of why these particular neighborhoods have poorer housing price performance. To investigate this measurement error possibility further, we reestimated model 4 of table 3 for two separate samples: ARMs in non-black neighborhoods (less than 40 percent black residents) and ARMs in black neighborhoods (more than 60 percent black residents). We find that the coefficients on current LTV for the black neighborhood sample are smaller than for the non-black, while statistical significance remains at one percent for both samples. Underestimated current LTV in black neighborhoods would have implied the opposite result, suggesting that this source of measurement error cannot be solely responsible for our findings.[22]

---

[21] For example, 78 percent of tracts that have over 80 percent black residents are located in just five community districts.

[22] A second source of possible measurement error could come from misreported borrower race, such that the neighborhood racial composition becomes a proxy for the individual's race and the coefficients on individual race become small and insignificant. In the HMDA data, race is usually self-reported, but if the individual does not specify a race, the lender can record a subjective assessment, which could lead to misreporting error. To the extent that black borrowers believe that there is racial discrimination in lending, they may have an incentive to misreport their race, especially if the loan application is not made in person. However, we know of no empirical evidence that suggests that HMDA race classifications are systematically in error.

In sum, we think that the most likely explanation of our neighborhood racial composition findings is that the non-prime mortgage industry (brokers, lenders, underwriters, etc.) treated black neighborhoods differently in terms of marketing and underwriting practices or loan terms in ways that are unobserved in our data. However, our data and results do not allow us to discern the reasons for or the nature of this differential treatment. One possibility is that some non-prime lenders and brokers "targeted" minority neighborhoods with inappropriately priced loans and improper underwriting.[23] Our results are also consistent with the existence of institutional pressures to extend more credit to these neighborhoods, to boost homeownership rates, or simply to tap new markets, by employing strong marketing tactics or more lenient underwriting standards.[24] Unfortunately, we do not have detailed data on the underwriting process or marketing campaigns. Without such crucial information, we are unable to directly examine the role of these factors.

6. Conclusion

Our rich data set allows us to improve upon the existing literature by assessing the impact that borrower characteristics, the type of loan and its terms, and the characteristics of the neighborhood (measured at the census tract level) have on the probability that a non-prime

[23] The targeting hypothesis has been stated most starkly in litigation some cities have filed against various lenders. Baltimore, for example, has accused Wells Fargo of "target[ing] these kinds of predatory practices ['charging excessive fees; charging excessively high interest rates that are not justified by borrowers' creditworthiness; requiring large prepayment penalties while deliberately misleading borrowers about the penalties; using deceptive sales practices to wrap insurance products into mortgages; convincing borrowers to refinance mortgages into new loans that only benefit Wells Fargo; deceiving borrowers into believing that they are getting fixed rate loans when they are really getting adjustable rate loans'] at African-American neighborhoods and residents." Mayor and City Council of Baltimore v. Wells Fargo N.A., Case No. 1:08-cv-00062-JFM (D.Md. April 7, 2010).

[24] Although some commentators have faulted federal policies like the Community Reinvestment Act (CRA) for creating pressure for imprudent lending in low-income neighborhoods, there is evidence that challenges this argument (e.g., see Laderman and Reid, 2008). Because the loans in our study were originated by non-depository institutions, they do not qualify for CRA credit and thus the CRA is not a plausible explanation for this finding.

mortgage will default. Similar to existing research on mortgage defaults using loan level data, we find that the current LTV, borrower credit scores and debt-to-income ratios at origination, interest rates, loan size, ARM margins and payment shocks upon rate adjustment are all significant in explaining default behavior. We further show that, other than LTV, the magnitude of the estimates on these loan and borrower risk characteristics do not change when we add a variety of census tract level controls to capture very local neighborhood characteristics such as foreclosure and REO activity, racial composition, and the share of existing non-prime mortgages. This is an important finding that confirms the validity of most existing research on mortgage defaults that only have data on the loan and borrower risk variables, and not the rich neighborhood level information that we are able to use here.

Regarding the role of neighborhood characteristics in non-prime mortgage default risk, we uncover several new facts. First, default rates increase as the rate of foreclosure notices and the number of REOs in the neighborhood increases, even after controlling for a rich set of loan and borrower risk characteristics. We argue that this likely reflects a contagion effect, or that foreclosure and REO rates may be serving as proxies for weak neighborhood housing market conditions that are not already captured in the community district level price indices. At first blush, it seems hardly surprising that defaults are higher in high foreclosure and REO areas. However, what we have shown is that this effect is happening at an extremely local level. Our estimates remain large and significant when we include fixed effects for calendar year interacted with the 56 community districts in our New York City sample. These findings suggest that efforts to target foreclosure avoidance programs such as loan modifications must take account of extremely local differences in the concentration and effects of foreclosures.

Second, we find that home purchase loans in census tracts with larger shares of black residents have higher default risk, even after controlling for an extensive set of loan and borrower characteristics, including the borrower's own race. After considering a variety of potential explanations, we argue that this result most plausibly reflects differential treatment of black neighborhoods by the mortgage industry. However, the exact nature or cause of this differential treatment is not something that our data can address. Perhaps black neighborhoods were "targeted" for unsustainable non-prime loans that were inappropriate in pricing or other terms, or perhaps these neighborhoods were subject to low quality or less stringent underwriting practices. Untangling these possibilities would require additional loan-level information on the marketing and underwriting process. This is an important issue for future research.

Our finding that neighborhood characteristics have significant effects on default suggests that policymakers should take neighborhood context into account in designing their responses to the foreclosure crisis, and in shaping the regulation of mortgage products and lending practices. Neighborhoods matter: the risk that a borrower will default is not just a function of the borrower's characteristics, the loan terms, and economic trends, but also depends significantly on the neighborhood in which the borrower lives. Accordingly, neighborhood level efforts may be necessary to reduce default risk, and to address the consequences of default.

Data Appendix

Using a hierarchical matching algorithm, we were able to match 93 percent of the LoanPerformance sample (as described in section 3) to deeds from the New York City Department of Finance (DOF)'s Automated City Register Information System (ACRIS).[25] This gave us the exact location of the property and allowed us to merge on: building characteristics from the DOF's tax assessment records, repeat sales house price indices for 56 different community districts from the Furman Center for Real Estate and Urban Policy[26], tract characteristics from the 2000 Census, tract shares of various types of mortgage originations and loan application denial rates using Home Mortgage Disclosure Act (HMDA) data coupled with LoanPerformance, the tract rate of mortgage foreclosure notices (*lis pendens*), the tract share of properties owned by lenders (REOs)[27], annual unemployment rates in each community district from the American Community Survey, and monthly tract level crime rates from the New York City Police Department. We also merged the deeds with HMDA data to get additional borrower characteristics for each loan.[28] Of the original LoanPerformance sample, 78 percent matched both the deeds records and HMDA.

---

[25] Our procedure for matching LoanPerformance to ACRIS is similar to the method used by Haughwout et al. (2009) to match LoanPerformance to HMDA. Our data from ACRIS do not include Staten Island and thus we had to drop this borough from our analysis. We merged LoanPerformance loans to ACRIS mortgage deeds using three common fields: origination or deed date, loan amount and zip code, using six stages of hierarchical matching. At the end of each stage, loans and deeds that uniquely matched each other were set aside and considered matched, while all other loans and deeds enter the next stage. Stage 1 matched loans and deeds on the raw values of date, loan amount and zip code. Stage 2 matched the remaining loans and deeds on the raw values of date and zip code, and the loan amount rounded to $1,000. Stage 3 matched on the raw values of date and zip code, and the loan amount rounded to $10,000. Stage 4 matched on the raw values of zip code and loan amount, and allowed dates to differ by up to 90 days. Stage 5 matched on the raw value of zip code, loan amount rounded to $1,000, and allowed dates to differ by up to 90 days. Stage 6 matched on the raw value of zip code, loan amount rounded to $10,000, and allowed dates to differ by up to 90 days. We believe it is valid to introduce a 90-day window because for a good fraction of LoanPerformance loans, the origination date is imputed by backdating the first payment date by one month, and in ACRIS, there may be administrative lags in the recording of the deeds data. The chance of false positive matching is low because we are matching loans to the full universe of deed records, and only considering unique matches.

[26] See Furman Center (2009) for a description. We transform quarterly into monthly series by linear interpolation.

[27] The *lis pendens* are from Public Data Corporation and the REOs are from DOF property sales data.

[28] We merged HMDA records to ACRIS deeds based on date, loan amount and census tract, using the same six stage hierarchical matching technique as for the LoanPerformance-ACRIS match. We then uniquely paired the LoanPerformance records with HMDA records based on the unique deed identification number from ACRIS. While other researchers have matched loan level data directly to HMDA by using the zip code as a common geographic identifier, our matching strategy is likely more reliable as it uses a more precise geographical identifier (census tract).

Acknowledgements

We thank Caitlin Coslett, Johanna Lacoe, Mary Weselcouch and Jaclene Begley for their assistance on this paper. This paper has benefited from comments by Ingrid Gould Ellen, Josiah Madar, Richard Revesz, and Joseph Tracy, as well as participants at the NYU Law School Faculty Brownbag Workshop, the 2010 Policy Summit at the Federal Reserve Bank of Cleveland, and the annual meetings of the American Law and Economics Association, the Association for Public Policy Analysis and Management, and the Eastern Economic Association. We are also grateful to the staff at the Federal Reserve Bank of New York for their help with the LoanPerformance data and their patience in answering our many questions, as well as to the Public Data Corporation and the New York City Department of Finance for their assistance in providing the additional data needed for this project.

References

Agarwal, Sumit, Brent W. Ambrose, Souphala Chomsisengphet and Anthony B. Sanders. 2011. "Thy Neighbor's Mortgage: Does Living in a Subprime Neighborhood Affect One's Probability of Default?" *Real Estate Economics*, forthcoming.

Amronim, Gene, and Anna L. Paulson. 2010. "Default Rates on Prime and Subprime Mortgages: Differences and Similarities." *Profitwise News and Views*, September.

Aragon, Diego, Andrew Caplin, Sumit Chopra, John V. Leahy, Yann LeCun, Marco Scoffier, and Joseph Tracy. 2010. "Reassessing FHA Risk." NBER Working Paper 15802.

Berkovec, James A., Glenn B. Canner, Stuart A. Gabriel, and Timothy H. Hannan. 1994. "Race, Redlining, and Residential Mortgage Loan Performance." *Journal of Real Estate Finance and Economics*, 9(3): 263-94.

Campbell, John Y., Stefano Giglio and Parag Pathak. 2009. "Forced Sales and House Prices." NBER Working Paper 14866.

Coulton, Claudia, Tsui Chan, Michael Schramm, and Kristen Mikelbank. 2008. "Pathways to Foreclosure: A Longitudinal Study of Mortgage Loans, Cleveland and Cuyahoga County, 2005-2008." *Center on Urban Poverty and Community Development, Case Western Reserve University.*

Courchane, Marsha J., Brian J. Surette, and Peter M. Zorn. 2004. "Subprime Borrowers: Mortgage Transitions and Outcomes." *Journal of Real Estate Finance and Economics*, 29(4): 365-92.

Danis, Michelle A., and Anthony N. Pennington-Cross. 2005. "The Delinquency of Subprime Mortgages." Federal Reserve Bank of St. Louis Working Paper 2005-022A.

Demyanyk, Yuliya. 2009. "Quick Exits of Subprime Mortgages." *Federal Reserve Bank of St. Louis Review*, 2: 79-93.

Ding, Lei, Roberto G. Quercia, Janneke Ratcliffe, and Wei Lei. 2008. "Risky Borrowers or Risky Mortgages: Disaggregating Effects Using Propensity Score Models." Center for Community Capital University of North Carolina Working Paper.

Firestone, Simon, Robert Van Order, and Peter M. Zorn. 2007. "The Performance of Low-Income and Minority Mortgages." *Real Estate Economics,* 35(4): 479-504.

Foote, Christopher L., Kristopher S. Gerardi, Lorenz F. Goette, and Paul Willen. 2009. "Reducing Foreclosures." Federal Reserve Bank of Boston Public Policy Discussion Paper 2009-02.

Furman Center for Real Estate and Urban Policy. 2009. "State of the City's Housing & Neighborhoods 2009." http://furmancenter.org/research/sonychan/2009-report

Furman Center for Real Estate and Urban Policy. 2010. "Foreclosed Properties in NYC: A Look at the Last 15 Years." http://furmancenter.org/files/publications/Furman_Center_Fact_Sheet_on_REO_Properties.pdf.

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

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2009. "Moral and Social Constraints to Strategic Default on Mortgages." NBER Working Paper 15145.

Harding, John, Eric Rosenblatt, and Vincent Yao. 2009. "The Contagion Effect of Foreclosed Properties." *Journal of Urban Economics*, 66(3): 164-78.

Hartley, Daniel A. 2010. "The Effect of Foreclosures on Owner-Occupied Housing Prices: Supply or Dis-Amenity?" Paper presented at the Federal Reserve Bank of Cleveland Policy Summit, Cleveland, OH.

Haughwout, Andrew, Christopher Mayer, and Joseph Tracy. 2009. "Subprime Mortgage Pricing: The Impact of Race, Ethnicity and Gender on the Cost of Borrowing." *Brookings-Wharton Papers on Urban Affairs*: 33-63.

Haughwout, Andrew, Richard Peach, and Joseph Tracy. 2008. "Juvenile Delinquent Mortgages: Bad Credit or Bad Economy?" *Journal of Urban Economics*, 64(2): 246-57.

Immergluck, Dan, and Geoff Smith. 2006. "The External Costs of Foreclosure: The Impact of Single-Family Mortgage Foreclosures on Property Values." *Housing Policy Debate*, 17(1): 57-79.

Ioannides, Yannis M. 2003. "Interactive Property Valuations." *Journal of Urban Economics*, 53(1): 145-70.

Jiang, Wei, Ashlyn Aiko Nelson, and Edward Vytlacil. 2009. "Liar's Loan? Effects of Origination Channel and Information Falsification on Mortgage Delinquency." Columbia University Business School Working Paper.

Laderman, Elizabeth, and Carolina Reid. 2008. "Lending in Low- and Moderate-Income Neighborhoods in California: The Performance of CRA Lending During the Subprime Meltdown." Federal Reserve Bank of San Francisco Working Paper 2008-05.

Lee, Kai-Yan. 2008. "Foreclosure's Price-Depressing Spillover Effects on Local Properties: A Literature Review." Federal Reserve Bank of Boston Community Affairs Discussion Paper 2008-01.

Leonard, Tammy, and James Murdoch. 2009. "The Neighborhood Effects of Foreclosure." *Journal of Geographical Systems*, 11(4): 317-32.

Lin, Zhenguo, Eric Rosenblatt, and Vincent W. Yao. 2009. "Spillover Effects of Foreclosures on Neighborhood Property Values." *Journal of Real Estate Finance and Economics,* 38(4): 387-407.

Mayer, Christopher, Karen Pence, and Shane M. Sherlund. 2009. "The Rise in Mortgage Defaults." *Journal of Economic Perspectives*, 23(1): 27-50.

Mian, Atif R. and Amir Sufi. 2009. "The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis." *Quarterly Journal of Economics*, 124(4): 1449-1496.

Morin, Rich. 2010. "Walking Away. A Third of the Public Says It's Sometimes OK to Stop Paying a Mortgage." Pew Research Center Social and Demographic Trends Report, September 2010.

Okah, Ebiere, and James Orr. 2010. "Subprime Mortgage Lending in New York City: Prevalence and Performance." Federal Reserve Bank of New York Staff Report 432.

Rogers, William, and William Winter. 2009. "The Impact of Foreclosures on Neighboring Housing Sales." *Journal of Real Estate Research,* 31(4): 455-479.

Schloemer, Ellen, Wei Li, Keith Ernst, and Kathleen Keest. 2006. "Losing Ground: Foreclosures in the Subprime Market and Their Cost to Homeowners." *Center for Responsible Lending.*

Schuetz, Jenny, Vicki Been, and Ingrid Gould Ellen. 2008. "Neighborhood Effects of Concentrated Mortgage Foreclosures." *Journal of Housing Economics,* 17(4): 306-319.

Van Order, Robert, and Peter Zorn. 2000. "Income, Location and Default: Some Implications for Community Lending." *Real Estate Economics*, 28(3): 385-404.

Table 1: Loan and Borrower Characteristics

	Adjustable Rate Mortgages	Fixed Rate Mortgages
Average loan amount	$352,461	$360,614
Owner-occupier	0.909	0.880
Average interest rate at origination	7.4	6.9
Average relative interest rate at origination[1]	4.3	0.9
Average ARM margin[2]	5.9	
Payment shock at first ARM adjustment[3]:		
<20%	0.975	
20-30%	0.014	
>30%	0.011	
Average FICO score at origination	621	665
Debt-to-income at origination > 45%[4]	0.424	0.361
Average combined LTV at origination	74.4	69.4
Took on additional mortgage debt[5]	0.051	0.108
Full documentation	0.382	0.419
Has coborrower	0.206	0.284
Primary borrower race/ethnicity[6]:		
White	0.224	0.311
Black	0.461	0.368
Hispanic white	0.152	0.141
Asian	0.093	0.111
Other	0.070	0.068
Number of loans	30,307	29,414

Source: LoanPerformance New York City sample, as described in the text.

Table indicates the fraction of records in each category, unless otherwise noted.

[1] For ARMs: interest rate minus the six-month LIBOR rate at origination. For FRMs: interest rate minus the Freddie Mac average interest rate for prime 30-year FRMs at origination. Expressed in percentage points.

[2] Percentage points added to the six-month LIBOR to determine the interest rate at future rate adjustments.

[3] The jump in monthly payments that 2/28 borrowers experience at month 25 and 3/27 borrowers experience at month 37.

[4] DTI is missing for 19% of ARMs and 38% of FRMs.

[5] Additional mortgage debt was secured that totaled at least 5% of the original loan amount.

[6] Race/ethnicity is missing for 15% of ARMs and 20% of FRMs.

Table 2: Neighborhood Characteristics

		Adjustable Rate Mortgages	Fixed Rate Mortgages
Census tract demographics from 2000 Census			
Median income:	<$30,000	0.230	0.211
	$30,000-$40,000	0.261	0.253
	>$40,000	*0.509*	*0.535*
% High school graduates:	*<70%*	*0.465*	*0.421*
	70-80%	0.342	0.326
	>80%	0.193	0.252
% Non-native born	*<20%*	*0.119*	*0.129*
	20-40%	0.466	0.445
	40-60%	0.351	0.353
	>60%	0.064	0.074
% Black	*0-20%*	*0.296*	*0.424*
	20-40%	0.105	0.097
	40-60%	0.104	0.082
	60-80%	0.162	0.136
	80-100%	0.333	0.261
% Hispanic	*0-20%*	*0.611*	*0.618*
	20-40%	0.190	0.192
	40-100%	0.199	0.190
% Asian	*0-20%*	*0.913*	*0.883*
	20-40%	0.076	0.097
	40-100%	0.011	0.020
Census tract loan composition at origination			
Share of non-prime loans[1]	*<10%*	*0.079*	*0.141*
	10-20%	0.332	0.359
	20-30%	0.529	0.445
	>30%	0.060	0.054
Number of loans		30,307	29,414
Census tract foreclosure and REO activity (dynamic)			
Recent foreclosure rate[2]	*<1%*	*0.526*	*0.555*
	1-2%	0.292	0.246
	2-3%	0.111	0.113
	>3%	0.071	0.085
Recent REO rate[3]	*<1%*	*0.648*	*0.675*
	1-2%	0.132	0.113
	2-3%	0.077	0.068
	>3%	0.142	0.144
Number of loan-months		503,579	946,922

Table indicates fraction of records in each category. Italics indicate category omitted from hazard models in tables 3-5.

[1] The number of non-prime loans originated in the Loan Performance database / Total loans originated in HMDA during the 6 months preceding the loan's origination.

[2] The number of *lis pendens* filed in the 6 months preceding the analysis month / The stock of buildings.

[3] The number of properties listed as REO in the 6 months preceding the analysis month / The stock of buildings.

Table 3: Hazard Models of Default for Adjustable Rate Mortgages

		Model 1			Model 2			Model 3			Model 4			Model 5		
Loan characteristics																
Relative interest rate at origination:																
	2-4	1.256	(0.063)	**	1.243	(0.063)	**	1.236	(0.063)	**	1.241	(0.063)	**	1.240	(0.063)	**
	4-6	1.777	(0.106)	**	1.723	(0.103)	**	1.699	(0.102)	**	1.707	(0.103)	**	1.713	(0.102)	**
	>6	3.484	(0.265)	**	3.329	(0.258)	**	3.279	(0.256)	**	3.294	(0.256)	**	3.299	(0.253)	**
ARM Margin:	5-6%	1.145	(0.047)	**	1.126	(0.047)	**	1.122	(0.046)	**	1.121	(0.047)	**	1.127	(0.047)	**
	6-7%	1.310	(0.050)	**	1.291	(0.050)	**	1.283	(0.050)	**	1.283	(0.050)	**	1.294	(0.050)	**
	>7%	1.503	(0.086)	**	1.473	(0.085)	**	1.468	(0.085)	**	1.465	(0.085)	**	1.477	(0.085)	**
3-6 months post-adjustment and:																
	payment shock <20%	1.169	(0.100)		1.173	(0.100)		1.177	(0.101)		1.177	(0.101)		1.177	(0.100)	
	payment shock 20-30%	1.216	(0.137)		1.199	(0.135)		1.204	(0.136)		1.206	(0.136)		1.207	(0.136)	
	payment shock >30%	1.631	(0.203)	**	1.625	(0.203)	**	1.627	(0.203)	**	1.626	(0.203)	**	1.623	(0.203)	**
7-12 months post-adjustment and:																
	payment shock <20%	1.397	(0.117)	**	1.400	(0.116)	**	1.406	(0.117)	**	1.405	(0.117)	**	1.405	(0.117)	**
	payment shock 20-30%	1.735	(0.171)	**	1.697	(0.167)	**	1.703	(0.167)	**	1.713	(0.168)	**	1.717	(0.168)	**
	payment shock >30%	1.939	(0.228)	**	1.919	(0.226)	**	1.920	(0.225)	**	1.928	(0.225)	**	1.924	(0.225)	**
FICO score at origination:	680-720	1.205	(0.058)	**	1.208	(0.057)	**	1.202	(0.057)	**	1.207	(0.058)	**	1.209	(0.058)	**
	650-680	1.312	(0.061)	**	1.311	(0.060)	**	1.307	(0.060)	**	1.304	(0.061)	**	1.304	(0.061)	**
	620-650	1.499	(0.069)	**	1.499	(0.070)	**	1.489	(0.070)	**	1.489	(0.070)	**	1.489	(0.070)	**
	590-620	1.490	(0.079)	**	1.505	(0.080)	**	1.494	(0.079)	**	1.492	(0.080)	**	1.491	(0.080)	**
	560-590	1.783	(0.099)	**	1.816	(0.101)	**	1.810	(0.101)	**	1.806	(0.101)	**	1.797	(0.100)	**
	530-560	1.889	(0.110)	**	1.937	(0.115)	**	1.928	(0.114)	**	1.920	(0.113)	**	1.913	(0.113)	**
	<530	1.864	(0.123)	**	1.919	(0.129)	**	1.902	(0.128)	**	1.899	(0.128)	**	1.895	(0.127)	**
Debt-to-income at origination >45%		1.127	(0.029)	**	1.130	(0.029)	**	1.132	(0.029)	**	1.129	(0.029)	**	1.128	(0.029)	**
Current combined LTV:	60-70%	1.178	(0.054)	**	1.127	(0.051)	**	1.111	(0.051)	*	1.113	(0.051)	*	1.126	(0.051)	**
	70-80%	1.423	(0.061)	**	1.327	(0.056)	**	1.302	(0.055)	**	1.307	(0.056)	**	1.328	(0.056)	**
	80-90%	1.618	(0.076)	**	1.478	(0.070)	**	1.443	(0.069)	**	1.447	(0.069)	**	1.475	(0.070)	**
	90-95%	2.201	(0.124)	**	1.980	(0.112)	**	1.935	(0.110)	**	1.938	(0.110)	**	1.975	(0.112)	**
	95-100%	2.340	(0.147)	**	2.094	(0.132)	**	2.048	(0.129)	**	2.053	(0.129)	**	2.089	(0.131)	**
	>100%	2.284	(0.142)	**	1.964	(0.125)	**	1.917	(0.123)	**	1.922	(0.123)	**	1.965	(0.125)	**

continued

		Model 1			Model 2			Model 3			Model 4			Model 5		
Took on additional mortgage debt		0.929	(0.039)		0.944	(0.039)		0.942	(0.039)		0.945	(0.039)		0.946	(0.039)	
Has coborrower		0.754	(0.024)	**	0.775	(0.024)	**	0.778	(0.024)	**	0.780	(0.025)	**	0.779	(0.025)	**
Home purchase		1.377	(0.097)	**	1.374	(0.095)	**	1.377	(0.096)	**	1.389	(0.097)	**	1.389	(0.096)	**
Owner-occupier		1.311	(0.054)	**	1.350	(0.057)	**	1.357	(0.057)	**	1.371	(0.058)	**	1.370	(0.057)	**
Full documentation		1.003	(0.156)		0.987	(0.151)		0.976	(0.148)		0.958	(0.143)		0.960	(0.144)	
Non-racial neighborhood characteristics																
Median income:	<$30,000				1.195	(0.044)	**	1.172	(0.046)	**	1.146	(0.045)	**	1.166	(0.043)	**
	$30,000-$40,000				1.077	(0.032)	*	1.067	(0.032)	*	1.062	(0.032)	*	1.074	(0.032)	*
% High school graduates:	70-80%				0.996	(0.030)		0.939	(0.030)	*	0.938	(0.030)	*	0.964	(0.029)	
	>80%				0.977	(0.036)		0.915	(0.037)	*	0.903	(0.037)	*	0.925	(0.035)	*
% Non-native born:	20-40%				0.914	(0.036)	*	0.924	(0.037)	*	0.917	(0.036)	*	0.918	(0.035)	*
	40-60%				0.870	(0.036)	**	0.893	(0.038)	**	0.886	(0.037)	**	0.871	(0.036)	**
	>60%				0.659	(0.044)	**	0.717	(0.049)	**	0.721	(0.048)	**	0.684	(0.044)	**
Recent foreclosure rate:	1-2%				1.220	(0.035)	**	1.151	(0.034)	**	1.149	(0.034)	**	1.191	(0.034)	**
	2-3%				1.303	(0.050)	**	1.207	(0.047)	**	1.206	(0.047)	**	1.263	(0.048)	**
	>3%				1.392	(0.062)	**	1.270	(0.058)	**	1.268	(0.058)	**	1.341	(0.060)	**
Recent REO rate:	1-2%				1.076	(0.039)	*	1.056	(0.038)		1.059	(0.039)		1.070	(0.039)	
	2-3%				1.070	(0.045)		1.047	(0.044)		1.047	(0.044)		1.061	(0.045)	
	>3%				1.140	(0.041)	**	1.116	(0.041)	**	1.119	(0.041)	**	1.136	(0.041)	**
Share of non-prime loans:	10-20%				1.135	(0.061)	*	1.091	(0.060)		1.093	(0.060)		1.127	(0.061)	*
	20-30%				1.182	(0.065)	**	1.069	(0.062)		1.066	(0.062)		1.142	(0.064)	*
	>30%				1.257	(0.083)	**	1.139	(0.079)		1.137	(0.079)		1.217	(0.081)	**

continued

		Model 1	Model 2	Model 3	Model 4	Model 5
Neighborhood racial composition						
% Hispanic:	20-40%			1.023 (0.041)	1.045 (0.041)	
	40-100%			0.945 (0.045)	0.993 (0.047)	
% Asian:	20-40%			0.931 (0.048)	0.948 (0.049)	
	40-100%			0.930 (0.137)	0.951 (0.139)	
% Black:	20-40%			1.111 (0.049) *		
	40-60%			1.189 (0.054) **		
	60-80%			1.209 (0.062) **		
	80-100%			1.268 (0.060) **		
Borrower race/ethnicity						
Black borrower and neighborhood is:						
	0-20% black				1.258 (0.113) **	
	20-40% black				1.291 (0.097) **	
	40-60% black				1.183 (0.073) **	
	60-80% black				1.255 (0.074) **	
	80-100% black				1.278 (0.067) **	
Non-black borrower and neighborhood is:						
	20-40% black				1.068 (0.055)	
	40-60% black				1.175 (0.076) *	
	60-80% black				1.228 (0.086) **	
	80-100% black				1.310 (0.082) **	
Black borrower						1.098 (0.037) **
Hispanic white borrower					0.829 (0.036) **	0.804 (0.033)
Asian borrower					0.937 (0.043)	0.919 (0.041)
Number of loan-months				503,579		
Number of loans				30,307		
Log pseudolikelihood		-76861.238	-76653.936	-76621.404	-76588.244	-76609.448

Cox proportional hazard models of 90-day default, hazard ratios reported. See tables 1 and 2 for notes on explanatory variables and the left-out categories.

Robust standard errors clustered by census tracts are reported in ( ). Statistical significance is indicated by: *5% and **1%.

All models include: a cubic in original loan balance; indicators for 3/27s, prepayment penalty in effect, cash-out refinance, second home, low documentation, property type (single & 2-4 family), building age (0-10, 11-50, >50 years & missing), >12 months post-adjustment and payment shock is <20, 20-30 & >30%; indicators for missing values of FICO, DTI, prepayment penalty; and fixed effects for origination year and calendar quarter. Models 4 and 5 also include indicators for Hispanic black, Hispanic other race, non-Hispanic other race, missing race, female and missing gender.

Table 4: Hazard Models of Default for Adjustable Rate Mortgages - Home Purchases and Refinances

		All ARMs			Home Purchases			Refinances		
Recent foreclosure rate:	1-2%	1.149	(0.034)	**	1.136	(0.045)	**	1.167	(0.050)	**
	2-3%	1.206	(0.047)	**	1.202	(0.064)	**	1.197	(0.066)	**
	>3%	1.268	(0.058)	**	1.244	(0.076)	**	1.296	(0.084)	**
Recent REO rate:	1-2%	1.059	(0.039)		1.133	(0.055)	**	0.986	(0.049)	
	2-3%	1.047	(0.044)		1.149	(0.070)	*	0.954	(0.057)	
	>3%	1.119	(0.041)	**	1.198	(0.058)	**	1.015	(0.055)	
% Hispanic:	20-40%	1.045	(0.041)		1.122	(0.064)	*	0.953	(0.053)	
	40-100%	0.993	(0.047)		1.072	(0.075)		0.913	(0.064)	
% Asian:	20-40%	0.948	(0.049)		0.984	(0.069)		0.925	(0.071)	
	40-100%	0.951	(0.139)		0.948	(0.197)		0.975	(0.185)	
Black borrower and neighborhood is:										
	0-20% black	1.258	(0.113)	**	1.463	(0.169)	**	1.143	(0.155)	
	20-40% black	1.291	(0.097)	**	1.570	(0.174)	**	1.072	(0.104)	
	40-60% black	1.183	(0.073)	**	1.566	(0.139)	**	0.910	(0.081)	
	60-80% black	1.255	(0.074)	**	1.652	(0.133)	**	0.943	(0.077)	
	80-100% black	1.278	(0.067)	**	1.691	(0.136)	**	0.980	(0.073)	
Non-black borrower and neighborhood is:										
	20-40% black	1.068	(0.055)		1.058	(0.073)		1.125	(0.087)	
	40-60% black	1.175	(0.076)	*	1.318	(0.113)	**	1.018	(0.091)	
	60-80% black	1.228	(0.086)	**	1.345	(0.138)	**	1.145	(0.113)	
	80-100% black	1.310	(0.082)	**	1.643	(0.143)	**	1.020	(0.096)	
Hispanic white borrower		0.829	(0.036)	**	0.896	(0.051)		0.789	(0.051)	**
Asian borrower		0.937	(0.043)		0.938	(0.057)		1.008	(0.075)	
Number of loan-months		503,579			198,680			304,875		
Number of loans		30,307			11,931			18,376		
Log pseudolikelihood		-76588.244			-36244.342			-34292.662		

Cox proportional hazard models of 90-day default, hazard ratios reported.

See tables 1 and 2 for notes on explanatory variables and the left-out categories.

Robust standard errors clustered by census tracts are reported in ( ). Statistical significance is indicated by: *5% and **1%.
All models include: a cubic in original loan balance; indicators for 3/27s, prepayment penalty in effect, second home, low documentation, property type (single & 2-4 family), building age (0-10, 11-50, >50 years & missing), >12 months post-adjustment and payment shock is <20, 20-30 & >30%, Hispanic black, Hispanic other race, non-Hispanic other race, female; indicators for missing values of FICO, DTI, prepayment penalty, race, gender; and fixed effects for origination year and calendar quarter. The refinance model also includes an indicator for cash-out refinance.

Table 5: Hazard Models of Default for Fixed Rate Mortgages

		Model 1			Model 2			Model 3			Model 4			Model 5		
Loan characteristics																
Relative interest rate at origination:																
	0-1	1.459	(0.087)	**	1.456	(0.087)	**	1.451	(0.087)	**	1.456	(0.087)	**	1.463	(0.087)	**
	1-2	2.327	(0.148)	**	2.286	(0.145)	**	2.262	(0.144)	**	2.273	(0.144)	**	2.300	(0.146)	**
	2-3	3.685	(0.281)	**	3.546	(0.267)	**	3.462	(0.260)	**	3.483	(0.264)	**	3.557	(0.269)	**
	>3	4.690	(0.439)	**	4.499	(0.423)	**	4.422	(0.417)	**	4.434	(0.418)	**	4.487	(0.423)	**
FICO score at origination:	680-720	1.732	(0.084)	**	1.722	(0.084)	**	1.725	(0.084)	**	1.727	(0.084)	**	1.722	(0.084)	**
	650-680	2.197	(0.113)	**	2.166	(0.111)	**	2.163	(0.111)	**	2.169	(0.112)	**	2.162	(0.111)	**
	620-650	2.943	(0.152)	**	2.918	(0.150)	**	2.902	(0.149)	**	2.897	(0.149)	**	2.900	(0.150)	**
	590-620	3.526	(0.216)	**	3.487	(0.213)	**	3.471	(0.212)	**	3.469	(0.213)	**	3.475	(0.213)	**
	560-590	4.179	(0.298)	**	4.168	(0.298)	**	4.152	(0.297)	**	4.105	(0.294)	**	4.098	(0.294)	**
	530-560	4.760	(0.397)	**	4.702	(0.392)	**	4.722	(0.392)	**	4.645	(0.389)	**	4.621	(0.388)	**
	<530	5.183	(0.522)	**	5.226	(0.532)	**	5.292	(0.540)	**	5.191	(0.533)	**	5.135	(0.526)	**
Debt-to-income at origination >45%		1.102	(0.040)	**	1.094	(0.040)	*	1.091	(0.040)	*	1.088	(0.040)	*	1.091	(0.040)	*
Current combined LTV:	60-70%	1.247	(0.065)	**	1.167	(0.061)	**	1.150	(0.060)	**	1.149	(0.060)	**	1.166	(0.061)	**
	70-80%	1.521	(0.082)	**	1.383	(0.075)	**	1.358	(0.073)	**	1.354	(0.073)	**	1.380	(0.074)	**
	80-90%	1.988	(0.107)	**	1.758	(0.096)	**	1.722	(0.093)	**	1.714	(0.093)	**	1.750	(0.095)	**
	90-95%	2.631	(0.176)	**	2.261	(0.155)	**	2.212	(0.151)	**	2.208	(0.151)	**	2.256	(0.155)	**
	95-100%	2.816	(0.201)	**	2.396	(0.175)	**	2.338	(0.172)	**	2.333	(0.171)	**	2.389	(0.175)	**
	>100%	3.532	(0.215)	**	2.819	(0.181)	**	2.732	(0.176)	**	2.720	(0.175)	**	2.807	(0.181)	**
Took on additional mortgage debt		1.118	(0.044)	**	1.127	(0.044)	**	1.131	(0.044)	**	1.132	(0.045)	**	1.132	(0.044)	**
Has coborrower		0.809	(0.027)	**	0.826	(0.027)	**	0.833	(0.028)	**	0.828	(0.028)	**	0.823	(0.028)	**
Home purchase		1.094	(0.066)		1.167	(0.071)	*	1.190	(0.073)	**	1.205	(0.074)	**	1.187	(0.072)	**
Owner-occupier		1.040	(0.048)		1.062	(0.049)		1.064	(0.049)		1.069	(0.050)		1.070	(0.050)	
Full documentation		0.770	(0.063)	**	0.753	(0.061)	**	0.735	(0.059)	**	0.732	(0.060)	**	0.747	(0.061)	**

continued

		Model 1	Model 2			Model 3			Model 4			Model 5		
Non-racial neighborhood characteristics														
Median income:	<$30,000		1.022	(0.046)		0.963	(0.045)		0.962	(0.045)		1.016	(0.045)	
	$30,000-$40,000		0.977	(0.036)	**	0.960	(0.035)	**	0.958	(0.035)	**	0.975	(0.035)	**
% High school graduates:	70-80%		0.956	(0.035)		0.921	(0.036)	*	0.925	(0.036)	*	0.943	(0.035)	
	>80%		0.960	(0.045)		0.925	(0.045)		0.929	(0.045)		0.943	(0.045)	
% Non-native born:	20-40%		0.971	(0.042)		0.954	(0.040)		0.953	(0.040)		0.968	(0.041)	
	40-60%		0.978	(0.044)		0.956	(0.043)		0.957	(0.043)		0.974	(0.043)	
	>60%		0.865	(0.057)	*	0.876	(0.063)		0.884	(0.064)		0.881	(0.059)	
Recent foreclosure rate:	1-2%		1.216	(0.047)	**	1.144	(0.045)	**	1.142	(0.045)	**	1.203	(0.047)	**
	2-3%		1.273	(0.062)	**	1.177	(0.058)	**	1.174	(0.058)	**	1.253	(0.061)	**
	>3%		1.344	(0.076)	**	1.228	(0.071)	**	1.223	(0.071)	**	1.320	(0.075)	**
Recent REO rate:	1-2%		1.096	(0.051)	*	1.073	(0.050)		1.079	(0.050)		1.096	(0.051)	*
	2-3%		1.123	(0.060)	*	1.098	(0.060)		1.105	(0.060)		1.125	(0.060)	*
	>3%		1.153	(0.048)	**	1.122	(0.047)	**	1.129	(0.047)	**	1.159	(0.048)	**
Share of non-prime loans:	10-20%		1.142	(0.069)	*	1.111	(0.067)		1.118	(0.068)		1.147	(0.069)	*
	20-30%		1.240	(0.081)	**	1.143	(0.077)	*	1.148	(0.077)	*	1.229	(0.081)	**
	>30%		1.152	(0.093)		1.057	(0.088)		1.056	(0.088)		1.137	(0.093)	

continued

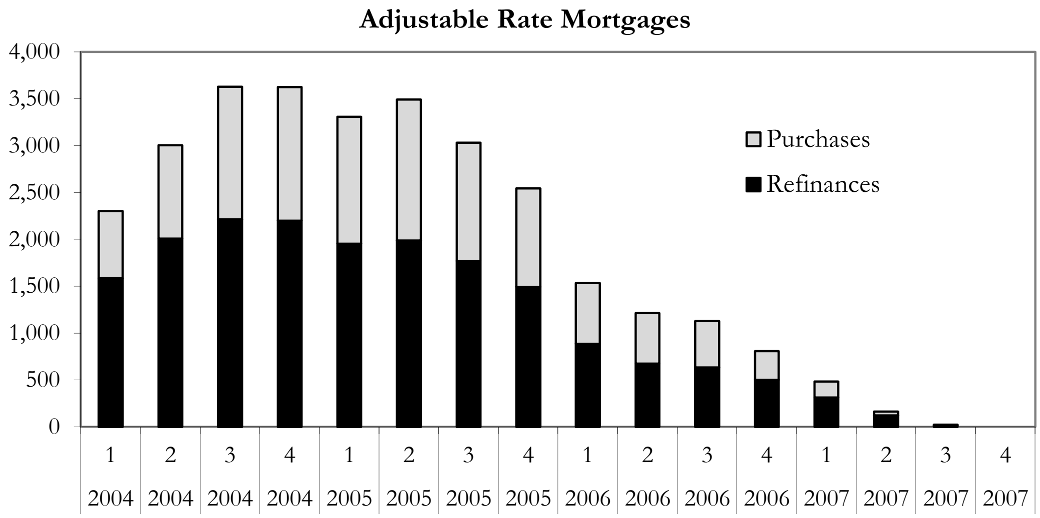
		Model 1	Model 2	Model 3	Model 4	Model 5
Neighborhood racial composition						
% Hispanic:	20-40%			1.011 (0.045)	1.018 (0.046)	
	40-100%			1.080 (0.059)	1.107 (0.063)	
% Asian:	20-40%			1.060 (0.058)	1.063 (0.058)	
	40-100%			1.160 (0.119)	1.182 (0.123)	
% Black:	20-40%			1.157 (0.062) **		
	40-60%			1.308 (0.073) **		
	60-80%			1.311 (0.077) **		
	80-100%			1.347 (0.076) **		
Borrower race/ethnicity						
Black borrower and neighborhood is:						
	0-20% black				1.267 (0.133) *	
	20-40% black				1.195 (0.128)	
	40-60% black				1.270 (0.118) **	
	60-80% black				1.290 (0.095) **	
	80-100% black				1.280 (0.083) **	
Non-black borrower and neighborhood is:						
	20-40% black				1.164 (0.075) *	
	40-60% black				1.333 (0.111) **	
	60-80% black				1.312 (0.110) **	
	80-100% black				1.522 (0.117) **	
Black borrower						1.043 (0.043)
Hispanic white borrower					0.902 (0.048)	0.902 (0.047) *
Asian borrower					0.963 (0.053)	0.967 (0.053)
Number of loan-months				946,922		
Number of loans				29,414		
Log pseudolikelihood		-52412.870	-52327.696	-52306.504	-52282.475	-52305.809

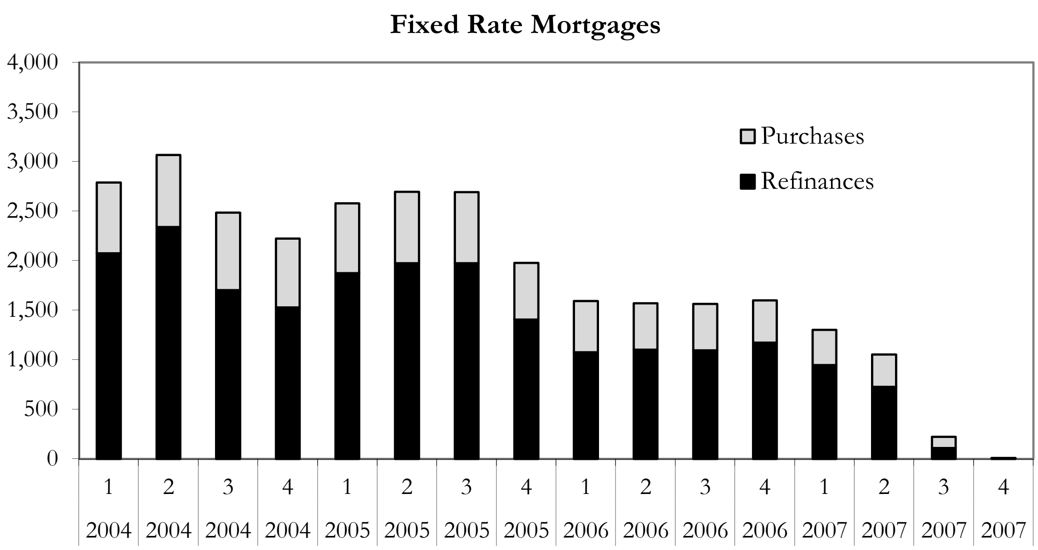
Cox proportional hazard models of 90-day default, hazard ratios reported. See tables 1 and 2 for notes on explanatory variables and the left-out categories.

Robust standard errors clustered by census tracts are reported in ( ). Statistical significance is indicated by: *5% and **1%.

All models include: a cubic in original loan balance; indicators for prepayment penalty in effect, cash-out refinance, second home, low documentation, property type (single & 2-4 family), building age (0-10, 11-50, >50 years & missing); indicators for missing values of FICO, DTI, prepayment penalty; and fixed effects for origination year and calendar quarter. Models 4 and 5 also include indicators for Hispanic black, Hispanic other race, non-Hispanic other race, missing race, female and missing gender.

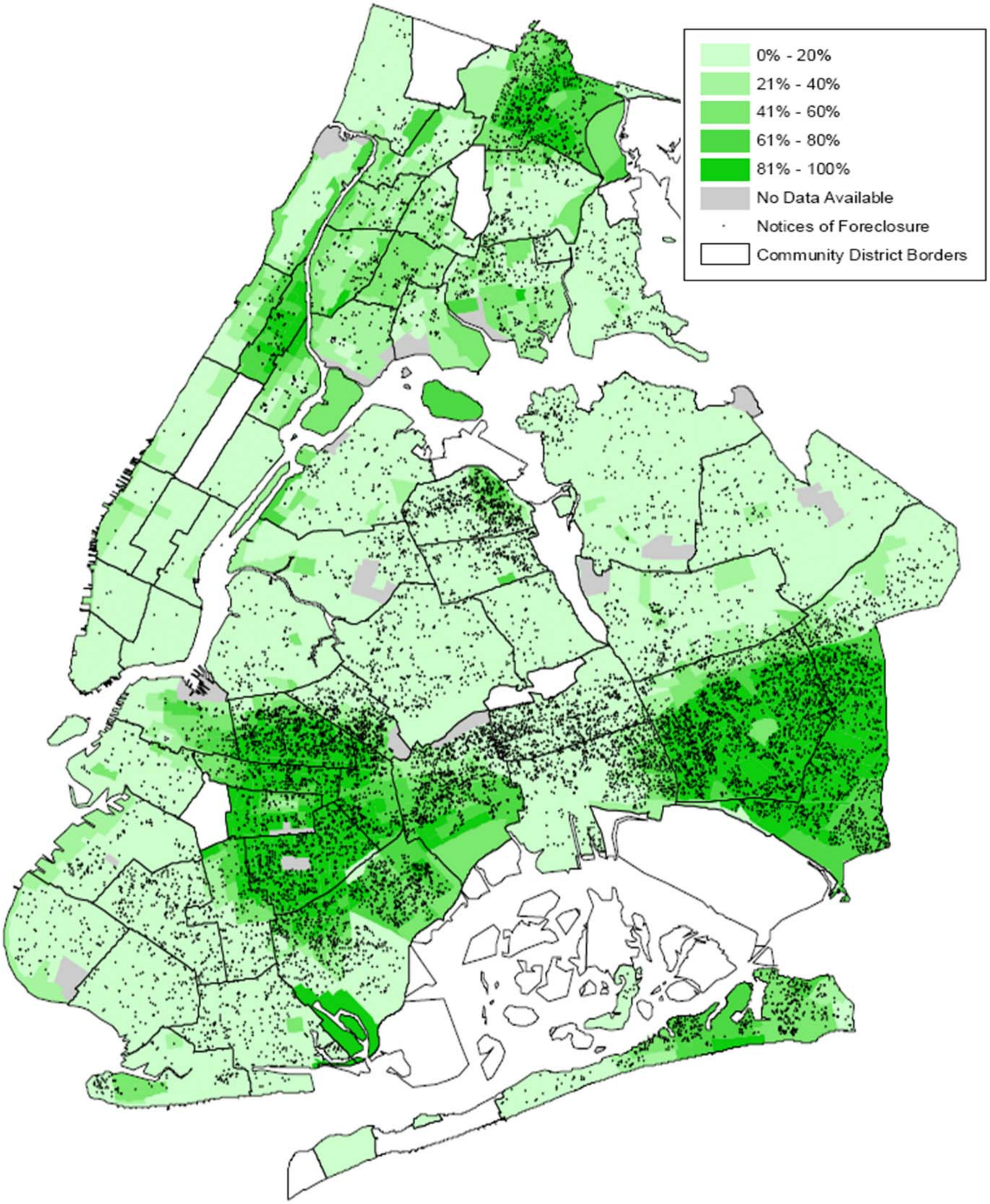
Figure 1: Originations in the Analysis Sample by Quarter





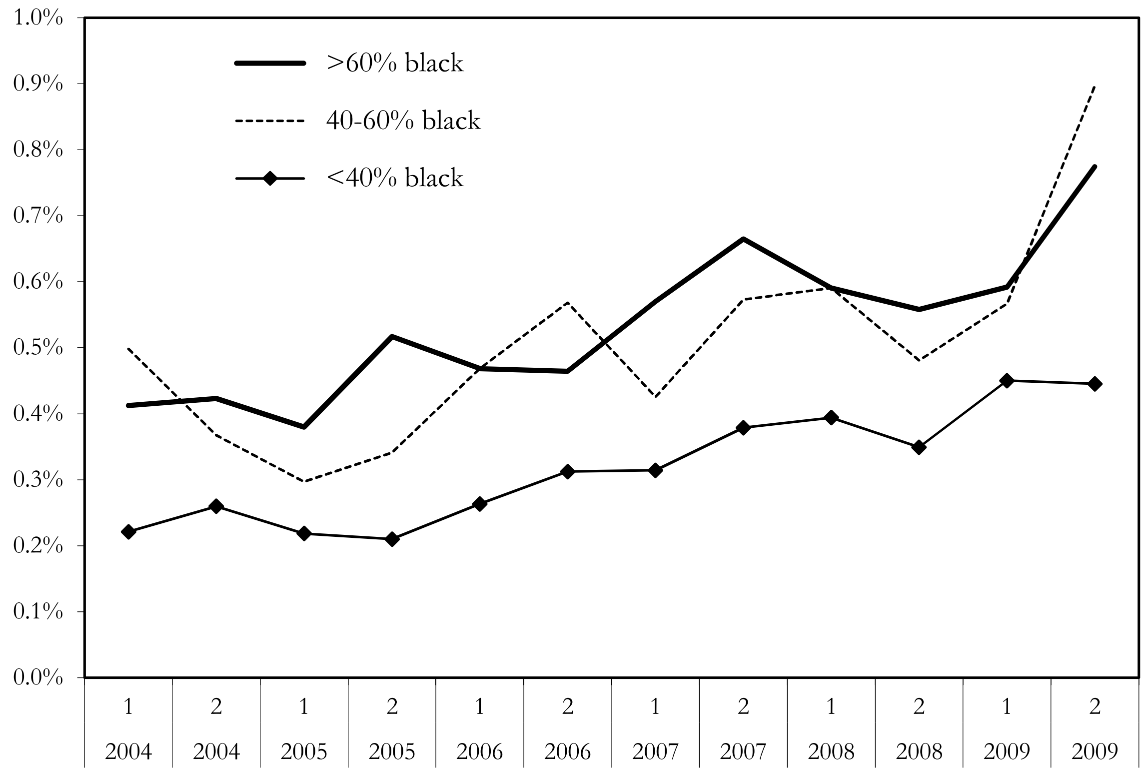
Source: LoanPerformance New York City sample, as described in the text.

Figure 2: Notices of Foreclosure and Percent Black Residents in New York City Census Tracts



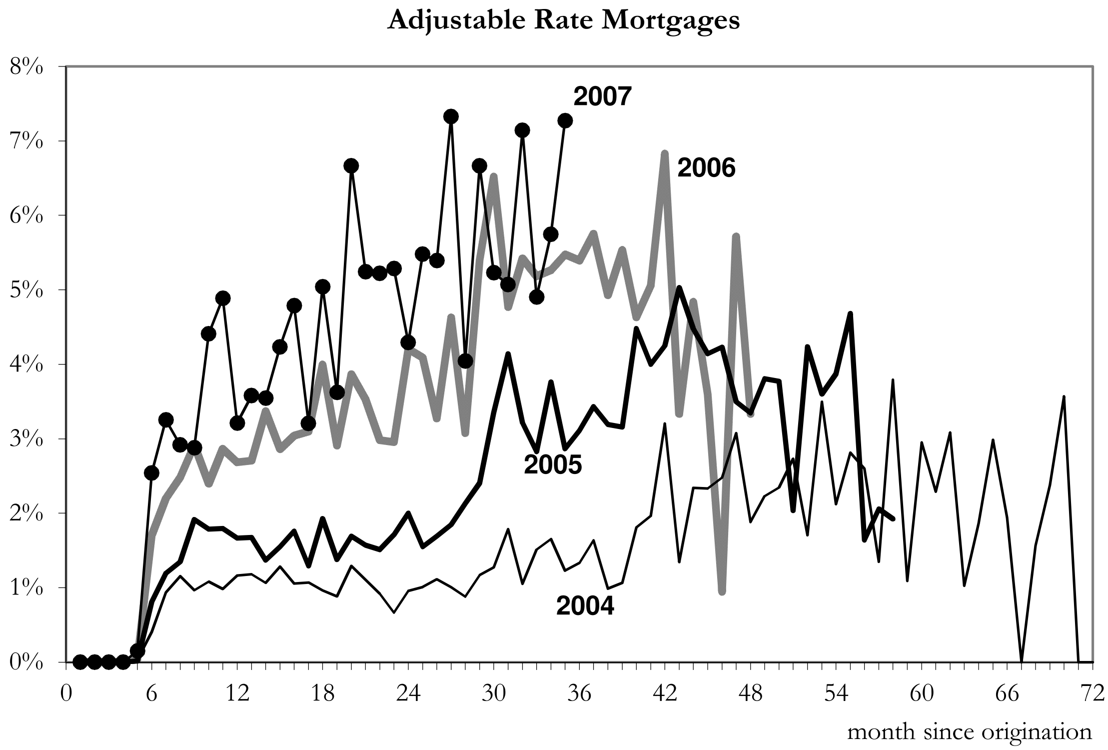
Each dot represents a notice of foreclosure (*lis pendens*) occurring in 2009. Source: Public Data Corporation. Census tracts are shaded based on the percent of black residents in the 2000 US Census.

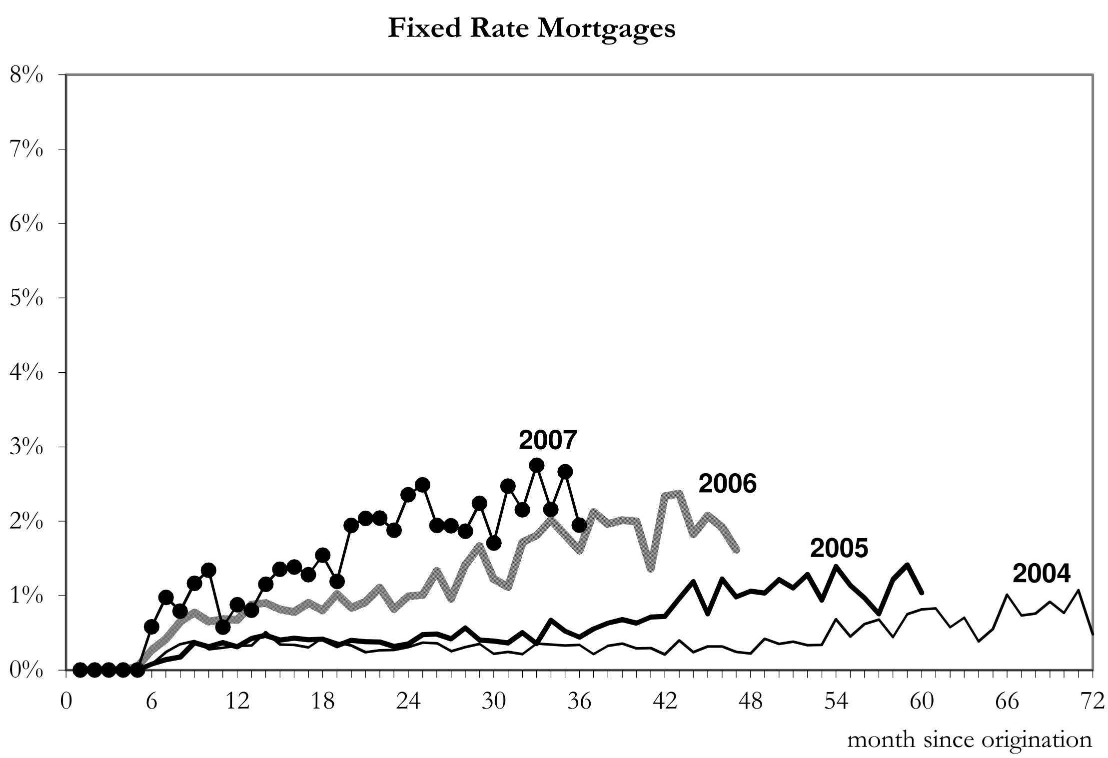
Figure 3: Census Tract Foreclosure Rates in New York City by Percent Black Residents



Percent of black residents in census tracts is from the 2000 US Census. The foreclosure rate is the number of foreclosure notices issued in a census tract within each 6-month period, divided by the stock of buildings in that tract. Source: Public Data Corporation.

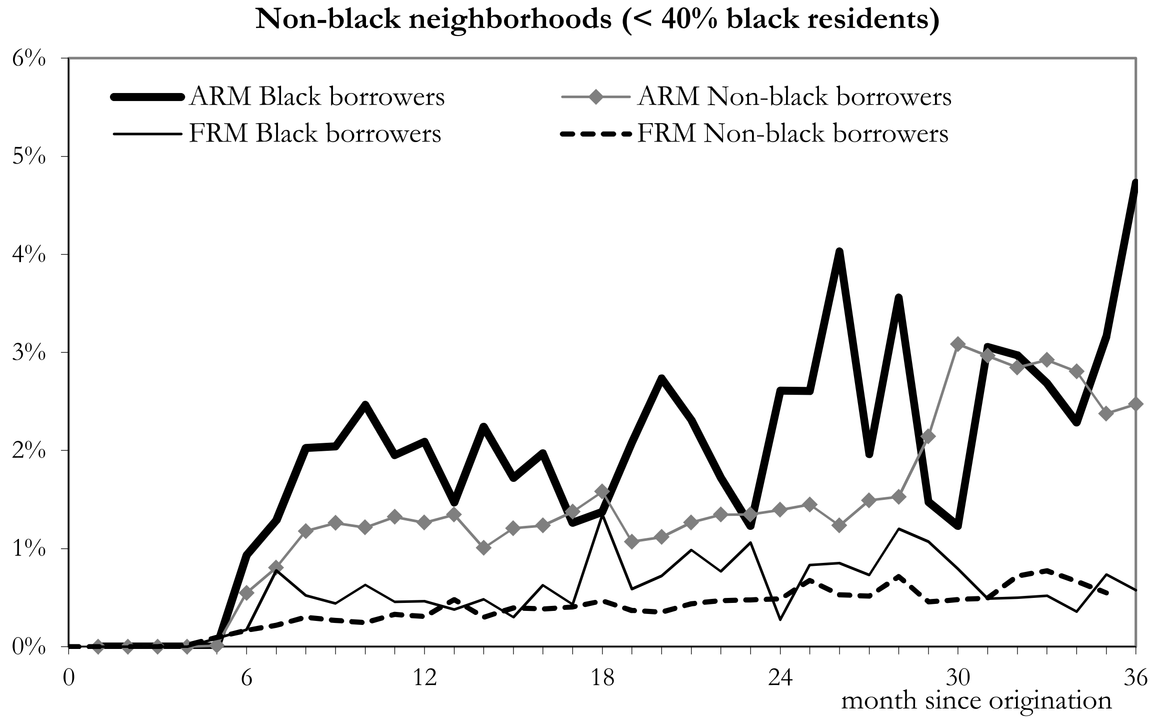
Figure 4: Default Hazards by Month Since Origination and Origination Year

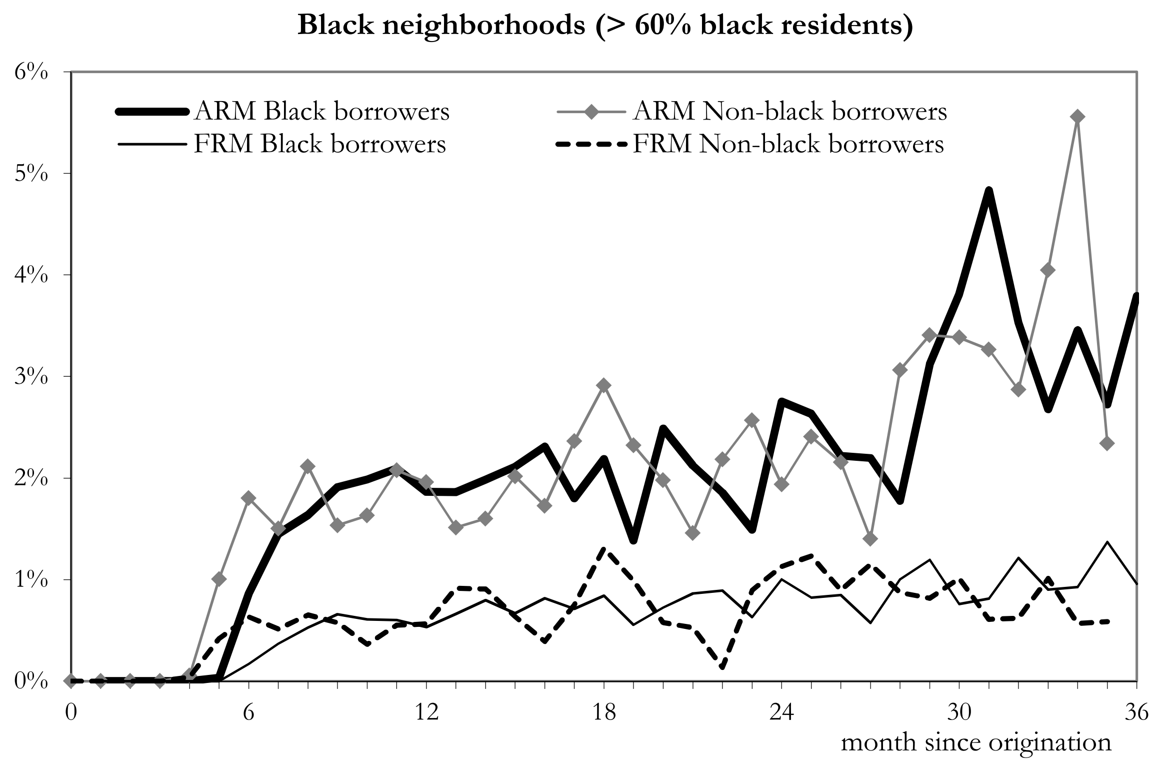




Source: LoanPerformance New York City sample, as described in the text.

Figure 5: Default Hazards by Month Since Origination, Borrower Race and Census Tract Percent Black





Source: LoanPerformance New York City sample, as described in the text. Percent of black residents in census tracts is from the 2000 US Census. "Black borrowers" includes only non-Hispanic blacks. "Non-black borrowers" includes all borrowers who reported a race, but did not report being black. Borrowers with missing race information are excluded.