PRELIMINARY AND INCOMPLETE
Do not cite or circulate without the authors' permission.
This Draft: October 28, 2009

# When Paths Diverge, Which Ones Lead to Ruin? Variation in Outcomes of Distressed Mortgage Borrowers (Part One)

**Sewin Chan**

**Vicki Been**

**Michael Gedal**

New York University's Furman Center for Real Estate and Urban Policy

**Andrew Haughwout**

Federal Reserve Bank of New York

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. We thank Caitlin Costlett, Johanna Lacoe and Mary Weselcouch for their assistance on this paper. The pathways project has benefitted enormously from the expertise, thinking and support of Amy Armstrong, Jackie Begley, Caroline Bhalla, Ingrid Gould Ellen, Solomon Green, Josiah Madar, Jenny Schuetz, Joseph Tracy, and participants in the Pathways Research Roundtable that launched the project. We are also grateful to Diego Aragón, Jenny Chan, Ebiere Okah and Jeff Seifert at the Federal Reserve Bank of New York for their help with the LoanPerformance data, and 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, and their patience in answering our many questions.
Correspondence should be sent to sewin.chan@nyu.edu.

## Project Abstract

What happens to distressed mortgage borrowers? Once a mortgage borrower falls behind on payments and enters financial distress, several different outcomes are possible. Some borrowers who become delinquent are able to obtain loan restructuring or forgiveness, allowing them to become current on their payments and remain in their homes. Others may sell the house and repay the loan (either in full, or with the shortfall forgiven by the lender) before foreclosure proceedings are completed. Although this results in their displacement, it allows them to avoid the damage to their credit rating from a completed foreclosure. Even among borrowers who move through the entire process and end up in a foreclosure auction, the outcome of the auction may vary: the property may be sold for more or less than the amount of the outstanding loan, the property may return to the lender (REO), or it may be sold to a third party.

For policymakers, predicting which outcomes are most likely in particular neighborhoods or among particular types of borrowers could have important consequences. If particular kinds of borrowers are less likely than others to secure or take advantage of loan modification programs, for example, the government and the financial industry may need to adjust the terms of those programs to allow or encourage those borrowers to participate. Or, if certain loan terms are associated with higher probabilities that a borrower in distress ends up in foreclosure, regulators may need to scrutinize the need for those loan terms more closely. Further, identifying the features of borrowers and neighborhoods that lead to more favorable outcomes should help public policy-makers better direct government efforts such as neighborhood stabilization to the areas likely to be hardest hit (or likely to be helped most by interventions). Finally, differences in outcomes of distress among types of borrowers, loans, and neighborhoods, should be useful to regulators both as they seek to understand the probability that the mortgage lenders will fail or need government assistance to survive, and as they implement long term regulatory solutions to the mortgage crises

Current research on foreclosures casts little light on the relative frequency of these outcomes or what factors influence the final result. We might expect that both the type of mortgage and borrower characteristics such as income and credit history will affect the ability of the borrower to negotiate more favorable loan terms and then resume payments. The strength of the local housing market and both property and neighborhood characteristics also likely affect the probability that a distressed borrower will be able to sell the house for enough to pay off the loan before final foreclosure.

We shed new light on these issues using a unique combination of data from New York City containing (i) individual mortgage data, with information on borrower characteristics and loan characteristics, (ii) dynamic mortgage performance data, including delinquencies, defaults, and loan modifications, (iii) foreclosure sales and auctions transaction data, and (iv) detailed information on property and neighborhood characteristics, including very local house price dynamics based on repeat-sales. We examine which factors are important in the eventual outcomes of distressed loans. The rich data allow us to incorporate the role of neighborhood dynamics, in particular, whether distressed borrowers who live in neighborhoods with many other distressed borrowers have different experiences from those who live in more stable neighborhoods.

In this first part of the analysis, we examine the role of borrower, loan and neighborhood characteristics on the probability that non-prime mortgages will default, or be 90 days or more delinquent. We find that several features of the neighborhood in which the property is located are significant predictors of the probability of default.

## 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 to examine the role of borrower, loan and neighborhood characteristics on default rates of non-prime mortgages, where "default" is defined as 90 days of delinquency. The dataset covers securitized mortgages in New York City, but its depth 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 effects while controlling for detailed individual and loan characteristics. The addition of neighborhood effects may cast some light on whether there is a need for different public policy strategies for different types of neighborhoods

This paper is the first part of a larger research project on mortgage distress and the choices made by mortgage holders in distress. In this paper, we launch the project at a natural starting point, by examining the determinants of mortgage distress. An understanding of how households get into mortgage distress is needed for an understanding of how households recover, or fail to do so.

Many other researchers have used the LoanPerformance database, a commercial database that is the major source of 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. The zip code level 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 2,100 census tracts, compared to 230 zip codes.

We have matched LoanPerformance records to actual parcels of land with a high level of precision, by matching loan records to real property deeds in New York City. The deeds data contain a unique tax lot identifier. This identifier allows us to merge information on the characteristics of the property and neighborhood to the loan records, based on the census tract. 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 borrowers' credit scores.

Another innovation in this paper is the use of quarterly calendar-time fixed effects to control for macroeconomic effects, including general city-wide house price movements. The remaining variation between quarters in our models is accordingly driven by differences in more local level house price dynamics. We are able to concentrate on very local house price movements because we have quarterly repeat-sales price indices at the community district level – a much finer level of geography than in the widely-used Case-Shiller and OFHEO indices. A "community district" is a political jurisdiction within New York City; there are 59 such districts in the City, and 56 are included in our sample.[1]

Importantly for our analysis, we also have precise information on the number of foreclosures at the census tract level. This allows us to examine whether there are contagion effects associated with foreclosure, an important policy concern given the geographic concentration of foreclosure across the country.

We report detailed results on the effects of borrower and loan characteristics on mortgage default. These results broadly confirm earlier results from the literature, while using the additional controls provided by our richer dataset. The primary contribution of the paper, however, is our finding that the neighborhood in which a mortgage holder lives has a powerful impact on the likelihood of default. Most importantly, mortgage holders living in poor or predominantly black neighborhoods have a substantially higher chance of falling behind on their mortgages, as do mortgage holders living in areas with high foreclosure rates. These effects exist even after conditioning on loan and individual characteristics, including the interest rates paid by borrowers. The poverty and race neighborhood effects may reflect that certain types of neighborhoods were either targeted for or were more likely to take out unsustainable loans, or they may reflect that certain neighborhoods are especially vulnerable to defaults because their residents have lower financial reserves or higher risk of job loss, uninsured expenses related to health crises, family instability, or other crises that may trigger default The effect of neighborhood foreclosure rates may reflect a local contagion effect or the generally negative local effects of substantial foreclosure activity.

[1] The community district is the smallest level of geography over which repeat-sales data can be constructed for NYC as there are often too few sales at the census tract level. As explained below, we drop Staten Island from our analysis, and this removes 3 community districts.

## 2. Related Literature

### Subprime Mortgage Distress

Empirical research has shown that subprime borrowers tend to default more frequently than prime borrowers, remain in delinquency for longer periods of time, and enter foreclosure more frequently, without controlling for loan terms or borrower characteristics. Subprime borrowers often have higher risk characteristics and less knowledge about the mortgage process than prime borrowers (Courchane, Surrette, & Zorn, 2004). As a result, subprime loans often are originated with junior liens, from banks that do not have local branches or from independent mortgage companies, and through brokers, all of which are characteristics associated with higher rates of default (Coulton, Chan, Schramm, & Mikelbank, 2008; Courchane, Surrette, & Zorn, 2004; Reid & Laderman, 2008). Subprime status is one of the strongest predictors of foreclosure (Coulton, Chan, Schramm, & Mikelbank, 2008; Ding, Quercia, & Ratcliffe, 2008; Gerardi, Shapiro, & Willen, 2007).

### Loan Characteristics, Borrower Characteristics and Mortgage Distress

There is an extensive and growing body of empirical literature on the relationship between loan terms, borrower financial characteristics, and loan performance (Capozza & Thomson, 2005, 2006; Cutts, & Van Order, 2005; Elmer & Seelig, 1998; Phillips & VanderHoff, 2004; Quercia & Stegman, 1992; Vandell, 1995). The emergence of two detailed, loan-level datasets – First American LoanPerformance and LPS Applied Analytics (formerly McDash Analytics) – has allowed recent studies to improve upon the existing literature by modeling the dynamics of the subprime mortgage market with increased power and precision (Danis & Pennington-Cross, 2005; Foote, Gerardi, Goette, & Willen, 2009; Haughwout, Mayer, & Tracy, 2009; Haughwout, Peach, & Tracy, 2008; Mayer, Pence, & Sherlund, 2009; Pennington-Cross & Ho, 2006; Quercia, Stegman, & Davis, 2007; Sherlund, 2008). Relying primarily on deeds data that were matched to data from the Home Mortgage Disclosure Act (HMDA), researchers at the Federal Reserve Bank of Boston also have done extensive work analyzing the causes and consequences of the subprime mortgage crisis (Gerardi, Lehnert, Sherland, & Willen, 2008; Gerardi, Shapiro, & Willen, 2007; Gerardi & Willen, 2008).

Historically, the loan terms found to influence mortgage defaults are: (1) the loan-to-value ratio, both at origination and over time as the loan matures, (2) the debt-to-income ratio (based on the monthly loan-service payments and income), (3) the credit score, (4) and whether income was documented at loan origination to the lender. In recent years, the mortgage interest rate has been added to this list, with particular focus on the performance of (5) adjustable-rate mortgages (ARMs) that are reset to market rates after a certain time period, compared with fixed-rate mortgages (FRMs).

The loan-to-value (LTV) ratio plays an important role in determining subprime mortgage default: the probability of delinquency and default increases with LTV for both subprime and Alt-A mortgages (Demyanyk, 2009; Haughwout, Peach, & Tracy, 2008). There is some evidence that it is current combined LTV, not initial combined LTV, that increases the probability of default (Sherlund, 2009). Among borrowers with negative equity, investors are more likely to default than borrowers who claim to be owner occupants (Haughwout, Peach, & Tracy, 2008). 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 (Danis & Pennington-Cross, 2005; Haughwout, Peach, & Tracy, 2008; Schloemer, Li, Ernst, & Keest, 2006). The effect of price appreciation on default may be asymmetric – when house prices increase, early defaults for subprime mortgages are reduced by a smaller margin than the margin by which they increase when house prices decrease (Haughwout, Peach, & Tracy, 2008). Schloemer, Li, Ernst, and Keest (2006) find that in the subprime market, strong housing price appreciation is related to increases in distressed prepayments (sale and prepayment to avoid foreclosure) and decreases in foreclosure rates, compared to housing markets experiencing less price appreciation where foreclosures are more likely (Schloemer, Li, Ernst, & Keest, 2006).

The debt-to-income (DTI) ratio is the debt share of a borrower's monthly gross income and a measure of the affordability of a mortgage. Several studies have found that higher initial DTI ratios 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 (Ding, Quercia, & Ratcliffe, 2008; Haughwout, Peach, & Tracy, 2008 ; Foote, Gerardi, Goette, & Willen, 2009).

The majority of studies on mortgage lending are not able to observe borrower credit scores or credit history, and researchers throughout the field routinely recognize this omission to be problematic. For those studies that do have this information, it appears that credit scores play a large role in predicting default (Demyanyk, 2009; Foote, Gerardi, Goette, & Willen, 2009; Haughwout, Peach, & Tracy, 2008).

Increasingly, research has focused on the effects of adjustable rates on the probability of mortgage default. Overall, adjustable rate subprime mortgages have a higher probability of default than fixed rate subprime mortgages (Haughwout, Peach, & Tracy, 2008; Quercia, Stegman, & Davis, 2007; Sherlund, 2009). However, Pennington-Cross and Ho (2006) assessed the effect of hybrid (2/28) subprime loans versus fixed rate subprime loans on the probability of default, and found a small, temporary increase in defaults when the ARM interest rate first resets, but found that after this point the propensity to terminate is less than that for subprime fixed-rate loans. In addition, Mayer, Pence, and Sherlund (2009) find that a change in interest rates two or three years after origination does not appear to be strongly associated with increased defaults, which they interpret to signify that default rates are high for hybrid mortgages not because of the loan type, but because of borrower risk. The mixed results may reflect the fact that despite the propensity of subprime borrowers with hybrid mortgages to prepay and default more than those with fixed rate mortgages, the fact that a loan is a hybrid is not a strong predictor of default standing alone (Demyanyk, 2009). The existence of an adjustable rate interacts with several other loan terms to affect default rates. Consistent with the findings cited above, higher credit scores often are related to lower probabilities of default for ARMs, and low or negative equity ARMs (measured by LTV) are more likely to default than higher equity ARMs (Pennington-Cross & Ho, 2006).

Although a variable indicating whether a borrower has low or no documentation of income is often included in default models, it is rarely cited as an important factor in predicting mortgage default. One study did find that low documentation of income is related to a 3 percent higher default rate for subprime borrowers, compared to like borrowers with documented income (Haughwout, Peach, & Tracy, 2008).

Few models of subprime mortgage default using detailed loan-level data are able to control for borrower race. In those studies able to identify borrower race, there is some evidence that black borrowers are more likely to default than white borrowers. Coulton, Chan, Schramm, and Mikelbank (2008) identify large racial disparities in the probability of receiving a subprime loan, a strong connection between subprime loan status and the probability of foreclosure, and high rates of foreclosure for black borrowers (40 percent of subprime loans to blacks will be in place in 3 years, while 60 percent of subprime loans to white borrowers will survive), although they are unable to control for borrowers' credit history.

### Neighborhood Racial Composition and Mortgage Distress

Research on the impact of neighborhood context, especially the racial composition of the neighborhood, on the probability of default is rather thin. Studies that measure the effects of borrower and loan characteristics on prime mortgage outcomes, controlling for neighborhood factors, find inconsistent evidence about whether neighborhood racial composition influences default (Berkovec, Canner, Gabriel, & Hannan, 1994; Firestone, Van Order, & Zorn, 2007).

There are a few studies that analyze aggregate default and foreclosure patterns instead of the behavior of individual borrowers. Neighborhood-level studies find a correlation between the census tract share of foreclosures or defaults and the share of minority residents or homeowners in the census tract (Apgar & Duda, 2005; Grover, Smith, & Todd, 2008; Pederson & Delgadillo, 2007; Van Order & Zorn, 2000)[2]. Importantly, however, these studies concede that minority share in a census tract may merely be a proxy for credit history, wealth, or economic conditions (Berkovec, Canner, Gabriel, & Hannan, 1994; Pederson & Delgadillo, 2007), and some find that when credit history is controlled for, the effect of the minority share disappears (Van Order & Zorn, 2000).

### Neighborhood Foreclosures and Mortgage Distress

The presence and extent of other defaults or foreclosures in a borrower's neighborhood may affect the probability that a borrower will default by decreasing the value of the collateral property. Neighborhoods may experience negative externalities from foreclosures, particularly in terms of the values of nearby housing, due to deterioration of the property in delinquency or foreclosure, disinvestment in the community, or decreased housing prices due to an increased supply of housing. Several studies have estimated the effect of foreclosures on nearby property values (summarized in detail by Schuetz, Ellen, and Been (2008)). In Chicago, Immergluck and Smith (2006) find that, on average, one additional 1-4 family foreclosure results in a one percent

[2] Van Order and Zorn (2000) focus on the effects of prime loans and Apgar and Duda (2005) focus on the effects of nonprime loans. Grover, Smith, and Todd (2008) have a sample of prime and subprime loans, while Pederson and Delgadillo (2007) are not able to determine the status of the loans in their analysis.

decrease in value for properties within a 1/8 mile radius in 1999. In Dallas, Leonard and Murdoch (2007) find that in neighborhoods with homeownership rates below 80 percent, each additional foreclosure within 250 feet of a sale results in a one percent decrease in sales price between 2003 and 2006. Lin, Rosenblatt and Yao (2009) find that foreclosures negatively affect sales prices up to 0.9 kilometers away from the sale and up to five years after the foreclosure in Chicago (sales in 2006). In St. Louis, Rogers and Winter (2009) find that the marginal impact of a single family home foreclosure on home sales between 2000 and 2007 is about one percent within 1/8 of a mile, and the marginal impact of foreclosures decline when the number of foreclosures increases. In New York City, Schuetz, Ellen, and Been (2008) find evidence that properties near foreclosures sell at a discount, although they find lower discounts than the other studies, perhaps because they control for price differences between properties that pre-existed the foreclosures (sales between 2000 and 2005). Further, the magnitude of the effect increases with the number of filings (although they are not linearly related). Using a repeat sales model that controls for property characteristics, Harding, Rosenblatt, and Yao (2009) estimate a contagion effect of the closest foreclosure to be less than 0.5 percent within 1/8th of a mile, and the cumulative effect of being near three or more foreclosed properties within 300 feet of a non-distressed sale is a 3 percent discount in market price. Further, the authors are able to determine the dates of the foreclosure sale and the REO sale, and they find that the strongest contagion effect occurs in the year leading up to the foreclosure sale.

## 3. Data Sources

To investigate the determinants of default, we rely on a unique database that combines rich loan-level information with data on neighborhood and property characteristics. We focus on a sample of non-prime loans originated in New York City between 2004 and 2008. The source of the loan-level data is the LoanPerformance database. Although LoanPerformance provides detailed information on borrower characteristics, loan terms, and payment history, it provides relatively little in terms of property or neighborhood characteristics. We therefore supplement the loan-level data with information from multiple sources.

First, we attach information on demographic characteristics of the neighborhood, at the census tract level. Second, we merge the data with a repeat home sales price index, which allows us to account for differential rates of house price appreciation in 56 different community districts of New York City over the study period. Third, we attach information on the concentration of notices of mortgage foreclosure (*lis pendens*) in the census tract. Finally, we merge information on the housing characteristics of the property.

**Loan Performance**

Our source of information on individual loans and borrowers is FirstAmerican CoreLogic's LoanPerformance data set, which, as of February 2009, provided loan-level information at a monthly frequency on approximately 4.8 million active, securitized subprime and alt-A loans in the United States. This represents almost 100 percent of securitized non-prime mortgages since 2003. However, it excludes all loans held in bank portfolios (Mayer and Pence 2008). Pennington-Cross (2002) argues that securitized subprime mortgages differ systematically from those retained in portfolio. Because our data are limited to securitized loans, any inferences should be limited to this set of loans.

Subprime mortgages are often made to borrowers with some blemish on their credit history, or who are willing to commit large shares of their incomes to debt service. Alt-A mortgages are typically larger value loans made to more credit-worthy borrowers who, for a variety of reasons, may choose not to provide the income or asset verification required to obtain a prime mortgage. Both types of nonprime mortgages are typically higher-cost than prime conforming loans.

The LoanPerformance data set is a rich source of information on the characteristics of these securitized loans. The data set includes information on the date of origination, the zip code in which the collateral property is located, details of the mortgage contract (including term, initial interest rate, and rate adjustment schedule), and underwriting information (including borrower credit score, debt–to-income ratio, the loan-to-value ratio for senior and junior liens at the time of origination, and the extent of income and asset verification provided by the borrower). Also included are monthly updates of "dynamic" information including the current interest rate, scheduled payment, mortgage balance and the borrower's payment record.

We limit our sample to loans originated in New York City[3] between 2004 and 2008.[4] For each loan, we observe changes in loan terms and delinquency status monthly through December 2008. We further limit the sample just first lien loans, and to 30-year fixed-rate mortgages (FRMs) and hybrid 2/28 and 3/27 adjustable-rate mortgages (ARMs). These three mortgage products represent almost two thirds of all the first lien LoanPerformance mortgages in New York City.[5] The hybrid ARMs' names indicate the rate adjustment schedules: 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 of the term, while on a 3/27 the initial rate is in effect for 3 years and floats for 27 years. Both types of hybrid ARMs reset rates are given by a "margin" which is added to the prevailing 6-month London Interbank Offered Rate (LIBOR), but individual loans may have periodic or lifetime rate floors or ceilings that may cause adjustments in a given period to differ from this simple formula.

**Matching Loans to a Tax Lot Identifier**

To match loans to a unique tax lot identifier, we relied on a database of property deeds in New York City. We obtained from the New York City Department of Finance the Automated City Register Information System (ACRIS) database, which lists all recorded deeds since 2003.[6] We rely on the following fields from ACRIS: deed date, mortgage amount, zip code, and a unique tax lot identifier. We merged LoanPerformance data to deeds based on three common

[3] Our sample excludes loans originated in Staten Island. See discussion below.
[4] Because loans enter the LoanPerformance database only once they have been securitized, our sample will not include any loans that were not securitized before December 2008.
[5] Other LoanPerformance mortgage products include other kinds of ARMs, interest only mortgages and balloon mortgages.
[6] The ACRIS database lists deed information for properties in Staten Island (Richmond County) in a different format than for the remaining four boroughs (counties) in New York City. We have not yet reformatted Staten Island records in a way that would allow for successful matching. Therefore, we were not able to include Staten Island properties in the match. All analyses presented here are restricted to loans in the four other New York City boroughs (Bronx, Brooklyn, Manhattan, and Queens) and exclude loans in Staten Island.

fields: date, mortgage amount, and zip code. The process included six stages of hierarchical matching.[7]

Ultimately, we are able to match approximately 95 percent of the LoanPerformance loans uniquely to a tax lot identifier. Our analysis focuses on this subset of 69,479 successfully matched loans. We are confident in the success of the matching algorithm for two reasons. First, we include as matches only loan and deed records that match uniquely. Second, because we are matching loans to the full universe of deed records, this greatly reduces the chance of false positive matching. We verified that the sample of matched loans does not significantly differ from the much smaller sample of unmatched loans in terms of borrower and loan characteristics.

### Additional Data Sources

We augment loan-level data with information from four sources. First, we merge neighborhood demographics from the 2000 Census, including race/ethnicity[8] and poverty rate, all at the census tract level. Second, we include a series of quarterly repeat-sales home price indices developed by the Furman Center for Real Estate and Urban Policy.[9] We use a separate repeat sales price index for each of the 56 community districts in our sample. Third, we attach

[7] ***Stage 1*** considered the full set of loans and the full set of deeds, matching on the raw values of date, loan amount and zip code; loans and deeds that uniquely matched each other in stage 1 were then set aside and considered matched. ***Stage 2*** considered only the subset of loans and deeds that went unmatched during stage 1, matching on the raw values of date and zip code, and the loan amount rounded to $1,000. ***Stage 3*** considered only the subset of loans and deeds that went unmatched during stage 2, matching 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 allowing 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 allowing 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 allowing dates to differ by up to 90 days. We believe it is valid to introduce a 90-day window for "fuzzy" matching for the following reasons. First, we know that for many loans in the LoanPerformance data the actual origination date is not available, but is instead imputed by backdating the first payment date by one month. Second, there may be administrative lags or data errors in the recording of the deeds data. Again, because we are matching loans to the full universe of deed records, we are confident that this matching algorithm is appropriate.

[8] We identify five mutually exclusive race/ethnicity categories: Hispanic; black non-Hispanic; white non-Hispanic; Asian non-Hispanic; other non-Hispanic. For the remainder of the paper, we refer to these race/ethnicity categories as follows: Hispanic; black; white; Asian; other.

[9] For our hazard models, we transform these quarterly indices into smooth monthly series by interpolation. For a description of the indices, see Armstrong et al. (2009). In this paper, we use the indices that include sales for all housing types, regardless of the dominant housing type in the neighborhood.

information on the number of notices of mortgage foreclosure (*lis pendens*) in the census tract; we obtained data on *lis pendens* filings from a private vendor, Public Data Corporation.[10]

Finally, we attach information for property characteristics using data provided by the New York City Department of Finance. This information is collected annually for the purposes of computing property tax assessments and is provided in the Real Property Assessment Data (RPAD) file. Our analysis relies on the following fields from RPAD: the unique tax lot identifier and building type (single-family home, 2-4 family home, 5+ family home, and condominium).[11]

## 4. Data Description

As discussed above, we limit the analysis to the 69,479 loan records that were successfully matched to a unique tax lot identifier in the New York City deeds data. Table 1 describes this sample. The sample is split roughly equally between FRMs and ARMs. Approximately three-quarters of the ARMs are 2/28 hybrids, with the remainder being 3/27 hybrids. As the table indicates, ARM borrowers have considerably higher default rates. This is not surprising because ARM borrowers also tend to have much lower credit (FICO) scores, somewhat higher initial interest rates, and higher initial loan-to-value ratios.

### Loan and Borrower Characteristics

Table 1 displays mean values for the measures of loan and borrower characteristics we analyze.[12] The margin for ARMs is the amount added to the six-month LIBOR rate to determine the adjustable rate in the future; the margin is between 6 and 7 percent for almost half of the adjustable-rate loans in our sample. The interest rate at origination for ARMs is more than half a percentage point higher than that for fixed-rate mortgages. The relative interest rate at origination is calculated as the interest rate minus the six-month LIBOR rate at origination for

[10] A *lis pendens* may be filed for a host of reasons unrelated to a mortgage foreclosure. The Furman Center uses a variety of screening techniques to identify only those *lis pendens* related to a mortgage. Further, if the same property received any additional *lis pendens* within 365 days of the initial *lis pendens*, the additional *lis pendens* are not included in our rate to avoid double-counting the same foreclosure. See Armstrong et al. (2009)

[11] Single-family homes include both detached and attached single-family structures. Our sample does not include co-ops.

[12] Note that for categorical variables, the reference category (categories) used in the hazard models are indicated in Table 1 in italics.

ARMs, and the interest rate minus the Freddie Mac average interest rate for prime 30-year fixed rate mortgages during the month of origination for FRMs. The average loan amount is slightly larger for FRMs ($360,000) than ARMs ($350,000). Payment shock at the time of the first ARM readjustment represents that jump in monthly payments that 2/28 borrowers experience at month 25 (after completing two years of payments) and 3/27 borrowers experience at month 37 (after three years of payments). The vast majority of loans experience a payment shock of less than 20 percent at this time. While payments will also adjust every six months following the first adjustment, these subsequent adjustments are tiny in comparison because of the wide-spread use of teaser rates for the initial fixed period.

Measures of borrower risk include FICO credit score[13], debt-to-income ratio (DTI) and combined loan-to-value ratio (LTV). The LTV measure combines the loan amounts for the first lien mortgage (that is the focus of this analysis) as well as any second lien existing at the time of origination. Any junior liens that are taken out subsequent to origination will not be reflected in this measure. Table 1 summarizes these variables as of origination. As noted above, ARM borrowers tend to have lower FICO scores and higher DTI and LTV at origination, compared to fixed-rate mortgage holders. It is important to note that in Table 1 LTV is reported as of *origination*. This is in contrast to our hazard models, where LTV is reported *contemporaneously* and is adjusted over time to reflect changes in the outstanding loan balance and the home value.[14]

Our analysis also relies on measures of loan purpose, owner type, and building type. The dominant loan purpose is cashout refinance (64 percent for FRMs and 57 percent for ARMs), followed by home purchase (30 percent and 40 percent) and non-cashout refinance (7 percent and 3 percent). In terms of owner type, about 90 percent of borrowers claim to be owner-occupants of mortgaged properties, both for ARMs and FRMs.[15] The distribution across building types is also similar for ARMs and FRMs. The majority of loans are collateralized by

[13] The Fair Isaac Corporation (FICO) credit score is the most widely used credit score model in the United States. The exact model for calculating the score is a trade secret, but it depends on payment history, credit utilization, length of credit history, types of credit used and recent searches for credit.

[14] The loan balance is adjusted using information from LoanPerformance. We do not have information on the dynamic loan balance of the second lien, and so we assume that this is an interest only mortgage where the principal does not change. Our hazard model results are not affected if instead we assume this second lien has the same amortization schedule as the first lien mortgage. Home values are adjusted using the community district level home price indices described above.

[15] These self-reported claims of owner-occupancy may be unreliable as, all else equal, owner-occupiers tend to get lower cost mortgages than investors.

2-4 family buildings (about 57 percent), followed by single-family (about 35 percent), condominiums (about 5 percent), and 5+ family buildings (about 1.5 percent).

**Neighborhood Characteristics**

Table 2 summarizes the distribution of loans across neighborhoods. In particular, we examine neighborhood poverty, race/ethnicity, and foreclosure concentration. All neighborhood data are reported at the census tract level.

The neighborhood foreclosure rate is defined as: the number of notices of foreclosure (*lis pendens*) issued in the census tract in the prior six months (restricted to 1-4 family buildings), divided by the total number of 1-4 family buildings in the census tract.[16] It is important to note that Table 2 reports the neighborhood foreclosure rate in the six month period preceding the origination month; in contrast, for the hazard models, we use the foreclosure rate in the six month period preceding the month of analysis.

It is also important to emphasize that our neighborhood foreclosure measure is based on a sample that differs in three important ways from the sample of loans we analyze. First, the listing of foreclosures includes only 1-4 family homes; our sample of loans, on the other hand, includes 1-4 family properties, 5+ family homes and condominiums. Second, unlike the sample of foreclosures (which is drawn from the full universe of active mortgages for 1-4 family homes in New York City), our sample of loans is drawn from the subset of non-prime mortgage loans that are 30-year fixed-rate, 3/27 ARMs, or 2/28 ARMs and that appear in the LoanPerformance database. Finally, the sample of foreclosures may include loans that were originated before 2004 while our sample includes only loans originated between 2004 and 2008.

Table 2 indicates that the majority of loans are in census tracts that experienced a foreclosure rate of less than one percent in the six months prior to loan origination. About 15 percent of FRMs and 12 percent of ARMs were in neighborhoods where the foreclosure rate was 2 percent or more.

[16] Note that we restrict to 1-4 family buildings only for the purpose of measuring foreclosure activity in neighborhoods, because this is the format in which we received the data. Our loan sample is *not* restricted to 1-4 family buildings.

Figure 1 provides a snapshot of the distribution of foreclosures across New York City community districts in 2008.[17] Each dot represents one notice of foreclosure for a 1-4 family building. The map shows that foreclosure activity is concentrated in a handful of community districts. It further shows that foreclosures tend to be heavily concentrated in majority non-white community districts.

In terms of poverty concentration, the majority of loans are located in census tracts where the poverty rate was between 10 and 30 percent in 2000. For context, the citywide poverty rate was 21 percent in 2000. Adjustable-rate mortgages are somewhat more likely to be in higher poverty tracts than loans with a fixed-rate.

Turning to neighborhood racial and ethnic composition, a considerable number of loans are in tracts that were at least 60 percent black; 38 percent of FRMs were in such areas, compared to almost half of ARMs. These figures may be surprising, considering that blacks made up just 25 percent of New Yorkers in 2000, but they likely reflect the influence of two important factors: first, there is evidence that black borrowers are more likely to receive non-prime loans that similarly situated non-black borrowers (Mayer and Pence, 2008); second, these figures also reflect relatively high levels of residential racial segregation in New York City. The pattern we observe is consistent with black borrowers being more likely both to receive non-prime loans and to live in neighborhoods that have a high concentration of black residents. Finally, FRMs are more likely than ARMs to be in tracts that are predominantly white, Hispanic or Asian; in predominantly black tracts, the pattern flips.

[17] For Figure 1, neighborhood racial composition is from the 2007 American Community Survey. For the purposes of this figure, neighborhood racial composition has been aggregated to the community district level. Note that in the main analysis, we examine racial composition at the census tract level. In this figure, we display racial composition at the community district level for the simple reason that census tracts are too small to display the information in a meaningful way.

## 5. Empirical Specification and Results

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:

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

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. We define a default as a mortgage's first 90 day delinquency. 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 $\alpha$, $\beta$, $\gamma$ and $\delta$ 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. We estimate separate hazard models for the fixed rate mortgages and the hybrid ARMs.

Our hazard model results are displayed in Table 3. 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. The first set of rows in the table show the hazard ratios associated with loan pricing terms. Virtually all of these terms are strongly statistically significant, as we would expect. For ARMs, having a margin of 7 percent is associated with a 36 percent higher default hazard compared with a margin that is less than 5 percent (the margin is the amount added to the six-month LIBOR rate to determine the adjustable rate in the future). An ARM with an initial interest rate that is over 6 percentage points higher than the LIBOR six-month index at the time of origination has over six times the default hazard as an ARM with an initial rate that is under 1 percent above LIBOR (the left out category). For FRMs, a rate that is 3 percentage points above average at origination is associated with a seven and a half times higher default hazard than a below average rate. For both fixeds and adjustables, higher balance loans have significantly higher default rates.

The next set of explanatory variables in Table 3 measure the size of the payment shock upon the first adjustment of ARMs. In the model, the percentage increase in payment is interacted with the number of months since the adjustment. Consistent with previous research (Pennington-Cross & Ho, 2006), we find that even hybrid ARMs without much of a payment shock (less than 20 percent increase in payments) experience a significantly elevated default hazard following the initial adjustment. From 7 months after the initial adjustment, the default hazard is 50 percent higher for this group compared with before the adjustment (the left out category). Larger payment shocks are associated with even larger increases in the default hazard post initial adjustment. For those with payment shocks greater than 30 percent, the default hazard is over twice as high 5 months post adjustment compared with before adjustment. We also find that the 3/27 ARMs have significantly lower default rates overall than do the 2/28 ARMs. These finding may be contrary to that of Gerardi, Shapiro, and Willen (2007) who determined that house price depreciation is the primary driver of foreclosures, even when controlling for the six month LIBOR rate (an index for adjustable rate mortgages).

Borrower risk characteristics generally have the expected effect on default behavior. Lower credit scores have a consistently positive effect on the default hazard. In the case of FRMs, a borrower with a FICO less than 560 has a default hazard that is five times larger than an FRM borrower with a FICO higher than 720 (the left out category). For ARMs, the credit score does not play as large a role in predicting default, probably because the payment shocks are more important. ARM borrowers with FICOs less than 590 have an over 70 percent higher default hazard than those with FICOs higher than 720.

Higher debt-to-income ratios at origination matter for FRMs, though the magnitude of the hazard ratio is small compared with many of the other covariates. DTIs higher than 50 percent are associated with a 16 percent higher default hazard compared with DTIs lower than 45 percent. For ARMs, the 45-50 percent DTI group has a 15 percent higher default hazard, while the effect is positive though insignificant for those with DTIs of 50 or more.

Consistent with virtually all prior research, the current combined LTV has large and significant effects on default. For FRMs, a current LTV higher than 95 is associated with a default hazard that is 3 and a half times higher than FRMs with LTVs lower than 60. For ARMs, the effect is more than twice as large. We also tried other model specifications, adding measures of community district level house price appreciation, both in the past year, and over the life of the mortgage. These variables were never significant when all the other displayed controls were included.

We find that home purchase loans have higher default rates than do refinances, possibly 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 for ARMs compared with investors, though as we note earlier, owner-occupancy is self-reported and maybe unreliable. This unexpected result might reflect differential location choices rather than risk attributes of the borrower. We also find that single family residences have higher default rates than buildings with 5 or more units, including condos. This could be because multi-unit buildings often provide some additional oversight on the financial health of buyers.

For the neighborhood characteristics, the results from our hazard models highlight important spatial patterns in default. We detect a contagion effect on the hazard of default for both fixed rate and adjustable rate loans. A 1-2 percent foreclosure rate in the surrounding census tract increases the default hazard by 14 percent for FRMs and by 17 percent for ARMs, compared with mortgages in neighborhoods where the foreclosure rate is less than 1 percent. As the foreclosure rate in the neighborhood increases, so does the default hazard – a foreclosure rate of over 4 percent increases the default hazard by 46 percent for FRMs and by 31 percent for ARMs, relative to tracts with the lowest foreclosure rates.

This may be the result of negative externalities resulting from surrounding foreclosures, such as the visible deterioration of properties, which lead to decreased property values in the neighborhood, and cause more borrowers to go underwater. Alternatively, as suggested by Harding, Rosenblatt, and Yao (2009), foreclosure activity in a neighborhood increases the housing supply, potentially driving down market prices. As noted earlier, we also tried to include measures of house price appreciation in the community district both in the year preceding the analysis month, and since origination of the mortgage, but none of these covariates were statistically significant after controlling for current LTV and the foreclosure rate. Because the foreclosure rates are at the census tract level, while our house price indices are at the larger community district level, it is quite possible that the foreclosure rates are a better indicator of very local housing values, especially because our calendar time fixed effects are already controlling for the effect of general house price movements in the city.

The importance of the foreclosure contagion effect may also be attributable to increased information. Neighbors may share information about the efficacy of default or the foreclosure process, leading other neighbors struggling with mortgage payments to enter into default. Or, high neighborhood foreclosure rates may reduce the stigma associated with defaulting on a mortgage. Alternatively, perhaps neighbors tend to suffer correlated income shocks especially if they have the same socioeconomic background or human capital, so that job losses in one industry or occupation affect many residents in the same neighborhood.

If racial and economic segregation contribute to systematic differences in the types of mortgages borrowers living in different neighborhoods receive, we may detect disparities in default rates across neighborhoods even after controlling for a variety of borrower and loan characteristics. Indeed, our analysis identifies a significant, nonlinear relationship between the poverty rate of a borrower's neighborhood and the likelihood of default. For all borrowers, living in census tracts with low poverty (10-20 percent in poverty) does not affect the default hazard compared to tracts with the lowest poverty (10 percent or less in poverty). Over one third of the borrowers in our sample reside in these low-poverty tracts. However, as the share of tract residents in poverty increases to 40-50 percent, the default hazard also increases by 22 percent for FRMs and 45 percent for ARMs, relative to the lowest poverty tracts. Highly concentrated poverty (over 50 percent of residents in poverty) has a large effect on the default hazard for FRM borrowers (44 percent), and a smaller effect for ARM borrowers (29 percent). Given the small number of loans originated to borrowers living in neighborhoods with highly concentrated poverty (approximately 1 percent of our sample), the results for ARMs may reflect the lack of a contagion effect in neighborhoods with very few homeowners.

While residing in a census tract with a proportion of black residents that is under 40 percent does not affect default rates for adjustable rate mortgages and only marginally increases the default hazard for FRMs, once the share is over 40 percent the default hazard significantly increases. A sizable share of our loans are made to borrowers living in tracts with large shares of black residents – one third of ARMs, and over one quarter of FRMs are originated in census tracts that are over 80 percent black. For borrowers in these tracts, the default hazard is about 30 percent higher compared to borrowers in tracts with the lowest shares of black residents (0-20 percent). Very few of the borrowers in our sample live in neighborhoods with high concentrations of Hispanics or Asians, and we do not see any effect of the share of the neighborhood population that is Hispanic or Asian on default outcomes. However, in New York City, many neighborhoods are home to both blacks and Hispanics, and often people of other backgrounds as well, therefore the impact of living in a largely minority neighborhood may be a significant predictor of mortgage default.

These spatial patterns suggest that neighborhood characteristics play a significant role in default outcomes beyond the effects of individual borrower and loan characteristics. To the extent that non-prime loans were targeted to neighborhoods with higher poor and minority populations, the contagion effect exacerbates the harm foreclosures in the targeted neighborhoods may be causing. Even in the absence of targeting, the contagion effect may require policy-makers to take into account the disproportionate harm that foreclosures may have in poor neighborhoods and black neighborhoods.

## 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, the characteristics of the property, and characteristics of the neighborhood (measured at the census tract level) have on the probability that a non-prime mortgage will default. Our findings reveal that several important results from prior research hold true even when credit scores and other borrower characteristics, detailed information about the property and neighborhood, and controls for macroeconomic trends in the housing market are added to the model. Lower credit scores are associated with a higher default hazard, as one would expect. Higher initial interest rates on both FRMs and ARMs, as well as higher margins on ARMs, increase the probability of default. Larger mortgages, whether FRMs or ARMs, have higher default rates. Hybrid 2/28 ARMs have higher default rates than 3/27 ARMs; for both types, default rates increase after adjustment, and larger adjustments are associated with more defaults. Higher debt-to-income ratios at origination have a relatively small effect on default rates, but the current LTV has a large effect on the default hazard of FRMs, and an even larger effect on ARMs. Home purchase loans have higher default rates than refinances.

But the most important finding is that neighborhood characteristics significantly affect the default rates. As the rate of foreclosure notices filed in the neighborhood increases, the hazard of default increases for both FRMs and ARMs. Living in tracts with high poverty rates has a large effect on the default hazard; the pattern is non-linear and differs somewhat for ARM and FRM borrowers. Residing in a census tract with less than 40 percent black residents does not affect default rates for ARMs, and increases the rate for FRMs only marginally, but the default hazard significantly increases when the share of black residents exceeds 40 percent.

Our finding that neighborhood characteristics have significant effects on the default hazard suggests that policy-makers must take neighborhood context into account in designing their responses to the foreclosure crisis, and in shaping the regulation of mortgage products and the financial industry. Part Two of our analysis will examine the effects borrower, loan, property and neighborhood effects have on the pathways borrowers follow once they have defaulted on their mortgages, and will assess further how differences among neighborhoods affect the way in which mortgage distress plays out.

## References


Ambrose, B. W. and Capone, C. A. (1998). Modeling the conditional probability of foreclosure in the context of single-family mortgage default resolutions. *Real Estate Economics, 26*, 391-429.

Anderson, R. and VanderHoff, J. (1999). Mortgage default rates and borrower race. *Journal of Real Estate Research,* 18(2), 279-288.

Apgar, W. C. and Duda, M. (2005). Mortgage foreclosures in Atlanta: Patterns and policy issues. Washington, D.C.: NeighborWorks America.

Armstrong, A. et al. (2009). State of New York City's housing and neighborhoods 2008. Furman Center for Real Estate and Urban Policy.

Been, V., Ellen, I. G., and Madar, J. (2009). The high cost of segregation: Exploring racial disparities in high cost lending. *Fordham Urban Law Journal, 36*(3), 361-393.

Berkovec, J., Canner, G., Gabriel, S., and Hannan, T. (1994). Race, redlining, and residential mortgage loan performance. *Journal of Real Estate Finance and Economics 9*(3), 263–294.

Capozza, D. R. and Thomson, T. A. (2005). Optimal stopping and losses on subprime mortgages. *The Journal of Real Estate Finance and Economics, 30*(2), 115-131.

Capozza, D. R. and Thomson, T. A. (2006). Subprime transitions: Lingering or malingering in default? *Journal Real Estate Finance and Economics, 33*(3), 241-258.

Clapp, J. M., Goldberg, G. M., Harding, J. P., and LaCour-Little, M. (2001). Movers and shuckers: Interdependent prepayment decisions. *Real Estate Economics, 29*(3), 411-450.

Coulton, C., Chan, T., Schramm, M., and Mikelbank, K. (2008). Pathways to foreclosure: A longitudinal study of mortgage loans, Cleveland and Cuyahoga County, 2005-2008. Center on Urban Poverty and Community Development, Mandel School of Applied Social Sciences, Case Western Reserve University, Cleveland, OH.

Courchane, M. J., Surrette, B. J., and Zorn, P. M. (2004). Subprime borrowers: Mortgage transitions and outcomes. *Journal of Real Estate Finance and Economics, 29*(4), 365-392.

Crews Cutts, A. and Merrill, W. A. (2008). Interventions in mortgage default: Policies and practices to prevent home loss and lower costs. Freddie Mac Working Paper #08-01.

Crews Cutts, A. and Van Order, R. A. (2005). On the economics of subprime lending. *The Journal of Real Estate Finance and Economics, 30*(2), 167-196.

Danis, M. H. and Pennington-Cross, A. (2005). The delinquency of subprime mortgages. St. Louis: Federal Reserve Bank of St. Louis Research Division, Working Paper 2005-022A.

Demyanyk, Y. S. (2009). Quick exits of subprime mortgages. *Journal Federal Reserve Bank of St. Louis Review, 91* (2), 79-93.

Deng, Y., Quigley, J. M., and Van Order, R. (2000). Mortgage terminations, heterogeneity, and the exercise of mortgage options. *Econometrica, 68* (2), 275-307.

Ding, L., Quercia, R. G., and Ratcliffe, J. (2008). Risky borrowers or risky mortgages: Disaggregating effects using propensity score models. Center for Community Capital Working Paper.

Elmer, P. J. and Seelig, S. A. (1998). The rising long-term trend of single-family mortgage foreclosure rates. Federal Deposit Insurance Corporation, Division of Research and Statistics, Working Paper 98-2.

Firestone, S., Van Order, R. and Zorn, P. (2007). The performance of low-income and minority mortgages. *Real Estate Economics, 35*(4), 479-504.

Foote, C., Gerardi, K., Goette, L., and Willen, P. (2009). Reducing foreclosures: No easy answers. National Bureau of Economic Research, Working Paper No. 15063.

Foster, C., and Van Order, R. (1984). An option-based model of mortgage default. *Housing Finance Review, 3*(4), 351-372.

Foster, C., and Van Order, R. (1985). FHA terminations: A prelude to rational mortgage pricing. *Journal of The American Real Estate and Urban Economics Association, 13*(3), 273-291.

Gerardi, K. S. and Willen, P. S. (2008). Subprime mortgages, foreclosures, and urban neighborhoods. Federal Reserve Bank of Boston, Working Paper No. 08-6.

Gerardi, K. S., Shapiro, A. H., and Willen, P. S. (2007). Subprime outcomes: Risky mortgages, homeownership experiences, and foreclosures. Federal Reserve Bank of Boston, Working Paper, No.07-15.

Grover, M., Smith, L., and Todd, R. M. (2008). Targeting foreclosure interventions: An analysis of neighborhood characteristics associated with high foreclosure rates in two Minnesota counties. *Journal of Economics and Business, 6I,* 91-109.

Harding, J. P., Rosenblatt, E., and Yao, V. W. (2009). The contagion effect of foreclosed properties. *Journal of Urban Economics*, *66* (3), 164-178.

Haughwout, A., Mayer, C., and Tracy, J. (2009). Subprime mortgage pricing: The impact of race, ethnicity, and gender on the cost of borrowing. Federal Reserve Bank of New York Staff Report No. 368, 33-63.

Haughwout, A., Peach, R., and Tracy, J. (2008). Juvenile delinquent mortgages: Bad credit or bad economy? *Journal of Urban Economics, 64*, 246-257.

Immergluck, D. and Smith, G. (2006). The external costs of foreclosure: The impact of single-family mortgage foreclosures on property values. *Housing Policy Debate*, 17 (1), 57-79.

Leonard, T., and Murdoch, J. (2007). Homeownership, foreclosure, and neighborhood quality. Working paper.

Lin, Z., Rosenblatt, E., and Yao, V. W. (2009). Spillover effects of foreclosures on neighborhood property values." *Journal of Real Estate Finance and Economics, 38*, 387-407.

Mayer, C., and Pence, K. (2008). Subprime mortgages: What, where, and to whom? Working Paper No. 14083, NBER.

Mayer, C., Pence, K., and Sherlund, S. M. (2009). The rise in mortgage defaults. *Journal of Economic Perspectives, 23*(1), 27-50.

Pederson, C. and Delgadillo, L. (2007). Residential mortgage default in low and high minority census tracts. *Family and Consumer Sciences Research Journal, 35*, 374-391.

Pennington-Cross, A. (2002). Subprime lending in the primary and secondary markets. *Journal of Housing Research, 13*(1), 31-50.

Pennington-Cross, A. and Ho, G. (2006). The termination of subprime hybrid and fixed rate mortgages. Federal Reserve Bank of St. Louis, Working Paper 2006-042A.

Pennington-Cross, A. and Ho, G. (2008). Predatory lending laws and the cost of credit. *Real Estate Economics, 36* (2), 175-211.

Quercia, R. G. and Stegman, M. A. (1992). Residential mortgage default: A review of the literature. *Journal of Housing Research, 3* (2), 341-379.

Quercia, Stegman, & Davis. (2007). The impact of predatory loan terms on subprime foreclosures: The special case of prepayment penalties and balloon payments. *Housing Policy Debate, 18* (2), 311-346.

Quigley, J. M. (1996). Mortgage performance and housing market discrimination. *Cityscape: A Journal of Policy Development and Research, 2* (1). Washington, D.C.: U.S. Department of Housing and Urban Development.

Reid, C. and Laderman, E. (2008). Lending in cow- and moderate-income neighborhoods in California: The performance of CRA lending during the subprime meltdown. Federal Reserve Working Paper presented at the Federal Reserve System Conference on Housing and Mortgage Markets, Washington DC.

Rogers, W. H. and Winter, W. (2009). The impact of foreclosures on neighboring housing sales.

Sherlund, S. M. (2008). The past, present, and future of subprime mortgages. Working Paper No 2008-63, Finance and Economics Discussion Series, Board of Governors of the Federal Reserve System (U.S.).

Schuetz, J., Been, V., and Ellen, I. G. (2008). Neighborhood effects of concentrated mortgage foreclosures. *Journal of Housing Economics, 17* (4), 306-319.

Schloemer, E., Li, W., Ernst, K., and Keest, K. (2006). Losing ground: Foreclosures in the subprime market and their cost to homeowners. The Center for Responsible Lending.

Van Order, R. and Zorn, P. (2000). Income, location, and default: Some implications for community lending. *Real Estate Economics, 28* (3), 385-404.

White, A. M. (2004). Risk-based mortgage pricing: Present and future research. Fannie Mae Foundation, *Housing Policy Debate, 15* (3).

**Table 1**
**Loan and borrower characteristics**

| | Fixed rate | 228 and 327 ARMs | Total |
|---|---|---|---|
| Number of loans | 35,151 | 34,328 | 69,479 |
| Percent of all loans | 50.6% | 49.4% | |
| Default rate | 0.149 | 0.250 | |
| | | | |
| ARM type | | | |
| 3/27 ARM | | 0.237 | |
| *2/28 ARM* | | *0.763* | |
| Margin | | | |
| *<5%* | | *0.156* | |
| 5-6% | | 0.300 | |
| 6-7% | | 0.474 | |
| 7%+ | | 0.070 | |
| Interest rate at origination | 6.922 | 7.441 | |
| Relative interest rate at origination | | | |
| <0 | *0.101* | | |
| 0-1% | 0.524 | *0.002* | |
| 1-2% | 0.260 | 0.037 | |
| 2-3% | 0.081 | 0.167 | |
| >3% | 0.034 | | |
| 3-4% | | 0.253 | |
| 4-5% | | 0.243 | |
| 5-6% | | 0.177 | |
| 6%+ | | 0.121 | |
| Original loan amount ($00,000) | 3.600 | 3.524 | |
| Payment shock at time of first ARM adjustment | | | |
| <20% | | 0.908 | |
| 20-30% | | 0.047 | |
| 30%+ | | 0.044 | |
| FICO score at origination | | | |
| Missing | 0.009 | 0.003 | |
| <530 | 0.026 | 0.097 | |
| 530-560 | 0.042 | 0.120 | |
| 560-590 | 0.065 | 0.126 | |
| 590-620 | 0.101 | 0.144 | |
| 620-650 | 0.161 | 0.169 | |
| 650-680 | 0.171 | 0.138 | |
| 680-720 | 0.198 | 0.117 | |
| *720+* | *0.228* | *0.085* | |
| Debt-to-income (DTI) at origination | | | |
| Missing | 0.396 | 0.194 | |
| *<45%* | *0.388* | *0.465* | |
| 45-50% | 0.143 | 0.255 | |
| 50%+ | 0.073 | 0.085 | |
| Combined LTV at origination | | | |
| *<60%* | *0.254* | *0.134* | |
| 60-70% | 0.183 | 0.128 | |
| 70-80% | 0.282 | 0.258 | |
| 80-90% | 0.187 | 0.247 | |
| 90-100% | 0.092 | 0.229 | |
| 100%+ | 0.002 | 0.003 | |
| Loan purpose | | | |
| Home purchase | 0.291 | 0.396 | |
| Cashout refinance | 0.641 | 0.572 | |
| *Non-cashout refinance* | *0.068* | *0.032* | |
| Owner type | | | |
| Owner-occupier | 0.875 | 0.905 | |
| Second home | 0.007 | 0.004 | |
| *Investor* | *0.119* | *0.091* | |
| Building type | | | |
| Single-family | 0.329 | 0.361 | |
| 2-4 family | 0.563 | 0.585 | |
| *5+family* | *0.018* | *0.013* | |
| *Condo* | *0.068* | *0.030* | |

*Notes:*
See text for more detailed variable descriptions. Observations are from LoanPerformance, 2004-2008.
For categorical variables, italics indicate the reference category (categories) for the hazard models.

## Table 2
## Neighborhood characteristics

| | Fixed rate | 2/28 and 3/27 ARMs |
|---|---|---|
| Number of loans | 35,151 | 34,328 |
| ***Distribution by neighborhood foreclosure concentration*** | | |
| Percent of units foreclosed in census tract[2] | | |
| *<1%* | *0.620* | *0.594* |
| 1-2% | 0.228 | 0.281 |
| 2-3% | 0.065 | 0.076 |
| 3-4% | 0.016 | 0.018 |
| 4%+ | 0.072 | 0.031 |
| ***Distribution by neighborhood demographics*** | | |
| Poverty rate in census tract | | |
| *<10%* | *0.250* | *0.220* |
| 10-20% | 0.387 | 0.384 |
| 20-30% | 0.206 | 0.216 |
| 30-40% | 0.101 | 0.114 |
| 40-50% | 0.047 | 0.056 |
| 50-100% | 0.009 | 0.010 |
| Percent white in census tract | | |
| *<20%* | *0.613* | *0.745* |
| 20-40% | 0.111 | 0.106 |
| 40-60% | 0.081 | 0.051 |
| 60-80% | 0.114 | 0.059 |
| 80-100% | 0.081 | 0.039 |
| Percent black in census tract | | |
| *<20%* | *0.441* | *0.303* |
| 20-40% | 0.096 | 0.105 |
| 40-60% | 0.081 | 0.103 |
| 60-80% | 0.125 | 0.152 |
| 80-100% | 0.257 | 0.337 |
| Percent Hispanic in census tract | | |
| *<20%* | *0.865* | *0.902* |
| 20-40% | 0.111 | 0.085 |
| 40-100% | 0.024 | 0.014 |
| Percent Asian in census tract | | |
| *<20%* | *0.871* | *0.906* |
| 20-40% | 0.105 | 0.081 |
| 40-100% | 0.023 | 0.013 |

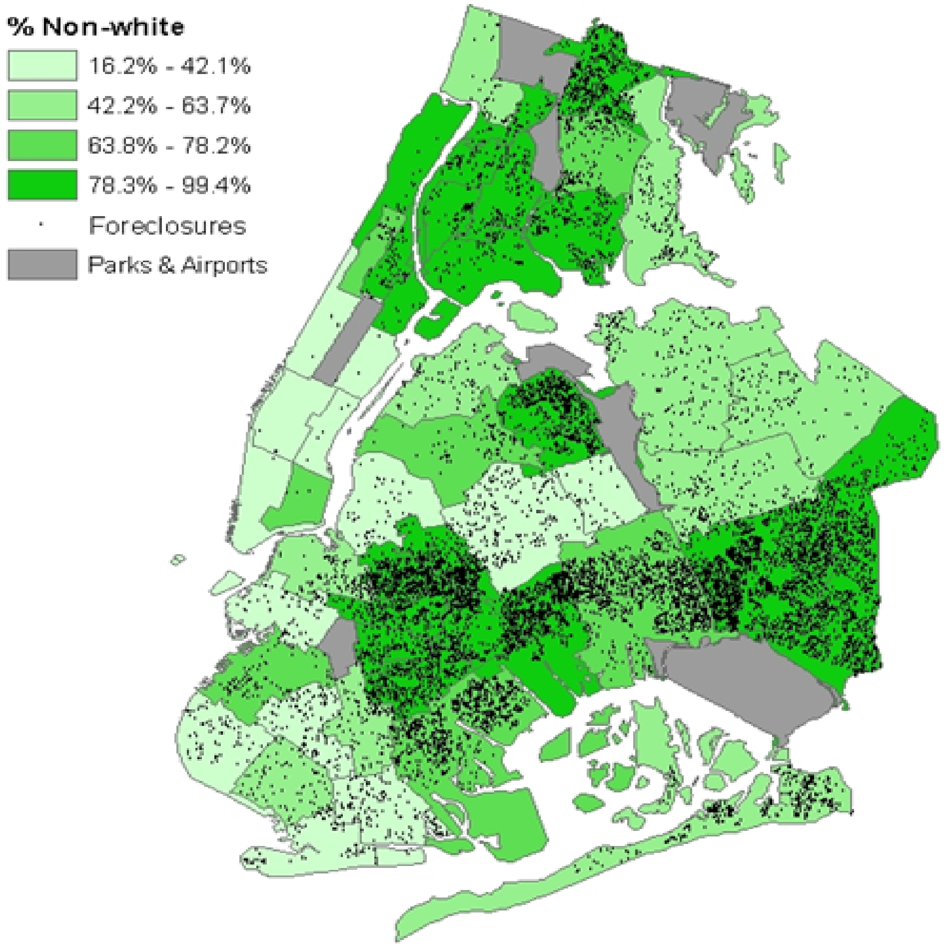
*Notes:*
See text for more detailed variable descriptions. Observations are from LoanPerformance, 2004-2008.
For categorical variables, italics indicate the reference category (categories) for the hazard models.

**Table 3**
**Hazard model results**

| | Fixed rate mortgages | | | 2/28 and 3/27 ARMS | | |
|---|---|---|---|---|---|---|
| | Hazard Ratio | | z | Hazard Ratio | | z |
| margin 5-6% | | | | 1.130 | ** | 3.08 |
| margin 6-7% | | | | 1.256 | ** | 5.92 |
| margin >7% | | | | 1.364 | ** | 5.7 |
| relative interest rate at origination 0-1% | 1.767 | ** | 6.32 | | | |
| relative interest rate at origination 1-2% | 3.353 | ** | 13.18 | 1.275 | | 1.03 |
| relative interest rate at origination 2-3% | 5.886 | ** | 18.2 | 1.701 | * | 2.29 |
| relative interest rate at origination >3% | 7.521 | ** | 18.78 | | | |
| relative interest rate at origination 3-4% | | | | 2.196 | ** | 3.39 |
| relative interest rate at origination 4-5% | | | | 2.668 | ** | 4.21 |
| relative interest rate at origination 5-6% | | | | 3.640 | ** | 5.5 |
| relative interest rate at origination >6% | | | | 6.154 | ** | 7.67 |
| original loan amount $00,000* | 1.428 | ** | 6.63 | 1.965 | ** | 8.24 |
| original loan amount squared | 0.974 | ** | -3.52 | 0.917 | ** | -5.52 |
| original loan amount cubed | 1.001 | * | 2.17 | 1.003 | ** | 3.62 |
| payment increase <20% X 3-4months post adjustment | | | | 1.260 | | 1.61 |
| payment increase <20% X 5-6months post adjustment | | | | 1.318 | | 1.92 |
| payment increase <20% X 7-12months post adjustment | | | | 1.611 | ** | 4.15 |
| payment increase <20% X 13-18 months post adjustment | | | | 1.520 | ** | 2.59 |
| payment increase <20% X 19+ months post adjustment | | | | 1.549 | * | 1.94 |
| payment increase 20-30% X 3-4months post adjustment | | | | 1.176 | | 0.91 |
| payment increase 20-30% X 5-6months post adjustment | | | | 1.580 | ** | 2.74 |
| payment increase 20-30% X 7-12months post adjustment | | | | 2.109 | ** | 6.08 |
| payment increase 20-30% X 13-18 months post adjustment | | | | 1.607 | ** | 2.78 |
| payment increase 20-30% X 19+ months post adjustment | | | | 1.769 | * | 2.31 |
| payment increase >30%% X 3-4months post adjustment | | | | 1.751 | ** | 3.06 |
| payment increase >30%% X 5-6months post adjustment | | | | 2.128 | ** | 4.26 |
| payment increase >30%% X 7-12months post adjustment | | | | 2.314 | ** | 6.32 |
| payment increase >30%% X 13-18 months post adjustment | | | | 1.681 | ** | 2.85 |
| payment increase >30%% X 19+ months post adjustment | | | | 2.413 | ** | 3.74 |
| 3/27 ARM | | | | 0.835 | ** | -5.95 |
| fico 680-720 | 1.873 | ** | 9.74 | 1.182 | ** | 3.33 |
| fico 650-680 | 2.435 | ** | 13.8 | 1.292 | ** | 5.28 |
| fico 620-650 | 3.325 | ** | 18.69 | 1.486 | ** | 8.4 |
| fico 590-620 | 3.787 | ** | 18.45 | 1.445 | ** | 7.12 |
| fico 560-590 | 4.896 | ** | 20.51 | 1.716 | ** | 9.72 |
| fico 530-560 | 5.330 | ** | 19.44 | 1.775 | ** | 9.74 |
| fico <530 | 5.528 | ** | 17.36 | 1.794 | ** | 9.12 |
| fico missing | 1.270 | | 0.91 | 1.732 | ** | 2.56 |
| debt-to-income missing | 1.015 | | 0.39 | 1.071 | * | 2.19 |
| debt-to-income 45-50% | 0.974 | | -0.59 | 1.146 | ** | 5.2 |
| debt-to-income 50% + | 1.158 | ** | 2.69 | 1.077 | | 1.84 |
| current combined LTV 60-70% | 1.225 | ** | 3.97 | 1.102 | * | 2.15 |
| current combined LTV 70-80% | 1.463 | ** | 7.01 | 1.298 | ** | 5.7 |
| current combined LTV 80-90% | 1.926 | ** | 10.63 | 1.541 | ** | 8.57 |
| current combined LTV 90-95% | 2.979 | ** | 13.69 | 1.981 | ** | 11.45 |
| current combined LTV 95-100% | 3.622 | ** | 14.45 | 2.548 | ** | 14.2 |
| current combined LTV 100%+ | 3.674 | ** | 13.92 | 2.040 | ** | 9.36 |
| home purchase | 1.176 | * | 2.17 | 1.440 | ** | 5.04 |
| cashout refinance | 0.982 | | -0.26 | 0.960 | | -0.58 |
| owner-occupier | 1.027 | | 0.52 | 1.358 | ** | 7.28 |
| second home | 0.925 | | -0.28 | 0.883 | | -0.57 |
| single family residence | 1.200 | ** | 2.77 | 1.158 | ** | 2.63 |
| 2-4 family residence | 1.068 | | 1.08 | 1.035 | | 0.63 |
| census tract foreclosure rate 1-2% | 1.137 | ** | 2.88 | 1.165 | ** | 5.19 |
| census tract foreclosure rate 2-3% | 1.296 | ** | 4.84 | 1.264 | ** | 6.22 |
| census tract foreclosure rate 3-4% | 1.157 | * | 2.01 | 1.344 | ** | 5.88 |
| census tract foreclosure rate >4% | 1.457 | ** | 4.89 | 1.306 | ** | 4.57 |
| census tract poverty 10-20% | 1.028 | | 0.65 | 1.011 | | 0.34 |
| census tract poverty 20-30% | 1.163 | ** | 3.05 | 1.193 | ** | 5.03 |
| census tract poverty 30-40% | 1.368 | ** | 5.17 | 1.271 | ** | 5.6 |
| census tract poverty 40-50% | 1.224 | ** | 2.61 | 1.449 | ** | 7.18 |
| census tract poverty 50%+ | 1.443 | ** | 2.6 | 1.286 | * | 2.44 |
| census tract black 20-40% | 1.140 | * | 2.19 | 1.060 | | 1.35 |
| census tract black 40-60% | 1.270 | ** | 3.82 | 1.191 | ** | 3.98 |
| census tract black 60-80% | 1.227 | ** | 3.71 | 1.176 | ** | 4.11 |
| census tract black 80%+ | 1.302 | ** | 5.45 | 1.319 | ** | 8.04 |
| census tract hispanic 20-40% | 0.666 | | -1.45 | 1.006 | | 0.04 |
| census tract hispanic 40%+ | 0.743 | | -0.46 | 0.687 | | -0.71 |
| census tract asian 20-40% | 1.433 | | 1.27 | 0.909 | | -0.55 |
| census tract asian 40%+ | 1.212 | | 0.29 | 1.324 | | 0.52 |
| Number of subjects | 35,151 | | | 34,328 | | |
| Number of observations | 925,116 | | | 521,686 | | |

*Notes:*
Cox proportional hazard models. The dependent variable is the hazard of default (90 days delinquent)
The model also includes quarterly calendar time fixed effects
* represents 5 percent significance level, ** represents 1 percent significance level

**Figure 1**
**Notices of foreclosure[1] (2008) and neighborhood racial composition[2] (2007) by New York City community district**




*Notes:*

[1] Each dot represents one notice of foreclosure (lis pendens). The sample is restricted to foreclosures occurring in 2008. These data were obtained from Public Data Corporation.

[2] Neighborhood racial composition is from the 2007 American Community Survey.