Social Dimensions of Subprime Mortgage Default [1]

Robert A. Connolly,[a] Lynn M. Fisher,[b] and Gary Painter[c]

[a] Kenan-Flagler Business School

University of North Carolina at Chapel Hill

[b] Mortgage Bankers Association

[c]Sol Price School of Public Policy

University of Southern California

This version: June 1, 2015

[1]Comments are most welcome. We gratefully acknowledge Tim Robinson for insights into these issues developed during his dissertation work, CLARC/CoreLogic for providing much of the data used in this paper, our discussant at the 2015 AREUEA meetings, Alvin Murphy, as well as Tom Davidoff, Andra Ghent, and other session participants. Please address comments to Bob Connolly (email: Robert_Connolly@unc.edu; phone: (919) 962-0053), Lynn Fisher (e-mail: Lynn_Fisher@mba.org; phone: (202) 557-2821), or Gary Painter (email: gpainter@usc.edu; phone: (213) 740-8754).

Social Dimensions of Subprime Mortgage Default

Abstract

The principal purpose of this study is to focus more carefully on the impact of social, or neighborhood, effects on mortgage decision making than has been done previously. Only a few studies pay direct attention to neighborhoods and none that we are aware of has asked whether neighborhood race and ethnicity and immigration are related to the decision to default *when the borrower experiences negative equity.* Our basic approach is to use negative equity to isolate conditions under which social context is more important for borrower decision making. Using a rich data set containing subprime loans from over 9,000 Census tracts in six gateway cities, we find evidence in favor of three mechanisms. First, interactions of tract-level ethnic composition with dynamic loan to value ratios reveal that information about neighborhood homogeneity may substitute (in part) for borrower-level information about wealth or liquidity constraints, as opposed to indicating a propensity to strategically default. Second, our index of cumulative local foreclosures is positively related to default consistent with prior studies of contagion. Finally, we find that loans in tracts that have greater proportions of recent immigrants are both less likely to default at low loan to value ratios and more likely to default at high loan to value ratios relative to communities with fewer recent immigrants. We speculate that the tendency to default when loan to value ratios are high may be related to lower default costs.

JEL Classification: G21

Keywords: mortgage, subprime, default, prepayment, gateway cities, race, ethnicity, immigrant status

1. Introduction

Because the design of mortgages and mortgage-related public policies depend critically on our understanding borrower behavior, researchers have recently focused on defining and quantifying the extent of strategic defaults. Nonetheless, while strategic defaults occur under certain conditions (Mayer, et al. (2014)), their occurrence is not that frequent (Bradley, et al. (2013)). Bhutta, et al. (2011) demonstrate that the median borrower in their study does not default until negative equity – the difference between the face value of a loan and house value when the former exceeds the latter – exceeds 60 percent of house value.

This is consistent with the findings in the extant literature on default that borrowers do not "ruthlessly" default in the manner prescribed by option theory (Vandell, (1995)), and for twenty years researchers have attempted to identify more complete explanations of heterogeneity in mortgage outcomes. Beyond option-related determinants, they have proposed variation in borrower default costs, liquidity and employment. Recent examples include Gerardi, et al. (2007), Elul, et al. (2010), Gerardi, et al. (2013), and Gyourko and Tracy (2014).

The principal purpose of this study is to focus more carefully on the impact of social, or neighborhood, effects on mortgage decision making. Like results from studies of homeownership, researchers have found gaps or differences in mortgage outcomes that are affiliated with borrower race and ethnicity but otherwise unexplained by observable factors (for example, Berkovec, et al. (1994) and more recently Chan, et al (2013) and Bayer, et al. (2013)). Blacks and Hispanics, in particular, have been estimated to default at higher rates than non-Hispanic whites, although this is not always the case. Several studies have found that borrower race and ethnicity is rendered less important than neighborhood demographics on the rate of default when the neighborhood factors are included (see Cotterman (2001), Firestone, et al. (2007), and Chan, et al. (2013)).

While local demographic characteristics are frequently used as controls in empirical studies of default, the extant literature has failed to offer satisfactory explanations of neighborhood effects. Few studies pay direct attention to neighborhoods and none that we are aware of has asked whether neighborhood race and ethnicity and immigration are related to the decision to default *when the borrower experiences negative equity.* This is the principal purpose of our study. The basic approach is to use negative equity to isolate conditions under which social context is more important for borrower decision making.

Social context may influence financial decision making through several channels. First, corporate finance studies have suggested that religiosity, or specific strains of religion, may influence social norms and attitudes towards risk and debt. Likewise, we think that other social institutions may influence the acceptability of certain actions. Second, social networks may be important for mortgage default decision making if households are unfamiliar with the process or the potential costs of default and foreclosure. Bhutta, et al. (2011) convincingly argue that pecuniary costs of foreclosure for the borrower are limited. However, this information may not have been readily understood by households at the outset of the housing crisis (see Seiler, et al. (2013)). Our (speculative) hypothesis is that dense social networks among different groups may be related to mortgage decision-making because they more effectively disseminate information about the consequences of foreclosure, increasing the rate of default. A competing and potentially off-setting hypothesis is that places with denser social networks have commensurately higher non-pecuniary default costs because of borrower attachment to the community.

To explore the social context of mortgage-related decisions, we assemble a rich data set containing loans from over 9,000 Census tracts in six gateway cities to study how decision making of subprime borrowers may be related to variation in tract-level measures of social composition. In particular, we investigate race and ethnicity (percent of population that is white, Asian, black or Hispanic) and recent immigration (percent of immigrants who have entered the U.S. since 2000). We also introduce entropy measures of neighborhood homogeneity borrowed from the literature on segregation; they are particularly useful in summarizing information about multiple racial or ethnic groups.

Thus, we utilize the geographic proximity of similar households to identify social groups.

Given this definition, there are at least two alternative hypotheses about why neighborhoods matter for the measurement of loan performance. The first is that the use of neighborhood characteristics proxy for omitted variables regarding borrowers, loans and housing markets. For example, the geographic concentration of subprime mortgage originations in low income and minority neighborhoods is well-documented (see Calem, et al. (2004), Mayer and Pence (2008), Mian and Sufi (2009), and Gerardi and Willen (2009)). Second, there is recent empirical evidence in support of foreclosure contagion (see Towe and Lawley (2013), Goodstein, et al. (2011), Guiso, et al. (2013), Bradley, et al. (forthcoming), and Munroe and Wilse-Samson (2013)). Our data and methodology, described below, allow us to control substantially for these issues, and we directly control for the cumulative subprime foreclosure experience of neighborhoods.

We focus our empirical analysis on subprime loans originated to owner-occupiers between 2004 and 2006. We obtain data from CoreLogic on six gateway cities with a great deal of both inter- and intra-city variation in neighborhood composition. In particular, we observe loan performance for more than 650,000 first-lien, subprime mortgages through the end of 2009 in Chicago, Los Angeles, Miami, New York City, San Diego and San Francisco. Our dataset (described in Section 5) allows us to capture variations in cumulative loan to value ratios over time and in places that vary, sometimes dramatically, in neighborhood composition.

An innovation in our modeling approach is that we do not constrain the impact of social factors to be identical across these diverse decision making settings. Rather, we focus on variation in decision making when the option to default is in the money. In this regard, our specification is a departure from common empirical approaches to modeling neighborhood race and ethnicity and nativity. Typically, neighborhood characteristics enter as a time invariant shifter of the hazard function, implying that a particular group is more or less likely to default in all situations. Our approach allows us to differentiate social impacts at high versus low loan to value ratios, and our empirical estimates show this variety exists.

With respect to the role of neighborhood effects on loan performance, we find evidence in favor of three mechanisms. First, interactions of tract-level ethnic composition with dynamic loan to value ratios reveal that information about neighborhood homogeneity may substitute (in part) for borrower-level information about wealth or liquidity constraints. In particular, loans in black and Hispanic tracts are more likely to default at low loan to value ratios when the strategic default option is not in the money. Several recent studies have identified the important role of unemployment and liquidity constraints on mortgage outcomes (see Elul, et al. (2010), and Gerardi, et al. (2013)). Complementing this finding is the fact that loans in minority tracts are less likely to default at high loan to value ratios relative to loans in predominantly white tracts. In other words, despite findings in the literature that black and Hispanic borrowers are more likely to default, we find evidence that in minority tracts such defaults are not strategic in nature.

Second, our index of cumulative local foreclosures is positively related to default consistent with prior studies of contagion. Despite the inclusion of this control, however, we still find an economically-substantial residual role for other neighborhood characteristics.

Finally, we find that loans in tracts that have greater proportions of recent immigrants are both less likely to default at low loan to value ratios and more likely to default at high loan to value ratios relative to communities with fewer recent immigrants. This result may reflect omitted variables if recent immigrants have less wealth and are more likely to become unemployed. However, loans in high immigrant tracts do not display the same behavior at low loan to value ratios as loans in minority tracts. We speculate that the tendency to default when loan to value ratios are high may be related to lower default costs.

Our findings suggest prior studies on individual borrower race should not be used to infer the impact of social or cultural norms. Our results also stand in contrast to the conclusions of survey evidence about racial attitudes towards strategic default in Guiso, et al. (2013). If knowledge about attitudes towards homeownership and borrower obligations can be better understood and modeled, then attempts to intervene in housing and mortgages markets may be better tailored to local conditions. That is, there may be a different socially-optimal supply of foreclosures depending on market demand characteristics.

In the next section, we review important parts of the literature relevant to the questions that we study here, and we introduce measures of social homogeneity in Section 3. Any effort to understand how social factors might affect mortgage default decisions must account for a number of empirical challenges, and we describe our approach in Section 4. We describe our data set and measurement methods in Section 5. The first part of Section 6 presents our workhorse models and discusses the results of these multinomial logit regressions; the second part of Section 6 demonstrates that our principal findings are robust to alternative modeling strategies. We conclude our paper in Section 7.

2. Relevant Literature

The issues that we address in this paper focus on the role of neighborhood characteristics in subprime mortgage-related decisions and loan performance. A review of the literature suggests that the overriding reason that neighborhood race and ethnicity and income measures have been included in models of default and prepayment is to proxy for otherwise unobserved attributes of borrowers, loans and local housing markets. The main neighborhood features studied in the extant literature are minority status, income, and more recently, foreclosure rates. Only a few studies set out to explore the role of neighborhoods outside of recent studies of foreclosure contagion. Most often, studies have used neighborhood characteristics as unspecified "controls" without elaborating on the channel connecting these characteristics and mortgage-related decisions.

Some studies are more deliberate, however. For example, neighborhood race and ethnicity is specifically used as a control for unobserved lender origination behavior and loan characteristics in some studies (see Berkovec, et al. (1994), Cotterman (2001), Foote, et al. (2008), and Chan, et al. (2013)). Origination behavior has been a particular concern in the recent literature since the geographic concentration of subprime mortgage originations in low income and minority neighborhoods has been well-documented (see, for example, Calem, et al. (2004), Mayer and Pence (2008), Mian and Sufi (2009), and Gerardi and Willen (2009)). Reid and

Laderman (2009) show that the channel through which minority borrowers in California obtain mortgages influences the price of the loan. Nonetheless, conditional on obtaining a subprime loan and controlling for origination effects, the authors find that borrowers in black, Hispanic and Asian neighborhoods are still more likely to default than borrowers in predominately white, non-Hispanic communities. With respect to unobserved loan characteristics, studies with both borrower and neighborhood information on race and ethnicity have shown that inclusion of greater loan detail diminishes the importance of borrower race and ethnicity for explaining loan performance, while a residual role for neighborhood race and ethnicity often remains (see Goldberg and Harding (2003), Firestone, et al. (2007), Cotterman (2001), and Chan, et al. (2013)). Thus, a neighborhood effect appears to persist in the literature despite improved specifications of clustered originations and loan characteristics over time.

Recent studies have also focused on contagion effects resulting from nearby foreclosures. Guiso, et al. (2013) find that a one standard deviation increase in the percentage of foreclosed properties in a respondent's ZIP code increases the stated willingness of a respondent to strategically default by twenty three percent. Towe and Lawley (2013) find evidence of increased default hazards based on the foreclosure experience of nearest neighbors, and Goodman, et al. (2011) and Bradley et al. (forthcoming) find that strategic default is sensitive to the presence of other nearby strategic defaulters. While Bradley et al. (forthcoming) and Cotterman (2001) speculate that foreclosures diminish the desirability of neighborhoods inducing subsequent default, Munroe and Wilse-Samson (2013) show that neighbors learn about court and lender behavior from nearby foreclosures.

In summary, the literature on mortgage default appears to find a residual role for neighborhood composition and foreclosures despite a wide range of individual borrower, loan and economic controls. Beyond foreclosure externalities, our focus in this paper is on whether other externalities related to social networks impact mortgage performance. We think that one clue is found within the literature on race and ethnicity and homeownership. For example, in a series of studies, Painter and co-authors (2001, 2003, 2008) describe heterogeneity in mobility and endowments (including income, education and immigrant status) among various groups and relate this variation to differences in homeownership rates. We speculate that each of these factors may be related to behavioral differences with respect to mortgage-related decisions, too.

While this evidence pertains to omitted borrower characteristics, other work suggests that ethnic ties may play a role in helping immigrants to "catch-up" to native born homeownership attainment over time, since the gap in ownership falls with length of residency. At least for some ethnic groups, living in areas with greater concentration of certain ethnicities raises the likelihood of homeownership. In Painter, et al. (2004), not only do English language skills matter for homeownership attainment, but speaking multiple languages in the home is also positively associated with homeownership in larger cities. Speaking multiple languages may increase an individual's access to both mainstream and group-related resources, including finance. In addition, greater connections to a particular ethnic group may result in "peer" effects influencing preferences for homeownership, and, we speculate, default.

3. Neighborhood Homogeneity

It is common for researchers to measure social network dimensions with a set of measures of the percent of a group in a jurisdiction, e.g., the percent of black residents in a census tract. When more than two groups are being considered, the social composition of neighborhoods may not be well described by group percentages. In addition, we wish to investigate the social networking hypothesis by describing the social homogeneity (or diversity) of an area. This approach distinguishes the hypotheses from the attitudes or beliefs of a particular group, but rather focuses on the homogeneity of any group as facilitating intra-group communication.

In this section we briefly describe a measure of diversity borrowed from the literature on segregation. In particular, we utilize a measure developed by Thiel and Finizza (1971) based on information theory. Let q_{jk} be the proportion of group k in tract j relative to tract total population. Given K groups, $\sum_{k=1}^{K} q_{jk} = 1$. A tract-level measure of diversity based on the entropy of a distribution is

$$E_j = \sum_{k=1}^{K} q_{jk} \log_K \left(\frac{1}{q_{jk}} \right). \tag{1}$$

If we allow the log base to be equal to the number of groups K, then this diversity measure falls between 0 and 1.

Because we want to compare this index to the percent of minority population in Census tracts, we re-frame the measure in terms of homogeneity for ease of exposition. Specifically, we will focus on an index of tract ethnic homogeneity defined as

$$H_j = 1 - E_j. \tag{2}$$

The maximum value of one indicates a perfectly homogenous population (of one ethnic group). The index H_j takes a value of zero under conditions of perfect diversity, with each group comprising $\frac{1}{K}$ of the tract's population.

The other characteristic of interest is the percent of a tract that has recently immigrated. The seventy-fifth percentile of the distribution of recent immigrant population across tracts in our data is 9.5 percent. Because no tract has 100 percent recent immigrants, more immigrants make the tract more diverse. Therefore, we utilize the basic entropy measure for our recent immigrant index and define it as

$$I_j = E_j. \tag{3}$$

We do this so that a larger value of the index coincides with more recent immigrants in a particular tract, thus making the interpretation of our results consistent between indexes.

For comparison across gateway cities we also define a city level measure of diversity. Let y_j be the population of tract j and Y the population of the city. The (weighted) average entropy of tract diversity in the city is:

$$\overline{E} = \sum_{j=1}^{N} \frac{y_j}{Y} \sum_{k=1}^{K} q_{jk} \log_K \left(\frac{1}{q_{jk}} \right). \tag{4}$$

4. Empirical Modeling Strategy

We derive our empirical strategy from a random utility set-up (see, for example, Marquez (1997)). A borrower in each time period chooses one of three alternatives: remain current, default or prepay. Let the utility from alternative $a = 0, 1, 2$ with respect to loan i in tract j at time t be

$$U_{ijt}^{a} = V_{ijt}^{a} + \varepsilon_{ijt}^{a}. \tag{5}$$

Alternative c is selected if

$$U_{ijt}^{c} \geq U_{ijt}^{b} \text{ for } b \neq c. \tag{6}$$

If we specify the error term ε_{ijt}^{a} as having an extreme value type I distribution, we can estimate the model as a multinomial logit in which the probability of observing outcome c is

$$\Pr\left(c_{ijt}\right) = \frac{\exp\left(V_{ijt}^{c}\right)}{\sum_{a=0}^{2} \exp\left(V_{ijt}^{a}\right)}. \tag{7}$$

To allow for heterogeneity at the loan level, we also consider a (parametric) random intercept. We specify the linear predictor of utility as

$$V_{it}^{a} = g^{a\prime} X_{ijt} + \gamma_i^{a} \tag{8}$$

where X_{it} is a vector of explanatory variables specific to loan i at time t, g^{a} is an alternative-specific vector of coefficients to be estimated, and γ_i^{a} is the loan level random effects for each alternative. In the interest of parsimony we estimate $\gamma_i^{2} = \alpha\gamma_i^{1}$ so that we only need to estimate one random effect despite having two alternatives for which we estimate coefficients (note that by making "current" our reference category, $\gamma_i^{0} = 0$).

5. Data and Summary Statistics

In this study we use data on non-agency, securitized subprime mortgages from CoreLogic.[1] Their loan-level data covers 97% of active, non-agency securitized mortgages, of which subprime loans are a subset. For first lien mortgages on single family houses, townhouses and condominiums that are owner-occupied, we obtain quarterly performance data on loans originated from 2004 to 2006 in six gateway cities: Chicago, Los Angeles, Miami, New York, San Diego and San Francisco.[2] We follow performance on a quarterly basis through 2009. As we noted earlier, our empirical design is structured to exploit differences across and within cities that uncover neighborhood race and ethnicity and immigration status effects on default and prepayment choice.

The typical loan information is supplemented by CoreLogic in two ways. First, CoreLogic performed a title search for additional liens that were open or active six months after first lien origination. From this data, we observe the amount of any additional liens and add these amounts to the first lien in order to obtain a cumulative loan to value ratio at origination. In addition, CoreLogic provides us with updated home values at the end of each quarter for the life of the loan. House price changes are calculated by CoreLogic based on the best combination of CoreLogic's zip code, county, Core-Based Statistical Area (city), and state house price indices. Therefore, we update cumulative loan to value ratios based on updated first loan balances and updated home prices. Junior loan balances are assumed to change at the same rate as the first lien over time. To this data, we are able to match county and Census tract characteristics (these are described shortly).

We define a default as occurring in the first quarter in which the borrower is 90 days delinquent, conditional on being at least 90 days delinquent in the following quarter, similar to Bhutta, et al. (2011).[3] We define default in this manner so as to focus on borrower decision-

[1] Specifically, we include all loans in BC securities and exclude loans in Alt A and jumbo securities.

[2] Painter and his co-authors have shown that there are important differences in housing choice in cities that are gateways into the United States for immigrants.

[3] It is worth noting that there is considerable variety in how studies have defined default on a mortgage. Our

making independent of lender choices about whether to renegotiate or foreclose on the loan.[4] A loan is prepaid in the quarter in which its status is indicated as paid-off (with a loan balance of zero). The loan leaves the panel data set in the quarter following either default or prepayment.

5.1. Loan and Property Characteristics

Data from the loan origination includes both the type and purpose of loan, property type (condo, single family house, townhouse), original loan balance, original loan-to-value ratio, loan term, borrower FICO score, degree of documentation, scheduled rate resets and the existence and length of prepayment penalties. Among dynamic characteristics of loans, we observe the current loan balance and interest rate in addition to the payment status: whether current, 30, 60 or 90 days delinquent, in foreclosure, REO or paid off.

To use the model specification described below, we undertake additional recoding of variables. The *current* cumulative loan to house value ratio (CLTV) varies over time according to changes in the first mortgage loan balance and estimated house price. We enter this as a categorical variable. Specifically, we re-label the categories in our results tables with *current* CLTV labels for ease of interpretation. The categories are for less than 80 percent CLTV (omitted category), $80-90$ percent, $90-100$ percent, $100-110$ percent, $110-120$, and greater than 120 percent CLTV.[5]

Among the potential fixed loan characteristics that we control for in our modeling, we include the initial loan balance as a control for size of the loan and the borrower FICO score entered as categories of greater than 725 (omitted), $621-725$, and less than 621. Loan purpose is also categorical with the omitted category of a purchase loan, and the other two categories reflecting refinancing with no cash out and cash out, respectively. We control for whether or not the loan is a fixed rate mortgage (FRM), interest-only (IO) or has a balloon payment. The type

approach is fairly conservative relative to the range that exists in the literature.

[4]In so doing, we are attempting to minimize the impact on our modeling of variations in the supply of foreclosures across jurisdictions. On this specific topic, see recent work by Favara and Gianetti (2014).

[5]In some estimation applications, convergence difficulties led us to collapse the categories slightly into $100-115$ percent and greater than 115 percent CLTV.

of documentation used for underwriting is coded as full (omitted category), low and none (no doc). A dummy is included for whether or not the loan has a 30 year term.

Besides *current* CLTV, other loan level controls may also vary over time. In particular, we code a dummy variable for the current quarter in which a prepayment penalty is in effect. We similarly dummy for whether an interest rate reset occurred in the prior quarter. An interest rate spread is calculated as the difference between the end of quarter loan interest rate and the ten-year Treasury rate.

5.2. Tract and Location Characteristics

To control for other dimensions of the borrower's choice, we also incorporate a number of location characteristics into our analysis. First, we incorporate the three-month change in unemployment rate, lagged one quarter, at the county level; the underlying data is from the Bureau of Labor Statistics, Local Area Unemployment Statistics. We also match into our loan data a set of tract level data items from the American Community Survey (ACS) averaged over the 2005-2009 period. In all specifications, we control for the natural logarithm of tract median household income. Dynamic controls include the lagged quarterly change in the county unemployment rate, a lagged quarterly change in tract level house prices, and an index of cumulative foreclosure by Census tract. For the latter measure, we calculate an index of the cumulative number of foreclosures of all subprime mortgages by tract from the first quarter of 2004 until the prior quarter as $ln(1 + cumforeclosures_{t-1})$.

5.3. Summary Statistics

Our sample is comprised of over 650,000 subprime loans originated between 2004 and 2006. Recalling that these are subprime originations made at the height of the housing cycle, 28 percent of these loans ultimately defaulted before the end of 2009 and 58 percent prepaid (see Table 1). The total rate of default ranges from 23 percent to 36 percent across cities, whereas prepayments range between 50 and 61 percent (also in Table 1). The prepayment rates are an important reminder of why we typically model mortgage performance in a competing risks framework.

As the data in Table 1 show, loans in our sample are predominantly taken out for the purpose of refinancing and to withdraw equity (59 percent of loans). Loans are mostly adjustable rate mortgages (72 percent) with a sizeable fraction described as interest-only ARMs, especially in the three California cities in our sample. While the majority of loans had low documentation at origination, so-called 'no doc' loans represent less than one percent of all loans. Most loans in the overall sample have prepayment penalties, although there is sizeable variation in the existence of prepayment penalties across cities.

Table 2 reports additional statistics about house value, loan amount, interest rate and borrower FICO scores at origination. We note that there is considerable variation in home values as well as value-to-loan ratios at origination. In this respect, our sample is not especially concentrated around particular house value, CLTV, or other category values.

Table 3 provides a rudimentary statistical comparison between the total number of Census tracts in the cities investigated and the tracts for which we observe loans in our sample. Sample loans represent 74 percent of all possible tracts in the six cities. Unsurprisingly, since our sample represents homeownership loans, our sample tracts have slightly higher homeownership rates and income as compared to all possible tracts in these cities. With respect to tract population percentages by race and ethnicity and recent immigration (population immigrated since 2000), the mean traits of our sample tracts are quite similar to the means for all tracts.

We also introduce two alternative social measures. The first is a categorical variable that equals 1 if a particular ethnic group represents more than 33 percent of tract population, or recent immigrant population comprises more than 10 percent.[6] This choice reflects a set of tradeoffs in the data. When we set the threshold below twenty five percent of the population for any ethnic group, the confounding fact is that Census tracts begin to be classified as "high" for multiple groups. On the other hand, using a cut off of 50 percent results in qualifying tracts

[6]For both raw percentages and the categorical variables we use ACS averages from 2005-2009 for population of ages 25-64.

that are mainly Hispanic. If we compromise at 33 percent, tracts with multiple "high" qualifiers are mainly those that are high in a racial category and also have more than 10 percent recent foreign immigrants (this is threshold for the 75th percentile of all tracts with respect to recent immigrants), for which we separately control. With these considerations in mind, we therefore present results using 33 percent thresholds for Asian, black and Hispanic. Throughout, we use 10 percent as a threshold for share of recent foreign immigrants. In Table 3, we show that 63 percent of tracts qualify as "High Hispanic." We also note that our sample tracts once again reflect well the population of all Census tracts.

Our second set of alternative measures are the homogeneity and recent immigrant measures introduced in Section 3. The unweighted mean ethnic homogeneity index of tracts in our sample is 0.44, and it is 0.32 for the recent immigrant index. Figures 1 and 2 depict the distribution of these measures by Census tract for each gateway city. There is considerable variation among tracts in each city. The weighted average entropy measure by city (whether using all tracts, or separately, sample tracts) ranks Chicago as the most homogenous city and San Francisco as the most diverse.[7] On the other hand, Los Angeles has the highest weighted average index for recent immigrant concentration, followed by San Francisco and New York City.[8]

Finally, in Table 4, we present tract characteristics organized by city. The typical census tract has a population of 4,000 to 6,000 people. As expected, there is considerable variation in homeownership rates, median household income and the ethnic composition. We note the high variability in the percent Asian and percent black. This variation is crucial since it helps to identify the relation between social dimensions of option exercise costs and decisions to default or prepay.

Figures 3 and 4 show the sample default and prepayment hazard rates for each of our six gateway cities, split by the year in which loans were originated. We observe that default hazard rates are relatively modest for loans originated in 2004 versus loans originated in 2006. Likewise,

[7]The rank order from most diverse to most homogenous is: San Francisco, San Diego, Miami, Los Angeles, New York City, and Chicago.

[8]Miami, San Diego and Chicago round out the recent immigrant tract ranking (with Chicago tracts being the least concentrated).

the sample prepayment hazards decline with each cohort. For our empirical model to be judged reasonably successful, it should be able to match the shape of these hazard functions over time and among cities.[9]

6. Results

6.1. Main Models

We focus initially on estimates of three models where loans are pooled across cities but without random intercepts. In each model, time since loan inception is entered as a cubic function interacted with cohort year in order to accommodate the baseline sample hazards depicted in Figure 3. Fixed effects are also included for each county, the top ten originators, and for calendar years. Unless otherwise noted, we report model coefficients and robust standard errors clustered at the loan level. By virtue of our sample size, most coefficients are precisely estimated. Before discussing our main results in Table 5, we note that virtually all loan and economic variables enter with the expected signs in the remainder of the specification.

The first model, whose estimates are reported in Panel A of Table 5, is intended to replicate the typical use of Census tract data in studies of loan performance: it enters the tract population percentage of each ethnic group and for recent immigrants as separate covariates. The estimates in Panel A indicate that greater concentrations of Asian and Hispanic households are associated with lower hazards of loan default. Neither the percent black nor percent recent immigrant are statistically different from zero.

The second model, whose estimates are reported in Panel B of Table 5, relies on our new measures of social homogeneity. These entropy measures summarize information about multiple ethnic groups and create an information-theory based index of homogeneity. A greater value of H_j indicates greater ethnic homogeneity (recall the index is bounded between zero and one), and larger values of I_j indicate a greater presence of recent immigrants..

[9] Since the figures indicate the importance of the mortgage vintage, we experimented with alternatives before settling on a cubic function of time since mortgage inception interacted with mortgage cohort. We are thus able to reproduce the sample prepayment hazards fairly closely.

Results are non-linear in CLTV for the ethnic homogeneity index. At low CLTVs, loans in more homogenous neighborhoods are more likely to default. Loans are less likely to default at CLTVs just over one in more homogenous tracts, but are more likely to default at very high CLTVs relative to loans in more diverse tracts. Higher immigrant entropy index values are associated with a lower likelihood of loan default at low CLTVs and a greater likelihood of default at high CLTVs in Panel B of Table 5. The coefficient of 0.1494 on the homogeneity index at CLTVs greater than 120 percent appears to be driven by loans in mainly homogenous white tracts because loans in both black and Hispanic tracts are less likely to default at the highest CLTVs.

We report estimates of the loan default equation in which we interact tract population percentages with categories of CLTV in our multinomial logit model in Table 5, Panel C.[10] We report estimates of the additional covariates in Table 7[11]. The baseline estimate on CLTV is positive for the full sample. For the Asian neighborhood interaction terms, we find that there are only two significant estimates, and both indicate a lower probability of default in two CLTV categories (one of them is the highest CLTV category). By contrast, default probabilities vary with CLTV in black neighborhoods: higher at the lowest three CLTV categories, but lower in the highest two CLTV categories. With respect to Hispanic neighborhoods, our estimates show that there is a lower probability of default at CLTV values above one. To summarize, our estimates strongly imply that strategic default is not the norm in minority census tracts. To the contrary, these neighborhoods are characterized by lower default probabilities in exactly those setting where strategic default would be expected based on the value of the default option.

By contrast, in the case of immigrant neighborhood effects, the full sample estimates indicate

[10]To save space and to focus the discussion on subprime default, estimates of the prepayment model coefficients are not reported here.

[11]We do not report the coefficient estimates in the table, but we also include controls for the year in which a mortgage was originated, a cubic function that captures times since origination, interactions between the origination year and the cubic function, and a set of dummy variables for the top 12 originators. Broadly, we find what we would expect based on the earlier literature. The probability of default is increasing in the size of the original loan balance, the interest rate spread, the cumulative foreclosure experience of a tract, rising unemployment in a tract, and as the FICO score is lower. Refinancings are associated with lower default probabilities, as are rising incomes and rising prices in a tract. By contrast, low documentation and no documentation loans are more likely to default.

lower default probabilities at low CLTV but higher default probabilities once CLTV is above one. Unlike our results using racial or ethnic measures (i.e., a lower probability of default at higher CLTV), in heavily immigrant neighborhoods, we find evidence in favor of the strategic default hypothesis. Finally, we note that after controlling for all these factors, we also find that the probability of default is lower at high CLTV levels in locations where lenders have recourse, which is in keeping with the findings in Ghent and Kudylak (2011).

Table 6 reports city-level estimates of our multinomial logit model, and for comparison purposes, we provide estimates from the same model applied to the pooled data in the left-most column. While the baseline estimate on CLTV is positive for the full sample and for each individual city, the magnitudes vary quite substantially across the six cities in our sample. For example, the estimate is 0.206 for Chicago, but 1.938 for San Diego. The estimate for Miami is about 3.5 times the Chicago estimate, and it is the smallest after Chicago.

Note that when we account for variations in default probabilities as a function of CLTV, the full sample estimates rise monotonically across the CLTV categories, from 0.448 to 1.289. Estimates in Table 6 show that there is considerable heterogeneity in this vector of estimates across the six cities that make up our data. For New York City with the third-lowest CLTV baseline estimate, the estimates for the categories begin at 0.201 and rise to 0.505. For Miami which has the second-lowest CLTV baseline estimate, the estimates for the categories begin at 0.113 and rise to only 0.708. By contrast, the estimates for San Francisco begin at 0.367 and increase to 2.195. That is, the sensitivity of default probabilities to higher CLTV values clearly varies depending on location.

We also find that there is considerable variability across cities in the estimates related to racial and social heterogeneity, however, the city level estimates largely uphold our pooled results. There is no difference in default probabilities in Asian neighborhoods for Chicago and Miami at any level of CLTV. However, there is a statistically-significant negative effect for the California cities and New York City in the highest CLTV category, and, interestingly, in the 90 - 100% CLTV for New York City, Los Angeles, and San Diego.

As we showed in Table 5, Panel C for the pooled sample, default probabilities vary with CLTV in black neighborhoods: higher at the lowest three CLTV categories, but lower in the highest two CLTV categories. This pattern is evident in Chicago, San Francisco, and to a somewhat smaller degree in New York City. There is no difference in default probabilities in black neighborhoods in Miami, and only at the highest levels of CLTV in Los Angeles and San Diego. Clearly, samples drawn from Chicago will produce a very different picture of black default on subprime mortgages relative to Miami. The 'national' picture of black default on subprime mortgages appears to be largely driven by Chicago and San Francisco.

With respect to Hispanic neighborhoods, it is clear from the full sample that there is a lower probability of default at CLTV values above one. This is also evident in the data from San Diego and San Francisco, where the estimated coefficients are large relative to the full sample estimates. There is more limited evidence consistent with this from the Los Angeles and New York City samples, but there is no Hispanic neighborhood effect on default probabilities in Chicago or Miami. Much as with the Asian and black neighborhood estimates, the full sample estimate appear driven by several locations.

Turning to immigrant neighborhood effects, we find that Los Angeles is the only city that mimics fully the strategic default findings in the pooled sample results, i.e., lower default probabilities at low CLTV but higher default probabilities once CLTV exceeds one. Chicago, Miami, New York City, and San Diego are largely consistent, but the estimated effects are not always statistically significant. By contrast, the estimates from San Francisco are exactly the opposite: higher default probabilities for the two lowest CLTV categories and a negative default probability estimate (although not statistically significant) for the higher CLTV categories.

We also included a number of covariates in our multinomial model, and we report most of their estimated effects in Table 7.[12] For many of the covariates, the full sample estimates are largely mimicked in the individual city samples. For example, the default probability is increasing in the natural log of the initial loan balance and the interest rate spread in each

[12]A handful of covariates are not presented, but identified in the notes to Table 7.

sample. The default probability is decreasing in the change in tract-level prices in every sample, too. We note that cumulative foreclosures in a tract raise the probability of default. The default probability is lower for refinancings with or without cash out compared to purchase mortgages. The estimated default probability for subprime mortgages with low documentation is higher than for full documentation loans. Table 7 reports city-by-city estimates of the other covariates in our multinomial subprime mortgage default model. We largely confirm what we expect to see based on the pooled sample estimates.

6.2. Discussion

We have conjectured that the role of neighborhood effects on loan performance are either due to omitted variables regarding borrower, loan or housing market characteristics, due to the contagion of foreclosures, or result from social networks proxied by neighborhood homogeneity. We find evidence of all three mechanisms.

First, interactions of tract ethnic composition with CLTV reveal that information about neighborhood homogeneity may in part substitute for information about borrower wealth or liquidity constraints. In particular, loans in black and Hispanic tracts are more likely to default at low CLTVs when the strategic default option is not in the money. Several recent studies have identified the important role of unemployment and liquidity constraints on mortgage outcomes (see Elul, et al. (2010), Gerardi, et al. (2013)). Complementing this finding is the fact that loans in minority tracts are less likely to default at high CLTVs relative to loans in predominantly white tracts. In other words, despite findings in the literature that black and Hispanic borrowers are more likely to default, we find evidence that such defaults are not strategic in nature.

Second, our index of cumulative local foreclosures is positively related to the incidence of default consistent with prior studies of contagion (detailed results are in the Appendix). Despite the inclusion of this control, however, we still find a residual role for neighborhood characteristics.

Finally, we find that loans in tracts that have relatively greater proportions of recent immigrants are both less likely to default at low LTVs and more likely to default at high CLTVs relative to communities with fewer recent immigrants. The pooled results appear mainly driven by Los Angeles and surely seems to be an important question for additional research.

6.3. Robustness

To establish the robustness of our results, we also estimated a random effects model. The pooled results found in Panel C of Table 5 are largely unchanged. Results are available upon requests.

We also estimate an xtlogit model in which we concentrate on that portion of the sample observations with mortgage defaults, and asks what factors affect the timing of an observed default. We report estimates of this model in Table 8.

CLTV enters here in percentage form. Because the model displayed some difficulties with convergence, we rely on indicator variables for whether the tract has a race/ethnicity/recent immigrant population percentage in the highest quartile of tracts in the pooled sample. We find that the conditional probability of default rises with higher CLTV at the mortgage origination. Notably, loans in more highly black and Hispanic tracts are less likely to default at higher CLTVs, confirming what we found in earlier model estimates. The immigrant and Asian effects are both also positive, but not statistically significant.

7. Conclusions

While local demographic characteristics are frequently used as controls in empirical studies of default, in this paper we ask whether social context systematically influences the decision to default when a subprime borrower experiences negative equity. We focus on gateway cities due to their size, density and social heterogeneity, and because these locations will allow us to study economic versus social factors that explain systematic variation in default behavior among communities.

We find evidence in favor of three mechanisms. First, interactions of tract-level ethnic composition with dynamic loan to value ratios reveal that information about neighborhood homogeneity may substitute (in part) for borrower-level information about wealth or liquidity constraints. Second, our index of cumulative local foreclosures is positively related to default consistent with prior studies of contagion. Finally, we find that loans in tracts that have greater proportions of recent immigrants are both less likely to default at low loan to value ratios and more likely to default at high loan to value ratios relative to communities with fewer recent immigrants. We speculate that the tendency to default when loan to value ratios are high may be related to lower default costs.

Each of our results are robust to the specification of the baseline hazard, disaggregation of estimation sample by city and cohort, the inclusion of a variety of controls and the specification of random effects in sub-samples. Although the literature to date is quite mixed with respect to the impact of borrower race and ethnicity and neighborhood racial composition on the probability of loan default, we think that our research design and empirical approach are informative and suggest that subprime mortgages in minority neighborhoods were less likely to default at high CLTVs as compared to more diverse and more white neighborhoods.

Bibliography

Ambrose, Brent W.. and Capone, Charles. (1998). Modeling the Conditional Probability of Foreclosure in the Context of Single-Family Mortgage Default Resolutions. *Real Estate Economics*, 26(3), 391-429.

Anderson, R. and VanderHoff, J. (1999). Mortgage Default Rates and Borrower Race. *Journal of Real Estate Research*, 18(2), 279.

Bajari, P., Chu, C. and Park, M. (2008). An Empirical Model of Subprime Mortgage Default. NBER Working Paper 14625.

Bayer, Patrick, Fernando Ferreira and Stephen Ross. 2013. The Vulnerability of Minority Homeowners in the Housing Boom and Bust. ERID Working Paper Number 145.

Berkovec, J., 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: 263 - 294.

Bhutta, N., Dokko, J., Shan, H., (2011). Consumer Ruthlessness and Mortgage Default During the 2007-2009 Housing Bust. Working paper.

Bradley, Michael G., Amy Crews Cutts, and Wei Liu. (2014) Strategic Mortgage Default: The Effect of Neighborhood Factors. *Real Estate Economics*, (forthcoming).

Cahill, Meagan E. and Franklin, Rachel S. (2013). The Minority Homeownership Gap, Home Foreclosure, and Nativity: Evidence from Miami-Dade County. *Journal of Regional Science*, 53(1), 91-117.

Calem, P., Gillen, K., and Wachter, S., (2004). The Neighborhood Distribution of Subprime Mortgage Lending. *Journal of Real Estate Finance and Economics*, 29(4), 393-410.

Chan, S., Gedal, M., Been, V., and Haughwout, A, (2011). The Role of Neighborhood Characteristics in Mortgage Default Risk: Evidence from New York City. *Journal of Housing Economics* 22(2): 100-118.

Clapp, John M. Clapp, Deng, Yongheng, and An, Xudong (2006). Unobserved Heterogeneity in Models of Competing Mortgage Termination Risks. *Real Estate Economics*, 34(2), 243-273.

Cotterman, Robert F. 2001. Neighborhood effects in mortgage default risk. U.S. Department of Housing and Urban Development, Office of Policy Development and Research. Available at *http://www.huduser.org/Publications/PDF/default_full.pdf*.

Deng, Yongheng and Stuart Gabriel. 2006. Risk-Based pricing and the Enhancement of Mortgage Credit Availability among Underserved and Higher Credit-Risk Populations. *Journal of Money, Credit and Banking* 38(6): 1431-1460.

Deng, Yongheng, Quigley, John M., and Van Order, Robert (2000). Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options, *Econometrica* 68 (2), 275-307.

Elul, R., Souleles, N., Chomsisengphet, S., Glennon, D., and Hunt, R., (2010). What "Triggers" Mortgage Default? Federal Reserve Bank of Philadelphia, Working Paper No. 10-13.

Favara, Giovanni and Giannetti, Mariassunta, Mortgage Concentration, Foreclosures and House Prices. working paper, 2014.

Gerardi, Kristopher and Willen, Paul. (2009). Subprime Mortgages, Foreclosures, and Urban Neighborhood. *The B.E. Journal in Economic Analysis and Policy*, 9(3), 1-37.

Ghent, Andra and Kudlyak, Marianna (2011). Recourse and Residential Mortgage Default: Evidence from U.S. States. *Review of Financial Studies* 24(9): 3139-3186.

Goodstein Ryan M., Paul Hanouna, Carlos D. Ramierz, Christof W. Stahel (2011). Are Foreclosures Contagious? FDIC Center for Financial Research Working Paper No. $2011-4$.

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

Gerardi, Kristopher, Lehnert, Andreas, Sherlund, Shane M., and Willen, Paul (2008). Making Sense of the Subprime Crisis. *Brookings Papers on Economic Activity*, Fall 2008, pp. 69-159.

Gerardi, Kristopher, Rosenblatt, Eric, Willen, Paul S. , and Yao, Vincent (2012). Foreclosure Externalities: Some New Evidence. NBER Working Paper 18353.

Goldberg, Gerson M. and John P. Harding (2003). Investment characteristics of low- and moderate-income mortgage loans. *Journal of Housing Economics* 12: 151-180.

Guiso, L., Sapienza, P., and Zingales, L. (2013), The Determinants of Attitudes towards Strategic Default on Mortgages. *Journal of Finance*, 68(4), 1473-1515.

Gyourko, Joseph and Joseph Tracy (2014). Reconciling theory and empirics on the role of unemployment in mortgage default. *Journal of Urban Economics*, 80(1), 87-96.

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 Reports: 368.

Hogg, M. and Abrams, D. (1988). *Social identifications: A social psychology of intergroup relations and group processes.* Florence, KY US: Taylor and Frances/Routledge.

Kau, J., and Keenan, D., (1995). An Overview of the Option-Theoretic Pricing of Mortgages. *Journal of Housing Research*, 6(2), 217-244.

Marquez, C., (2007). Modeling Residential Mortgage Performance with Random Utility Models. University of California, Berkeley, Program on Housing and Urban Policy, Dissertation No. D07-002.

Mayer, C. and Pence, K. (2009). "Subprime Mortgages: What, Where, and to Whom?" in Glaeser, E. and Quigley, J. eds., *Housing Markets and the Economy: Risk, Regulation, and*

Policy. Cambridge, MA: Lincoln Institute of Land Policy.

Mayer, Christopher, Edward Morrison, Tomasz Piskorski and Arpit Gupta. (2014) Mortgage Modification and Strategic Behavior: Evidence from a Legal Settlement with Countrywide. *The American Economic Review*, forthcoming.

Mian, A., and Sufi, A., (2009). The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis. *Quarterly Journal of Economics*, 124 (4), 1449-96.

Painter, G., Yang, L., and Yu, Z., (2003). Heterogeneity in Asian-American Homeownership: The Role of Household Endowments and Immigrant Status. *Urban Studies*, 40(3), 505-530.

Munroe, David J. and Laurence Wilse-Samson. 2013. Foreclosure Contagion: Measurement and Mechanisms. Working Paper.

Painter, G., Yang, L., and Yu, Z., (2004). Homeownership Determinants of Chinese Americans: Assimilation, Ethnic Concentration, and Nativity. *Real Estate Economics*, 32 (3), 509-539.

Painter, G., and Yu, Z., (2008). Leaving Gateway Metropolitan Areas in the United States: Immigrants and the Housing Market. *Urban Studies*, 45 (5&6), 1163-1191.

Reid, R. and Laderman, E. (2009). The Untold Costs of Subprime Lending: Examining the Links among Higher-Priced Lending, Foreclosures and Race in California. Federal Reserve Bank of San Francisco Working paper.

Saiz, A. and Wachter, S. (2011). Immigration and the Neighborhood. *American Economic Journals: Policy*, 3(2), 169-188.

Towe, Charles, and Chad Lawley. 2013. The Contagion Effect of Neighboring Foreclosures. *American Economic Journal: Economic Policy* 5(2): 313-35.

Vandell, Kerry D., (1995). How Ruthless is Mortgage Default? A Review and Synthesis of the Evidence. *Journal of Housing Research* 6(2), 245-264.

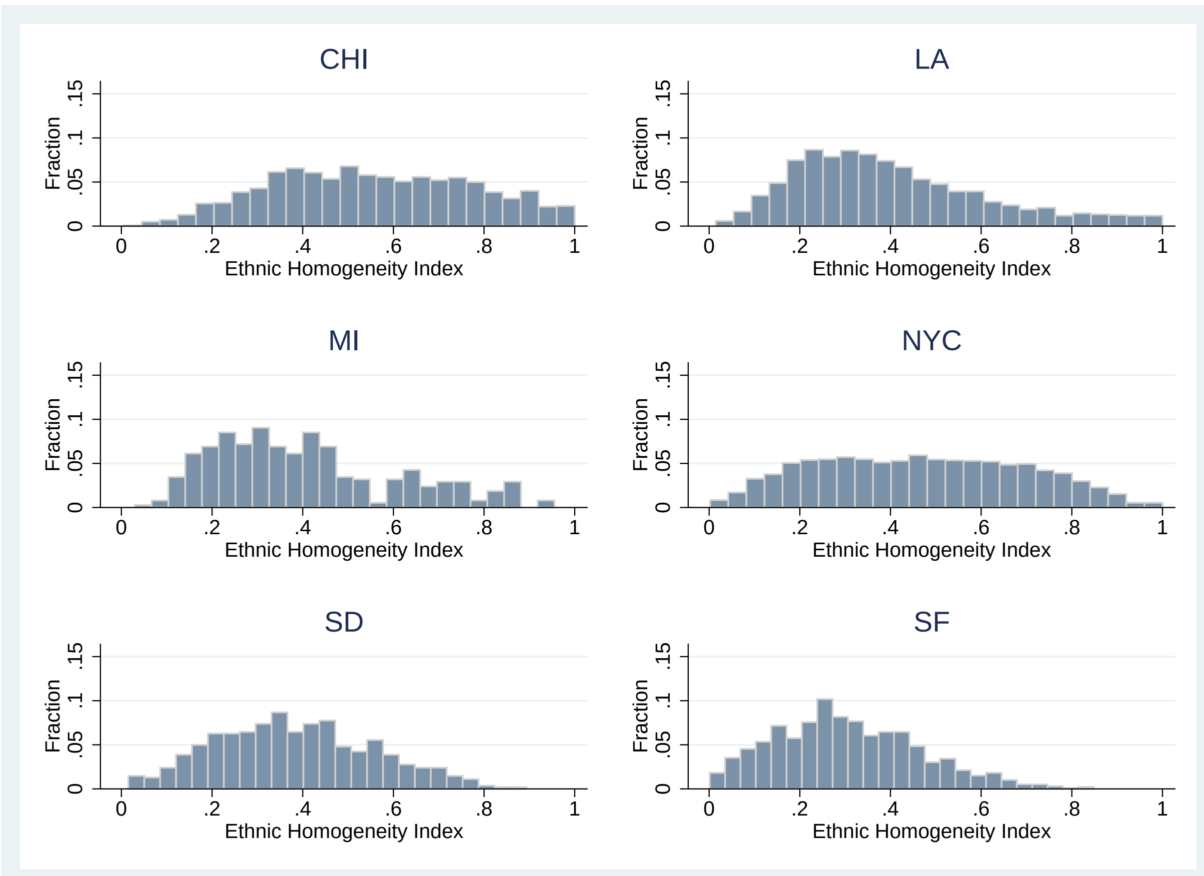

Figure 1: Distribution of Sample Census Tracts by Ethnic Homogeneity.

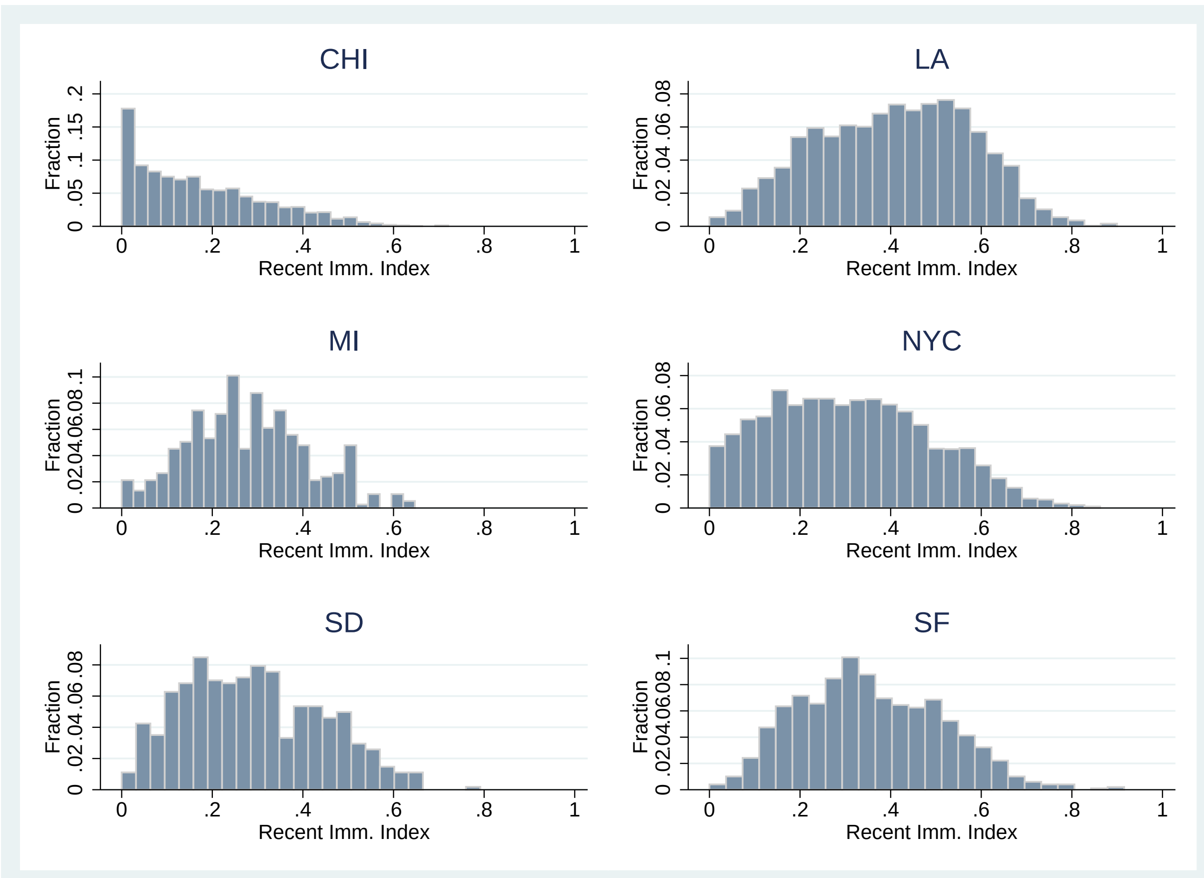

Figure 2: Distribution of Sample Census Tracts by Concentration of Recent Immigrants.

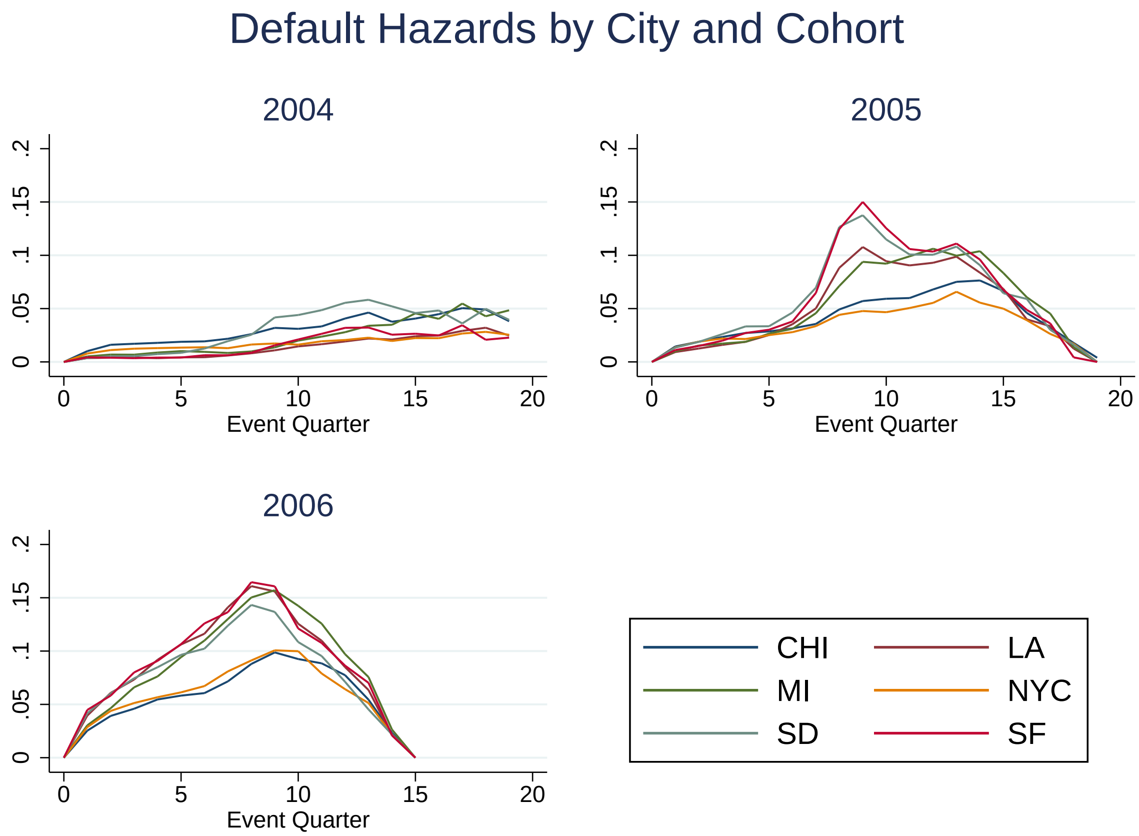

Figure 3: Sample Default Hazard Rates by City for Origination Cohorts 2004 - 2006.

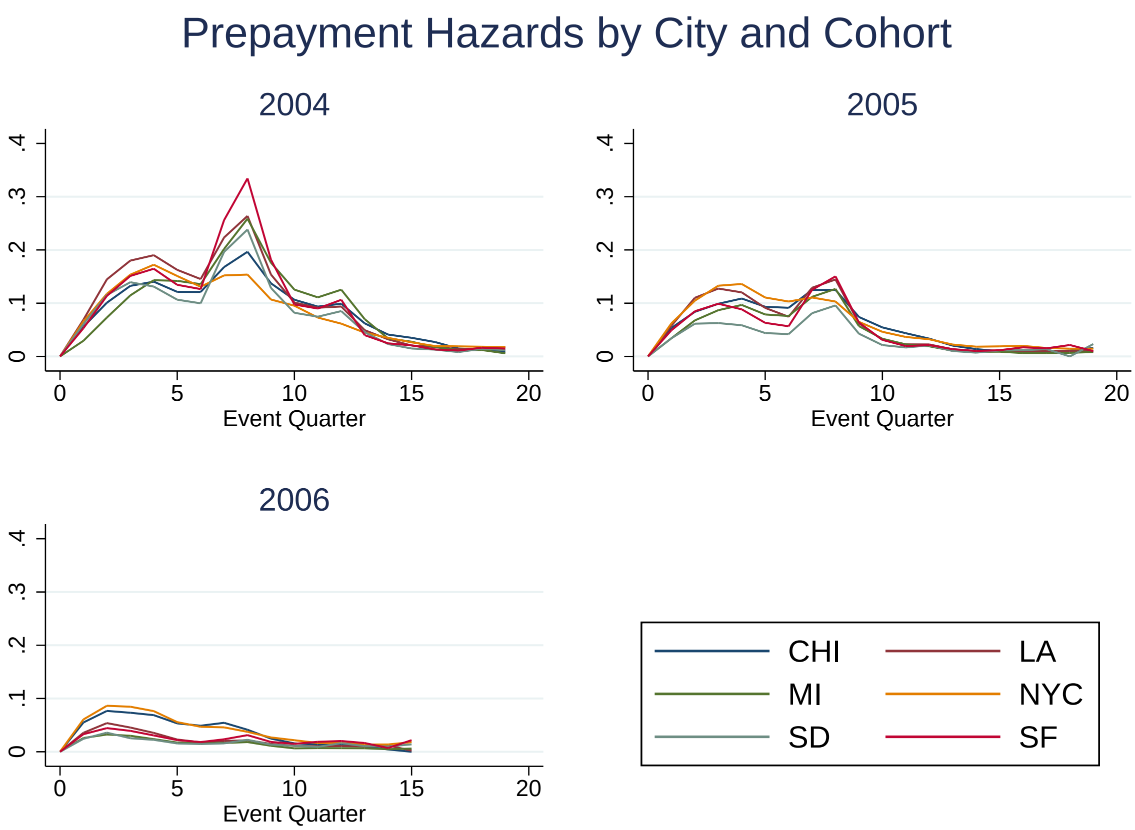

Figure 4: Sample Prepayment Hazard Rates by City for Origination Cohorts 2004 - 2006.

Table 1: Summary Statistics on Loan Characteristics

	Share of Loans						
	Total	**Chicago**	**Los Angeles**	**Miami**	**New York City**	**San Diego**	**San Francisco**
Ever Default	0.28	0.27	0.28	0.36	0.23	0.36	0.31
Ever Prepay	0.58	0.57	0.60	0.50	0.61	0.50	0.57
Loan Purpose							
Purchase	0.36	0.37	0.36	0.39	0.29	0.40	0.45
Refinance	0.05	0.08	0.04	0.03	0.04	0.05	0.05
Refi - Cash Out	0.59	0.55	0.60	0.58	0.67	0.55	0.50
Type of Loan							
FRM	0.16	0.12	0.16	0.19	0.23	0.12	0.09
Fixed - IO	0.01	0.00	0.01	0.01	0.01	0.02	0.01
ARM	0.48	0.70	0.38	0.57	0.56	0.32	0.32
ARM - IO	0.24	0.09	0.31	0.12	0.10	0.43	0.44
Balloon	0.12	0.09	0.13	0.11	0.11	0.11	0.14
Documentation							
Full Doc	0.53	0.62	0.50	0.51	0.52	0.50	0.51
Low Doc	0.47	0.38	0.50	0.49	0.48	0.50	0.49
No Doc	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Not 30 yr term	0.03	0.03	0.03	0.03	0.03	0.02	0.05
Prepay Penalty	0.69	0.27	0.94	0.95	0.24	0.93	0.91
Condo	0.12	0.10	0.12	0.21	0.06	0.23	0.13
No. Loans	654797	116080	274728	50956	114827	41109	57097

Table 2: Summary Statistics of Loans by City

City	Variable	Mean	SD	P25	Median	P75
Chicago	Initial House Val.	220099	121803	144385	190000	261905
	CLTV at Orig.	0.88	0.14	0.80	0.90	1.00
	Initial Loan Amt.	179395	95954	118230	157176	214837
	Initial Spread	3.45	1.08	2.66	3.38	4.14
	FICO	614	65	575	615	652
Los Angeles	Initial House Val.	414554	175105	298788	385000	500000
	CLTV at Orig.	0.83	0.18	0.72	0.85	1.00
	Initial Loan Amt.	312433	136245	215920	291360	384967
	Initial Spread	2.59	0.99	1.85	2.50	3.14
	FICO	629	69	588	631	670
Miami	Initial House Val.	239229	120910	165000	219090	283820
	CLTV at Orig.	0.84	0.16	0.75	0.85	1.00
	Initial Loan Amt.	184906	91875	126529	167801	221314
	Initial Spread	3.30	1.09	2.61	3.14	4.06
	FICO	611	67	570	612	650
New York City	Initial House Val.	392829	169375	295000	365000	450000
	CLTV at Orig.	0.80	0.18	0.70	0.80	0.94
	Initial Loan Amt.	294035	122549	211262	279032	353603
	Initial Spread	3.12	1.17	2.36	3.06	3.85
	FICO	615	67	569	614	655
San Diego	Initial House Val.	468510	158095	372705	450000	530000
	CLTV at Orig.	0.84	0.19	0.74	0.90	1.00
	Initial Loan Amt.	354000	126102	270929	344000	417450
	Initial Spread	2.44	0.98	1.76	2.38	3.06
	FICO	637	68	598	639	678
San Francisco	Initial House Val.	546092	195997	417971	525000	645000
	CLTV at Orig.	0.86	0.19	0.79	0.90	1.00
	Initial Loan Amt.	415026	146552	314394	401510	499999
	Initial Spread	2.44	1.01	1.76	2.36	3.06
	FICO	642	69	602	644	685
TOTAL	Initial House Val.	377485	189940	246000	351351	472727
	CLTV at Orig.	0.84	0.18	0.75	0.88	1.00
	Initial Loan Amt.	287254	142957	181287	264813	365986
	Initial Spread	2.87	1.12	2.06	2.76	3.61
	FICO	624	69	582	625	665

Table 3: Summary Statistics Comparison

Panel A: All Census Tracts in Sample CBSAs

ACS Variable (05-09)	**Mean**	**P25**	**Median**	**P75**
Population	4,812	3,145	4,452	6,018
Pct Homeownership	57.3	35.3	60.5	81.5
Med HH Inc	65,889	42,010	60,379	83,646
Pct Asian	10.7	1.5	5.6	13.7
Pct Black	15.5	1.0	4.0	15.2
Pct Hispanic	46.0	14.3	48.0	75.0
Pct Immigrant	6.5	2.5	5.4	9.5
High Asian	8.0	0.0	0.0	0.0
High Black	15.0	0.0	0.0	0.0
High Hispanic	61.0	0.0	100.0	100.0
High Immigrant	23.0	0.0	0.0	0.0
Ethnic Homogeneity	0.44	0.26	0.41	0.60
Immigrant Index	0.31	0.17	0.30	0.45
No. Census Tracts	*12,389*			

Panel B: Census Tracts with Sample Loans

ACS Variable (05-09)	**Mean**	**P25**	**Median**	**P75**
Population	4,491	3,293	4,365	5,555
Pct Homeownership	60.1	39.4	63.3	82.9
Med HH Inc	68,350	44,637	62,721	85,833
Pct Asian	11.2	1.8	6.1	14.6
Pct Black	14.4	1.0	3.6	13.4
Pct Hispanic	47.3	16.5	50.4	75.8
Pct Recent Immigrant	6.5	2.6	5.5	9.5
High Asian	9.0	0.0	0.0	0.0
High Black	14.0	0.0	0.0	0.0
High Hispanic	63.0	0.0	100.0	100.0
High Immigrant	23.0	0.0	0.0	0.0
Ethnic Homogeneity	0.44	0.26	0.41	0.60
Immigrant Index	0.32	0.17	0.31	0.45
No. Census Tracts	*9,198*			

Table 4: Summary Statistics Over Census Tract Characteristics (Equally-Weighted)

	Chicago (1401 Census Tracts)				Los Angeles (2542 Census Tracts)				Miami (376 Census Tracts)			
ACS Variable (05-09)	**Mean**	**P25**	**Median**	**P75**	**Mean**	**P25**	**Median**	**P75**	**Mean**	**P25**	**Median**	**P75**
Population	4,455	3,221	4,379	5,541	4,685	3,580	4,554	5,668	4,861	3,655	4,792	5,989
Pct Homeownership	67.1	50.7	70.6	86.3	56.0	35.0	58.4	78.1	69.8	56.1	73.6	87.6
Med HH Inc	62,644	42,632	57,350	76,972	63,168	40,708	56,975	79,076	56,129	38,339	50,199	70,383
Pct Asian	4.6	0.0	2.0	6.0	13.7	2.9	8.5	17.4	2.7	0.7	2.0	4.3
Pct Black	22.5	0.8	4.1	28.6	6.9	0.7	2.8	7.0	24.1	4.8	12.3	30.9
Pct Hispanic	54.3	21.2	65.5	84.0	36.0	8.7	32.9	61.1	52.2	33.2	57.2	74.2
Pct Immigrant	3.0	0.6	2.1	4.5	8.9	4.7	8.5	12.3	5.2	2.7	4.7	6.9
High Asian	1.0	0.0	0.0	0.0	11.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
High Black	24.0	0.0	0.0	0.0	4.0	0.0	0.0	0.0	24.0	0.0	0.0	0.0
High Hispanic	69.0	0.0	100.0	100.0	50.0	0.0	0.0	100.0	75.0	100.0	100.0	100.0
High Immigrant	4.0	0.0	0.0	0.0	39.0	0.0	0.0	100.0	10.0	0.0	0.0	0.0
Ethnic Homogeneity	0.6	0.4	0.5	0.7	0.4	0.2	0.4	0.5	0.4	0.2	0.4	0.5
Immigrant Index	0.2	0.1	0.1	0.3	0.4	0.3	0.4	0.5	0.3	0.2	0.3	0.4

	New York City (3345 Census Tracts)				San Diego (542 Census Tracts)				San Francisco (992 Census Tracts)			
	Mean	**P25**	**Median**	**P75**	**Mean**	**P25**	**Median**	**P75**	**Mean**	**P25**	**Median**	**P75**
Population	4,241	2,991	4,051	5,329	4,785	3,492	4,521	5,775	4,585	3,461	4,472	5,662
Pct Homeownership	59.4	34.6	62.6	85.7	58.9	41.1	63.0	79.0	60.3	42.1	63.1	80.2
Med HH Inc	71,816	47,500	66,528	90,000	66,780	46,825	62,734	82,019	83,490	59,222	79,041	103,170
Pct Asian	9.3	1.5	5.2	11.9	10.4	3.5	6.6	12.0	24.4	10.4	18.8	35.0
Pct Black	19.4	1.1	5.0	24.1	4.9	0.9	2.8	6.4	7.2	1.0	2.8	7.4
Pct Hispanic	51.8	16.1	60.0	83.3	53.8	32.4	59.8	75.4	46.1	24.9	46.6	67.7
Pct Immigrant	6.3	2.5	5.3	9.0	5.6	2.5	4.8	8.2	7.3	3.9	6.3	10.2
High Asian	6.0	0.0	0.0	0.0	5.0	0.0	0.0	0.0	27.0	0.0	0.0	100.0
High Black	21.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	5.0	0.0	0.0	0.0
High Hispanic	66.0	0.0	100.0	100.0	75.0	0.0	100.0	100.0	67.0	0.0	100.0	100.0
High Immigrant	20.0	0.0	0.0	0.0	15.0	0.0	0.0	0.0	26.0	0.0	0.0	100.0
Ethnic Homogeneity	0.5	0.3	0.5	0.7	0.4	0.3	0.4	0.5	0.3	0.2	0.3	0.4
Immigrant Index	0.3	0.2	0.3	0.4	0.3	0.2	0.3	0.4	0.4	0.2	0.3	0.5

Table 5: Default Coefficient Estimates from Multinomial Logit Model of Mortgage Choice Using Pooled City Sample

Panel A: Tract Population Percentage

	Main Effect	**Asian**	**Black**	**Hispanic**	**Immigrant**
		-0.0014***	0.0003	-0.0017***	0.0002
		[0.000]	[0.000]	[0.000]	[0.001]
CLTV90	0.4423***				
	[0.010]				
CLTV100	0.7208***				
	[0.010]				
CLTV110	0.9863***				
	[0.011]				
CLTV120	1.1958***				
	[0.013]				
CLTV120P	1.3607***				
	[0.013]				

Panel B: Entropy Measure

		Social Homogeneity Measure	
	Main Effect	**Ethnicity**	**Immigrant**
CLTV80		0.1596***	-0.9561***
		[0.030]	[0.040]
CLTV90	0.2719***	0.1501***	-0.3847***
	[0.031]	[0.033]	[0.042]
CLTV100	0.4574***	0.0922***	0.007
	[0.029]	[0.030]	[0.037]
CLTV110	0.7015***	-0.0801***	0.2578***
	[0.029]	[0.029]	[0.036]
CLTV120	0.7462***	0.0174	0.6029***
	[0.035]	[0.042]	[0.050]
CLTV120P	0.7845***	0.1494***	0.7836***
	[0.030]	[0.033]	[0.040]

Panel C: CLTV Category Interations with Tract Population Percentage

	Main Effect	**Asian**	**Black**	**Hispanic**	**Immigrant**
CLTV80		0.0001	0.0016***	0.0001	-2.0226***
		[0.001]	[0.000]	[0.000]	[0.204]
CLTV90	0.4479***	-0.0035***	0.0024***	0.0001	-0.7039***
	[0.046]	[0.001]	[0.000]	[0.000]	[0.217]
CLTV100	0.7127***	0.0000	0.0019***	-0.0006	-0.1322
	[0.044]	[0.001]	[0.000]	[0.000]	[0.196]
CLTV115	0.9876***	0.0004	-0.0002	-0.0010***	0.5060***
	[0.042]	[0.000]	[0.000]	[0.000]	[0.165]
CLTV115P	1.2879***	-0.0042***	-0.0057***	-0.0045***	1.6529***
	[0.046]	[0.001]	[0.000]	[0.000]	[0.187]

Notes: CLTV80 indicates a category where CLTV is less than .80, and **CLTV90** indicates a category where CLTV is greater than or equal to .80 but less than .90. The other categories are defined analogously. **CLTV120P** indicates a category where CLTV is at or above 1.20. Robust standard errors, clustered at the loan level, are provided in brackets. Statistical significance of estimates is indicated as follows: *** p<0.01, ** p<0.05, * p<0.1. Estimates for prepayment alternative are available upon request from the authors.

Table 6: Estimated Racial, Social and Social Homogeneity Effects on Borrower Default Choice - Full vs. City Samples

RHS Variables	U.S.	CHI	LA	MI	NYC	SD	SF
CLTV90	0.4479***	0.1344	0.3512***	0.1127	0.2005*	0.3589	0.367
	[0.046]	[0.153]	[0.080]	[0.229]	[0.116]	[0.246]	[0.235]
CLTV100	0.7127***	0.2063	0.5461***	0.289	0.3470***	0.5587**	1.1226***
	[0.044]	[0.145]	[0.074]	[0.228]	[0.116]	[0.230]	[0.207]
CLTV115	0.9876***	0.4503***	0.7840***	0.5576***	0.6837***	0.8664***	1.4885***
	[0.042]	[0.149]	[0.068]	[0.201]	[0.120]	[0.206]	[0.193]
CLTV115P	1.2879***	0.9826***	1.1302***	0.7079***	0.5054*	1.1511***	2.1948***
	[0.046]	[0.217]	[0.071]	[0.197]	[0.271]	[0.209]	[0.208]
Asian * CLTV80	0.0001	-0.0039	0.0000	-0.0109	-0.0007	-0.001	-0.0015
	[0.001]	[0.003]	[0.001]	[0.008]	[0.002]	[0.003]	[0.002]
Asian * CLTV90	-0.0035***	-0.0013	-0.0024**	0.0081	-0.0054***	-0.0055*	-0.0024
	[0.001]	[0.003]	[0.001]	[0.010]	[0.002]	[0.003]	[0.002]
Asian * CLTV100	0.0000	-0.0005	0.0007	-0.0014	-0.0041**	-0.001	0.0018
	[0.001]	[0.002]	[0.001]	[0.010]	[0.002]	[0.002]	[0.001]
Asian * CLTV115	0.0004	0.0034	0.0008	-0.0059	-0.0015	-0.0038**	-0.0008
	[0.000]	[0.002]	[0.001]	[0.007]	[0.002]	[0.002]	[0.001]
Asian * CLTV115P	-0.0042***	0.0009	-0.0053***	-0.0079	-0.0120*	-0.0065***	-0.0067***
	[0.001]	[0.004]	[0.001]	[0.006]	[0.006]	[0.001]	[0.001]
Black * CLTV80	0.0016***	0.0033***	0.0002	0.0019	0.0018**	0.0057	0.0106***
	[0.000]	[0.001]	[0.001]	[0.001]	[0.001]	[0.004]	[0.003]
Black * CLTV90	0.0024***	0.0044***	0.0003	0.0015	0.0014	0.0017	0.0085***
	[0.000]	[0.001]	[0.001]	[0.002]	[0.001]	[0.004]	[0.003]
Black * CLTV100	0.0019***	0.0051***	-0.0001	-0.0002	0.0020**	-0.004	0.002
	[0.000]	[0.001]	[0.001]	[0.002]	[0.001]	[0.004]	[0.002]

Table 6: (continued)

RHS Variables	U.S.	CHI	LA	MI	NYC	SD	SF
Black * CLTV115	-0.0002	0.0029***	-0.0019***	0	-0.0014	-0.0024	-0.0042***
	[0.000]	[0.001]	[0.001]	[0.001]	[0.001]	[0.003]	[0.001]
Black * CLTV115P	-0.0057***	-0.0044**	-0.0080***	-0.0014	-0.0027	-0.0068***	-0.0126***
	[0.000]	[0.002]	[0.001]	[0.001]	[0.003]	[0.003]	[0.002]
Hispanic * CLTV80	0.0001	0.0012	0.0007	0.0007	0.0009	0.0002	0.0036*
	[0.000]	[0.001]	[0.001]	[0.002]	[0.001]	[0.002]	[0.002]
Hispanic * CLTV90	0.0001	0.0011	0.0011	0.0021	-0.0008	0.0009	0.0021
	[0.000]	[0.001]	[0.001]	[0.002]	[0.001]	[0.002]	[0.002]
Hispanic * CLTV100	-0.0006	0.0021**	0.0008	0.0021	-0.0016	-0.0013	-0.0039**
	[0.000]	[0.001]	[0.001]	[0.002]	[0.001]	[0.002]	[0.002]
Hispanic * CLTV115	-0.0010***	0.0006	0.0002	0.0009	-0.0035***	-0.0024**	-0.0045***
	[0.000]	[0.001]	[0.000]	[0.002]	[0.001]	[0.001]	[0.001]
Hispanic * CLTV115P	-0.0045***	-0.0027	-0.0046***	0.0004	-0.0002	-0.0063***	-0.0131***
	[0.000]	[0.002]	[0.001]	[0.001]	[0.003]	[0.001]	[0.001]
Immigrant * CLTV80	-2.0226***	-2.0773**	-1.1352***	-1.3816**	-1.6560***	-0.4143	1.8375*
	[0.204]	[0.965]	[0.362]	[0.661]	[0.396]	[1.174]	[1.023]
Immigrant * CLTV90	-0.7039***	-0.6358	-0.3643	-1.0144	-0.2649	0.7744	1.8212*
	[0.217]	[0.837]	[0.390]	[0.829]	[0.459]	[1.144]	[0.969]
Immigrant * CLTV100	-0.1322	1.3482*	0.0273	0.0095	0.4439	0.7392	-1.0085
	[0.196]	[0.722]	[0.328]	[0.811]	[0.455]	[0.975]	[0.715]
Immigrant * CLTV115	0.5060***	2.3574***	0.4343*	-0.1816	0.4209	0.2559	0.0077
	[0.165]	[0.736]	[0.247]	[0.597]	[0.483]	[0.660]	[0.494]
Immigrant * CLTV115P	1.6529***	1.6849	1.5456***	1.1415**	0.6111	1.6442**	-0.1322
	[0.187]	[1.380]	[0.251]	[0.487]	[1.506]	[0.691]	[0.620]

Notes: **CLTV80** indicates a category where CLTV is less than .80, and **CLTV90** indicates a category where CLTV is greater than or equal to .80 but less than .90. The other categories are defined analogously. **CLTV115P** indicates a category where CLTV is at or above 1.15. We also included an additional set of controls in our model, although we do not report estimates here: a cohort variable indicating the year in which a mortgage was originated, interactions between the origination year and a cubic function that captures times since origination, and a set of dummy variables for the top 12 originators. Robust standard errors, clustered at the loan level, are provided in brackets. Statistical significance of estimates is indicated as follows: *** p<0.01, ** p<0.05, * p<0.1.

Table 7: Covariate Effects on Borrower Default Choice - Full vs. City Samples

RHS Variables	U.S.	CHI	LA	MI	NYC	SD	SF
Rec. * CLTV80	0.4107***						
	[0.039]						
Rec. * CLTV90	0.0365						
	[0.039]						
Rec. * CLTV100	-0.1552***						
	[0.039]						
Rec. * CLTV115	-0.3077***						
	[0.039]						
Rec. * CLTV115P	-0.2527***						
	[0.040]						
ln(Init. Balance)	0.5467***	0.5020***	0.5598***	0.6811***	0.4175***	0.4345***	0.4552***
	[0.010]	[0.019]	[0.019]	[0.033]	[0.026]	[0.043]	[0.049]
Spread	0.3032***	0.2375***	0.3352***	0.2487***	0.2799***	0.3539***	0.3595***
	[0.002]	[0.005]	[0.003]	[0.006]	[0.005]	[0.007]	[0.007]
620 < FICO < 725	0.4629***	0.4788***	0.4428***	0.3608***	0.4750***	0.4816***	0.4632***
	[0.011]	[0.040]	[0.016]	[0.044]	[0.033]	[0.034]	[0.029]
FICO < 620	0.7619***	0.7877***	0.7385***	0.5948***	0.8208***	0.8212***	0.6837***
	[0.012]	[0.041]	[0.018]	[0.047]	[0.035]	[0.040]	[0.035]
Refi - no cash	-0.1290***	-0.2287***	-0.1225***	-0.1867***	0.0200	0.0142	-0.1983***
	[0.014]	[0.025]	[0.025]	[0.063]	[0.036]	[0.050]	[0.051]
Refi - cash out	-0.1284***	-0.2639***	-0.1225***	-0.1011***	0.0016	0.0331	-0.1334***
	[0.008]	[0.015]	[0.013]	[0.036]	[0.018]	[0.027]	[0.029]
ARM/Hybrid	0.5822***	0.5123***	0.6534***	0.4616***	0.5566***	0.6815***	0.8547***
	[0.010]	[0.023]	[0.018]	[0.030]	[0.022]	[0.048]	[0.055]
IO	-0.0219***	-0.0060	-0.0404***	-0.0175	-0.0642***	-0.0145	-0.0209
	[0.007]	[0.021]	[0.012]	[0.024]	[0.021]	[0.028]	[0.028]
Balloon	-0.0163**	-0.0112	-0.0572***	-0.0621***	-0.0200	0.0101	-0.0176
	[0.008]	[0.020]	[0.014]	[0.024]	[0.020]	[0.035]	[0.032]
Int. Rate Reset$_{t-1}$	0.0385***	0.0479***	0.0354***	0.0108	0.0434***	0.0360*	-0.0074
	[0.005]	[0.013]	[0.009]	[0.018]	[0.014]	[0.019]	[0.017]
Low Doc	0.2405***	0.2512***	0.2021***	0.1919***	0.3040***	0.3416***	0.2027***
	[0.006]	[0.014]	[0.009]	[0.018]	[0.014]	[0.021]	[0.019]
No Doc	0.1409**	0.1964	0.0720	0.2979**	0.3684**	-0.0753	-0.2396
	[0.059]	[0.162]	[0.086]	[0.151]	[0.144]	[0.247]	[0.239]
Not 30yr.	-0.0640***	-0.0958**	-0.0753***	-0.1382***	-0.1359***	0.0103	-0.0670*
	[0.014]	[0.038]	[0.025]	[0.046]	[0.039]	[0.060]	[0.039]
Prepay Penalty	-0.1955***	-0.1490***	-0.1965***	-0.1625***	-0.0991***	-0.2505***	-0.3152***
	[0.007]	[0.020]	[0.012]	[0.024]	[0.027]	[0.026]	[0.024]
Cum. Foreclosure	0.1230***	0.0865***	0.1621***	0.0856***	0.0731***	0.0932***	0.1384***
	[0.004]	[0.010]	[0.006]	[0.015]	[0.010]	[0.016]	[0.014]
ΔUnemRate	0.2128***	0.1212***	0.3887***	0.3085***	0.2259***	0.5100***	0.1924***
	[0.006]	[0.009]	[0.013]	[0.026]	[0.016]	[0.036]	[0.022]
ΔPrices - Tract	-0.0436***	-0.0125**	-0.0287***	-0.0243***	-0.0423***	-0.0442***	-0.0135***
	[0.001]	[0.005]	[0.002]	[0.005]	[0.007]	[0.006]	[0.005]
ΔLn(Income) - Tract	-0.2021***	-0.2846***	-0.2057***	-0.1715***	-0.1540***	-0.0845**	-0.1672***
	[0.011]	[0.027]	[0.017]	[0.039]	[0.029]	[0.039]	[0.039]

Notes: **Rec.** is a dummy variable that indicates whether a jurisdiction permits recourse in the case of default. We also included an additional set of controls in our model, although we do not report estimates here: a cohort variable indicating the year in which a mortgage was originated, interactions between the origination year and a cubic function that captures times since origination, and a set of dummy variables for the top 12 originators. Robust standard errors, clustered at the loan level, are provided in brackets. Statistical significance of estimates is indicated as follows: *** p<0.01, ** p<0.05, * p<0.1.

Table 8: Estimates from Random Effects Model

RHS Variables	
CLTVO	29.881***
	[0.649]
AsianQ4 * CLTV	-0.6243
	[0.865]
BlackQ4 * CLTV	-6.7317***
	[0.853]
HispanicQ4 * CLTV	-3.3999***
	[0.949]
ImmigrantQ4 * CLTV	-1.136
	[0.827]
Spread	0.4470***
	[0.016]
Prepay Effect	-44.648
	[2.131][a]
Int. Rate Reset	0.2977***
	[0.016]
Cum. Foreclosure	4,233.2
	[1.567][a]
ΔUnemRate	0.7575***
	[0.020]
ΔPrices - Tract	-0.0490***
	[0.008]

Notes: **AsianQ4 * CLTV**, **BlackQ4 * CLTV**, **HispanicQ4 * CLTV**, and **ImmigrantQ4 * CLTV** are interaction terms formed using dummy variables that indicate only the top quartile values of the indicated social group. **CLTVO** indicates the loan to value at the origination of the mortgage. Robust standard errors, clustered at the loan level, are provided in brackets. [a] indicates the original standard error has been divided by 10,000. Statistical significance of estimates is indicated as follows: *** p<0.01, ** p<0.05, * p<0.1.