A New Look at the U.S. Foreclosure Crisis: Panel Data Evidence of Prime and Subprime Borrowers from 1997 to 2012*

June 5, 2015

Fernando Ferreira and Joseph Gyourko
The Wharton School
University of Pennsylvania and NBER

Abstract

Utilizing new panel micro data on the ownership sequences of all types of borrowers from 1997-2012 leads to a reinterpretation of the U.S. foreclosure crisis as more of a prime, rather than a subprime, borrower issue. Moreover, traditional mortgage default factors associated with the economic cycle, such as negative equity, completely account for the foreclosure propensity of prime borrowers relative to all-cash owners, and for three-quarters of the analogous subprime gap. Housing traits, race, initial income, and speculators did not play a meaningful role, and initial leverage only accounts for a small variation in outcomes of prime and subprime borrowers.

*We appreciate the excellent research assistance of Matt Davis, Lindsay Relihan, Yitong Wang, and Chen Zheng. We thank Raphael Bostic, Keith Head, Vernon Henderson, Henry Overman, Melissa Prado, Olmo Silva, Nancy Wallace, Paul Willen and seminar participants at the Nova School of Business and Economics, London School of Economics, NBER Real Estate Summer Institute, Trinity College Dublin, Inter-American Development Bank, Lubrafin Conference, and Homer Hoyt Institute for comments and suggestions. We are grateful to the Research Sponsors Program of the Zell/Lurie Real Estate Center at Wharton for financial support.

I. Introduction

Most economic analysis of the recent American housing market bust and the subsequent default and foreclosure crises focuses on the role of the subprime mortgage sector.[1] Roughly three-quarters of the papers on the crisis reviewed in the next section use data only from the subprime sector and typically include outcomes from no later than 2008. For example, Mian & Sufi (2009) use mortgage defaults aggregated at the zip code level from 2005 to 2007 to conclude that a "salient feature of the mortgage default crisis is that it is concentrated in subprime ZIP codes throughout the country." However, subprime loans comprise a relatively small share of the complete housing market--about 15% in our data and never more than 21% in a given year. In addition, we document that most foreclosures in the United States occurred after 2008. These two issues raise questions about the representativeness of results based on selected subprime samples.

In this paper we provide new stylized facts about the foreclosure crisis and also empirically investigate different proposed explanations for why owners lost their homes during the last housing bust. We use micro data that track outcomes well past the beginning of the crisis and cover all types of house purchase financing – prime mortgages, Federal Housing Administration (FHA)/Veterans Administration (VA)-insured loans, loans from small or infrequent lenders, and all-cash buyers -- not just the subprime sector. The data (described below in Section III)) contain information on over 33 million unique ownership sequences in just over 19 million distinct owner-occupied housing units in 96 metropolitan areas (MSAs) from 1997(1)-2012(3), resulting in almost 800 million quarterly observations. It also includes information on up to three loans taken out at the time of home purchase, and all subsequent refinancing activity. Thus, we are able to create owner-specific panels with financing information from purchase through sale or other transfer of the home.

These data show that the crisis was not solely, or even primarily, a subprime sector event. It started out that way, but quickly morphed into a much bigger and broader event dominated by prime borrowers losing their homes. Figure 1 reports the raw number of homes lost via

[1] There is no legal definition of what constitutes a subprime mortgage. Researchers have used rules based on lender lists and credit score cutoffs, and in all cases found very high rates of subprime distress. We discuss the methods used by other researchers in the next section, with section III detailing how we distinguish prime from subprime borrowers.

foreclosure or short sale for the five different types of owners we track each year across all 96 metropolitan areas in our sample. There are only seven quarters, 2006(3)-2008(1) at the beginning of the housing market bust, in which there were more homes lost by subprime borrowers than by prime borrowers, although the gap is small as the figure illustrates. Over this time period, which is the focus of much of the previous literature in this area, 39,094 more subprime than prime borrowers lost their homes. This small difference was completely reversed by the beginning of 2009, as 40,630 more prime borrowers than subprime borrowers lost their homes just in the 2nd, 3rd, and 4th quarters of 2008. An additional 656,003 more prime than subprime borrowers lost their homes from 2009(1)-2012(3), so that twice as many prime borrowers lost their homes than did subprime borrowers over our full sample period.

One reason for this pattern is that the number of prime borrowers dwarfs that of subprime borrowers (and the other borrower/owner categories we consider). Table 1 lists the absolute number and share of all our borrower/owner categories over time. The prime borrower share varies around 60% over time and did not decline during the housing boom. Subprime borrower share nearly doubled during the boom, but only up to 21%. Subprime's increasing share came at the expense of the FHA/VA-insured sector, not the prime sector.

This helps put Figure 2's plot of foreclosure/short sale rates by borrower/owner type in proper perspective. Sharply higher subprime distress rates became evident early in the housing bust, just as the previous literature shows.[2] However, those high rates never affected anything close to a majority of the market. Moreover, loss rates among the much larger group of prime borrowers started to increase shortly thereafter—within a year. The jump in the foreclosure rate for prime owners become even more relevant empirically over time, as the market share of subprime borrowers dramatically declined after 2008 as shown in Table 1.

After documenting basic facts about the housing bust, we turn to estimating panel data models of the probability of losing a home in foreclosure or via short-sale as a function of prime and subprime status and other factors. That our micro data allows us to create panels of full ownership sequences provides a potentially important advantage relative to earlier research that relied on loan-level data sets. Our ability to track borrowers/owners using different types of debt

---

[2] The 'Small' lender sector also has a sudden and sharp early spike upward around the same time as Subprime. This group includes owners who financed their homes from nontraditional 'small' sources that issued no more than 100 mortgages throughout our sample period. This group itself is very small in number, never amounting to more than 2%-3% of all owners. See below for more on them.

(such as subprime versus prime mortgages) over time means that we can also use our panel to estimate whether there are common factors that can explain foreclosure activity across mortgage labels. For example, we can measure negative equity conditions (i.e., when the current loan-to-value (LTV) ratio is greater than one) for each quarter in each ownership sequence.

Current LTV is a powerful predictor of home loss, regardless of borrower type. This is consistent with the implications of a traditional home mortgage default literature which shows that negative equity changes an owner's incentive to keep current on one's loan (see Foster and Van Order (1984), Kau, Kennan & Kim (1994), and Deng, Quigley & Van Order (2000), for example).[3] Controlling flexibly for current LTV almost fully explains the spike and continued elevated rate of foreclosures and short sales by prime borrowers during the housing bust. Thus, prime borrowers do not lose their homes at appreciably higher rates during the crisis than do all cash owners after controlling for negative equity. Because the incentives arising from the presence of negative equity do not vary by type of mortgage contract to a first approximation, traditional mortgage default models imply this variable should be influential in accounting for home losses in the subprime sector, too. That is what we find. Current LTV explains about three-quarters of all home losses among subprime borrowers on average, including about one-half in the spike of the first year of the crisis.

The traditional mortgage default literature also posits a potentially important role for borrower illiquidity (typically thought of as arising from negative income shocks) in the decision to default, which is a precursor to losing one's home.[4] Because there are no large U.S. data sets with individual measures of on-going unemployment status, we are like the rest of the literature in being unable to test directly for this effect.[5] However, our data provide proxies for borrower

[3] While we observe the precise date the owner lost her home to foreclosure or short sale, we cannot tell when the initial default occurred. There are industry rules of thumb that can be applied to impute the start of the distress sequence, but they vary by jurisdiction and over time. In any event, we are more interested in home loss, which is measured with accuracy in our data.

[4] See Foote, et. al. (2010) for more on this 'double trigger' hypothesis. That terminology arises as follows. Negative equity is one trigger, but it is a necessary, not sufficient, condition for default. If the owner subsequently suffers a negative income shock that renders the household illiquid and unable to make monthly debt service payments in a timely manner, that is the second trigger which guarantees default because not even a sale of the property can pay off the outstanding balance when there is negative equity.

[5] The Panel Survey on Income Dynamics (PSID) does provide information on the employment status of household members, but its samples are too small for our purposes. We also experimented with aggregate employment measures, but they had little or no impact on further reducing the gap between subprime and prime foreclosure rates. Large attenuation bias is to be expected when an aggregate measure is used to proxy for individual unemployment status (Gyourko and Tracy (2014)). For example, if the local unemployment rate doubles from 5% to 10% in a

illiquidity such as census tract-by-quarter indicators (that control for local economic conditions over time at the neighborhood level) or ownership indicators (that precisely control for individual conditions which are permanent, but have no time variation) in our specifications. These proxies show an economically modest impact on the probability of losing one's home, but the role of negative equity remains very powerful. That said, it could be that our better measured LTV control is reflecting some of the impact related to unobserved individual income shocks. For example, only 18% of our owners who ever experienced negative equity ended up losing their homes.[6] That is a large number, given that 40% of our ownership sequences had negative equity for at least one quarter. However, if only negative equity mattered, all of them presumably would have defaulted and ultimately lost their homes. While we remain agnostic on the precise strength of each of these two mechanisms, our results show that their combined effect is to eliminate almost entirely the empirical importance of the Prime and Subprime labels in explaining differential probabilities of foreclosure during the housing bust.

We also estimate whether a host of other 'initial conditions' affect the probability of home loss or weaken the impact of negative equity. These include owner demographics such as race, initial self-reported income and whether she is a speculator, housing unit traits such as the number of bedrooms and square footage of living space, and other financial factors such as initial LTV, whether there is a second loan or a refinancing, and the loan cohort. Neither borrower traits nor housing unit traits appear to have played a meaningful role in the foreclosure crisis. Initial LTV has been shown to account for about 60% of the foreclosure crisis based on simulations of a macro model (Corbae and Quentin (2015)), and Bayer, Ferreira and Ross (forthcoming) and Palmer (2014) find that the cohort of the most recent purchase or refinancing is influential in predicting defaults. However, our data reveal only modest impacts for these factors on whether owners ultimately lose their homes. We argue that these variables are best thought as helping measure current LTV more accurately.

While this study is focused on the foreclosure crisis, it obviously is related to a large literature on how the U.S. housing boom started and evolved. In other work, we document

---

quarter, 90% of the labor force still is employed, so regressing whether an individual owners lost a home to foreclosure or engaged in a short sale on that aggregate variable is not likely to be statistically significant.

[6] That is quite close to an independent estimate made by the firm CoreLogic which reported that 85% of owners in negative equity as of the second quarter of 2012 still were current on their mortgages (http://www.dsnews.com/articles/borrowers-in-negative-equity-slowly-declining-as-home-values-gain-report-2012-09-12).

substantial variation in when housing booms began across metropolitan areas and show that the initial jumps in local prices tend to coincide with jumps in local area income (Ferreira and Gyourko (2013)). Shiller (2005) and Case, Shiller and Thompson (2012) famously argued that the subsequent sharp increases in price-to-income and price-to-rent ratios were based on unrealistic expectations of house price growth. Soo (2015) uses local newspaper coverage to try to quantify those animal spirits, with DeFusco, et al. (2014) examining how heterogeneity in local market price increases might have generated spillover effects on price growth in nearby areas. The role of speculators in pushing up prices in certain markets has been analyzed by Chinco and Mayer (2014) and Haughwout , et al. (2011). Mian and Sufi (2011) and Bhutta and Keys (2013) respectively study the roles of house prices and interest rates in increasing equity extraction during the housing boom and on future default rates. In addition to the host of research on the subprime sector discussed in the next section, there is a recent debate about changes in buyer composition during the run up of the housing boom (Adelino, Schoar and Severino (2015); Mian and Sufi (2015)). We do not directly address a question related to this latter debate—namely, whether prime borrower foreclosures would have happened in the absence of the initial increase in subprime distress. We suspect the answer is 'yes', as Figure 1 shows a somewhat similar timing in the surge of foreclosures across borrower types. However, presenting a complete theory and empirical analysis that links the beginning of local housing booms, how they evolved, and their respective busts, all within the broader context of the economic cycle which included a global financial crisis, requires a separate analysis that is left to future research.

The paper proceeds as follows. Section II discusses the related mortgage market literature focusing on the consequence of the bust. This is followed by a detailed description of our data in Section III. Section IV reports the empirical results. There is a brief conclusion.

II. Related Literature: Implications of the Focus on Subprime

Even though prime and subprime borrowers were losing their homes in roughly equal numbers as the crisis began (Figure 1), because loss rates initially spiked so sharply among subprime borrowers (Figure 2), researchers paid particular attention to that sector. Appendix Table 1 reports a list of published papers from 2008(1) to 2014(2) on the housing bust. While not exhaustive, it is representative of primarily empirically-oriented work on the fallout from bust.[7]

There are a number of noteworthy patterns in this research. First, three-quarters of the papers focus exclusively on the subprime sector in their data and analysis (see column 2). There is no legal definition of what distinguishes a subprime from a prime loan. Previous research uses one of two methods to categorize loan type. Papers using loan level data typically use a credit score cutoff (the FICO score range used runs from 600-660) to distinguish subprime from prime borrowers. Other research relies on lender lists compiled annually by HUD since 1997 or industry publications such as *Inside Mortgage Finance*, which reports the top 20 subprime lenders each year from 1990-onward, to make this distinction.[8] Lender lists focus on subprime loans that are securitized in the private sector, while low FICO scores capture mortgages that could be kept by banks and mortgage issuers and also mortgages securitized by the government. Both sets of papers find large default rates for subprime mortgages.

Second, the vast majority of the studies analyzing borrower payment behavior use loan level data (see column 3), with the predominant data source being First American Core Logic Loan Performance (LP, hereafter; see column 4 for the specific data provider). This source is based primarily on subprime mortgages which were used to collateralize private label mortgage-backed securities (MBS). The strength of the LP data is that they are rich in detail on loan traits.[9] A countervailing weakness is that it cannot be used to generate panels of ownership sequences, unless the owner never changes the debt it uses. Hence, these studies generally cannot control for the addition of new debt or link refinancing of original debt across a unique owner. Thus, cumulative LTVs cannot be known with accuracy unless these data are merged

[7] The articles listed here are published pieces in three urban/real estate journals (*Journal of Urban Economics*, *Real Estate Economics*, *Journal of Housing Economics*), select finance journals (*Journal of Finance*, *Review of Financial Studies*, *Journal of Financial Economics*), select banking-related journals (*Journal of Banking and Finance*, *Journal of Monetary Economics*) and general interest economics journals (*American Economic Review*, various *American Economic Journals*, *Quarterly Journal of Economics*, and the *Review of Economics and Statistics*). This list does not contain unpublished working papers. It also excludes a host of published work on related topics including local spillovers of foreclosures (e.g., Biswas (2012), Campbell, Giglio and Pathak (2011), Chan, et. al. (2013), Cheung, Cunningham and Meltzer (2014), Harding, Rosenblatt and Yao (2009), Schuetz, Been and Ellen (2008), and Whitaker and Fitzpatrick (2013)) and macroeconomic effects of the housing boom and bust (e.g., Mian and Sufi (2011, 2014)).
[8] *Inside Mortgage Finance*, which previously was called *B&C Mortgage Finance*, claims to capture up to 85% of all subprime originations in most years. See Chomsisengphet & Pennington-Cross (2006) for more detail on this source.
[9] See Mayer, Pence and Sherlund (2009) for an excellent overview of this data source. Similar data providers listed in the table include LPS Applied Analytics (which includes data on prime and subprime mortgages), Black Box Logic LLC, the HMDA files (which cover prime and subprime borrowing), and OCC-OTS Mortgage Metrics (which also covers prime and subprime mortgages).

with a credit bureau panel. As Appendix Table 1 documents, researchers have merged these data with credit bureau files (from Equifax typically) that provides detailed credit risk information on the borrower. This combination provides the foundation for much of the empirical work listed in the table on the role of loan and borrower traits in explaining subprime default patterns in particular.[10]

Most of these studies also use data from a relatively short window of time (see column 5). The private label subprime mortgage-backed securities market did not really boom until late in the last housing cycle, so these data do not become nationally representative until the middle of the last decade. The vast majority of studies also do not use originations from later than 2008. It is rare for borrower outcomes to be tracked past that year, too, which is two years prior to the peaking of prime borrower distress rates.

Geographic coverage is varied. Many studies use multi-state samples, and some have made considerable effort to control for differences in economic (and housing price) conditions across metropolitan areas. However, there has not yet been an extensive effort to control for detailed location effects within a metropolitan area.

While mostly limited to the subprime sector and the very beginning of the housing bust, these data have been useful in examining a host of interesting topics. Appendix Table 1 highlights that there have been a number of studies on special topics or features of the subprime market such as the impact of securitization on the all-on costs of origination, on lender screening incentives, and on incentives to renegotiate or work out distressed loans. There also have been separate studies on predatory lending and the prevalence and impact of untruthful data (e.g., 'liar loans'). This earlier research generally could not address the potential role of common factors across the prime and subprime sectors. That requires panels of complete ownership sequences combined with detailed financing information. It is to the creation of such data that we now turn.

III. Data Description

The home purchase and financing transactions files compiled by the data vendor DataQuick are the foundation of the rich micro data used in this paper. They permit us to

[10] The closest anybody comes to our goal of analyzing distress in the broader mortgage or housing market is Ascheberg, et. al. (2013). It notes the rise borrower defaults as the housing bust unfolded, but takes a very different research approach from ours. Those authors construct a dynamic simulation model to impute the spillover effects of subprime defaults on prime defaults.

observe sales transactions of single family units and homes in condominium or multi-unit structures. We also observe the financing associated with those purchases, as well as subsequent refinancings and subordinate mortgages. Our sample includes this information for the 96 metropolitan areas listed in the Appendix Table 2, along with their start and end dates. As the appendix table notes, different metropolitan areas enter the sample at different times, some as early at 1993(1), so homes purchased before these dates do not enter our study sample (unless they are resold later). We report results based on data from 1997(1)-2012(3) since we have coverage on virtually all MSAs over this time span.[11]

Detailed information is provided on the following variables (among others): (a) transaction date; (b) name of the buyer if the observation is for a purchase; name of the owner if the observation is for a refinancing or other debt; (c) name of the seller if the observation is for a home purchase; (d) names of up to three lenders for any type of transaction involving new debt; (e) sales price for all home purchases; (f) mortgage amounts for up to three loans on all observations using any type of financing; (g) street address and census tract of the underlying home; (f) various home characteristics including age of the home, size as reflected in the number of bedrooms, bathrooms, and square footage, etc.; and (g) codings provided by DataQuick indicating whether a transaction involves a home being foreclosed by a creditor, as well as whether the home is being sold out of foreclosure to a new owner; in both cases, names of the principals are reported, along with a purchase price for the latter type of transaction.

Because individual owners and all their financings can be tracked over time, we use these data to create a panel of individual ownership sequences. An ownership sequence is the complete span of time a unique owner owns a given residence. Our final panel contains 33,545,252 ownership sequences on 19,648,475 homes. There are just fewer than 780,000,000 quarterly observations on these ownership sequences from 1997(1)-on.

A. The Number and Types of Transactions

The predominant type of transaction is an arms-length purchase of an existing home. These constitute 80.2% of all our home sales transactions. Arms-length sales of new homes

[11] In practice, we have an unbalanced panel since more houses that are newly built or resold enter the sample over time. Fortunately, the impact on our panel estimates is not likely to be major because the foreclosure crisis starts in 2006-2007 when sample sizes have stabilized. See Table 1 for more detail.

from the builder (or other entity) to a household make up another 11.2% of all purchases.[12] The remaining sales observations are comprised of purchases out of foreclosure (8.6%). DataQuick does not code these as arms-length trades between two disinterested parties, but they are readily identifiable from another variable categorizing 'distress' transactions.[13]

We also observe about 48 million financings not associated with a home purchase. These include refinancings and the taking on of junior debt. First, second and third loans at purchase are clearly identified. However, DataQuick does not identify whether a subsequent financing within a unique ownership sequence represents a refinancing of existing debt or the taking on of an additional loan. We adopt the following rule to distinguish between the two cases. If a new mortgage taken out subsequent to purchase has an initial loan balance that is more than 50% of the total mortgage balance taken out at purchase or is more than 50% of the imputed current price of the home, we assume the new loan is a refinancing that replaces the prior debt; otherwise, it represents junior debt, which is added to the outstanding loan balance. Using this rule, we observe about 34 million refinancings and just over 14 million second loans.

B. Classifying Owners

Each ownership sequence is classified as one of five types based on the type of financing used by the owner. The most straightforward is those who buy their housing unit without using any debt. These are referred to as Cash owners in all tables and figures. They constitute a relatively stable 10%-11% share of our sample until 2010, after which their share increases to over 16% in 2012 (Table 1). If an owner purchases a house with no debt, but subsequently takes out a mortgage, that owner is no longer considered a Cash owner as of the quarter of the loan origination.

All other ownership sequences involve the use of some type of debt. We divide each of these owners into one of four groups of borrowers: (a) Prime; (b) Subprime; (c) FHA/VA-insured; or (d) 'Small'. Lender lists are used to define subprime mortgages because we do not have access to credit score micro data. More specifically, we define a borrower as subprime if it

[12] We can confirm new home sales by analyzing another variable identifying the year the home was built, as well as the name of the seller. The former allows us to exclude land sales (which occur prior to the time the structure was built). For new homes, the seller usually is a home builder.

[13] The seller in these cases typically is some type of financial entity, while the buyer usually is a household. See the discussion below for more on these transactions. Some do not consider these 'normal' sales, but they certainly are home purchases, and we count them as such. Their transaction prices also are included in the price series described below, although we can do all our analysis excluding them.

obtained its loan(s) from a lender on either the HUD or *Inside Mortgage Finance* lists, but the loan was not insured by FHA or VA. This group is called Subprime in all tables and figures.[14] As Figure 2 above and the data reported below show, our Subprime group has very high rates of home loss, which is consistent with the rest of the literature regardless of their data and procedure for distinguishing subprime from prime.

However, we do not categorize all other borrowers as Prime. Two other categories are included to help ensure we do not conflate subprime and prime owners. The first is comprised of borrowers whose loans were guaranteed by FHA or VA (regardless of lender identity). They are labeled FHA/VA owners in all tables and figures.[15] These loans often are considered of subprime quality because of the very high initial loan-to-value ratios usually involved, but we treat them separately from the 'private' subprime group. As shown above, the time series on their shares in our panel almost are the mirror images of one another.

We also distinguish another category of owners who were financed by individuals, households, or firms that issued less than 100 loans throughout our sample period. Our reasoning is that those owners who obtain financing from individuals or other entities that do not appear to be traditional banks and financial institutions could be riskier, and thus more subprime-like. We label them as 'Small' owners because they obtained their debt from entities that issued a very small number of loans.[16] Their temporal pattern of foreclosures/short sales is much more like that for Subprime than for Prime as shown above in Figure 2. That said, they always constitute a small share of our sample, never amounting to more than 2%-3% of all observations in any one year.

[14] The entities on the subprime lender list generally distinguish among the several units of a lender. For the HUD list in particular, identification was based on the HMDA identification number of the entity, and different subsidiaries of a large bank typically had different ID numbers. Thus, having a subsidiary of (say) Bank of America that HUD believes specializes in subprime lending on the list does not mean that all of Bank of America's mortgage issuance gets classified as subprime. Banks and subsidiaries also enter and leave the HUD list over time. The HUD list also ends in 2005. The *Inside Mortgage Finance* publication also lists specific units of some large financial institutions, but we also consider those units as subprime if they ever show up on that publication's list.

[15] Ten metropolitan areas in the northeastern part of the country do not report data for this particular variable. They are Barnstable Town, MA, Boston-Cambridge-Quincy, MA-NH, Bridgeport-Stamford-Norwalk, CT, Harford-West Hartford-East Hartford, CT, New Haven-Milford, CT, Pittsfield, MA, Providence-New Bedford-Fall River, RI-MA, Springfield, MA, and Worcester, MA. We still include observations from these metropolitan areas in our regression analysis, but code this variable for them so that it is estimated separately from that for the other MSAs. Hence, coefficient estimates for this group of borrowers reported below in Table 3 are based on the 86 MSAs that have full information on FHA/VA loan status.

[16] That said, some of these small lenders could arise from measurement error in the way DataQuick reports the names of lenders in the data.

All remaining owners with debt are Prime borrowers by definition. Their share always exceeds 50%, and it rose, not fell, as the boom built, from a low of 54.9% in 2000(1) to a high of 65.6% in 2008(1). Thus, the rough doubling of Subprime share over the same period is at the expense of the FHA/VA-insured sector, not the Prime sector (Table 1).

C. Distress: Losing One's Home Via Foreclosure or Short Sale

We define distress as the home being lost to foreclosure or short sale. Foreclosed homes are explicitly identified in the DataQuick files by a distress code that indicates the exact date when the home was lost by the previous owner. We are able to confirm this by looking at the name of the new owner, which typically is some type of financial entity (e.g., bank, RMBS pool number, special servicer), not a household. We define a short sale as a transaction in which the sales price is no more than 90% of the outstanding balance on all existing debt. DataQuick has a variable that indicates when a sale is considered a short-sale, but our conversations with the company revealed that such information is based on a proprietary model. Our proxy for short-sales matches the DataQuick indicator 90% of the time. We use our version of short-sales because the DataQuick variable is only populated since 2004. Of the 2.971 million cases of owners losing their homes depicted in Figure 1, two-thirds were due to foreclosure (2.071 million) with the rest (0.899 million being short sales). We report results below using only foreclosures that yield very similar findings.

D. Constant Quality House Prices

Constant quality nominal house price series are used throughout our analysis. We use hedonic price indexes, rather than repeat sales indexes popularized by Case and Shiller (1987, 1989) because of their much less onerous data requirements.[17] This is relevant because we create semi-annual price indexes for groups of census tracts that are intended to proxy for neighborhoods within a metropolitan area.[18] There is significant variation in price growth over time across tract groups, and that heterogeneity is exploited when creating loan-to-value ratios for individual ownership panel sequences.

[17] That said, at the metropolitan area level, the correlation between our hedonic price indexes and repeat sales indexes typically is higher than 0.95.
[18] Because there are few home sales within a given tract in any period, we aggregate tracts into groups of 4-6, with the average being 4.5 tracts per group. The grouping is done to make the tracts as contiguous as possible.

We do not include observations for which the reported sales price is less than $10,000 or greater than $5 million. Nominal price in logarithmic form for units in each neighborhood is modeled as a function of the square footage of the home entered in quadratic form, the number of bedrooms, the number of bathrooms, and the age of the home. We also include a dummy for condominiums or houses located in subdivisions and interact these dummies with the linear and quadratic terms for square footage. The hedonic index values are derived from the coefficients on the semi-annual dummies included in the model – the actual equation and additional details are available upon request. The estimated indexes are then normalized to 100 in 2000(S1) for all neighborhoods.

Figure 3 reports the graphs of the neighborhood-level semi-annual hedonic price series for four of our 96 metropolitan areas: Boston-Cambridge-Quincy, Las Vegas-Paradise, Phoenix-Mesa-Scottsdale, and San Francisco-Oakland-Fremont. Tract groups in a given metropolitan area tend to move together over time, but there is considerable variation both on the upside and downside of the cycle. For example, the mean nominal price appreciation from 2000(1) to the cyclical peak across the 154 tract groups for which we estimated neighborhood-level hedonic price indexes in the Boston metropolitan area was 87.1%, with a standard deviation of 24.3 percentage points. The lowest neighborhood-level price growth to the peak was 48.6%, compared to 167.2% for the highest. There is less spatial heterogeneity in the bust, where the mean price decline from the peak to trough was 69.1%, with the spread from lowest to highest decline only 20 percentage points (from -61.1% to -82.6%). Naturally, this means that some submarkets in the Boston metro performed materially better than others since 2000(1). The mean price growth since 2000(1) was 45.8%, with the full range across neighborhoods running from 1.5% to 92.4%. This type of spatial heterogeneity is typical across tract groups within a given metropolitan area (larger ones in particular).

E. Leverage at Purchase and Over Time

Loan and purchase price data are combined to compute loan-to-value (LTV) ratios. Doing so at purchase is straightforward: divide the sum of all mortgages taken out at purchase by the purchase price recorded by DataQuick. Figure 4 shows how initial LTV varies over time by the different types of borrowers/owners who used debt. FHA/VA-insured loans have much higher initial LTVs (close to 1) than both prime and subprime loans throughout our full sample period, and actually fell slightly over our sample period. Subprime borrower average initial LTVs did increase from about 81% to 85% as the boom built in the mid-2000s. There is a more modest increase in Prime borrower initial LTVs over the same time period. Thus, there was not a dramatic surge in initial leverage ratios for the typical borrower in any sector of the mortgage market while the long boom in house prices built.

Current LTV by quarter must be estimated. Fortunately, in addition to having panels of ownership sequences that make its estimation feasible, two features of our data allow for a more accurate estimation than exists in other research: (a) the complete history of home financings, including refinancings and second loans; and (b) neighborhood-level house price indexes.[19] In imputing the numerator, we presume that all new debt taken on is fully amortizing, 30-year, fixed rate product. This is a conservative assumption that almost certainly leads us to understate true LTV, particularly on subprime product which the literature suggests more often involved adjustable rate mortgages (ARMs) and terms that did not require immediate amortization of principal. To impute current house value in the denominator, we start with house price at purchase, and update it on a half-year basis using our neighborhood-level price indexes. Noise in the denominator can arise in different ways. For example, values for distressed properties are likely to be overstated because they probably were receiving lesser maintenance and repair-related investment. This provides another reason why current LTV could be underestimated. However, we suspect that variation provided by refinancings, second loans and the local price index likely overshadow the measurement error due to this factor.

Leverage ratios at purchase may not have spiked during the run-up in prices as the boom built, but Figure 5 shows that average current LTV steadily declined by about 20 percentage points from 1997 until 2005-2006 near the peak of the housing price cycle. This fall in leverage, which is due to the extraordinary rise in house prices during the long boom, occurred across all borrower types. This pattern then reversed itself by the end of 2006, after which house prices fell dramatically and current LTVs increased rapidly to unprecedentedly high levels by 2009. The average current LTV for prime borrowers was just above 1.0 in the first quarter of 2009, while that for Subprime and FHA/VA borrowers was above 1.2. The average Prime owner

[19] Some private data vendors have begun creating cumulative LTVs on observations in their loan-level data sets. Essentially, they do it as we do, by linking to deeds records (which is what DataQuick does) so they can track a given observation over time. To our knowledge, this has not yet shown up in current or published research.

continued to have no equity in its home for the three following years, while that for all other owner types with debt remained in negative equity.

F. Identifying Speculators

Researchers and popular commentators have argued that speculators may have played an important role in the building of the last housing boom, thereby helping make its ultimate demise worse (e.g., Haughwout, et. al. (2011); Chinco and Mayer (2014)). We identify speculators in one of two ways. First, we follow Chinco and Mayer (2014) who reasoned that since speculators would not be living in the purchased unit, they would have their tax bills sent to another address. We compare the precise street address of the housing unit with the address to which the tax bill is sent – the 'Tax Address' in the DataQuick files. Whenever the two are appreciably different, we call that purchaser a speculator.[20] The second way we identify whether a purchaser is a speculator is by whether the buyer has a name that is a business. This includes corporate or commercial names that include LLC or INC in them, homebuilders, or trusts (especially mortgage-backed securities trusts which are typically identified by a four-digit number in their names).[21] Appendix Figure 1 shows that the share of speculators by type of borrower increases for all categories until 2002, but then remains stable for Prime, Subprime, and FHA/VA borrower/owners, while it keeps escalating for Cash owners and Small borrowers.

G. Demographics and Income of Borrowers

A weakness of the DataQuick files is that they do not contain any information on the owners beyond their names. To gain more insight into borrower demographic characteristics (race and gender of the head of the household) and the self-reported income levels, we match individual sales transactions to loan application data in the Home Mortgage Disclosure Act (HMDA) files. Observations are merged as follows. In the first step, each transaction was matched to a loan using the year in which the transaction occurred, the full 11 digit Census tract number, the lender name, and the exact loan amount. In cases where there were multiple matches, one of them was randomly assigned as being a true match while the rest were

[20] By appreciably different, we generally mean that more than one number in the street address before the zip code differs.
[21] Other academic research has identified speculators by whether the 'flip' properties quickly (e.g., Bayer, Geissler, Magnum and Roberts (2011)). We also investigated those cases, but more than 99% of them were already encompassed by our measures of tax address and names of business.

considered unmatched. The remaining unmatched observations were then merged based only on year, Census tract and exact loan amount with multiple matches being randomly assigned as in the first step. This two-step process was repeated several times allowing for the loan amounts to differ from each other in increments of $1,000 up to a total allowable difference of $10,000. Any observations remaining after this process then went through an identical matching procedure using 9 digit Census tract numbers. Observations surviving that procedure are considered to be unmatched. In total, 92.7% of the sales transactions in DataQuick were matched at some point in the procedure. Of those, approximately 60% were matched in the first step. Because we are unsure about the quality of the matches in subsequent steps, in the empirical work below we always distinguish the demographics in two groups – perfect and imperfect matches – and include both in the estimation. Reported regression coefficients are for the perfect matches only. Finally, the demographic data for Cash buyers is missing by definition because they never took out a loan, and hence, cannot be matched with any HMDA observation.

H. Summary Statistics

Table 2 reports summary statistics for a number of variables of interest in our empirical work for the full sample and also for each of the five owner categories, with their shares in our overall sample in parentheses: Prime (61%), Subprime (15%), FHA/VA (10%), Cash (11%), and Small (2%). The top panel notes distress rates. Foreclosures always are at least twice as prevalent as short sales in leading to the loss of a home. And, the unconditional probability of a Subprime owner losing its home is well more than double the rate a Prime owner does. The sample-wide mean is 0.73% for Subprime owners (1-in-137) versus only 0.34% (1-in-294) for Prime owners, but Figure 2 above shows that masks substantial heterogeneity over time. The group of owners that borrowed from 'small' lenders has a distress rate of 0.42%, above that for Prime owners, but well below that for Subprime owners. The same is true of borrowers using FHA/VA-insured mortgages, who have a distress rate of 0.38%. Not surprisingly, all cash owners have the lowest rate of home loss at 0.14% or 1-in-714. By definition, these owners cannot lose their homes to lenders. Examination of the names of the parties taking over these homes in foreclosure indicates that it is property taxes that are not being paid for the most part, as a local taxing authority or municipality often takes ownership.

Panel 2 reports data on housing traits. There is no evidence here that Subprime owners purchased systematically smaller or appreciably older units. Demographics are reported in Panel 3.[22] There are only modest differences in the fraction of male borrowers (defined as the head of household) across the four borrower types. Racial differences are quite large. The White share of Subprime owners is 10 percentage points below that of Prime owners, at 63% versus 73%. The FHA/VA-insured category has a similarly low White share. Reported income on the loan application recorded in the HMDA files shows Subprime owners to be less rich than Prime owners, but the difference is only 6%. There were many claims about misstatements on 'liar loans' in the subprime sector, so this modest difference may reflect some of that misreporting. Note that FHA/VA-insured borrowers, who are required to provide documentation about their earnings, have appreciably lower incomes on average. Panel 3 also reports aggregate data on speculators. One-quarter of all ownership sequences are classified as speculators, and they are more prevalent among the all Cash owner group at 61%.

The fourth panel reports summary statistics on house prices and LTV at the time of purchase. Transaction prices are higher and similar for both prime and subprime borrowers, and smaller for owners who took FHA/VA-insured loans. However, subprime owners are more likely to use high leverage than prime owners at purchase. But by no means are Subprime owners the most leveraged owners in our sample. Borrowers of FHA/VA loans tend to use much more debt at purchase. Nearly two-thirds (64%) of them have 95%-100% LTVs at origination, and another 23% have no equity at purchase.[23]

The fifth panel of Table 2 breaks down the financings by whether they are for a purchase or any subsequent financing. For the overall sample, 51% of all ownership sequences never altered the debt they took on at purchase (see column 1). Another one-third (34%) refinanced and another 15% took out a second mortgage. Across owner types, the Subprime group was the least likely to not refinance or take on a second mortgage. Finally, the last panel shows averages for current LTV, which has been discussed above and in Figure 5.

---

[22] Statistics are reported here for perfect matches only.

[23] This does not appear to reflect data error. FHA rules allow home buyers to borrow the upfront fee that FHA charges for guaranteeing the mortgage and add it to the mortgage balance. That, plus other loan sources, leads to over one-fifth of FHA/VA-insured borrowers having 100% (or slightly higher) loan-to-value ratios at origination.

IV. Empirical Results

A. Panel Structure Estimates: Negative Equity and Borrower Illiquidity

We estimate various panel specifications, where the dependent variable is a binary distress outcome y_{ht} which equals one if the owner lost the home via foreclosure or short sale that quarter and equals zero otherwise, and it is indexed by housing ownership sequence *h* and year-quarter *t*. We always control for the different type of borrowers/owners (O_k): Prime, Subprime, FHA/VA, Small and Cash, with Cash being the omitted category in all reported results. A given owner is classified as one (and only one) of these types each quarter, with *k* indexing the five ownership types.[24]

We first investigate the importance of the two key factors suggested by the traditional home mortgage default literature: negative equity and borrower illiquidity. Negative equity is directly measured for each housing unit each quarter by the current LTV variable discussed in the previous section. Current LTV enters the model split into two dozen intervals (denoted by the vector D_{ht}), typically of 10 percentage points in terms of leverage (i.e., from 30-39% LTV, 40-49%, etc.) in order to see whether entering negative equity status or being more and more underwater increases the probability of losing one's home. Because there is no large micro data source that tracks homeowner-level employment status, we proxy for the presence of a negative income shock that would render a household illiquid by including census tract-by-quarter fixed effects (denoted S_{tn} in the model below) that allow us to capture very local (but still not household-specific) economic conditions in each quarter *t*. There are nearly 1.6 million of these dummy variables. Given the extremely large number of observations, this does not pose a statistical problem. Rather, the primary challenge is computational, which is why we estimate linear probability models of the following type:

$$(1)\ y_{ht} = \alpha_k O_{htk} + S_{tn} + \rho D_{ht} + \mu_{ht}\ ,$$

where μ_{ht} is the standard error term. Estimates are clustered at the census tract level.

---

[24] Classification is straightforward for any quarter in which an owner has no more than one mortgage. If there are multiple loans and at least two have different classifications, we use the following order to determine owner type: Subprime, FHA/VA, Prime, and Small. Thus, if the owner has at least one Subprime loan and any number of other types of loans, it is classified as Subprime. If it does not have any subprime loans, but has at least one FHA/VA-insured loan and any other loans, it is classified as FHA/VA, and so forth.

Table 3 reports results. The first column only includes the financing type dummies with no other covariates. Recall that Cash owners always are the omitted category, so that the estimated coefficients must be interpreted as relative to the 0.14% home loss rate for that group of owners (see the means in the top panel of Table 2). These baseline results confirm that Prime borrowers lose their homes at less than half the rate at which Subprime borrowers do. Subprime loss rates are larger than the other owner categories, too, which tend to be more similar to that for Prime owners. Hence, those we identify as Subprime borrowers certainly do look riskier unconditionally than the other groups of owners.

Including the census tract-by-quarter fixed effects vector in the second specification means that Prime versus Subprime loss rates are now being made for ownership sequences in the same neighborhood that faced similar average local economic shocks. Those results show increases in the Prime and Subprime coefficients of roughly the same magnitude. This means the Prime – Subprime gap in home loss rates largely is unaffected by these controls.[25] This suggests that Subprime and Prime owners are dispersed across tracts within a metropolitan area rather than spatially concentrated in a select few. Hence, we can rule out a dense spatial concentration of subprime loans in select neighborhoods that experienced some type of strong negative economic shock as explaining why subprime distress rates were so high.[26]

The results in the next column (#3) show that current LTV is much more influential. The point estimates on each ownership category fall substantially. All but the Subprime borrowers are no longer appreciably more at risk of losing their homes via foreclosure or short sale than are all Cash owners. The Prime owners' coefficient is negative (as is that for FHA/VA borrowers), indicating that once current LTV is controlled for, they are less likely to lose their homes than all Cash owners. The difference is quite small in absolute terms, but it is not unexpected or illogical. Prime borrowers should be very good credit risks, and they do not include as many speculators as the Cash owner group does. And, the 'distress gap' between Subprime and all Cash owners also falls by about three-quarters.

[25] The Prime/Subprime gap narrows by only 4%.

[26] The first column of Appendix Table 3 supports this assertion with its listing of the average share of tracts within a MSA that never had a Prime owner (top panel) or a Subprime owner (bottom panel) within its boundaries at any point in our sample period. On average, virtually no tracts never had a Prime owner in them, while only 3.5% of tracts have never had a Subprime owner. The remaining three columns of that table report the share of tracts containing 25%, 50% and 75% of each owner type, respectively. Neither Prime nor Subprime owners are randomly dispersed geographically, but both are pretty widespread. For example, 45.3% of tracts contain 75% of Prime owners on average; the analogous figure for Subprime owners is 46.6%.

It is noteworthy that these results are robust to a couple of important alternative specifications. First, the findings are not much altered if we restrict the measure of home loss to foreclosures only. In that case, the coefficients on the Prime, Subprime, FHA/VA, and Small categories of borrowers/owners are -0.00088, 0.00121, -0.00126, and 0.00008, respectively. Thus, it is negative equity, not any distinction between foreclosures and short sales that is driving the findings. We also estimated specifications adding lagged values of Current LTV to the specification in column 3 of Table 3. In that case, the coefficients on Prime, Subprime, FHA/VA and Small become -0.00097, 0.00173, -0.00205, and -0.00078, respectively. The similarity in results is not surprising, as it is not obvious that LTV at the time of initial delinquency is more relevant than LTV at the time of foreclosure. Non-strategic defaulters only go through foreclosure if they cannot sell for more than the outstanding mortgage balance when they actually lose the home. Even strategic defaulters that first miss a payment because they know they will give up the house ultimately might care more about LTV in the future.

Figure 6 plots each individual current LTV interval estimate for specification 3 of Table 3. As expected, increasing leverage from very low levels has little or no impact on the probability of losing one's home until negative equity is approached. Loss rates continue to increase the deeper underwater the owner becomes. For example, the impact of being from 30-40% underwater (current LTV bin of 1.3-1.4) is ten times larger than that of being barely above water (current LTV bin of 0.9-1.0). At 1.1%, it is very large economically, too, given the low average loss rates reported in Table 2. Even so, the probability of losing the home doubles again by the time one has a current LTV of 1.8-1.9, and goes above 3% in absolute value for the most highly leveraged owners.[27] This indicates that a large fraction of the distress in residential real estate markets, regardless of the type of mortgage finance, was concentrated among borrowers living in homes that were underwater.[28]

This large average impact is not being driven solely by the subprime sector, as shown in Appendix Figure 2. This plot, which is based on a specification that interacts our current LTV variable with borrower/owner type shows that a given amount of leverage is associated with a higher probability of home loss for a Subprime versus a Prime borrower. However, the shapes of

[27] Unconditionally, 30% of all owners who ever experienced a current LTV>2.0 subsequently lost their homes. That is 1.67 times the 18% share among those who ever experienced negative equity.
[28] The plot in Figure 6 looks very similar when other covariates are added ranging from household features to other components of leverage discussed below. This suggests that our negative equity variable both is well measured and that its effects on foreclosures and short sales are unlikely to be due to omitted factors.

the plots are similar and the impact of being underwater on Prime borrowers is very large relative to the average probability of foreclosure or short sale in that sector of the market.

Figure 7 documents heterogeneity in the Prime and Subprime effects over time, and in doing so, highlights how powerful negative equity is in accounting for homeowner distress after the global financial crisis began. These plots are from augmented versions of the first three specifications reported in Table 3 that allow for the impact of owner type to vary by quarter and borrower/owner type. There are a number of noteworthy features of these results. First, controlling for negative equity via our current LTV variable completely accounts for the spike in Prime borrower home losses relative to all Cash owners that begins in 2007 (left panel). Second, negative equity also is influential in explaining Subprime sector foreclosures/short sales. It accounts for about one-half of the sharp early spike in home losses in that sector before 2008. Following that, Subprime sector foreclosures are about three-quarters lower once the owner's negative equity position is controlled for (right panel). Third, prior to the global financial crisis, borrowers in the Prime and Subprime sector were no more likely to lose their homes than all Cash owners once we know their current LTV ratio. This is not unexpected, as sound underwriting should lead to groups of non-speculator owner-occupiers who are good credit risks. What is striking is that this remains true for Prime borrowers throughout the housing bust period. Given that the vast majority of foreclosures occurred in that sector from 2009-on, this suggests the crisis was largely one of sound borrowers falling into negative equity because of very large declines in house prices. Robustness tests reported below will confirm that initial conditions such as purchase quarter LTV and loan cohort effects do not change this conclusion.

Figure 8 documents heterogeneity in the Prime and Subprime effects by market. Each mark represents one of the 96 metropolitan areas in our sample that are arrayed in ascending order starting from the lowest Prime or Subprime coefficient in each specification (thus, the order of markets can and does differ by specification plotted). As the plot in the left panel for Prime borrower/owners shows, controlling for current LTV eliminates the gap in home loss rates with respect to all Cash owners in virtually all MSAs. The top five markets, which still have very small coefficients ranging from 0.00058-0.00163, are Memphis, TN-MS-AR, and four Rustbelt markets in Ohio (Springfield, Columbus, Dayton, and Cleveland-Elyria-Mentor). The next five markets also are relatively small and in economic decline. The right panel shows that current leverage explains much, but not all, of the gap between Subprime borrower and all Cash owner foreclosures/short sales in the typical market, but there are bigger outliers at the top end of the distribution. They also are smaller and older industrial metropolitan areas in economic decline. Eight of the top ten Prime markets are also among the top ten Subprime markets.[29] Thus, at the local market level, relatively high foreclosure/short sale rates after controlling for current LTV are concentrated in declining areas where defaults and ultimately, home losses, are less dependent of individual negative equity conditions.

We also estimated MSA-level models with heterogeneity over time to help gain insight into what might account for the sharp spike in subprime defaults beginning in 2007 and 2008. We know from Figure 7 that negative equity and other variables can account for no more than one-half of that initial jump. It turns out that this surge is associated with a spatially concentrated group of markets in central California. For example, the top ten metropolitan areas in terms of loss rates among subprime borrowers/owners in 2006(3) are all Rust Belt or declining industrial areas, as discussed above. One year later in 2007(3), the Detroit and Cleveland markets still have the two highest home loss rates among subprime borrowers, but seven of the other eight markets are Stockton, CA, Modesto, CA, Merced, CA, Sacramento-Arden-Arcade-Roseville, CA, Yuba City-Marysville, CA, Vallejo-Fairfield, CA, and Riverside-San Bernardino-Ontario, CA. Six months later at the end of the first quarter of 2008(1), Detroit still remains, but only has the seventh highest subprime sector home loss rate, Stockton has the highest loss rate, and other central California markets of Bakersfield and Salinas have joined the top ten. It is not until the beginning of 2009 that we see the Las Vegas-Paradise, NV, Phoenix-Mesa-Scottsdale, AZ, and small Florida markets join the top ten list. Thus, the initial increase in the subprime sector distress was driven by an array of central California markets in a way that only partially can be accounted for by our measure of current LTV.

B. *Panel Structure Estimates: Other Potential Factors*

The popular press and much of the previous scholarly literature have also focused on other factors to explain the foreclosure crisis. Nonacademic commentators often wrote about homeowners stretching to buy bigger and better homes during the boom in a 'keeping up with

[29] The two in the Subprime top ten list not in the Prime top ten are Detroit-Warren-Livonia, MI, and Cincinnati-Middletown, OH-KY-IN. The two in Subprime list not in the Prime top ten are Baltimore-Towson, MD, and Yakima, WA. Hence, all 22 MSAs are industrial markets in economic decline.

the Jones's' mentality.[30] It is true that the size of the typical new home rose substantially during the boom[31], but typical unit size is similar across borrower types except for the FHA/VA group, who bought smaller homes on average (panel 2, Table 2). The findings reported in column 4 of Table 3 show that adding housing units traits including the square footage of living area (in quadratic form), the number of bedrooms and bathrooms, and age of the unit to the specification with census tract-by-quarter fixed effects does not change the coefficient on the borrower/owner category variables virtually at all, much less to the extent that adding current LTV did. Adding it to the third specification that also includes current LTV does not alter any of our aforementioned conclusions either. Nor do these variables have an economically large independent impact on the probability of home loss.

Next we look at household traits, which include the race and gender of the owner, the self-reported initial income of the owner and our imputation for whether the owner is a speculator. These variables could impact foreclosures in a number of ways. Race, for example, could be important since minorities have a larger share of subprime mortgages relative to prime (panel 3, Table 2), and usually have less wealth than non-minorities (Bayer, Ferreira, Ross, forthcoming). Speculators could react faster to the first sign of negative equity and stop making monthly payments early in order to avoid future bigger losses. One quarter of our owners are categorized as speculators, but there is only a small difference in their share of the Prime and Subprime groups (panel 3, Table 2). Low self-reported income could indicate a lower likelihood to sustain mortgage payments in the future. These are all plausible mechanisms, but adding these household traits to the specification including census tract-by-quarter fixed effects is barely more impactful than adding housing unit traits was (column 5, Table 3). Thus, owner demographics, reported income and speculator status cannot account for differences in foreclosure/short sale outcomes across borrower/owner types, and they do not vitiate the influence of negative equity in explaining those differences.

This is not to say that factors such race do not matter at all. It does, but not in a way that can materially explain outcomes across borrower categories. For example, Whites do lose their

[30] One example is the Dr. Housing Bubble Blog which wrote frequently on this issue as far back as 2006 (http://www.doctorhousingbubble.com/category/keeping-up-with-the-joneses/). It is not hard to find examples in the mainstream media such as this New York Times's article "Housing Costs Rise. So Does Life on the Edge' from October 8, 2006 (http://query.nytimes.com/gst/fullpage.html?res=9C03E5D91530F93BA35753C1A9609C8B63).

[31] The median square footage of a new constructed single family home in the United States rose by nearly 11% from 2000 to 2007 according to U.S. Census data (see the chart "Median and Average Square Feet of Floor Area in New Single-Family Houses Completed by Location" at https://www.census.gov/const/C25Ann/sftotalmedavgsqft.pdf).

homes less frequently than Blacks as expected (*ceteris paribus*), but the absolute magnitudes of their impacts are relatively small and they are virtually uncorrelated with borrower type. [The racial/ethnic group with the highest rate of home loss is Hispanics.] Female heads are more likely to lose their homes than are male heads, but once again, this outcome is not strongly correlated with borrower type. The economic impact of self-reported initial income is quite modest in size, but this could be due at least partially to the variable being noisy. Speculators do lose their homes at slightly higher rates than non-speculators, but the coefficient is relatively small (0.00019) and does not change the relative impacts across borrower types.

The next three columns of Table 3 investigate whether other measures of leverage that influence current LTV can explain the foreclosure probabilities. The first component is initial LTV, which mechanically corresponds to the first current LTV observation in any ownership sequence. We look at this variable individually because recent work by Corbae and Quintin (2015) concludes that about 60% of the foreclosure crisis can be explained by higher initial LTV based on simulations of a macro model of housing markets they developed. This variable is transformed into five intervals for estimation purposes: 0.0-0.8, 0.8-0.9, 0.9-0.95, 0.95-1.0, and 1.0+. The regression results reported in column 6 show that initial leverage is more influential than the housing units and household trait vectors, but its impact is substantially less than that of current LTV for all but the FHA/VA-insured borrower groups. The extremely strong impact on this category of borrowers probably is due to their extremely high initial leverage being the salient fact about them. Controlling for initial leverage does account for over half the gap in the rate of home loss for Prime borrowers relative to all Cash owners (i.e., the relevant coefficient falls by 55% from 0.00329 in column 2 to 0.00147 in column 6), but that is far from fully accounting for the Prime – Cash gap, which current LTV does. Controlling for initial LTV yields a Subprime sector coefficient that is three times larger than when current LTV is controlled for (contrast column 3 versus column 6).

Column 7 then separately controls for whether refinancing a prior lien or taking on a second loan can account for foreclosure/short sale outcomes. Either change can directly contribute to variation in current LTV by discretely altering the mortgage balance during an ownership sequence. We know from Table 2 (columns 1 and 2, fifth panel) that nearly one-half of all ownership sequences in our sample contained a refinancing or second mortgage, and that this share was even higher among Subprime borrowers (columns 5 and 6, fifth panel). That this could prove important is suggested by Mian and Sufi's (2011) conclusion that home equity-based borrowing may explain one-third of mortgage defaults between 2006 and 2008. However, our findings imply no material economic role for this factor in accounting for foreclosure and short sales outcomes across different types of borrowers. In absolute terms, the loss rates from foreclosures or short sales for all types of borrowers are slightly higher, not lower, relative to all Cash owners if there is a refinancing or junior lien (compare with column 2). This could be due to positive selection in the sense that it was the better credit risks that were able to refinance and/or take on a second loan. Relatively speaking, the gap between outcomes for Prime versus Subprime borrowers also is changed only slightly, presumably because both types of borrowers refinanced a lot. In any event, there is no evidence that foreclosures or short sales can be accounted for by refinancing or second loans in general or in the subprime sector specifically.

The next column reports results for testing whether cohort dummies based on the quarter of the last purchase or mortgage transaction within an ownership sequence impact the probability of foreclosure or short sale. These cohorts affect current LTV because houses bought or refinanced near the peak of the housing boom had the largest declines in prices after the beginning of the recession, and therefore suffered the largest increases in current LTV. As shown above in Figure 3, these cohort effects may be quite relevant in MSAs where all neighborhood prices moved in sync with the rest of the market during the housing boom. Also, Bayer, Ferreira, and Ross (forthcoming) and Palmer (2014) report that cohort effects may explain some of the movements in defaults and prices respectively. Their impacts tend to be slightly less influential than those for initial LTV (compare column 8 with column 6), and thus are not a substitute for the impact of negative equity conditions as reflected in our current LTV control.

Column 9 includes all loan trait variables—initial LTV, whether there was a refinance or 2nd mortgage, and origination cohort. The results indicate that the latter two variables have some influence independent of initial LTV (compare to column 6), but the combination still is not as impactful as controlling for current LTV (column 3).[32]

[32] Computational constraints arising from estimating standard errors clustered at the census tract level required using a 50% random subsample (of each MSA) for this specification. The point estimates reported are not materially different from those arising from using the full sample, but not clustering. See the notes to Table 3.

Column 10 then includes current LTV and all other loan traits.[33] The pattern of results is similar to that for current LTV in column 3. Outcomes for Prime borrowers, who suffer the most losses of their homes to foreclosure and short sale, are virtually indistinguishable from those for all Cash owners. The gap between Subprime borrowers and all Cash owners remains positive, and is actually larger than when only current LTV is controlled for. As before, FHA/VA-insured borrowers do not lose their homes at higher rates once one controls for current or initial LTV, and small borrowers still are slightly more likely to lose their homes, but negative equity reduced that gap substantially. Finally, the estimated coefficients for the underlying negative equity controls are only 10-20% lower in this specification than in column 3, while the estimated coefficients for initial LTV are 50-70% lower when compared to the respective estimates in column 6.[34]

In sum, controlling for current LTV accounts for virtually all of the increase in foreclosures among Prime borrowers and a substantial fraction of the surge in Subprime home losses. Other variables that help predict current LTV are useful, but they reasonably can be interpreted as reducing the noise in our measure of negative equity at the individual homeowner level. As noted above, it is true that current LTV, which is measured at the household level, could proxy for micro-level borrower illiquidity conditions that our census tract-by-quarter fixed effects may not capture well for the reasons outlined in Gyourko and Tracy (2014). Much better, micro-level data on employment status, is necessary to provide a convincing test of the impact of desirability to pay (negative equity) versus ability to pay (unemployment). We did experiment with specifications that included ownership-specific fixed effects and quarter dummies (as opposed to the neighborhood-by-quarter fixed effects) to try to get around this problem, but these specifications also show a limited role for this type of permanent individual level factor.[35]

---

[33] Computational constraints also required estimating this specification on a smaller subsample (25%) – see the notes to Table 3.

[34] For example, current LTV bin [1.5,1.6) point estimates dropped from 0.017 to 0.014, and bin [2.4,2.5) point estimates dropped from 0.03 to 0.026. Meanwhile, initial LTV bins (0.9,0.95], (0.95,1.0], and (1.0,max] had drops from point estimates (0.0025, 0.0062, 0.0046) in column 6, to (0.0011, 0.0033, 0.0015) in column 10.

[35] Ownership fixed effects in principle deal with the permanent component of borrower illiquidity. In practice, the inclusion of ownership fixed effects did not change the impact of current LTV over time across borrower/owner groups – see Appendix Figure 3. Models with ownership fixed effects are identified by variation in households that switched mortgage type (e.g., from Prime to Subprime) within a given ownership sequence at the moment of a refinancing. That subsample is large, as nearly one-third of all our ownership sequences involved some type of switch. However, 'switchers' are obviously not a random subsample. Temporally, there is a tripling in the amount of switching as the boom built in the first half of the 2000s (which means virtually none of that subsample lost their homes during that period). Subprime owners switch at more than double the rate of Prime owners (62% versus

V. Conclusion

The housing bust and its consequences are among the defining economic events of the past quarter century. Constructing and analyzing new and very large micro data spanning the cycle and all sectors of the mortgage market leads us to reinterpret the ensuing foreclosure crisis as something much more than a subprime sector issue. Many more homes were lost by prime mortgage borrowers, and their loss rates not only increased relatively early in the crisis, but stayed high through 2012. This new characterization of the crisis motivates a very different empirical strategy from previous research on this topic. Rather than focus solely on the subprime sector and subprime traits, we turn to the traditional home mortgage default literature that explains outcomes in terms of common factors such as negative equity and borrower illiquidity.

The key empirical finding is that negative equity conditions can explain virtually all of the difference in foreclosure and short sale outcomes of Prime borrowers compared to all Cash owners. This is true on average, over time (including the spike in their foreclosure rate beginning in 2009), and across metropolitan areas. Given the predominance of this group in terms of foreclosures and short sales, this is tantamount to explaining the crisis itself. We can explain much, but not all, of the variation in Subprime borrower outcomes in terms of negative equity or borrower illiquidity conditions, so something potentially 'special' about the subprime sector still is unaccounted for. That said, it also could be that a less noisy measure of borrower illiquidity would be able to account for this residual variation. That remains for future research.

None of the other 'usual suspects' raised by previous research or public commentators change this conclusion. Housing quality traits, household demographics (race or gender), buyer income, and speculator status do not have a material influence on outcomes across borrower types. Certain loan-related attributes such as initial LTV, whether a refinancing occurred or a second mortgage was taken on, and loan cohort origination quarter do have some independent influence, but they are much weaker than that of current LTV.

While ours is not a normative economic analysis, our findings have potentially important implications for public policy. Regulatory issues are much more challenging when the economic

---

27%) One also can envision two-sided selection leading to higher distress rates in the sample of Subprime switchers. That is, households with declining credit quality could be switching from prime to subprime mortgages. Nonetheless, estimates from Appendix Figure 3 make any detailed discussion about the potential pitfalls of the ownership fixed effects somewhat moot, as the specification including them does not outperform that with current LTV and census tract-by-quarter fixed effects.

cycle itself plays a large role. That is the implication of our finding that large numbers of Prime borrowers who did not start out with extremely high LTVs still lost their homes to foreclosure. In that context, effective regulation is not just a matter of restricting certain exotic subprime contracts associated with extremely high default rates. We do not have detailed loan trait data, but it turns out that we do not need it to account for much of the difference in the propensity to lose one's home across prime and subprime mortgage borrowers.

Our findings also can help inform homeowner bailout policy. We are not able to provide a definitive recommendation one way or another, but we can rule out one noteworthy reason offered for not aiding homeowners—namely, that the crisis was mostly about irresponsible subprime sector actors (both lenders and borrowers) who were undeserving of transfers. Of course, this is not to say that there was no such behavior. The evidence from other research and serious journalists is that there was. However, it is clear from the passage of time (and the accumulation and analysis of new data that provides) that the problem was much more widespread and systemic. That is the meaning of a common factor playing such an influential role in determining foreclosure losses across all types of borrowers. That knowledge may or may not have affected policy makers' and the public's perspectives on bailouts. What we do know is that significant distress in the housing market which dramatically weakened household sector balance sheets had very large negative macroeconomic effects (Mian and Sufi (2014)).

In terms of research needed to make progress in understanding this past housing bust, and perhaps more importantly, the next one to come, there is one area in urgent need of more work: combining micro-level labor market data with housing data. That will allow for stronger tests of the impact of borrower illiquidity on defaults and foreclosures. This likely will take much effort and a change in policy among government data collectors, but it is a useful goal.

References

Adelino, Manuel, Kristopher Gerardi, and Paul Willen. "Why Don't Lenders Renegotiate More Home Mortgages? Redefaults, Self-Cures, and Securitization", *Journal of Monetary Economics*, Vol. 60, no. 7 (2013): 835-853.

Adelino, Manuel, Antoinette Schoar, and Felipe Severino. "Changes in Buyer Composition and Expansion of Credit During the Boom", National Bureau of Economic Research Working Paper w20848, January 2015.

Agarwal, Sumit, Gene Amromin, Itzhak Ben-David, Souphala Chomsisengphet, and Douglas Evanoff. "The Role of Securitization in Mortgage Renegotiation", *Journal of Financial Economics*, Vol. 102, no. 3 (2011): 559-578.

_______. "Predatory Lending and the Subprime Crisis", *Journal of Financial Economics*, Vol. 113, no. 1 (2014): 29-52.

Aschberg, Marius, Robert Jarrow, Holger Kraft, and Yildiray Yildirim, "Government Policies, Residential Mortgage Defaults and the Boom and Bust Cycle of Housing Prices", *Real Estate Economics*, forthcoming (2013 web version available).

Bayer, Patrick, Fernando Ferreira, and Stephen Ross, "The Vulnerability of Minority Homeowners in the Housing Boom and Bust", *American Economic Journal: Economic Policy*, forthcoming.

Bayer, Patrick, Christopher Geissler, Kyle Magnum and James Roberts. "Speculators and Middlemen: The Strategy and Performance of Investors in the Housing Market", NBER Working Paper 16784, February 2011.

Ben-David, Itzhak. "Financial Constraints and Inflated Home Prices During the Real Estate Boom", *American Economic Journal: Applied Economics*, Vol. 3, no. 3 (2011): 55-87.

Bhardwaj, Geetesh and Rajdeep Sengupta. "Subprime Cohorts and Loan Performance", *Journal of Banking & Finance*, Vol. 41 (2014): 236-252.

Bhutta, Neil, and Ben Keys. "Interest Rates and Equity Extraction During the Housing Boom", mimeo, University of Chicago, 2013.

Biswas, Arnab. "Housing Submarkets and the Impacts of Foreclosures on Property Prices", *Journal of Housing Economics*, Vol. 21, no. 3 (2012): 235-245.

Bubb, Ryan and Alex Kaufman. "Securitization and Moral Hazard: Evidence from Credit Score Cutoff Rules", *Journal of Monetary Economics*, Vol. 63 (2014): 1-18.

Campbell, John, Stefano Giglio, and Parag Pathak. "Forced Sales and House Prices", *American*

Economic Review, Vol. 101, no. 5 (2011): 2108-2131.

Capozza, Dennis and Robert Van Order. "The Great Surge in Mortgage Defaults 2006-2009: The Comparative Roles of Economic Conditions, Underwriting, and Moral Hazard", *Journal of Housing Economics*, Vol. 20, no. 2 (2011): 141-151.

Case, Karl and Robert Shiller. "Prices of Single Family Homes Since 1970: New Indexes for Four Cities", *New England Economic Review*, September/October (1987): 45-56.

________________________. "The Efficiency of the Market for Single Family Homes", *American Economic Review*, Vol. 79, no. 1 (1989): 125-37.

Case, Karl, Robert Shiller and Anne Thompson. "What Have They Been Thinking? Homebuyer Behavior in the Hot and Cold Markets", *Brookings Papers on Economic Activity*, Fall (2012): 265-298.

Chan, Sewin, Michael Gedal, Vicki Been, and Andrew Haughwout. "The Role of Neighborhood Characteristics in Mortgage Default Risk: Evidence from New York City", *Journal of Housing Economics*, Vol. 22, no. 2 (2013): 100-118.

Cheung, Ron, Chris Cunningham, and Rachel Meltzer. "Do Homeowners Associations Mitigate or Aggravate Negative Spillovers from Neighboring Homeowner Associations?", *Journal of Housing Economics*, Vol. 24, no. 1 (2014): 75-88.

Chinco, Alex and Christopher Mayer. "Misinformed Speculators and Mispricing in the Housing Market", NBER Working Paper 19817, January 2014.

Chomsisengphet, Souphala and Anthony Pennington-Cross. "Evolution of the Subprime Mortgage Market", *Federal Reserve Bank of St. Louis Review*, Vol. 88, no. 1 (2006): 31-56.

Coleman IV, Major, Michael LaCour-Little, and Kerry Vandell. "Subprime Lending and the Housing Bubble: Tail Wags Dog?", *Journal of Housing Economics*, Vol. 17, no. 4 (2008): 272-290.

Corbae, Dean, and Erwan Quintin, "Leverage and the Foreclosure Crisis", *Journal of Political Economy*, Vol.. 123, n.1 (2015): 1-65.

DeFusco, Anthony, Wenjie Deng, Fernando Ferreira and Joseph Gyourko. "The Role of Contagion in the Last American Housing Cycle", Zell/Lurie Real Estate Center at Wharton Working Paper, 2014.

Demyanyk, Yuliya and Otto Van Hemert. "Understanding the Subprime Mortgage Crisis", *Review of Financial Studies*, Vol. 24, no. 6 (2011): 1848-1880.

Deng, Yongheng, John Quigley, and Robert Van Order. "Mortgage Terminations, Heterogeneity, and the Exercise of Mortgage Options", *Econometrica*, Vol. 68, no. 2 (2000): 275-307.

Elul, Ronel, Nicholas Souleles, Souphala Chomsisengphet, Dennis Glennon, and Robert Hunt. "What "Triggers" Mortgage Default?", *American Economic Review*, Vol. 100, no. 2 (2010): 490-494.

Ferreira, Fernando and Gyourko, Joseph. "Anatomy of the Beginning of the Housing Boom: U.S. Neighborhoods and Metropolitan Areas, 1993-2009", Zell/Lurie Real Estate Center at Wharton Working Paper, March 2013.

Foote, Christopher, Kristopher Gerardi, Lorenz Goette, and Paul Willen. "Just the Facts: An Initial Analysis of Subprime's Role in the Housing Crisis", *Journal of Housing Economics*, Vol. 17, no. 4 (2008): 291-305.

____. "Reducing Foreclosures: No Easy Answers", *NBER Macroeconomics Annual 2009*. University of Chicago Press, April 2010.

Foote, Christopher, Kristopher Gerardi, and Paul Willen. "Negative Equity and Foreclosure: Theory and Evidence", *Journal of Urban Economics*, Vol. 64, no. 2 (2008): 234-245.

Foster, Chester and Robert Van Order. "An Option-based Model of Mortgage Default", *Housing Finance Review*, Vol. 3, no. 4 (1984): 351-372.

Goodman, Allen and Brent Smith. "Residential Mortgage Default: Theory Works and So Does Policy", *Journal of Housing Economics*, Vol. 19, no. 4 (2010): 280-294.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. "The Determinants of Attitudes toward Strategic Default on Mortgages", *The Journal of Finance*, Vol. 68, no. 4 (2013): 1473-1515.

Gyourko, Joseph and Joseph Tracy. "Reconciling Theory and Empirics on the Role of Unemployment in Mortgage Default", Journal of Urban Economics, Vol. 80, no. 1 (2014): 87-96.

Harding, John, Eric Rosenblatt, and Vincent Yao. "The Contagion Effect of Foreclosed Properties", *Journal of Urban Economics*, Vol. 66, no. 3 (2009): 164-178.

Haughwout, Andrew, Donghoon Lee, Joseph Tracy and Wilbert van der Klaauw, "Real Estate Investors, the Leverage Cycle and the Housing Market Crisis", Federal Reserve Bank of New York Staff Report 514, 2011.

Haughwout, Andrew, Richard Peach, and Joseph Tracy. "Juvenile Delinquent Mortgages: Bad Credit or Bad Economy?", *Journal of Urban Economics*, Vol. 64, no. 2 (2008): 246-257.

Jiang, Wei, Ashlyn Aiko Nelson, and Edward Vytlacil. "Liar's Loan? Effects of Origination

Channel and Information Falsification on Mortgage Delinquency", *The Review of Economics and Statistics*, Vol. 96, no. 1 (2014): 1-18.

Kau, James, Donald Keenan, and Taewon Kim. "Default Probabilities for Mortgages", *Journal of Urban Economics*, Vol. 35, no. 3 (1994): 278-296.

Kau, James, Donald Keenan, Constantine Lyubimov, and V. Carlos Slawson. "Subprime Mortgage Default", *Journal of Urban Economics*, Vol. 70, no. 2 (2011): 75-87.

Keys, Benjamin, Tanmoy Mukherjee, Amit Seru and Vikrant Vig. "Financial Regulation and Securitization: Evidence from Subprime Loans", *Journal of Monetary Economics*, Vol. 56, no. 5 (2009): 700-720.

____, "Did Securitization Lead to Tax Screening? Evidence From Subprime Loans", *The Quarterly Journal of Economics*, Vol. 125, no. 1 (2010): 307-362.

Keys, Benjamin, Amir Sufi, and Vikram Vig. "Lender Screening and the Role of Securitization: Evidence from Prime and Subprime Mortgage Markets", *Review of Financial Studies*, Vol. 25, no. 7 (2012): 2071-2108.

LaCour-Little, Michael, Charles Calhoun, and Wei Yu. "What Role Did Piggyback Lending Play in the Housing Bubble and Mortgage Collapse", *Journal of Housing Economics*, Vol. 20, no. 2 (2011): 81-100.

Mayer, Christopher and Karen Pence. "Subprime Mortgages: What, Where, and to Whom?", NBER Working Paper 14083 (2008).

Mayer, Christopher, Karen Pence, and Shane Sherlund. "The Rise in Mortgage Defaults", *Journal of Economic Perspectives*, Vol. 23, no. 1 (2009): 27-50.

Mian, Atif and Amir Sufi. "The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis", *Quarterly Journal of Economics*, Vol. 124, no. 4 (2009): 1449-1496.

______________________. "House Prices, Home Equity-Based Borrowing, and the US Household Leverage Crisis", *American Economic Review*, Vol. 101, no. 5 (2011): 2132-2156.

______________________. *House of Debt*. The University of Chicago Press: Chicago and London, 2014.

______________________. "Fraudulent Income Overstatement on Mortgage Applications during the Credit Expansion of 2002 to 2005", National Bureau of Economic Research Working Paper w20947, 2015.

Moulton, Shawn. "Did Affordable Housing Mandates Cause the Subprime Mortgage Crisis?", *Journal of Housing Economics*, Vol. 24, no. 1 (2014): 21-38.

Nadauld, Taylor and Shane Sherlund. "The Impact of Securitization on the Expansion of Subprime Credit", *Journal of Financial Economics*, Vol. 107, no. 2 (2013): 454-476.

Palmer, Christopher. "Why Did So Many Subprime Borrowers Default During the Crisis: Loose Credit or Plummeting Prices?", MIT Working Paper, July 2014.

Piskorski, Tomasz, Amit Seru, and Vikrant Vig. "Securitization and Distressed Loan Renegotiation: Evidence from the Subprime Mortgage Crisis", *Journal of Financial Economics*, Vol. 97, no. 3 (2010): 369-397.

Rose, Morgan. "Geographic Variation in Subprime Loan Features, Foreclosures, and Prepayments", *The Review of Economics and Statistics*, Vol. 95, no. 2 (2013): 563-590.

Schuetz, Jenny, Vicki Been, and Ingrid Gould Ellen. "Neighborhood Effects of Concentrated Mortgage Foreclosures", *Journal of Housing Economics*, Vol. 17, no. 4 (2008): 306-319.

Shiller, Robert. *Irrational Exuberance*. Princeton University Press: Princeton , NJ (2nd Edition): 2005.

Smith, Brent. "Stability in Consumer Credit Scores: Level and Direction of FICO Score Drift as a Precursor to Mortgage Default and Prepayment" *Journal of Housing Economics*, Vol. 20, no. 4 (2011): 285-298.

Soo, Cindy. "Quantifying Housing Spirits: News Media and Sentiment in the Housing Market", University of Michigan working paper, 2015.

Whitaker, Stephan and Thomas Fitzpatrick IV. "Deconstructing Distressed-Property Spillovers: The Effects of Vacant, Tax-Delinquent, and Foreclosed Properties in Housing Submarkets", *Journal of Housing Economics*, Vol. 22, no. 2 (2013): 79-91.

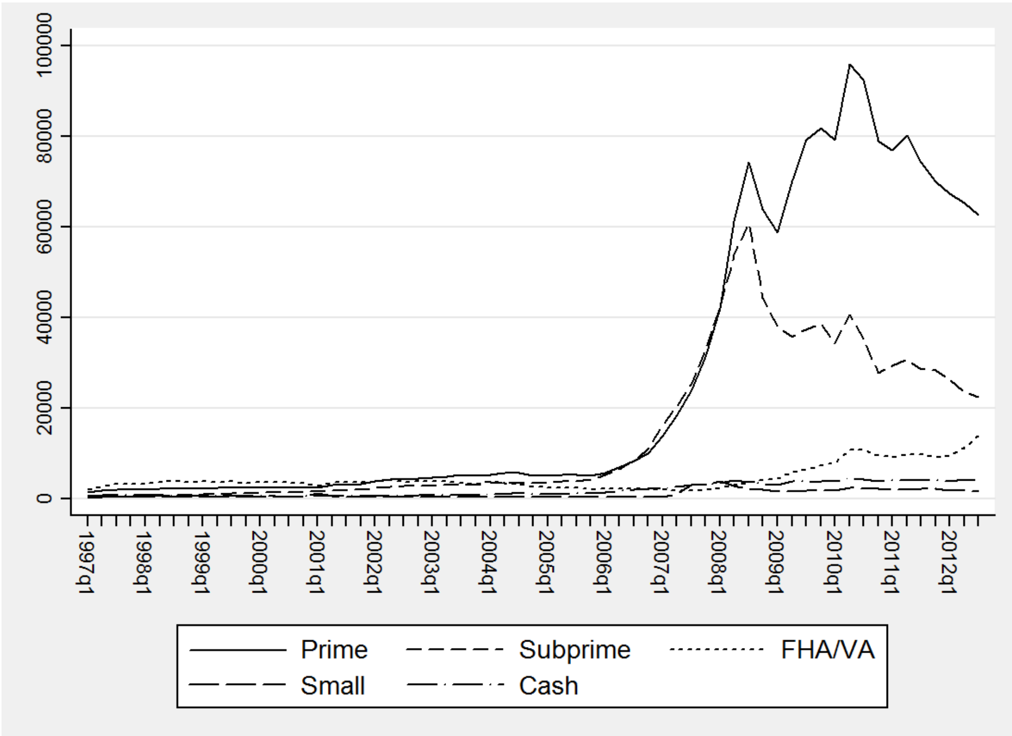


Figure 1: Total Foreclosures + Short Sales Over Time by Owner Type

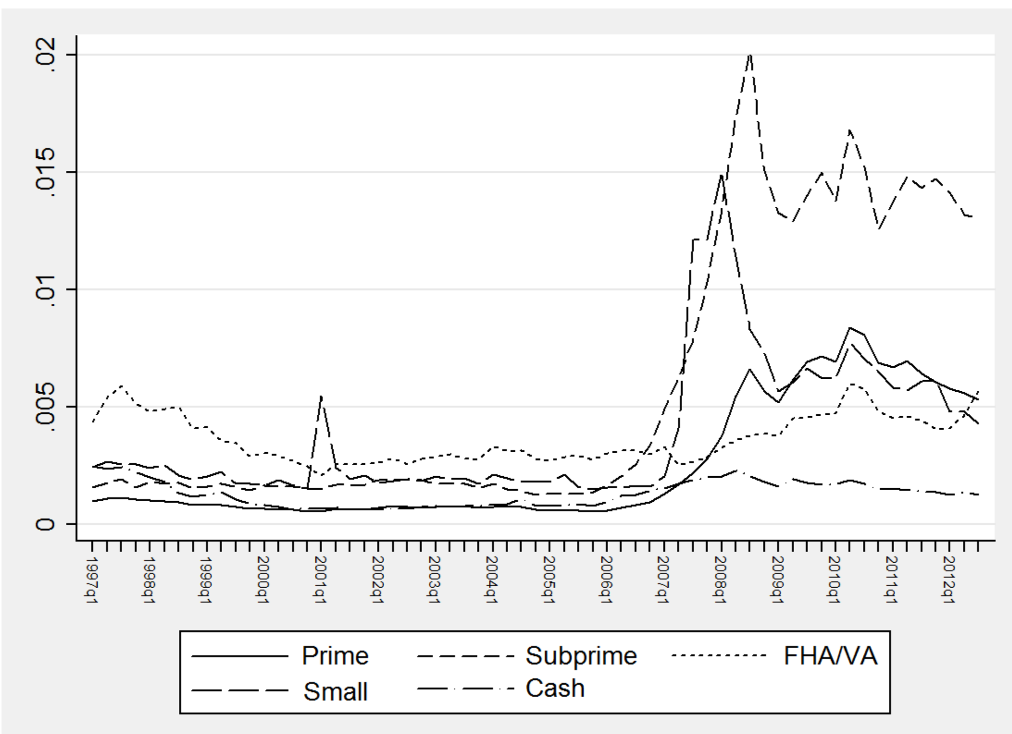


Figure 2: Unconditional Distress Probabilities (Foreclosures + Short Sales) Over Time by Owner Type

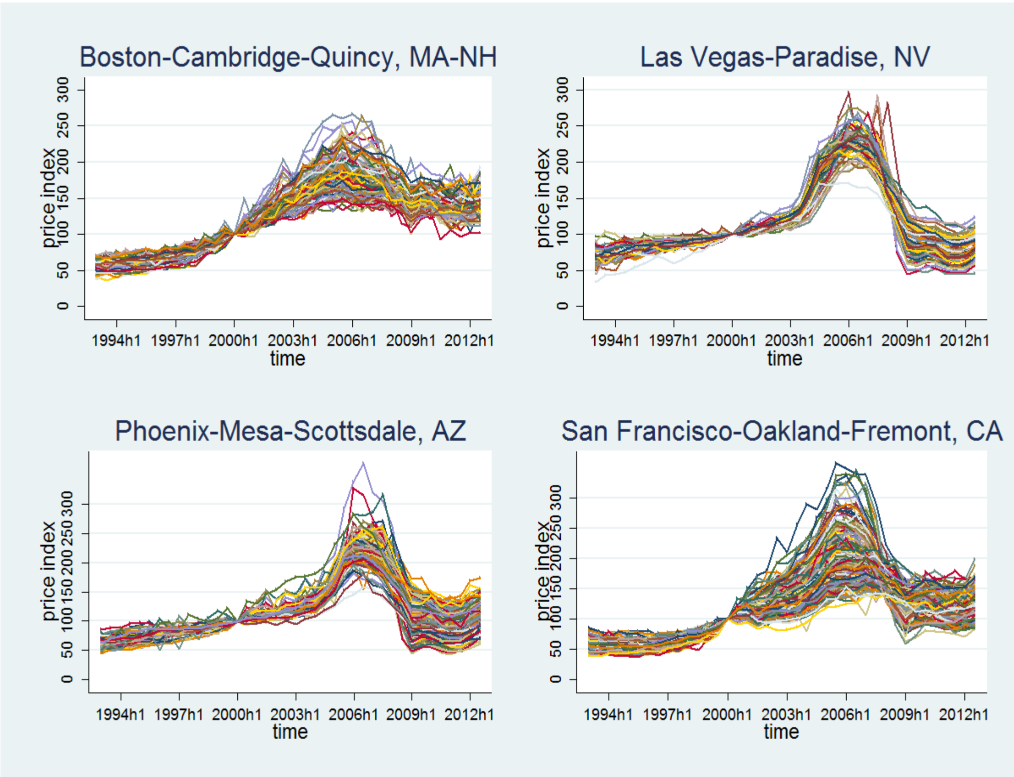


Figure 3. Hedonic Price Indexes by Neighborhood; Selected MSAs

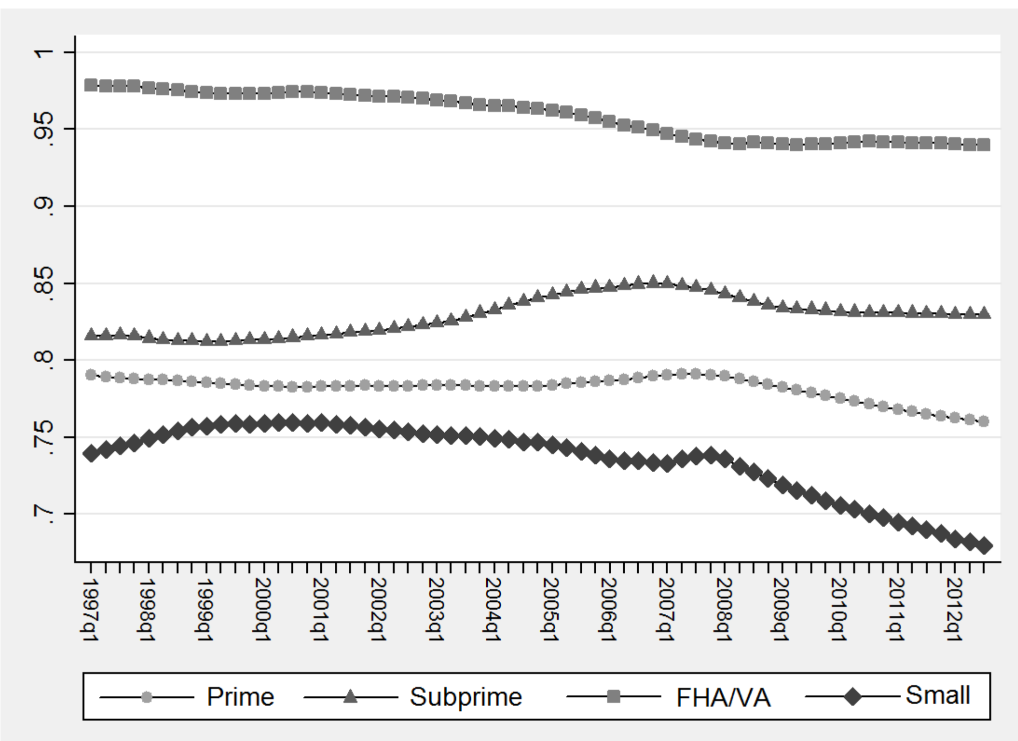


Figure 4. Average Initial LTV by Borrower Type Over Time

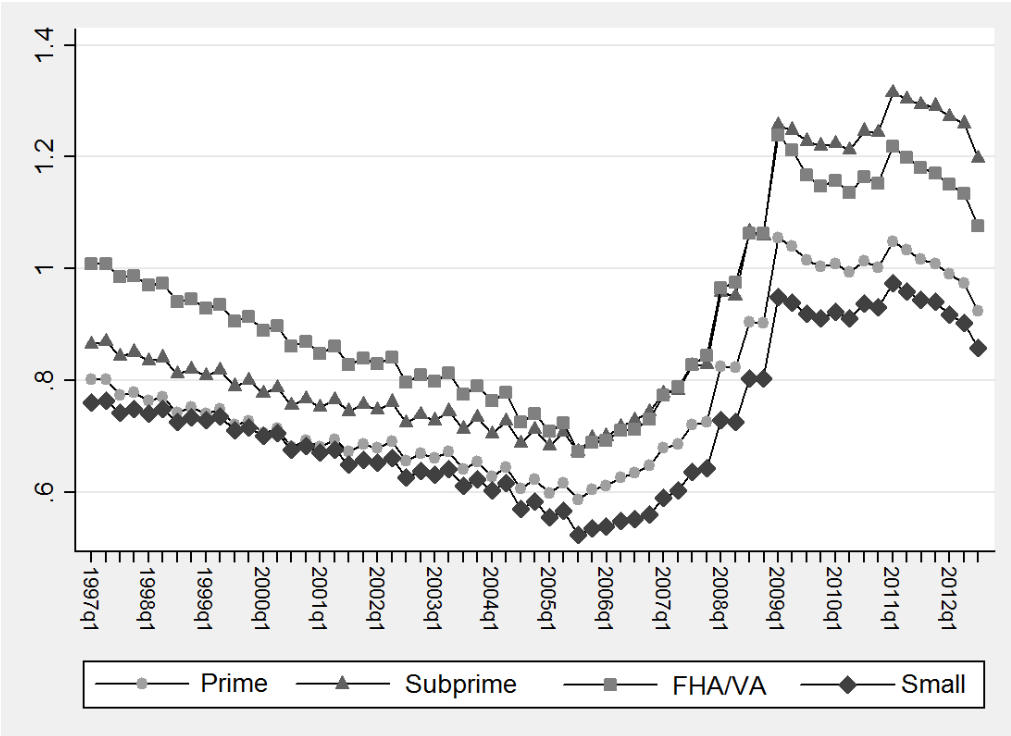


Figure 5. Average Current LTV by Borrower Type Over Time

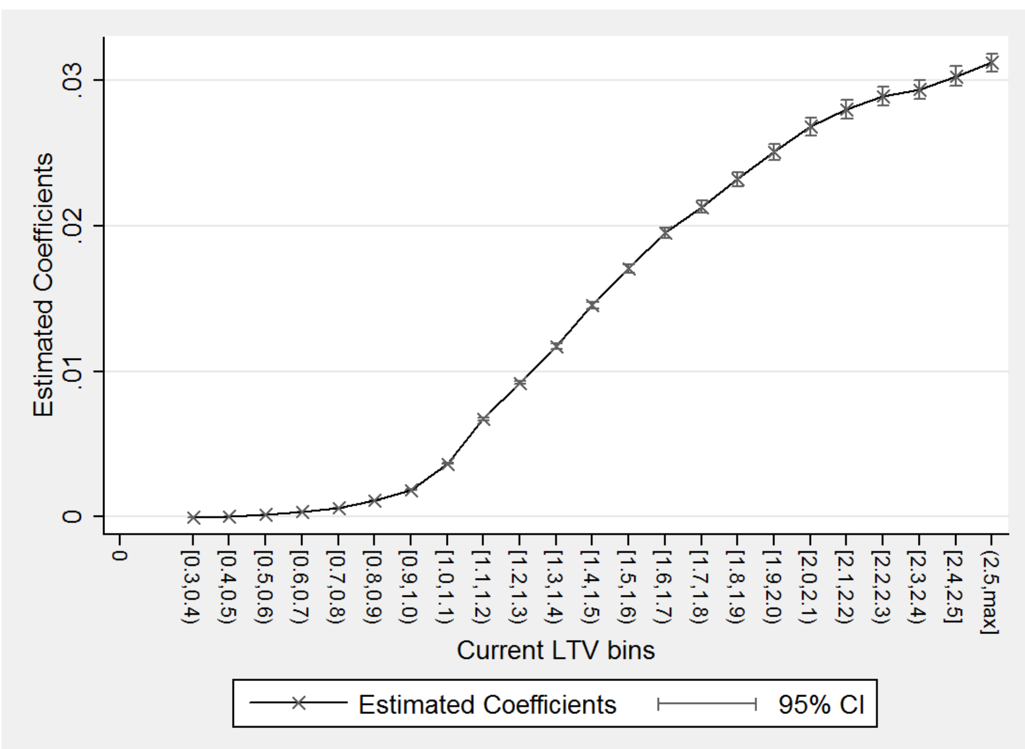


Figure 6. Current LTV Estimated Coefficients

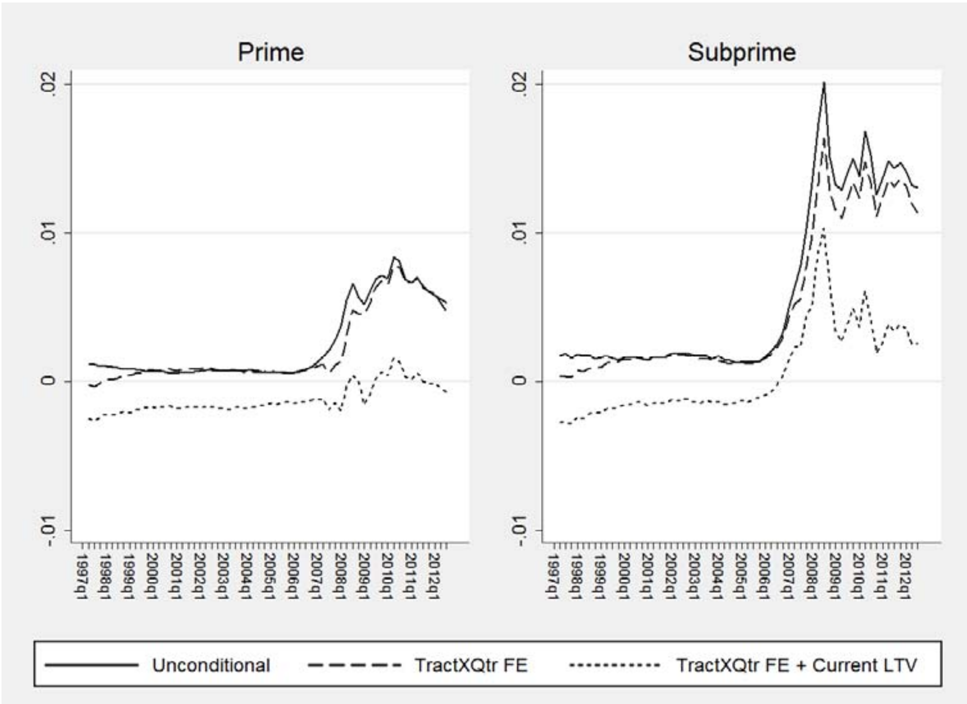


Figure 7: Heterogeneity – Prime and Subprime Borrowers Estimates by Quarter

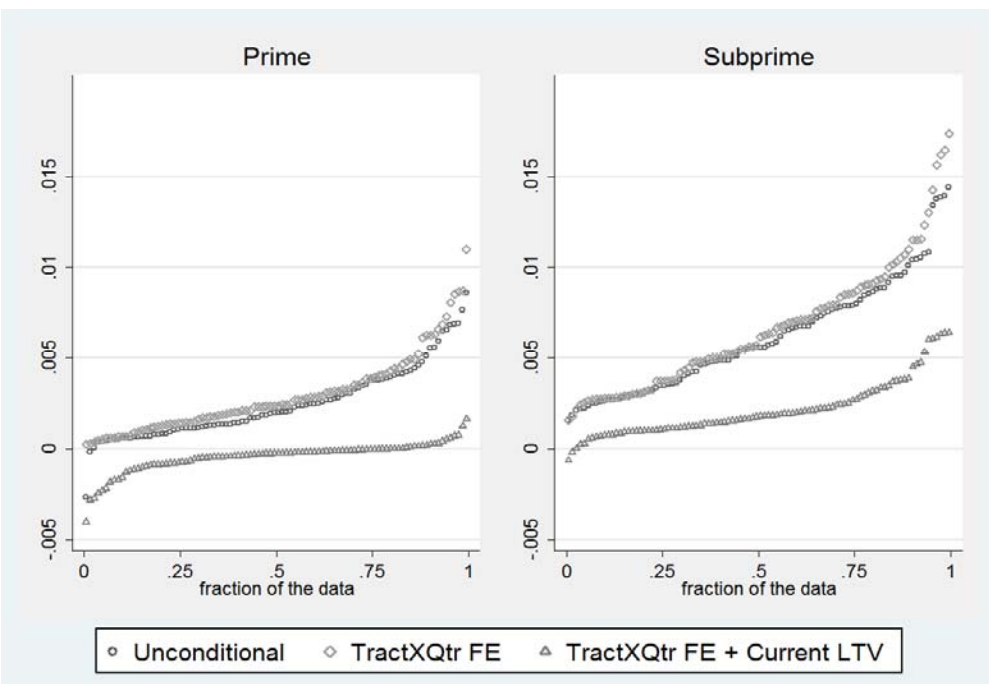


Figure 8: Heterogeneity – Prime and Subprime Borrowers Estimates by Metropolitan Area

Table 1. Number and Share of Owner types, 1997-2012

	Prime		Subprime		FHA/VA-Insured		Small Loans		All Cash		Total
Year - Qtr	Number	Share	Number	Share	Number	Share	Number	Share	Number	Share	Number
1997q1	1,461,113	0.581	197,036	0.078	460,444	0.183	113,492	0.045	283,463	0.113	2,515,548
1997q2	1,601,454	0.575	225,934	0.081	514,150	0.185	123,575	0.044	317,616	0.114	2,782,729
1997q3	1,746,683	0.572	259,032	0.085	568,961	0.186	132,789	0.043	347,115	0.114	3,054,580
1997q4	1,884,392	0.568	297,260	0.090	619,116	0.187	142,184	0.043	375,923	0.113	3,318,875
1998q1	2,020,950	0.562	340,021	0.095	671,277	0.187	150,742	0.042	411,399	0.114	3,594,389
1998q2	2,237,271	0.558	401,265	0.100	748,359	0.187	161,356	0.040	459,574	0.115	4,007,825
1998q3	2,460,573	0.557	465,821	0.106	820,667	0.186	170,777	0.039	495,943	0.112	4,413,781
1998q4	2,646,038	0.556	525,096	0.110	882,623	0.186	176,964	0.037	525,795	0.111	4,756,516
1999q1	2,807,548	0.555	570,725	0.113	940,982	0.186	182,342	0.036	559,505	0.111	5,061,102
1999q2	3,033,496	0.553	635,547	0.116	1,024,511	0.187	188,342	0.034	602,489	0.110	5,484,385
1999q3	3,252,421	0.552	705,968	0.120	1,101,525	0.187	195,192	0.033	640,539	0.109	5,895,645
1999q4	3,433,274	0.551	767,064	0.123	1,161,743	0.186	200,864	0.032	672,278	0.108	6,235,223
2000q1	3,583,428	0.549	821,455	0.126	1,209,190	0.185	206,093	0.032	706,451	0.108	6,526,617
2000q2	3,798,948	0.549	880,720	0.127	1,273,038	0.184	211,631	0.031	751,073	0.109	6,915,410
2000q3	4,009,833	0.551	938,001	0.129	1,324,346	0.182	215,892	0.030	785,107	0.108	7,273,179
2000q4	4,204,026	0.554	991,128	0.131	1,361,690	0.179	218,800	0.029	811,791	0.107	7,587,435
2001q1	4,388,714	0.558	1,035,416	0.132	1,380,689	0.176	221,910	0.028	835,262	0.106	7,861,991
2001q2	4,660,792	0.564	1,099,295	0.133	1,410,123	0.171	222,360	0.027	865,928	0.105	8,258,498
2001q3	4,910,388	0.569	1,167,156	0.135	1,435,779	0.166	222,729	0.026	891,335	0.103	8,627,387
2001q4	5,153,605	0.576	1,223,976	0.137	1,438,893	0.161	222,447	0.025	908,636	0.102	8,947,557
2002q1	5,359,427	0.581	1,282,624	0.139	1,435,207	0.155	222,112	0.024	931,928	0.101	9,231,298
2002q2	5,593,419	0.583	1,361,775	0.142	1,443,144	0.150	223,803	0.023	966,983	0.101	9,589,124
2002q3	5,853,513	0.588	1,446,270	0.145	1,428,856	0.144	225,841	0.023	992,892	0.100	9,947,372
2002q4	6,113,947	0.595	1,537,179	0.150	1,385,864	0.135	226,632	0.022	1,014,814	0.099	10,278,436
2003q1	6,315,580	0.599	1,623,965	0.154	1,339,169	0.127	226,843	0.022	1,036,251	0.098	10,541,808
2003q2	6,591,889	0.605	1,734,411	0.159	1,285,920	0.118	227,112	0.021	1,064,574	0.098	10,903,906
2003q3	6,864,625	0.609	1,869,740	0.166	1,227,857	0.109	228,976	0.020	1,089,676	0.097	11,280,874
2003q4	7,019,910	0.608	2,000,874	0.173	1,180,814	0.102	230,064	0.020	1,107,984	0.096	11,539,646
2004q1	7,215,483	0.607	2,146,787	0.181	1,138,374	0.096	232,244	0.020	1,153,076	0.097	11,885,964
2004q2	7,502,324	0.608	2,326,683	0.188	1,081,179	0.088	233,768	0.019	1,200,967	0.097	12,344,921
2004q3	7,737,105	0.606	2,510,437	0.197	1,027,257	0.081	233,172	0.018	1,251,342	0.098	12,759,313
2004q4	7,989,115	0.608	2,667,153	0.203	971,788	0.074	232,298	0.018	1,290,326	0.098	13,150,680
2005q1	8,251,956	0.611	2,770,546	0.205	922,916	0.068	232,437	0.017	1,337,543	0.099	13,515,398
2005q2	8,586,093	0.614	2,901,339	0.208	870,531	0.062	233,122	0.017	1,388,454	0.099	13,979,539
2005q3	8,954,585	0.620	3,017,225	0.209	813,351	0.056	232,700	0.016	1,427,138	0.099	14,444,999
2005q4	9,245,242	0.625	3,097,420	0.209	767,653	0.052	232,311	0.016	1,457,176	0.098	14,799,802
2006q1	9,492,134	0.629	3,149,586	0.209	739,391	0.049	233,065	0.015	1,484,070	0.098	15,098,246
2006q2	9,797,580	0.633	3,208,443	0.207	719,916	0.047	233,832	0.015	1,511,982	0.098	15,471,753
2006q3	10,065,708	0.637	3,254,400	0.206	702,920	0.045	236,052	0.015	1,531,561	0.097	15,790,641
2006q4	10,314,375	0.642	3,279,998	0.204	689,966	0.043	236,264	0.015	1,548,453	0.096	16,069,056
2007q1	10,533,191	0.646	3,277,257	0.201	681,157	0.042	238,261	0.015	1,568,706	0.096	16,298,572
2007q2	10,769,338	0.651	3,251,769	0.197	682,095	0.041	243,671	0.015	1,591,468	0.096	16,538,341
2007q3	10,933,035	0.654	3,232,601	0.193	691,431	0.041	251,404	0.015	1,616,606	0.097	16,725,077
2007q4	11,051,114	0.655	3,197,634	0.190	715,065	0.042	254,332	0.015	1,643,855	0.097	16,862,000
2008q1	11,132,898	0.656	3,159,711	0.186	754,146	0.044	256,489	0.015	1,678,459	0.099	16,981,703
2008q2	11,211,491	0.654	3,098,625	0.181	845,959	0.049	257,308	0.015	1,733,458	0.101	17,146,841
2008q3	11,251,476	0.651	3,015,692	0.174	964,310	0.056	261,744	0.015	1,799,066	0.104	17,292,288
2008q4	11,264,286	0.647	2,931,262	0.168	1,074,430	0.062	266,235	0.015	1,865,205	0.107	17,401,418
2009q1	11,283,744	0.644	2,858,075	0.163	1,170,126	0.067	270,759	0.015	1,938,607	0.111	17,521,311
2009q2	11,345,222	0.640	2,763,922	0.156	1,299,844	0.073	278,157	0.016	2,030,913	0.115	17,718,058
2009q3	11,385,042	0.636	2,667,633	0.149	1,436,689	0.080	288,357	0.016	2,124,824	0.119	17,902,545
2009q4	11,406,100	0.632	2,574,823	0.143	1,573,098	0.087	297,004	0.016	2,210,494	0.122	18,061,519
2010q1	11,405,386	0.628	2,488,628	0.137	1,663,575	0.092	304,504	0.017	2,295,678	0.126	18,157,771
2010q2	11,434,825	0.623	2,405,749	0.131	1,796,217	0.098	314,084	0.017	2,395,555	0.131	18,346,430
2010q3	11,445,388	0.620	2,308,470	0.125	1,886,678	0.102	322,442	0.017	2,486,839	0.135	18,449,817
2010q4	11,469,356	0.618	2,211,624	0.119	1,967,556	0.106	330,939	0.018	2,567,826	0.138	18,547,301
2011q1	11,485,234	0.615	2,134,384	0.114	2,036,482	0.109	339,208	0.018	2,667,933	0.143	18,663,241
2011q2	11,505,664	0.611	2,064,259	0.110	2,124,136	0.113	349,058	0.019	2,777,830	0.148	18,820,947
2011q3	11,537,139	0.609	1,993,259	0.105	2,204,347	0.116	356,184	0.019	2,867,608	0.151	18,958,537
2011q4	11,575,053	0.606	1,924,148	0.101	2,270,468	0.119	362,275	0.019	2,953,200	0.155	19,085,144
2012q1	11,602,503	0.604	1,856,461	0.097	2,329,389	0.121	368,892	0.019	3,042,104	0.158	19,199,349
2012q2	11,664,543	0.602	1,783,657	0.092	2,399,558	0.124	376,924	0.019	3,140,058	0.162	19,364,740
2012q3	11,724,049	0.601	1,712,800	0.088	2,449,919	0.126	383,474	0.020	3,221,830	0.165	19,492,072
Total	478,575,034		119,647,214		76,536,854		15,264,834		89,252,534		779,276,470

Notes: Our calculations based on our final data described in section III.

Table 2. Summary Statistics, 96 Market Aggregate and by Owner Type

	Overall		Prime		Subprime		FHA/VA		Cash		Small	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
VARIABLES	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
1. Distress Outcomes												
Foreclosure	0.0027	0.0515	0.0022	0.0464	0.0054	0.0732	0.0029	0.0542	0.0014	0.0369	0.0030	0.0543
Short Sale	0.0012	0.0340	0.0012	0.0349	0.0020	0.0442	0.0008	0.0285	0.0000	0.0000	0.0013	0.0357
Total	0.0038		0.0034		0.0073		0.0038		0.0014		0.0042	
2. House Characteristics												
Age of House (Years)	33	27	33	28	34	27	30	24	33	27	38	27
Square Footage	1788	14196	1863	13723	1758	12530	1463	3140	1720	21941	1665	16990
Bedrooms	2.3	5.8	2.4	3.1	2.5	10.2	2.2	10.8	1.9	2.5	2.1	1.9
Bathrooms	2.0	2.9	2.1	2.9	2.1	3.1	1.9	2.7	1.8	2.8	1.8	2.3
3. Borrower Characteristics												
Male	0.73	0.45	0.73	0.44	0.71	0.45	0.72	0.45	.	.	0.74	0.44
White	0.70	0.46	0.73	0.44	0.63	0.48	0.62	0.48	.	.	0.76	0.43
Black	0.06	0.24	0.05	0.21	0.08	0.27	0.12	0.32	.	.	0.05	0.22
Hispanic	0.15	0.36	0.12	0.33	0.20	0.40	0.22	0.41	.	.	0.11	0.31
Asian	0.09	0.29	0.10	0.30	0.09	0.28	0.04	0.20	.	.	0.08	0.27
Real Income (thousands)	117.1	185.6	125.1	198.7	117.5	168.3	68.6	110.8	.	.	141.5	250.1
Speculators												
a. Different Tax Address	0.24	0.42	0.20	0.40	0.18	0.39	0.15	0.35	0.55	0.50	0.44	0.50
b. Commerical Firms (exc diff. tax address)	0.01	0.11	0.01	0.09	0.01	0.07	0.00	0.04	0.06	0.24	0.03	0.16
c. Total	0.25		0.21		0.19		0.15		0.61		0.46	
4. Purchase Price and LTV												
Nominal Sale Price	270,311	271,629	312,457	274,304	307,621	251,678	168,073	101,066	220,042	321,270	211,272	272,555
Initial LTV < 80%	0.39	0.49	0.36	0.48	0.25	0.43	0.05	0.22	1.00	0.00	0.42	0.49
80% <= Initial LTV <= 90%	0.25	0.43	0.32	0.47	0.31	0.46	0.04	0.19	0.00	0.00	0.23	0.42
90% < Initial LTV <= 95%	0.11	0.31	0.13	0.34	0.13	0.34	0.05	0.21	0.00	0.00	0.10	0.30
95% < Initial LTV <= 100%	0.20	0.40	0.15	0.36	0.26	0.44	0.64	0.48	0.00	0.00	0.18	0.39
Initial LTV > 100%	0.05	0.22	0.03	0.18	0.05	0.21	0.23	0.42	0.00	0.00	0.06	0.24
5. Debt Repositioning												
No Debt Change Since Purchase	0.51	0.50	0.43	0.49	0.36	0.48	0.66	0.47	1.00	0.00	0.74	0.44
Refinance	0.34	0.47	0.41	0.49	0.37	0.48	0.24	0.43	0.00	0.00	0.24	0.43
Second Mortgage	0.15	0.36	0.16	0.37	0.27	0.44	0.11	0.31	0.00	0.00	0.02	0.12
6. Current LTV												
Current LTV < 80%	0.57	0.50	0.56	0.50	0.46	0.50	0.26	0.44	1.00	0.00	0.61	0.49
80% <= Current LTV <= 100%	0.22	0.41	0.22	0.41	0.25	0.43	0.41	0.49	0.00	0.00	0.22	0.41
100% < Current LTV < 120%	0.09	0.29	0.09	0.29	0.12	0.32	0.20	0.40	0.00	0.00	0.08	0.27
Current LTV >= 120%	0.12	0.32	0.13	0.33	0.18	0.39	0.13	0.34	0.00	0.00	0.09	0.29
Number of Obervations	779,276,470		478,575,034		119,647,214		76,536,854		89,252,534		15,264,834	

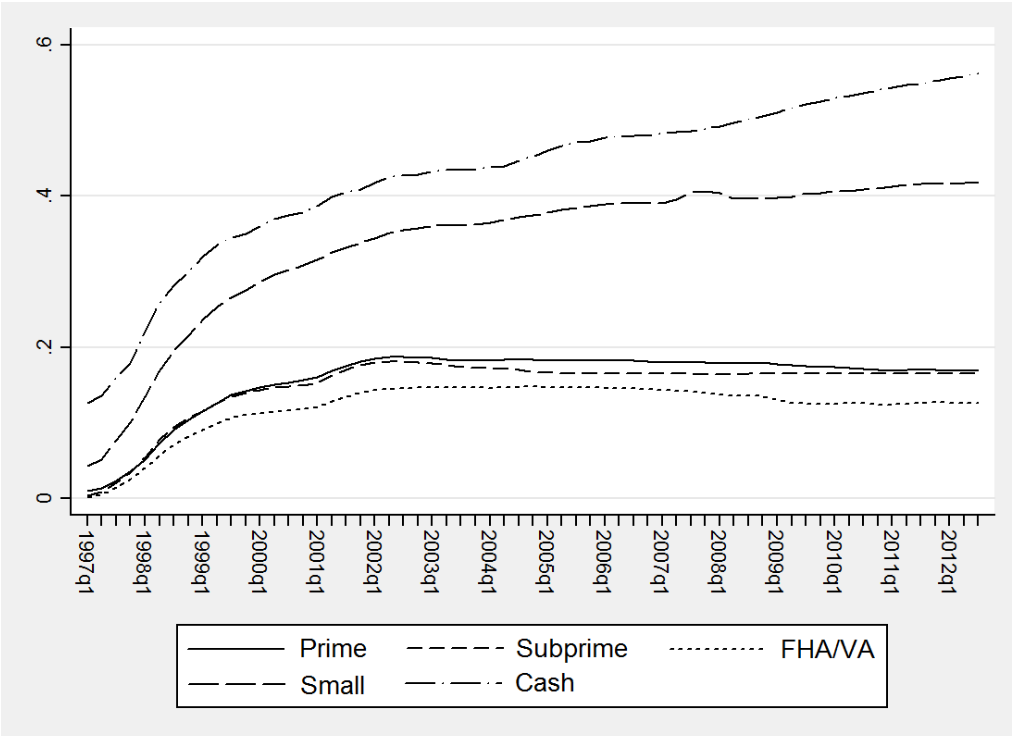
Notes: Calculations based on final data described in section III.

Table 3. Panel Estimates

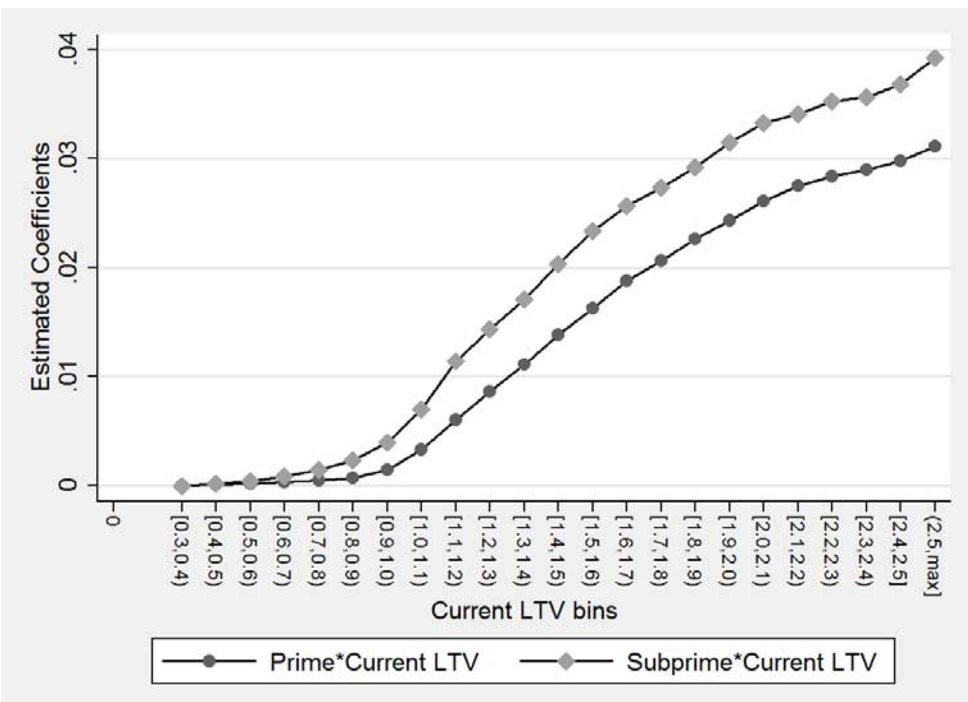
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Prime	0.00213	0.00329	-0.00085	0.00329	0.00333	0.00147	0.00390	0.00231	0.00138	-0.00001
	(0.00004)	(0.00005)	(0.00002)	(0.00005)	(0.00005)	(0.00003)	(0.00005)	(0.00003)	(0.00002)	(0.00002)
Subprime	0.00594	0.00696	0.00159	0.00696	0.00688	0.00463	0.00748	0.00529	0.00392	0.00223
	(0.00006)	(0.00007)	(0.00002)	(0.00007)	(0.00007)	(0.00004)	(0.00007)	(0.00004)	(0.00003)	(0.00003)
FHA/VA	0.00216	0.00210	-0.00194	0.00210	0.00231	-0.00253	0.00251	0.00293	-0.00058	-0.00161
	(0.00004)	(0.00004)	(0.00003)	(0.00004)	(0.00004)	(0.00006)	(0.00004)	(0.00004)	(0.00004)	(0.00004)
Small	0.00273	0.00367	0.00051	0.00366	0.00372	0.00181	0.00409	0.00343	0.00227	0.00102
	(0.00008)	(0.00009)	(0.00008)	(0.00009)	(0.00009)	(0.00008)	(0.00009)	(0.00008)	(0.00008)	(0.00008)
Tract*Quarter FE	NO	YES	YES	YES	YES	YES	YES	YES	YES	YES
Current LTV	NO	NO	YES	NO	NO	NO	NO	NO	NO	YES
Housing Traits	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO
Household Traits	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO
Initial LTV	NO	NO	NO	NO	NO	YES	NO	NO	YES	YES
Refi/2nd Mortgage	NO	NO	NO	NO	NO	NO	YES	NO	YES	YES
Loan Origination Cohort	NO	NO	NO	NO	NO	NO	NO	YES	YES	YES
Observations	779,276,470	779,276,470	779,276,470	779,275,663	779,276,470	779,276,470	779,276,470	779,276,470	389,638,265	194,819,128

Notes:

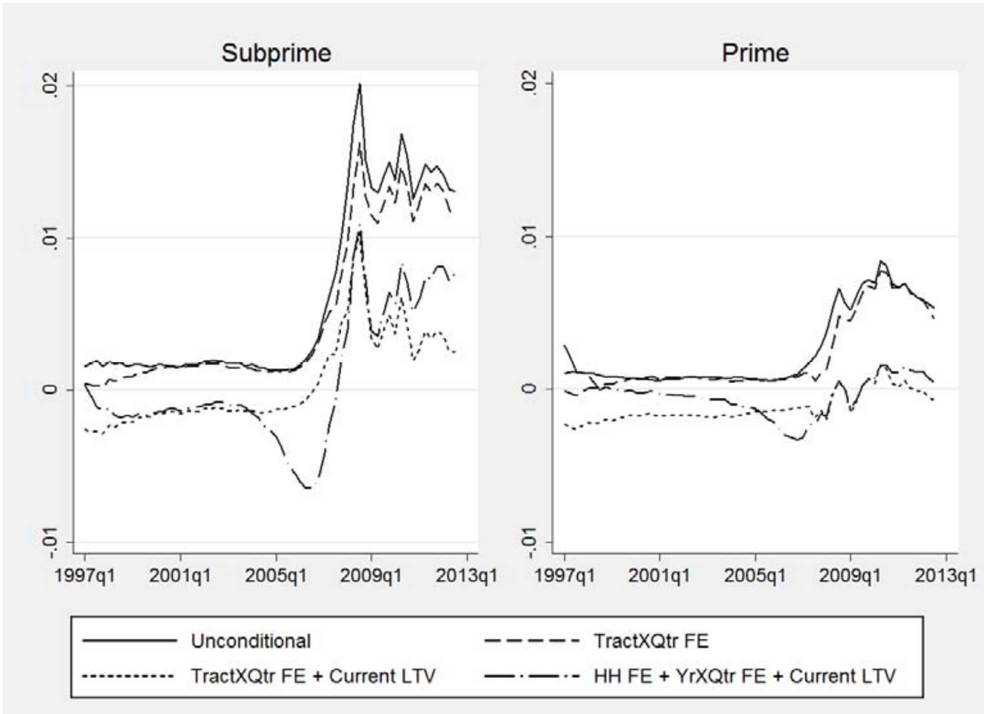
1. Estimates of the relative probability of foreclosures or short-sales for different home owner types based on equation (1) in the text. The omitted category is all specifications is all-Cash owners.
2. Standard errors are based on clustering at the census tract level. Because of computational constraints, the results in Column (9) are from a 50% random subsample of each MSA, whille those for Column (10)'s specification require a smaller 25% subsample to estimate clustered standard errors. Point estimates are not changed in any material way relative to estimating on full samples without clustering.
3. The Current LTV variable is entered in discrete form via the 25 bins depicted in Figure 6. They range from 0.0-0.3, 0.3-0.4, …, 2.4-2.5, 2.5+.
4. The Housing Trait vector includes square footage (entered quadratically), the number of bedrooms, and the number of bathrooms.
5. The Household Trait vector includes self-reported income, race and gender of the head of the household, and a dummy for speculators. To surmount computational constraints, this variable is made discrete by each owner being coded by an indicator variable as being in one (and only one) bin.
6. The Refi/2nd Mortgage dummy variable and Loan Origination Cohort vector of dummy variables are as described in Section III.



Appendix Figure 1: Share Over Time of Owner Types Classified As Speculators



Appendix Figure 2: Heterogeneity -- Current LTV Coefficients for Prime and Subprime Borrowers



Appendix Figure 3. Prime and Subprime Borrowers Estimates by Quarter with Ownership Fixed Effects (HH FE)

Appendix Table 1: Existing Literature--Data, Coverage, Issue Focus

Select Published Papers on Mortgage Performance From 2008(1) to 2014(2)

Data Coverage Characteristics

Paper	Sector	Type	Source	Time Period	Geographic Coverage	Aggregation	Issue Focus
Haughwout, Peach and Tracy 4(2008)	Subprime	Loan Level	First American Core Logic Loan Performance (LP)	1/01-4/07 originations; tracked through 4/08	National	Micro; 1% subsample; 117,000 loans	2nd leins; early defaults; underwriting standards
Mayer and Pence (2008)	Subprime	Loan Level	First American Core Logic Loan Performance (LP), HUD, HMDA	1998-2006	National	Micro	descriptive; geographic concentration; different data sources
Foote, et. al. (2008)	Subprime	Panel	Warren Group	1987-March 2008	Massachusetts	Micro	Underwriting interacted with deteriorating economy
Coleman, LaCour-Little & Vandell (2008)	Subprime	Aggregate Cross-Section, Time Series	HMDA, LP	1998-2006	20 S&P/Case-Shiller MSAs	MSA	Price dynamics
Mayer, Pence and Sherlund (2009)	Subprime	Loan Level	First American Core Logic Loan Performance (LP)	1/03-6/07 originations	National	Micro	Descriptive; comparison across cohorts; detail on loan traits
Adelino, Gerardi and Willen (2009a); now forthcoming	Subprime	Loan Level	First American Core Logic Loan Performance (LP)	1/05-6/07 originations; tracked through 8/08	National	Micro	Securitization and default
Keys, , et. al. (2009)	Subprime	Loan Level	First American Core Logic Loan Performance (LP)	1/01-12/06 originations	National	Micro	financial regulation and securitization
Foote, et. al. (2010)	Prime + Subprime	Loan Level	LPS Applied Analytics	originations through 2008; good coverage from 2005-on	National	Micro	Double Trigger hypothesis; subprime defaults before resets
Mian and Sufi (2009)	Subprime	Loan Level	Equifax (originations for purchase)	1991-2007 originations; tracked through 2007	40 MSAs	Zip code-level	Onset of subprime defaults; role of supply vs. demand factors
Elul, et. al. (2010)	Subprime	Loan Level	First American Core Logic Loan Performance (LP); merged with Equifax credit panel	2005-06 originations; tracked through 4/09	National	Micro	Double Trigger hypothesis
Piskorski, Seru and Vig (2010)	Subprime	Loan Level	Lender Processer Services (LPS; formerly McDash Analytics)	2005-2006 originations; tracked through 3/08	National	Micro	Securitization and renegotiation
Keys, et. al. (2010)	Subprime	Loan Level	First American Core Logic Loan Performance (LP)	1/01-12/06 originations	National	Micro	Securitization and lax screening
Goodman and Smith (2010)	Subprime	Loan Level	LPS Applied Analytics	1991-2008 originations	National	Zip code-level	State-level policy effects
Demyanyk & Van Hemmert (2011)	Subprime	Loan Level	First American Core Logic Loan Performance (LP)	2001-2007 originations; tracked through 6/08	National	Micro	Role of underwriting vs. economic conditions in default
Ben David (2011)	House Prices	Home Sales	Chicago MLS	2005-2008	Chicago	Micro	Role of inflated prices on default
Agarwal, et. al. (2011)	Prime + Subprime	Loan Level	OCC-OTCC Mortgage Metrics	originations tracked between October 2007-May 2009	National	Micro	Role of securitization in workouts

Appendix Table 1 Cont'd: Existing Literature--Data, Coverage, Issue Focus

Select Published Papers on Mortgage Performance From 2008(1) to 2014(2)

Data Coverage Characteristics

Paper	Sector	Type	Source	Time Period	Geographic Coverage	Aggregation	Issue Focus
Capozza and Van Order (2011)	Subprime	Loan Level	Mortgage Bankers Association	2000-2009 originations; tracked though 2009	National	Micro	Estimate default model; role of underwriting
Kau, et. al. (2011)	Subprime	Loan Level	Black Box Logic LLC	1997-2008(1) originations; tracked through 2009(2)	20 MSAs	Micro	Default model; role of borrower and loan traits
Smith (2011)	Prime + Subprime	Loan Level	LPS Applied Analytics	2001-2008 originations	Florida	Micro	Documents stability of FICO scores
LaCour-Little, Calhoun and Yu (2011)	Prime + Subprime	Loan Level	HMDA	2001-2008 originations	National	State and Zip codel-aggregation	impute rise of simultaneous second mortgages
Keys, Sufi and Vig (2012)	Prime + Subprime	Loan Level	Lender Processer Services (LPS; formerly McDash Analytics)	2001-2006 originations	National	Micro	Securitization and screening
Rose (2013)	Subprime	Loan Level	First American Core Logic Loan Performance (LP)	2002-2006 originations; tracked through October 2008	10 MSAs	Micro	Role of underwriting, loan traits, and aggregate location traits
Guiso, Sapienza and Zingales (2013)	Household Borrowers	Survey	Chicago Booth Kellogg School Financial Trust Index	12/08-9/10 survey waves	National	NA	Imputes strategic defaults from survey answers
Adelino, Gerardi and Willen (2013)	Borrowers-Prime + Subprime	Loan Level	Lender Processer Services (LPS; formerly McDash Analytics)	2005-2011	National	Micro	Documents loan modification types; estimates model of lender behavior
Nadauld and Sherland (2013)	Subprime	Loan Level	First American Core Logic Loan Performance (LP)	1997-2007 originations	National	Zip code-level	supply side effects through lower costs and screening incentives
Aschberg, et. al. (2013)	Prime + Subprime	NA	NA	macro-type model	match moments	NA	dynamic simulation model of spillovers onto prime market from subprime default
Jiang, Nelson and Vytacl (2014)	Proprietary—not fully revealed	Loan Level	Single mortgage bank (anonymous)	1/04-2/08 originations; tracked through 1/09	National	Micro	role of 3rd party originations (outside brokers) vs. bank originations in default
Bhandwaj & Sengupta (2014)	Subprime	Loan Level	First American Core Logic Loan Performance (LP)	2000-2006 originations	National	Micro	compares early vs. late cohorts; focus on role of macro effects
Bubb & Kaufmann (2014)	Borrowers—Prime + Subprime	Loan Level	Lender Processing Services Applied Analytics (LPS)	2003-2009 originations	National	Micro	Moral hazard and securitization
Agarwal, et. al. (2014)	Subprime	Loan Level	HMDA	2005-2007 originations	Chicago	Zip code-level	Impact of predatory lending legislation in IL
Moulton (2014)	Subprime	Loan Level	HMDA	2006-07 originations; tracked through 2008	National	Census tract-level	RD on affordable housing goal targets

Appendix Table 2. Sample Coverage

	Start	End		Start	End
Akron, OH	1997q2	2012q3	Norwich-New London, CT	1993q1	2012q3
Atlantic City, NJ	1998q1	2012q3	Ocala, FL	1996q1	2012q3
Bakersfield, CA	1993q1	2012q3	Oklahoma City, OK	1998q2	2012q3
Baltimore-Towson, MD	1997q2	2012q3	Olympia, WA	1995q1	2012q3
Barnstable Town, MA	1994q1	2012q3	Orlando, FL	1997q1	2012q3
Bellingham, WA	1995q1	2012q3	Oxnard-Thousand Oaks-Ventura, CA	1993q1	2012q3
Boston-Cambridge-Quincy, MA-NH	1993q1	2012q3	Palm Bay-Melbourne-Titusville, FL	1998q1	2012q3
Bremerton-Silverdale, WA	1994q1	2012q3	Panama City-Lynn Haven, FL	1996q2	2012q3
Bridgeport-Stamford-Norwalk, CT	1993q1	2012q3	Pensacola-Ferry Pass-Brent, FL	1998q1	2012q3
Cape Coral-Fort Myers, FL	1998q1	2012q3	Peoria, IL	1996q3	2012q3
Carson City, NV	1996q1	2012q3	Phoenix-Mesa-Scottsdale, AZ	1995q1	2012q3
Chicago-Naperville-Joliet, IL-IN-WI	1997q1	2012q3	Pittsfield, MA	1993q1	2012q3
Chico, CA	1995q1	2012q3	Portland-Vancouver-Beaverton, OR-WA	1993q4	2012q3
Cincinnati-Middletown, OH-KY-IN	1997q1	2012q3	Port St. Lucie-Fort Pierce, FL	1998q1	2012q3
Cleveland-Elyria-Mentor, OH	1997q1	2012q3	Prescott, AZ	1995q1	2012q3
Colorado Springs, CO	1998q1	2012q3	Providence-New Bedford-Fall River, RI-MA	1993q1	2012q3
Columbus, OH	1997q2	2012q3	Punta Gorda, FL	1998q1	2012q3
Corvallis, OR	1997q1	2012q3	Redding, CA	1995q1	2012q3
Dayton, OH	1997q1	2012q3	Reno-Sparks, NV	1995q1	2012q3
Deltona-Daytona Beach-Ormond Beach, FL	1998q1	2012q3	Riverside-San Bernardino-Ontario, CA	1993q1	2012q3
Denver-Aurora, CO	1998q1	2012q3	Sacramento--Arden-Arcade--Roseville, CA	1993q1	2012q3
Detroit-Warren-Livonia, MI	1998q1	2012q3	Salem, OR	1994q1	2012q3
Eugene-Springfield, OR	1995q1	2012q3	Salinas, CA	1993q1	2012q3
Flagstaff, AZ	1996q1	2012q3	San Diego-Carlsbad-San Marcos, CA	1993q1	2012q3
Fort Collins-Loveland, CO	1998q1	2012q3	San Francisco-Oakland-Fremont, CA	1993q1	2012q3
Fort Walton Beach-Crestview-Destin, FL	1996q2	2012q3	San Jose-Sunnyvale-Santa Clara, CA	1993q1	2012q3
Fresno, CA	1993q1	2012q3	San Luis Obispo-Paso Robles, CA	1994q1	2012q3
Gainesville, FL	1998q1	2012q3	Santa Barbara-Santa Maria-Goleta, CA	1993q1	2012q3
Grand Junction, CO	1998q2	2012q3	Santa Cruz-Watsonville, CA	1993q1	2012q3
Hanford-Corcoran, CA	1996q1	2012q3	Santa Rosa-Petaluma, CA	1993q1	2012q3
Hartford-West Hartford-East Hartford, CT	1993q1	2012q3	Sarasota-Bradenton-Venice, FL	1998q1	2012q3
Honolulu, HI	1997q4	2012q3	Seattle-Tacoma-Bellevue, WA	1993q1	2012q3
Jacksonville, FL	1996q1	2012q3	Spokane, WA	1995q1	2012q3
Kingston, NY	1994q1	2012q3	Springfield, MA	1993q1	2012q3
Lakeland-Winter Haven, FL	1998q1	2012q3	Springfield, OH	1997q2	2012q3
Las Vegas-Paradise, NV	1995q1	2012q3	Stockton, CA	1993q1	2012q3
Los Angeles-Long Beach-Santa Ana, CA	1993q1	2012q3	Tallahassee, FL	1998q1	2012q3
Madera, CA	1994q1	2012q3	Tampa-St. Petersburg-Clearwater, FL	1997q1	2012q3
Medford, OR	1995q1	2012q3	Tucson, AZ	1995q1	2012q3
Memphis, TN-MS-AR	1998q2	2012q3	Tulsa, OK	1998q2	2012q3
Merced, CA	1993q1	2012q3	Vallejo-Fairfield, CA	1993q1	2012q3
Miami-Fort Lauderdale-Miami Beach, FL	1997q1	2012q3	Vero Beach, FL	1997q1	2012q3
Modesto, CA	1993q1	2012q3	Visalia-Porterville, CA	1993q1	2012q3
Mount Vernon-Anacortes, WA	1996q1	2012q3	Washington-Arlington-Alexandria, DC-VA-MD	1998q1	2012q3
Napa, CA	1993q1	2012q3	Worcester, MA	1993q1	2012q3
Naples-Marco Island, FL	1998q1	2012q3	Yakima, WA	1994q1	2012q3
New Haven-Milford, CT	1993q1	2012q3	Yuba City-Marysville, CA	1995q1	2012q3
NYC-Northern NJ-Long Island, NY-NJ-PA	1998q1	2012q3	Yuma, AZ	1996q1	2012q3

Appendix Table 3. Spatial Concentration of Borrower Types, MSA Level, Aggregate Across Years

	Prime Borrower Category			
	Share of Tracts with No Prime Loans Ever	Share of Tracts with 25% of All Prime Ever	Share of Tracts with 50% of All Prime Ever	Share of Tracts with 755 of All Prime Ever
Minimum	0.0	3.4	10.3	30.4
25th Percentile	0.0	6.2	19.0	40.4
75th Percentile	1.3	10.8	26.7	50.0
Maximum	0.8	20.0	40.0	60.0
Mean	0.1	8.9	23.2	45.3
Standard Deviation	0.1	3.1	5.3	6.7
nobs	96	96	96	96
	Subprime Borrower Category			
	Share of Tracts with No Subprime Loans Ever	Share of Tracts with 25% of All Subprime Ever	Share of Tracts with 50% of All Subprime Ever	Share of Tracts with 75% of All Subprime Ever
Minimum	0.0	3.2	6.9	33.7
25th Percentile	0.1	7.3	20.7	42.2
75th Percentile	4.6	11.7	27.5	51.0
Maximum	21.9	20.0	34.0	60.0
Mean	3.5	9.5	24.2	46.6
Standard Deviation	4.0	3.0	4.9	6.1
nobs	96	96	96	96