Villains or Scapegoats?
The Role of Subprime Borrowers in Driving the U.S. Housing Boom*

James Conklin† W. Scott Frame‡ Kristopher Gerardi§ Haoyang Liu¶

December 20, 2018

Abstract

An expansion in mortgage credit to subprime borrowers is widely believed to have been a principal driver of the 2002-2006 U.S. house price boom. Contrary to this belief, we show that the house price and the subprime booms occurred in different places. Counties with the largest home price appreciation between 2002 and 2006 had the largest *declines* in the share of purchase mortgages to subprime borrowers. We also document that the expansion in speculative mortgage products and underwriting fraud was not concentrated among subprime borrowers.

JEL classification: D14, D18, D53, G21, G38

*We thank Deeksha Gupta (discussant), Brian Melzer, Pedro Gete, Daniel Cooper, Michael Reher, David Echeverry, Randal Verbrugge, Wayne Archer (discussant), Manuel Adelino, Brent Ambrose, Zahi Ben-David, Shaun Bond, John Clapp, Matteo Crosignani (discussant), Richard Crump, Fernando Duarte, Andreas Fuster, John Griffin, Andrew Haughwout, David Humphrey, Michael Lee, Zhenguo Lin, David Lucca, Ronan Lyons, Gonzalo Maturana, Christopher Palmer, Amiyatosh Purnanandam, Eva Steiner, James Vickery, Nancy Wallace, Paul Willen, and seminar participants at the Federal Reserve Bank of New York, the 2018 WFA Conference, the 2018 Notre Dame Roundtable on Housing and Mortgage Markets, and the 2018 Cambridge-UF-NUS Real Estate Finance & Investment Symposium for their valuable comments. We especially thank Daniel Sexton and Akhtar Shah for outstanding research assistance. Haoyang Liu also thanks the generous support from the Fisher Center for Real Estate and Urban Economics at UC Berkeley while part of the paper was written. James Conklin gratefully acknowledges research support from the University of Georgia's Terry-Sanford award. The views expressed in this article are those of the authors and not those of the Federal Reserve Bank of New York, the Federal Reserve Bank of Atlanta or the Federal Reserve System. First version: October 2017.

†Conklin: University of Georgia, jnc152@uga.edu.

‡Frame: Federal Reserve Bank of Atlanta, scott.frame@atl.frb.org

§Gerardi: Federal Reserve Bank of Atlanta, kristopher.gerardi@atl.frb.org

¶Liu: Federal Reserve Bank of New York, haoyang.liu@ny.frb.org.

1 Introduction

The U.S. housing boom and bust of the 2000s has generated an enormous amount of research into its causes and consequences. A central question, the answer to which there remains considerable academic debate, is the role of subprime mortgage lending in that housing cycle. This debate is vitally important because it has significant policy implications along many dimensions, including: access to mortgage credit to facilitate homeownership for marginally qualified borrowers, the regulation of financial institutions that specialize in lending to risky segments of the mortgage market, and macroprudential policies designed to prevent a future crisis from occurring.

There is widespread agreement that subprime mortgage lending increased dramatically during the U.S. housing boom (e.g., Gerardi, Lehnert, Sherlund, and Willen, 2008; Mayer, Pence, and Sherlund, 2009; Mian and Sufi, 2009). There is also agreement about the principal role of rapidly increasing defaults of loans backing privately issued, subprime mortgage-backed securities in provoking the financial crisis that first emerged in late 2007. The rapid deterioration of these subprime mortgage securities and their derivatives resulted in severe disruptions to short-term dollar funding markets and ultimately the global financial system (Brunnermeier, 2009; Dwyer and Tkac, 2009).

The points of disagreement in the academic literature are over the exact causes of the subprime mortgage credit expansion and the role of this expansion in driving the housing boom. This paper focuses on the latter issue and presents new empirical evidence that challenges the prevailing narrative in the literature. That narrative, largely based on the findings presented in Mian and Sufi (2009) and termed the *credit supply view* in Mian and Sufi (2017b), holds that the large expansion in the supply of mortgage credit to subprime borrowers in the early-to-mid 2000s inflated the housing bubble, and thus bears direct culpability to the subsequent financial crisis and deep recession that followed.

Our primary piece of evidence shows that the house price boom and the growth in subprime purchase mortgage lending occurred in completely different parts of the country. Hence, it cannot be the case that the expansion in subprime mortgage credit was a first-order driver of the U.S. housing bubble. Figure 1 illustrates this point graphically. The top panel maps county-level U.S. house price appreciation between 2002 and 2006 using data from the Federal Housing Finance Agency (FHFA), while the bottom panel plots the growth in the share of purchase mortgages to subprime borrowers over the same period. The contrast between the two panels is striking. House price growth was highest in the western part of the country, Florida, and the Northeast Corridor, while the highest growth in subprime purchase lending occurred in areas like the Midwest and Ohio River Valley. Regression analysis performed at the county-level confirms the negative correlation between house price appreciation and growth in subprime purchase market share over this period after conditioning on a rich set of controls and fixed effects. Here we find that a one standard deviation increase in the subprime share of purchase mortgages between 2002 and 2006 is associated with an approximately 4% *decrease* in house price appreciation over the same time period. This negative correlation is shown to be robust to different specifications, time periods, house price measures, and credit score thresholds.

We complement our stylized fact of a negative spatial correlation between the subprime lending boom and house price boom with new evidence showing that subprime borrowers were not an important source of speculative or fraudulent activity. Recent evidence suggests that speculative behavior by real estate investors played a significant role in driving house price growth in many areas of the country during this time period (Haughwout, Lee, Tracy, and Van der Klaauw, 2011; Chinco and Mayer, 2015; Albanesi, De Giorgi, and Nosal, 2017). We show that the dramatic rise of investor purchases during the boom period was almost entirely driven by borrowers with relatively high credit scores (not subprime borrowers). We also look at mortgage fraud during the housing boom, specifically income exaggeration on low documentation loans, inflated valuations on appraisals, and misrepresentation of occupancy status which have all been well-documented in the literature (Piskorski, Seru, and Witkin, 2015; Griffin and Maturana, 2016b; Kruger and Maturana, 2017; Ambrose, Conklin, and Yoshida, 2016). Griffin and Maturana (2016a) present evidence that mortgage fraud played a significant role in driving the housing boom and bust. We show that these forms of mortgage fraud were not concentrated in the subprime purchase mortgage market. Taken together, this evidence reinforces the idea that mortgage lending to subprime borrowers did not play an important role in the types of activities (speculation and mortgage fraud) that have been associated with the U.S. housing boom.

Throughout the paper, we follow Mian and Sufi (2009) and Adelino, Schoar, and Severino (2016) by defining subprime borrowers as those who have FICO scores below 660. However, our results are robust to adopting different subprime FICO thresholds. A primary advantage of using FICO scores, as opposed to income, is that credit scores are not subject to misreporting. Income misrepresentation has previously been shown to have been prevalent during the U.S. housing boom (Jiang, Nelson, and Vytlacil, 2014; Adelino, Schoar, and Severino, 2015; Ambrose, Conklin, and Yoshida, 2016; Mian and Sufi, 2017a). By using credit scores to define subprime borrowers, our intent is to focus on marginal borrowers in terms of credit risk.[1]

Our main analysis focuses on mortgage originations used to finance home purchases. In theory, house prices are determined by marginal buyers in the market and should not be directly affected by individuals refinancing existing mortgages. Thus, we use the term "subprime boom" to refer to an expansion in purchase mortgage originations to subprime borrowers relative to the total amount of home purchase lending (i.e., an increase in subprime purchase market share). While refinances may not have direct effects on house prices, they could exert indirect effects through general equilibrium forces. Thus, we also consider the relationship between growth in the share of subprime refinance loans and house price growth. Consistent with our findings for subprime purchase shares, we find no evidence of a positive correlation between home price growth and growth in subprime borrower refinance shares.

We believe that the lack of positive correlation between house price growth and the share of

[1] In robustness checks, we also use income to define marginal borrowers. Consistent with our subprime results, we find that growth in the share of purchase mortgages to low-income borrowers is negatively correlated with house price appreciation.

purchase mortgages to subprime borrowers sheds doubt upon the *credit supply* view of the U.S. housing boom. However, our results show a robust, *negative* correlation. One potential explanation for the negative correlation is that prospective subprime purchase borrowers, which tend to have lower incomes on average, became increasingly priced out of the boom markets.[2] Recent studies provide suggestive evidence of this "pricing out" effect. For example, Laeven and Popov (2017) show that the housing boom slowed down young households' conversion to homeownership, while Bhutta (2015) and Foote, Loewenstein, and Willen (2016) document that first-time home buying dropped disproportionately for low credit score borrowers during this time. Although these latter two papers find that the share of home purchases by subprime borrowers declined at the national level during the boom, this does not necessarily contradict the credit supply view. Subprime mortgage credit could still explain the housing boom if subprime expansion was concentrated in areas with high house price growth. But, the negative spatial correlation between suprime share growth and house price growth we document in this paper suggests that the expansion of subprime mortgage credit was not a first order driver of the housing boom.

The paper is also related to a recent debate in the literature about the nature of the expansion in mortgage credit during the mid-2000s. In a highly influential study, Mian and Sufi (2009) argued that credit growth was concentrated principally among subprime borrowers. A series of more recent papers has shown, however, that credit growth occurred in a uniform manner across the entire income and credit score distributions (Adelino, Schoar, and Severino, 2016; Foote, Loewenstein, and Willen, 2016; Albanesi, De Giorgi, and Nosal, 2017). In other words, the credit boom was not strictly a subprime phenomenon. While this is an important debate, it does not speak to whether the subprime expansion played a pivotal role in driving home prices to precipitous levels. Thus, the empirical results presented in this paper should be relevant to economists and policy-makers

[2] Some parts of the U.S. saw home prices rise over 100% between 2002 and 2006. For instance, the median home price in Los Angeles rose from \$262,000 in January 2002 to \$600,500 in December 2006. The median home price in Miami rose from \$156,300 in January 2002 to \$353,700 in December 2006. The median home price in Las Vegas rose from \$153,400 in January 2002 to \$303,100 in December 2006.

irrespective of the outcome of the debate over the exact nature of the credit expansion.

In addition, the paper is related to recent empirical evidence showing that an expansion of mortgage credit to marginal borrowers causes house prices to rise. Adelino, Schoar, and Severino (2013) use variation in the conforming loan limits to identify a causal link between the availability of cheaper financing and increasing house prices. Favara and Imbs (2015) show that deregulation stemming from the passage of the Interstate Banking and Branching Efficiency Act in the mid-1990s increased the supply of mortgage credit and put upward pressure on house prices. Di Maggio and Kermani (2017) exploit variation in mortgage credit supply induced by the federal preemption of national banks in the mid-2000s from state-level anti-predatory lending laws and find a significant positive effect of increased credit supply on house prices and employment in the short-run. We do not view our results as being inconsistent with the results of these empirical studies, which indicate that increases in credit supply can cause increases in housing prices. Rather, our evidence suggests that the expansion of credit to subprime borrowers was not a first-order driver of the U.S. housing boom of the mid-2000s.

Recent quantitative models of housing markets have been developed with the intention of trying to explain the U.S. housing boom and bust (e.g., Sommer, Sullivan, and Verbrugge, 2013, Kaplan, Mitman, and Violante, 2015, and Favilukis, Ludvigson, and Van Nieuwerburgh, 2017). These models include credit supply shocks, such as decreasing down payment requirements or increasing debt-to-income ratios. However, the magnitude of the effect of loosening lending standards on house prices is a matter of considerable debate. Our findings help to inform this literature by highlighting the importance of distinguishing between a credit expansion to marginal borrowers from a credit expansion of risky products. Credit expansion to marginal borrowers is unlikely to explain the rapid house price growth experienced in many markets in the early-to-mid 2000s.

Finally, a few other studies have cast doubt over the conventional wisdom of the role played by subprime mortgage lending during the boom/bust period. For example, Ferreira and Gyourko (2011) show that the U.S. foreclosure crisis was characterized by far more prime mortgage foreclosures than subprime foreclosures, and that both types of defaults were principally generated by house price declines rather than mortgage or socio-demographic characteristics. Berkovec, Chang, and McManus (2012) document that the geographic correlation between house price growth and growth in the share of interest-only and negative amortization mortgages is stronger than the correlation between price growth and growth in alternative lending channels. Using quarterly time series data from 20 metropolitan areas, Coleman IV, LaCour-Little, and Vandell (2008) find no evidence that the subprime intensity of recent mortgage lending is related to subsequent housing returns. Brueckner, Calem, and Nakamura (2012) analyzes a model of house prices and lending which shows that favorable lender house price expectations can lead to increased subprime lending, which in turn, can have a feedback effect on house prices.

The rest of the paper is organized as follows. Section 2 describes the data used in our analysis. Section 3 presents evidence that the housing boom and the subprime boom happened in different areas. Section 4 shows that several types of speculative mortgage products were not biased towards subprime borrowers. Section 5 concludes.

2 Data and descriptive evidence

Our primary data comes from two large national loan-level mortgage datasets that each include a large number of borrower and loan characteristics, as well as ongoing loan performance information. The primary dataset comes from McDash Analytics, which is constructed using information from mortgage servicers and covers between 60% and 80% of the U.S. residential mortgage market, including loans securitized by government agencies (Ginnie Mae, Fannie Mae, and Freddie Mac), loans held in bank portfolios, and loans that were packaged into privately issued mortgage-backed securities (PLS). While the dataset is broadly representative, it has somewhat limited coverage of PLS loans that were marketed to investors as "subprime" (Adelino, Schoar, and Severino, 2016). To address this, we supplement the McDash data by adding PLS loans from ABSNet, which covers virtually the entire PLS market. We combine the ABSNet data and McDash data using a matching algorithm to identify and drop all duplicates mortgages. The Online Appendix contains details about the algorithm.

While the majority of our analysis uses the combined dataset, we also conduct some robustness tests using the McDash and ABSNet datasets individually to verify that out results are not driven by the merging procedure. Another potential concern about the McDash data is that its coverage improves over time (Fuster and Vickery, 2014). This poses a particular challenge when analyzing *growth* variables. To verify that our results are not driven by any particular year of data, we vary both the start and end dates of our analysis.

Besides the two mortgage servicing datasets, we also collect standard county level economic data from various sources: home prices (Federal Housing Finance Agency), average wages (Internal Revenue Service), and unemployment rates (Bureau of Labor Statistics). One non-standard county level variable, the subprime share of the population (renters and owners), comes from GeoFRED. This variable, which is derived from the Federal Reserve Bank of New York's Consumer Credit Panel, provides the share of adults in a given county with a credit score below 660.

The main variable of focus in our analysis is the growth in the share of purchase mortgage originations to subprime borrowers at the county level during the U.S. housing boom. Following the recent literature, we classify a borrower as subprime if his/her FICO credit score is below 660 (although we study the sensitivity of our results to this threshold). By using credit score to define subprime, as opposed to income or mortgage contract characteristics, our intent is to focus on marginal borrowers in terms of credit risk. We also follow the literature and focus on the subprime share of home purchase mortgages, which we calculate by dividing the total number of subprime purchase originations in a county by the total number of purchase originations in the county. Our focus on an area's subprime share *growth*, however, represents an important departure from the existing literature. Previous studies have shown that subprime ZIP Codes—defined based on the initial *level* of subprime share in the ZIP Code—experienced greater mortgage credit growth in the boom period (Mian and Sufi, 2009, 2017b). However, as others have pointed out, mortgage credit growth in these subprime areas may actually be driven by prime borrowers (Adelino, Schoar, and Severino, 2016; Albanesi, De Giorgi, and Nosal, 2017). Our subprime share growth rate avoids this issue by directly measuring the *change* in subprime mortgage lending share, and in doing so, better captures the notion of subprime expansion put forward in the credit supply view.

We conduct the bulk of our analysis at the county level for several reasons. First, the vast majority of variation in house prices is across counties rather than within counties.[3] Second, while house prices are available at the ZIP Code level, this is the case for only large ZIP Codes. Thus, a ZIP Code level analysis is unlikely to be representative. Our county level analysis includes approximately 80% of U.S. counties while a ZIP Code level analysis would only include 13%–35% of all U.S. ZIP Codes depending on the house price index used. Nonetheless, performing the analysis at the ZIP Code level does not change the main conclusion of the paper.

Although evidence suggests that there is substantial variation in the start and end of the U.S. housing boom across metropolitan areas (Ferreira and Gyourko, 2011; Bhutta and Keys, 2016), we choose to focus on the period from 2002 to 2006 for two reasons. First, this time frame captures reasonably well the housing boom across markets (see Figure 3 in Ferreira and Gyourko, 2011). Second, this period is consistent with two recent influential papers related to our own (Mian and Sufi, 2009; Adelino, Schoar, and Severino, 2016). We consider slight variations in the sample period and show that the results are robust.

Table 1 reports averages and standard deviations for all variables used in our regressions below. The first column in the table displays statistics for the full sample of counties. We can clearly see that the 2002–2006 period witnessed unprecedented average home price growth of over 42% on a national level.[4] We also observe that, on average, the share of purchase mortgages to sub-

[3] Appendix Table A.1 reports the between- and within-county variation in U.S. house price growth during the 2002–2006 period. The table shows that the overwhelming majority of the variation in house price growth is between counties rather than within counties.

[4] The variable we report here and use in the regressions below is the change in the logarithm of the home price index. A log change of 0.35 corresponds to a 42% increase in home prices.

prime borrowers declined slightly during the housing boom by a little over two percentage points. Not surprisingly, the average county unemployment rate declined and the average county wage increased during the 2002–2006 period. Columns (2)–(5) in the table display summary statistics broken down by county level cumulative home price appreciation (HPA) between 2002 and 2006, with four categories considered: HPA $\geq$ 70%, 40% $\leq$ HPA $<$ 70%, 20% $\leq$ HPA $<$ 40, and HPA $<$ 20%. The share of purchase mortgages to subprime borrowers declined, on average, in the counties that experienced the largest home price gains during the period, while it increased in counties that experienced the smallest gains. This pattern is consistent with the message from Figure 1 and, as we will show below, is quite robust. Unsurprisingly, counties with the strongest house price growth also experienced the strongest wage growth and the largest declines in unemployment.

Figure 2 displays average values of our main variables of interest for county-years broken down by cumulative home price growth between 2002 and 2006. We can see that the subprime share of the underlying *population* (dotted blue line) was approximately constant within each of the house price growth categories during our sample period. In addition, the levels of the subprime population shares are very similar across the categories. For example, the counties with the highest house price growth ($\geq$ 70%) had a subprime population share of about 31%, while the counties with the lowest house price growth ($<$ 20%) had a share of about 33%. The solid red lines in the figure show that counties in the top two house price appreciation categories actually experienced slight declines in the share of purchase mortgages to subprime borrowers, while modest increases in the subprime share of purchases occurred in counties with slower house price appreciation. We also examine *refinance* mortgages to subprime borrowers in the figure (dotted yellow line) and see that they grew markedly in all areas beginning in 2004. In our analysis below we consider this pattern in more detail, although it is very unlikely that it was a major driver of the house price boom. House prices are determined, in part, by housing demand, which is reflected by home purchase activity (not refinance activity). To the extent that the volume of subprime refinance activity grew, it did so likely as a *response* to the house price boom.

3 House price growth and subprime growth

In this section we consider house price growth and its relationship to the growth in the county level share of purchase mortgage originations to subprime borrowers. We begin by looking at the distribution of the share nationally over time. Figure 3 reproduces Figure 2 from Adelino, Schoar, and Severino (2016) using our data. The figure shows the annual share of purchase mortgages originated to high FICO (> 720), medium FICO (between 680 and 720), and low FICO borrowers (< 660). The shares are remarkably constant over the boom period, a conclusion also reached by Adelino, Schoar, and Severino (2016). Put differently, any boom in purchase mortgages to subprime borrowers occurred simultaneously with a boom in purchase lending to prime borrowers at the *national* level.

Although the subprime share of purchase mortgages did not increase over time at the national level, the possibility remains that the subprime share of purchases grew disproportionately in areas that experienced high house price growth. To examine this issue, we return to Figure 1 . The top panel of the figure maps county level house price appreciation in the U.S. between 2002 and 2006 using data from the Federal Housing Finance Agency (FHFA). House price growth was highest in the western part of the country, Florida, and the Northeast Corridor, while the highest growth in subprime purchase lending occurred in areas like the Midwest and Ohio River Valley. The bottom panel of Figure 1 maps growth in the share of subprime purchase lending over the same period. The contrast between the top and bottom panels of Figure 1 is striking. Generally speaking, the areas that experienced house price booms did not experience large increases in the subprime share of purchase originations. In fact, many of the house price boom areas experienced declines in the subprime purchase share over this period. Although house price boom areas experienced an overall expansion in the flow of mortgage credit from 2002 to 2006, Figure 1 provides suggestive evidence that this increase was concentrated in the prime market. Since prime borrowers were becoming a larger — not a smaller — share of buyers in boom-markets, this casts doubt on the hypothesis that subprime borrowers were driving price increases in those markets.

One potential concern with our interpretation of Figure 1 is that we do not control for the initial share of purchase mortgages to subprime borrowers. For example, 30 percent growth in the share of purchase mortgages to subprime borrowers in a market that starts with a 20% initial share (from 20% to 26%), is very different from the same 30 percent growth in a market with initial share of 5% (from 5% to 6.5%). If house price appreciation is positively correlated with initial subprime share, this could drive our finding that subprime purchase share growth was low in high house price appreciation counties. In fact, we find the exact opposite. Figure C.1 in the Online Appendix plots the initial subprime shares for all U.S. counties in 2002 and shows that these shares tended to be lower in the areas that experienced the largest house price increases.

To shed additional insight on these patterns, in Figure 4 we plot the time series of the house prices and subprime purchase shares for the most populous county in each of the four sand states: (i) Maricopa County, AZ; (ii) Los Angeles County, CA; (iii) Miami-Dade County, FL; and (iv) Clark County, NV. Across each of these markets a similar pattern emerges. As house prices increase during the boom period, the subprime share of home purchases decreases. Although we only include time series plots for these four counties, this pattern is common among high house price appreciation counties during this period. Interestingly, there does appear to be an uptick in the subprime purchase share towards the end of the boom period. By this time, however, house prices had plateaued in many of these markets and hence cannot have been the driver of house price growth earlier in the cycle.

While the previous graphical analysis is illuminating, we now turn to regression analysis in order to control for potential confounding factors. We estimate models of the following form:

$$g_i^{02-06}(HPI) = \beta_0 + \beta 1 * g_i^{02-06}(SubShare) + \beta_2 * X_i + \gamma_{state} + \epsilon_i \tag{3.1}$$

where $g_i^{02-06}(HPI)$ is the growth in the FHFA house price index in county i between 2002 and

2006 and $g_i^{02-06}(SubShare)$ is the growth over the same period in the county share of first lien purchase mortgages to borrowers with a FICO score less than 660.[5]

The vector X_i includes level and growth variables that are likely to be correlated with the growth in home prices and purchase lending to subprime borrowers. First, to control for the credit quality of the underlying population of the county (owners and renters), we include the share of the county population with a FICO score less than 660 in 2002. We also include the initial share of purchase mortgages to subprime borrowers in the county. To account for the initial level of overall mortgage activity in the county, we include the total number of purchase loans originated in the county in 2002. Also, since both house prices and the subprime purchase share are likely related to employment and wages, we control for the county level unemployment rate in 2002 using BLS data and county level average wages in 2002 from the IRS.

X_i also includes the following variables that capture *changes* in county economic conditions between 2002 and 2006: the growth in the share of subprime individuals, wage growth, and the change in the unemployment rate. In some specifications we include state fixed effects in order to determine if the correlation between subprime share growth and house price growth differs when utilizing only within-state, county level variation. We weight most regressions by the total number of purchase mortgages observed in our data for a given county in 2002 and 2006 (summing over both years). The weights are included to make our county level observations representative of the underlying loan sample, so that we do not put too much emphasis on rural counties that do not have many loan originations. We also show results from unweighted regressions in our analysis below. Finally, to address potential serial and spatial correlation in the residuals, we cluster standard errors at the state level.

Table 2 presents estimates from equation (3.1). Columns (1)–(3) include regression weights while columns (4)–(6) display unweighted regression results. The first column of Table 2 shows

[5]The $g^{02-06}(\cdot)_i$ variables in equation (3.1) are measured as the natural logarithm of the ratio of the variables over the two time periods. For example, $g^{02-06}(HPI)_i$ is calculated as $\ln\left(\frac{HPI_{2006}}{HPI_{2002}}\right)$.

that county level growth in purchase mortgage originations to subprime borrowers between 2002 and 2006 is negatively correlated with local house price appreciation over the same period. The coefficient is large in economic magnitude as well: A one standard deviation increase in the growth of the subprime purchase share is associated with an 8% *decrease* in house price appreciation (0.179×-0.444). In column (2) we add the additional covariates and find that the conditional correlation between local house price growth and the growth in subprime purchase share falls in absolute magnitude but remains significantly negative. Column (3) adds state fixed effects so that the correlation is estimated using only within-state variation in county level house price growth and subprime purchase share growth. The conditional correlation between local house price growth and the growth in subprime purchase share falls further but remains significantly negative. Estimating the same regressions without using weights for the total number of loans in the county yields qualitatively similar results. The (absolute) magnitudes of the coefficients are lower, but remain negative and significantly different from zero.[6] The key takeaway from Table 2 is that U.S. house price appreciation was associated with a *decline* in the subprime purchase market share between 2002 and 2006. This evidence is inconsistent with the narrative that an expansion of mortgage credit to subprime borrowers fueled the U.S. housing boom.

We estimate a series of regressions to explore the robustness of the negative relation between house price appreciation and subprime purchase share growth. First, as pointed out by Ferreira and Gyourko (2011) and Bhutta and Keys (2016), there was significant geographic variation in the timing of the boom-bust period. Therefore, in Table 3 we vary the sample period over which the growth variables are measured. The first column reproduces the results in column (3) of Table 2 , which corresponds to the specification with covariates and state fixed effects. The remaining columns show results for the same specification but change the period over which house price appreciation and growth in the subprime purchase share is measured. Regardless of the period

[6] Although we weight our regressions by the number of loans in a given county, a potential concern is that our results are being driven by small counties. As a robustness check, we restrict our analysis to counties with at least 50,000 tax returns in 2002 and find that the results remain qualitatively unchanged.

considered, the relation between county level subprime share growth and house price appreciation is consistently negative and significantly different from zero.

Second, we check if our results are sensitive to our definition of subprime borrowers. Although a FICO score below 660 is commonly used to identify subprime borrowers, in Table 4 we adopt alternative credit score cutoffs. Columns (1)–(3) re-estimate equation (1) with subprime share calculated as the fraction of purchase mortgages in a county-year where the primary borrower's FICO score is less than 620; and columns (4)–(6) use a credit score cutoff of 580. In all columns of Table 4 house price appreciation is negatively related to subprime purchase share growth.

Third, in Table 5 we consider two alternative measures of the growth of subprime purchase mortgage activity. A concern about focusing on the growth rates in the shares of subprime originations is that the results could be driven by small initial rates of originations in small, rural counties. We attempted to address this issue by including county level subprime purchase shares in 2002 in our covariate set, but this may not completely address the issue. Thus, we consider the change (rather than the growth) in subprime purchase shares (columns (1)–(3)), and the growth in the number of subprime purchase loans rather than the growth in the share of subprime purchase mortgages (columns (4)–(6)). Both measures of subprime purchase mortgage activity are negatively related to house price appreciation at the county level, although the coefficient estimate for the latter measure is not statistically different from zero.[7]

Fourth, in Table 6 we consider the growth in the share of subprime refinance originations. As explained above, house prices are not directly affected by individuals refinancing existing mortgages but instead are determined by marginal homebuyers. However, there could be general equilibrium effects such that a boom in subprime refinance loans indirectly led to house price growth.[8]

[7]Note, though, that growth in prime lending is implicitly contained in the dependent variable in all regressions that involve subprime share. However, in columns (4) – (6) we do not control for growth in prime lending. If we also condition on growth in prime loans (Δ log(# Prime Purchase Loans)), the coefficients on Δ log(# Subprime Purchase Loans) are negative and become statistically significant.

[8]For example, the huge increase in mortgage equity withdrawal via cash-out refinances during the 2000s that has been documented in the literature could have led to an increase in consumption and an increase in local economic activity including household employment and income, which could have then put upward pressure on house prices.

The results in Table 6 suggest that this was not the case. We find little evidence of a significant, positive correlation between growth in the share of subprime refinance loans and house price appreciation at the county level.

Finally, we use an alternative method of defining marginal borrowers. In using credit scores to define subprime borrowers, our intent is to focus on marginal borrowers in terms of credit risk. But, income is another natural way to define marginal borrowers. Indeed the debate between the credit supply view and the more recent narrative of the housing boom focuses on marginal borrowers as measured using both credit scores and income levels. In Table 7 , we investigate the relationship between house price appreciation and low income share growth in purchase originations. Since borrower income is not available in our primary mortgage dataset, we turn to loan level information disclosed by financial institutions under the Home Mortgage Disclosure Act (HMDA). The HMDA data is widely considered to have the most comprehensive coverage of home lending activity in the United States, and is crucial for our purposes because borrower income is disclosed for each loan.[9] However, the drawback of the data is that it includes little information on loan level risk characteristics (e.g., FICO score), so we are unable to use it in our main analysis. Table 7 reports coefficient estimates from regression specifications where we replace subprime share growth with three alternative measures of low income share growth from the HMDA data (each measure is described in the notes that accompany the table). In all three columns, growth in the share of purchase mortgages to low income borrowers is significantly negatively correlated with house price appreciation. Thus, we conclude that an increase in lending to marginal borrowers, defined in terms of both credit scores and income levels, does not appear to be a first order driver of the housing boom.

Taken together, the analysis indicates that areas that saw an expansion of home purchase fi-

[9] As noted by others (Avery, Brevoort, and Canner (2007), Mian and Sufi (2017a), and Adelino, Schoar, and Severino (2016)), income is likely measured with error in the HMDA data. For example, lenders do not verify the accuracy of income on low-documentation loans. Alternatively, a borrower need not document all sources of income on the mortgage application. Noting these limitations, we proceed with the analysis using income information from HMDA.

nancing to subprime borrowers did not experience large house price booms. This conditional correlation has important implications for the narrative surrounding the role of subprime borrowers in the recent financial crisis. The traditional credit supply narrative posits that a reallocation of credit to subprime borrowers was responsible for the boom in house prices.[10] However, this requires a positive correlation between house price growth and the share of purchase mortgages to subprime borrowers. Our analysis, by contrast, uncovers a robust negative correlation.

4 Subprime borrowers, speculative activity, and mortgage fraud

In this section we investigate whether subprime borrowers played a major role in the speculative activity that has been linked to the housing boom. Specifically, we examine whether mortgage debt used to finance investment properties flowed disproportionately to subprime borrowers, particularly in housing boom areas. A boom (non-boom) area is defined as a county that experienced at least (less than) 20% house price appreciation from 2002 to 2006, which corresponds approximately to the median county house price growth in our sample. Our results show that in both boom and non-boom areas, subprime borrowers played a minor role in speculative activity. Recent studies also provide evidence of a positive spatial correlation between the incidence of mortgage fraud and house price appreciation. Using three different proxies for mortgage fraud, we find no evidence that mortgage fraud was concentrated among subprime borrowers. The results in this section reinforce our findings above that an expansion in lending to subprime mortgage borrowers was not a first-order driver of the U.S. housing bubble.

[10]Several studies in the related literature conduct their analysis at the ZIP Code level. As we discussed in Section 2 , we perform our analysis at the county level because there are drawbacks associated with a ZIP level analysis in the context of our study. But, to ensure our results are not driven by our unit of analysis, we also display regression estimates using ZIP Code level variation rather than county level variation in Table A.2 in the Online Appendix. Consistent with our county level results, we find no evidence that subprime share growth is positively related to house price growth.

4 .1 Investor mortgages

Recent evidence suggests that real estate investors, particularly speculators, played a large role in the U.S. housing boom of the mid-2000s. Haughwout, Lee, Tracy, and Van der Klaauw (2011) identify investors using the number of first-lien mortgages on an individual's credit report and find that investors comprised roughly 50% of mortgage purchase originations in boom states. Moreover, the authors provide evidence that mortgage durations decreased significantly for investors during the boom period, which suggests the composition of investors shifted from buy-and-hold investors to flippers. The authors also find that the investor share of delinquencies spiked during the housing bust, particularly in states that experienced high house price appreciation during the boom. Albanesi, De Giorgi, and Nosal (2017) identify investors in the same manner and find that much of the increase in mortgage defaults during the financial crisis was attributable to real estate investors. In addition, Chinco and Mayer (2015) show that "out-of-town" speculators played an important role in causing house prices to appreciate in the hottest markets during the boom period, including Phoenix, Las Vegas, and Miami.

Our definition of an investment property includes both investment properties and second homes. However, several studies document the incidence of mortgage fraud through misrepresentation of owner-occupancy status on mortgage applications (see Section 4 .3 below). Thus, our measure of the investment share, which is calculated based on information reported on the loan application, should be considered a lower-bound for the true market share of real estate investors. As Figure 5 shows, our investor share estimate is considerably lower than that reported by Haughwout, Lee, Tracy, and Van der Klaauw (2011). This is most likely due to owner occupancy misreporting on the loan application.

We examine whether purchase mortgages financing investment properties were made disproportionately to subprime borrowers. In the top two panels of Figure 5 we plot the share of mortgage originations financing investment properties across boom and non-boom areas, respectively (black solid line). Here, a county is defined as a boom area if it experienced at least 20% house price appreciation between 2002 and 2006. We also plot the percentage of total originations that are for prime and subprime investors (dotted red and blue lines), respectively. Over time, and consistent with prior research, we see that the investor share of purchase mortgages increased significantly during the boom and was greater in areas that experienced higher house price growth. By the end of 2006, approximately 15% of purchase originations in non-boom areas and 18% of originations in boom areas were for investment properties. Furthermore, the overall increase in the investor share is attributable almost entirely to buyers with higher credit scores, regardless of area house price appreciation. In the bottom panels of Figure 5 , we plot the prime and subprime investor shares separately across boom and non-boom areas.[11] In both panels, the subprime investor share is flat over time, while the prime investor share increases markedly. Although recent evidence suggests investors played a large role in the housing boom and bust, Figure 5 suggests that these investors tended to be prime borrowers. This casts further doubt on the idea that an increase in the supply of lending to marginal borrowers fueled the housing boom.

4 .2 Low-documentation mortgages

The role of income misrepresentation during the housing boom has received considerable attention in the literature. Full-documentation (full-doc) loans involve the lender meticulously documenting the borrower's source of income and assets to determine the borrower's ability to repay the debt. However, low-documentation (low-doc) mortgages, which became very prevalent during the housing boom, require little (if any) documentation of the borrowers' income and assets. Thus, a low-doc loan could have been potentially used to inflate borrower income on loan applications to obtain a larger loan than would otherwise have been available. Indeed, recent studies suggest that

[11]The key difference between the top and bottom panels is the denominator. In the top left (right) panel, the denominator is the total number of PLS purchase originations in boom (non-boom) areas. In the bottom left (right) panel, the denominator is either the total number of subprime originations or the total number of prime originations in boom (non-boom) areas.

mortgage fraud related to the misrepresentation of borrower income was a common occurrence during the mid-2000s (Blackburn and Vermilyea, 2012; Jiang, Nelson, and Vytlacil, 2014; Ambrose, Conklin, and Yoshida, 2016; Mian and Sufi, 2017a,b). Moreover, Mian and Sufi (2017a) claim that fraudulently overstated income in the boom was more severe for "marginal borrowers that were traditionally denied credit" (p. 1833).

We examine whether low-doc purchase mortgages flowed disproportionately to subprime borrowers and whether this varied with local house price growth. The top panel of Figure 6 plots the total proportion of low-doc loans in boom and non-boom counties (black, solid lines). We also break the low-doc share into its prime and subprime components (dotted blue and red lines). Several important facts emerge. First, there is a large expansion in the low-doc share of mortgage originations in both boom and non-boom areas. Second, although the rapid growth in the low-doc share is not confined to boom areas, the low-doc share of purchase originations is clearly higher in boom areas (by approximately 10 percentage points). Third, in both boom and non-boom areas, the rapid expansion in low-doc share is driven by prime borrowers.

In the bottom two panels of Figure 6 we plot the low-doc share of originations to prime and subprime borrowers separately. Note that the low-doc share of both prime and subprime purchase mortgage originations increased over time in both boom and non-boom markets. But because subprime loan originations were a relatively small share of the overall market, the sharp increase in overall low-doc share in the top panels is primarily driven by prime borrowers. Assuming that low-doc loans are sometimes used to misrepresent income, this does not support the conjecture by Mian and Sufi (2017a) that income overstatement was more severe for marginal borrowers that were traditionally denied credit.

4 .3 Owner occupancy fraud

Figure 5 is consistent with recent empirical evidence documenting that real estate investors played a large role in the U.S. housing boom and bust. However, the figure likely understates the importance of investors and speculative behavior in the market, since it assumes truthful reporting of occupancy status. A couple of recent papers have documented systemic misreporting by mortgage borrowers about their intentions to occupy the property in order to obtain more favorable loan terms during the mid-2000s housing boom (Piskorski, Seru, and Witkin, 2015; Griffin and Maturana, 2016b).

Here we investigate whether owner occupancy fraud was more prevalent for subprime purchase mortgages across boom (HPA $\geq$ 20%) and non-boom (HPA $<$ 20%) markets. Following Griffin and Maturana (2016b), we use Lewtan's Homeval data, which includes an indicator for suspected occupancy misreporting. To create this variable, loans in the ABSNet data are matched to public records data for property sales. As a reminder, ABSNet provides extensive coverage of the PLS market. The occupancy status reported in ABSNet is compared to the occupancy status reported in the public records.[12] There are some limitations to using the occupancy misreporting flag. First, this field is only available for a subset of the ABSNet loans due to difficulties merging mortgage originations with public records. Second, the occupancy fraud indicator is only available for loans that were still being serviced in 2012, potentially creating some survival bias.[13]

Figure 7 presents the rate of estimated occupancy fraud across boom and non-boom counties using the PLS sample. There are a few notable patterns. First, the figure shows that the rate of occupancy fraud trended down over time in both boom and non-boom areas. This downward trend in both boom and non-boom areas is consistent with the downward trend at the national level

[12]Although Griffin and Maturana (2016b) use ABSNet mortgage data in their analysis, they perform their own merge with public record files using DataQuick's Assessor and History files to identify occupancy misreporting. Thus, our measure is not identical to theirs, even though they are created in a similar manner.

[13]While evidence exists suggesting that subprime and prime loans end in foreclosure at similar rates once current LTV and calendar time are accounted for (Ferreira and Gyourko, 2015), prime borrowers may have been more likely to exit the sample through refinancing in the post-boom period.

reported in Griffin and Maturana (2016b).[14] Second, the top panels of the figure show that prime borrowers contributed much more to occupancy fraud than did subprime borrowers in both boom and non-boom counties. In the bottom panels, we see that with the exception of 2002, there was actually a higher rate of occupancy fraud among prime borrowers compared to subprime borrowers in both boom and non-boom areas. We are careful not to interpret these results too strongly due to the data limitations, which may explain why the incidence of occupancy misreporting in our data is significantly higher than in previous studies. However, this does provide some suggestive evidence that occupancy misreporting was more common among prime loans.

4 .4 Appraisal inflation

Although appraisals are supposed to be unbiased estimates of market value, an overwhelming amount of evidence from the boom-bust period suggests that significant appraisal inflation took place and may have played a role in inflating home prices (Ben-David, 2011; Agarwal, Ambrose, and Yao, 2014; Shi and Zhang, 2015; Ding and Nakamura, 2016; Calem, Lambie-Hanson, and Nakamura, 2017; Conklin, Coulson, Diop, and Le, 2017; Kruger and Maturana, 2017, among others). Moreover, research has shown that appraisal inflation is more prevalent for mortgages to financially constrained borrowers (Agarwal, Ben-David, and Yao, 2015).

In this section we ask whether appraisal inflation was more concentrated among subprime borrowers, who are more likely to be financially constrained. Following Kruger and Maturana (2017), we identify an appraisal as fraudulent if the difference between the appraised value and the estimated value at origination from Lewtan's (ABSNet) proprietary automated valuation model (AVM) is at least 20% above the average of these two value estimates.[15] Because both the appraisal

[14]As discussed above, our measure of occupancy fraud is slightly different from the one used in Griffin and Maturana (2016b). The incidence of occupancy fraud in Lewtan's Homeval data is higher than the level reported in Griffin and Maturana (2016b).

[15]Griffin and Maturana (2016b) define an appraisal as overstated if the appraisal is more than 20% above the AVM value. Our results are materially unchanged if we use their measure (as opposed to the average of the AVM and the appraisal). See Griffin and Maturana (2016b) for a more detailed discussion of Lewtan's automated valuation model.

and the AVM are *estimates* of the true value of the collateral, an AVM estimate above the appraisal may not actually be indicative of collateral misreporting. In fact, Demiroglu and James (2016) argue that comparisons of AVM estimates relative to appraisals should not be used as indicators of collateral misreporting because: (i) both appraisals and AVMs contain estimation errors and (ii) because appraisals and AVMs are not generally observed for non-funded loans in standard mortgage datasets. However, Kruger and Maturana (2017) provide strong evidence suggesting that intentional misrepresentation is likely to explain high appraisals relative to AVM values. Thus, we believe that our measure of appraisal fraud is generally capturing intentional inflation (fraud) by the appraiser.

The top panel of Figure 8 plots the share of PLS purchase mortgages which we flag as having fraudulent appraisals in boom and non-boom areas. The incidence of appraisal fraud does not appear to increase over time. In boom-areas, the overall share remains steady over time, while in non-boom areas the share decreases through the end of 2004 and then picks back up some. In the bottom two panels of Figure 8 , we delineate the shares of mortgages with appraisal fraud to prime and subprime borrowers separately. In both boom and non-boom areas the appraisal fraud rate for prime and subprime loans tracks very closely. This finding suggests that appraisal fraud was not concentrated in loans to subprime borrowers.

5 Conclusion

A widely held narrative of the U.S. housing boom and bust, termed the *credit supply view* by Mian and Sufi (2017b), holds that this cycle resulted from a credit expansion to marginal borrowers, which fueled an unsustainable rise in housing prices that ultimately ended in the mortgage and broader financial crises. This paper presents empirical evidence that is inconsistent with this view. The key finding is that the housing price boom and the subprime purchase mortgage boom occurred in different locations. Specifically, counties that experienced high house price growth were those that experienced a disproportionate *decline* in credit to subprime borrowers. One explanation for this finding, which is supported by recent research findings, is that lower-income and/or subprime borrowers were largely priced out of the boom markets.

We complement our main finding of a negative spatial correlation between the subprime lending boom and house price boom with new evidence showing that subprime borrowers were not an important source of speculative or fraudulent activity. Specifically, we show that the dramatic rise of investor purchases during the boom period was almost entirely driven by borrowers with relatively high credit scores. In addition we show that subprime purchase mortgage borrowers were not an important source of three types of fraudulent activity, income exaggeration on low documentation loans, owner occupancy fraud, and appraisal inflation, which have been tied to the housing boom by several influential recent studies.

Our paper contributes to the "new narrative" that rapid U.S. house price appreciation during the 2000s was mainly driven by prime borrowers. Thus, policy prescriptions intended to limit access to credit for marginal borrowers are unlikely to prevent a future housing boom.

6 Appendix

Appendix A ZIP code level analysis

In the main text, we conduct the analysis at the county level. There are a few reasons for this. First, the vast majority of variation in house prices is across counties rather than within counties. Table A.1 reports the between- and within-county variation in house price growth. Regardless of the house price index used, the overwhelming majority of the variation in house price growth is between-county rather than within-county. In contrast, between-county variation in subprime share growth is similar in magnitude to within-county variation. Second, while house prices are available at the ZIP Code level, this is the case for only large ZIP Codes. Thus, the representativeness of the sample would be questionable if we conducted the analysis at the level of the ZIP Code. Whereas our county level analysis includes approximately 80% of the counties in the United States, the limited availability of ZIP Code level house price information reduces our coverage to 13% - 35% of ZIP Codes depending on the index used.

Despite the drawbacks of a ZIP code level analysis, in this appendix we display results from regression specifications at the ZIP Code level. Many of the previous papers in the literature have focused on ZIP Code level variation. For example, Mian and Sufi (2009), Adelino, Schoar, and Severino (2016), and Foote, Loewenstein, and Willen (2016) investigate the relationship between credit growth, FICO scores, and income at the ZIP Code level. Thus, it is important to make sure that our results are robust to this change. In addition, by estimating regressions at the level of the ZIP Code we are able to include county fixed effects and estimate the relationship between house price appreciation and the growth in subprime purchase shares using only within-county variation. For the ZIP Code level analysis, we use loan level data from CoreLogic, another leading source of PLS mortgage data. We merge the CoreLogic data with the McDash data and eliminate duplicates using the same procedures described in Appendix B. Note that we use the Lewtan ABSNet data, as opposed to the CoreLogic data, in our main analysis because Lewtan's HomeVal data allows us to create the fraud indicators used in Section 4 , which are not available in the CoreLogic dataset.

We display regression results at the level of the ZIP Code below in Table A.2. We show results for two different, commonly used, ZIP Code level house price indexes, CoreLogic (columns (1)–(4)) and FHFA (columns (5)–(8)). The coverage of each index differs, as the FHFA index is populated for over 14 thousand ZIP Codes while the CoreLogic index covers a little over 5 thousand ZIP Codes.[16] For each HPI, we display four specifications: A simple univariate specification (columns (1) and (5)); a specification with county level controls (columns (2) and (6)), which corresponds exactly to the controls included in the county level regressions displayed in the main text; a specification with county level controls and state fixed effects (columns (3) and (7)); and a specification with county fixed effects (columns (4) and (8)). The only control variable included in the county fixed effects specification is the average annual wage in a ZIP Code.

The results from the ZIP Code level analysis are broadly consistent with the county level results presented in the main text. The coefficients reported in columns (1)–(3) and (5)–(7) are of the same sign and similar in magnitude to the corresponding specifications reported in Table 2 . The relationship between the growth in subprime purchase shares and house price appreciation is negative and statistically significant. The specifications that include county fixed effects (columns (4) and (8)) and thus, use only within-county variation to estimate the relationship, yield mixed results. The sign of the coefficients is negative, although the magnitude is small, and only statistically significant in the specification that uses the CoreLogic house price index. But again, we stress that there is relatively little variation in house price appreciation within counties, which likely explains the reduced magnitude and significance of the coefficients. Overall, using ZIP Code level variation, there is no evidence of a positive relationship between the growth in subprime purchase

[16]There are some important methodological differences between the indexes. The CoreLogic index is a repeat-sales index estimated at the monthly frequency and is constructed using all single-family, residential properties. The ZIP Code level FHFA repeat-mortgage transaction index is constructed annually based on mortgages purchased or securitized by Fannie Mae and Freddie Mac. The methodology is the same as a repeat-sales index, however, it includes single-family residence valuations on both purchase and refinance mortgage transactions.

shares and house price appreciation, but there is evidence of a negative relationship.

Table A.1. Between- and Within-County Standard Deviation of House Price and Subprime Share Growth, 2002 - 2006

	Zip Codes	Counties	Mean	SD	Between county (SD)	Within county (SD)
	(1)	(2)	(3)	(4)	(5)	(6)
Δ log(home price index) [CoreLogic]	5,350	799	0.392	0.225	0.192	0.050
Δ log(home price index) [FHFA]	14,481	2,324	0.289	0.191	0.146	0.048
Δ log(subprime share of purchase loans)	14,499	2,325	-0.014	0.423	0.354	0.349

Notes: This table reports descriptive statistics at the Zip Code level for house price growth and subprime share growth. Zip Code level variation is broken into between and within county variation. Column (1) reports the number of Zip Codes included in the calculations of the descriptive statistics, while column (2) is the number of counties that encompass those Zip Codes. Columns (3) and (4) report the mean and standard deviation of the variables at the Zip Code level. Columns (5) and (6) break the Zip Code level variation into its between county and within county components, respectively.

Table A.2. Growth of U.S. ZIP Code Level House Prices and the Share of Purchase Mortgages to Subprime Borrowers

Dependent Variable:	Δ log(HPI) 2002-2006							
HPI Source:	CoreLogic				FHFA			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ log(Subprime Share)	-0.33***	-0.18*	-0.21**	-0.02***	-0.21***	-0.12***	-0.03***	-0.01
	(0.093)	(0.095)	(0.101)	(0.005)	(0.043)	(0.049)	(0.010)	(0.009)
Covariates	N	Y	Y	Y	N	Y	Y	Y
State FE	N	N	Y	N	N	N	Y	N
County FE	N	N	N	Y	N	N	N	Y
Observations	5,325	5,325	5,326	5,092	14,326	14,313	14,315	13,744
Adjusted R^2	0.07	0.2	0.33	0.96	0.08	0.32	0.88	0.95

Notes: This table reports estimates from a regression of change in the log ZIP Code level house price index from 2002 to 2006 on the contemporaneous change in log ZIP Code subprime share of purchases. Observations are at the ZIP Code level and regressions in columns (1)–(4) use the CoreLogic ZIP Code level house price indices while columns (5)–(8) use the FHFA ZIP Code level house price indices. All regressions are weighted by the total number of loans in the ZIP Code (i.e. summed across both years, 2002 and 2006). Subprime share is calculated based on the fraction of first-lien purchase mortgage originations in a ZIP Code with a FICO score less than 660. The covariates in columns (2)–(3) and (6)–(7) include both level and change variables. The level variables (measured in 2002) are the percentage of the county population that is subprime, the county subprime share of mortgages, the number of loans originated, county average wages (from the IRS), and the county level unemployment rate. The change variables include: change in county level log wages, change in log(% of county subprime population), and the change in county level unemployment. The only covariate included in columns (4) and (8) is the ZIP Code average wage (from the IRS) Robust standard errors are in parentheses and are clustered at the county level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Appendix B Procedure for merging ABSNet with LPS

The following is the procedure used to construct a combined dataset from ABSNet and LPS, removing all duplicates:

1) Remove duplicates within LPS. There are a smaller number of duplicates within the LPS data. We identify them by loans with identical closing month, loan purpose (refinance or purchase), loan amount, credit score, and ZIP Code. For non-GSE loans with valid 5-digit ZIP Codes, a match of all 5 digits is required. For GSE loans with 3-digit ZIP Codes, a match of the first 3 digits is required[17].

2) Similar to step 1), remove duplicates within ABSNet. The matching criteria are also identical closing months, loan purposes, loan amounts, credit scores, and ZIP Codes.

3) Remove duplicates between ABSNet and LPS.

   3-A) Identify and remove duplicates with identical closing months, loan purposes, loan amounts, credit scores, and ZIP Codes.

   3-B) Among all criteria in 3-A), matching credit scores is perhaps the strictest one given that different credit bureaus have different formulae. The recorded credit scores in LPS and ABSNet could be different simply because they are from different sources. However, if we completely drop the matching credit score condition, the constraints are somewhat too weak. Thus in this step, we replace the identical credit score condition with matching LTVs and ARM/FRM. In other words, the criteria are identical closing months, loan purposes, loan amounts, ZIP Codes, ARM/FRM, and LTVs.

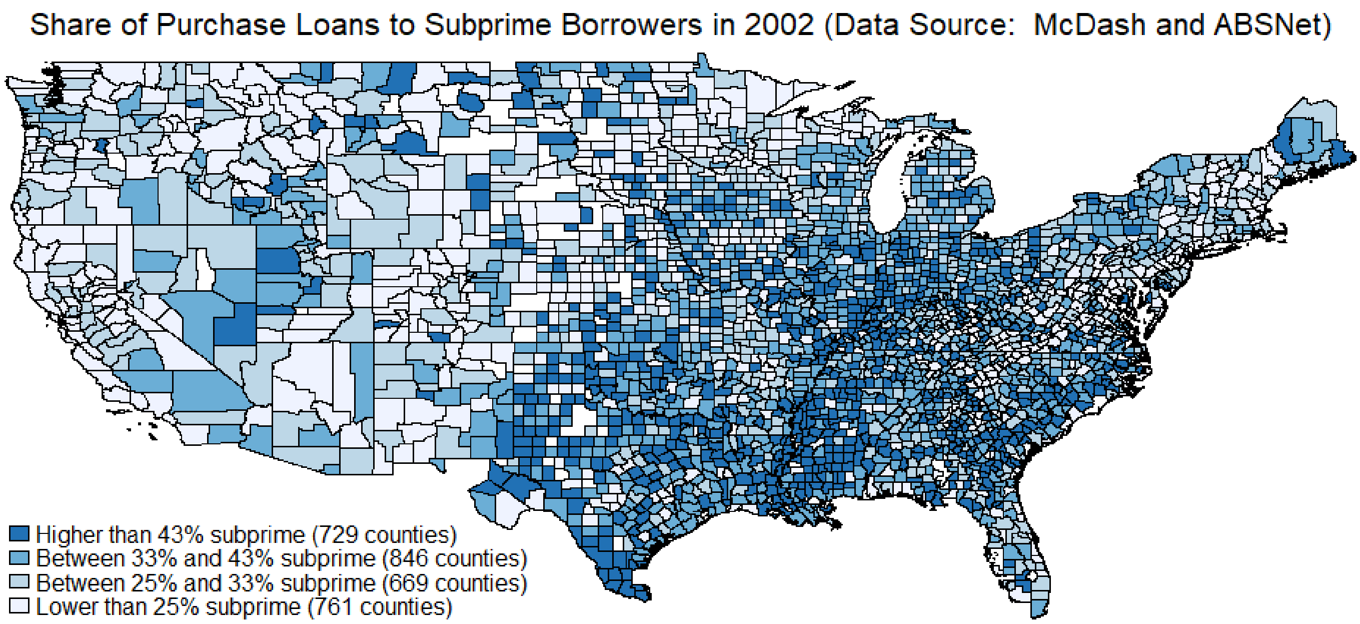
   3-C) Another variable in 3-A) or 3-B) that might require a fuzzy match is closing month. There are multiple important dates in a real estate transaction. The closing date could

[17] The LPS data that is available to us only has 3-digit ZIP Codes for GSE loans.

be misrecorded. To address this, we allow a three month window to identify closing month. After relaxing the closing month condition from 3-A), the full criteria are matching loan purposes, loan amounts, credit scores, ZIP Codes, and closing month one month before or after the other loan. Similarly, to make closing month a fuzzy match for 3-B), the full criteria are matching loan purposes, loan amounts, ARM/FRM, LTVs, ZIP Codes, and closing month one month before or after the other loan.

Appendix C   Additional figures and tables

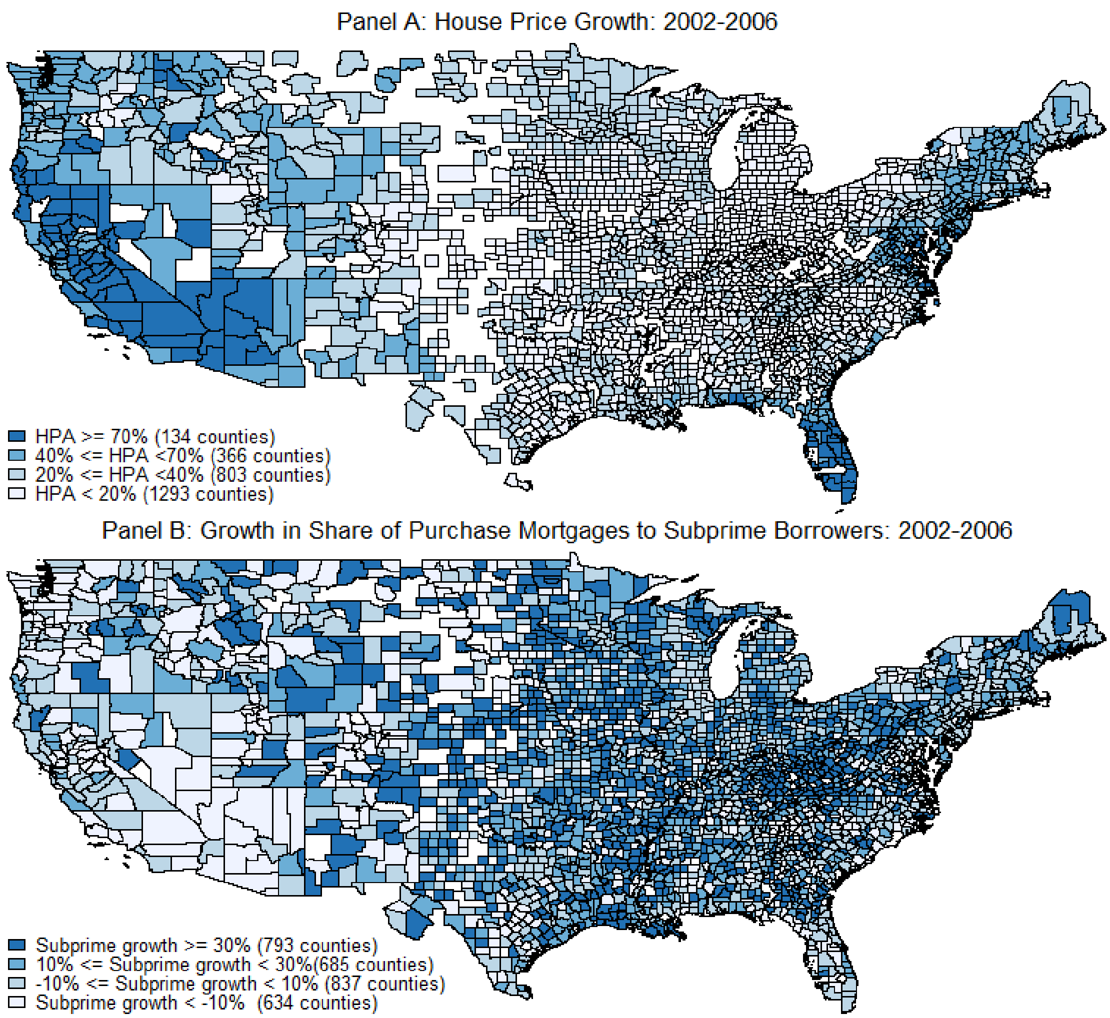
Figure C.1 . Purchase Subprime Share in 2002 (McDash + ABSNet)



Source: McDash, ABSNet, authors' calculations. This figure uses different colors to illustrate variations in initial share of purchase mortgages to subprime borrowers in 2002. The loan sample is a merged sample of first-lien purchase mortgages between McDash and ABSNet, excluding all duplicates. Subprime expansion is defined as the percentage changes in share of loans to subprime borrowers, defined as borrowers with FICO scores under 660.

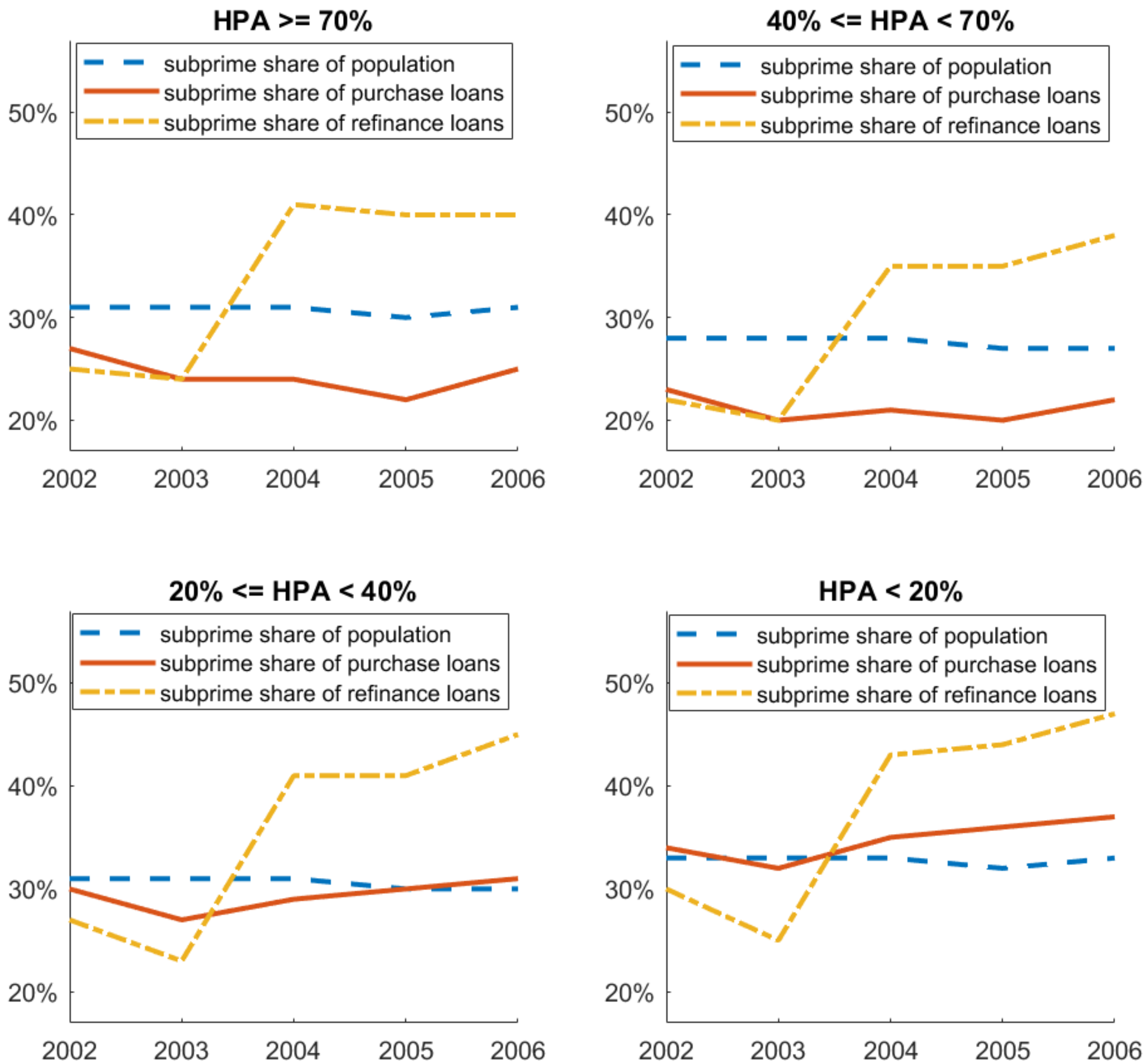
7 Figures and tables

Figure 1 . Growth of U.S. County-Level House Prices and the Share of Purchase Mortgages to Subprime Borrowers



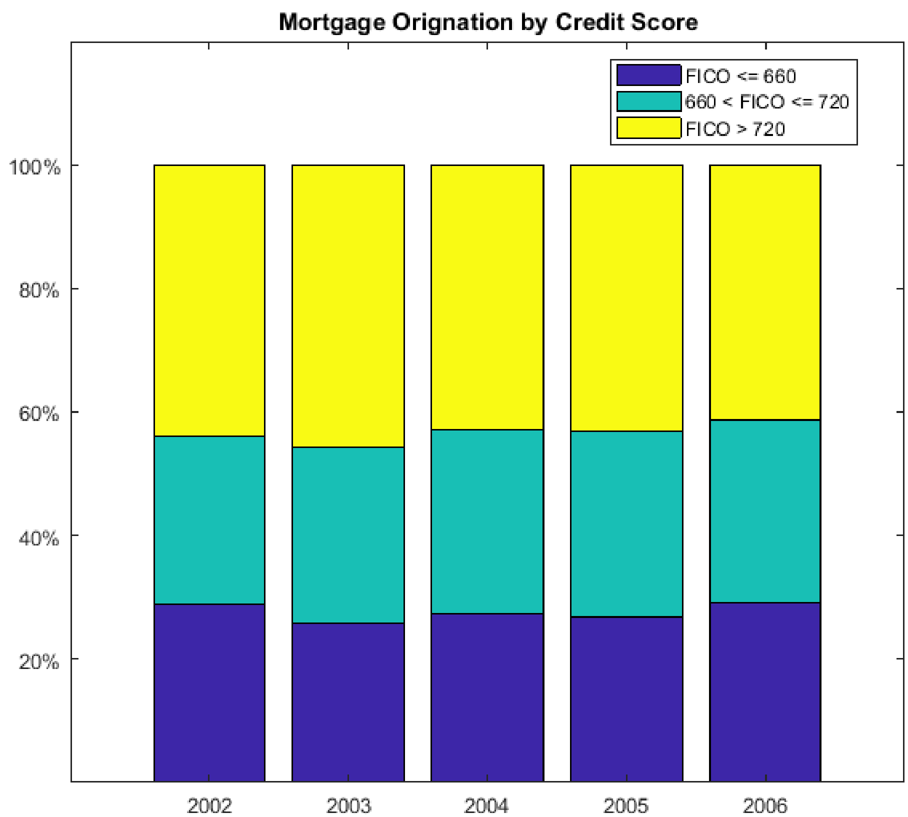
Source: FHFA, McDash, ABSNet, authors' calculations. *Top Panel:* home price appreciation between 2002 and 2006. Color indicates HPA ranging from light blue (low HPA) to dark blue (high HPA). *Bottom Panel:* change in share of first-lien purchase mortgages to subprime borrowers. Color indicates subprime growth ranging from light blue (subprime contraction) to dark blue (subprime expansion). The loan sample from the bottom panel is a merged sample of first-lien purchase mortgages from the McDash and ABSNet datasets after excluding all duplicates between the two data. The detailed merging procedure is described in Appendix B.

Figure 2 . Share of U.S. County Subprime Population and Home Purchase Mortgage Originations by House Price Growth: 2002–2006



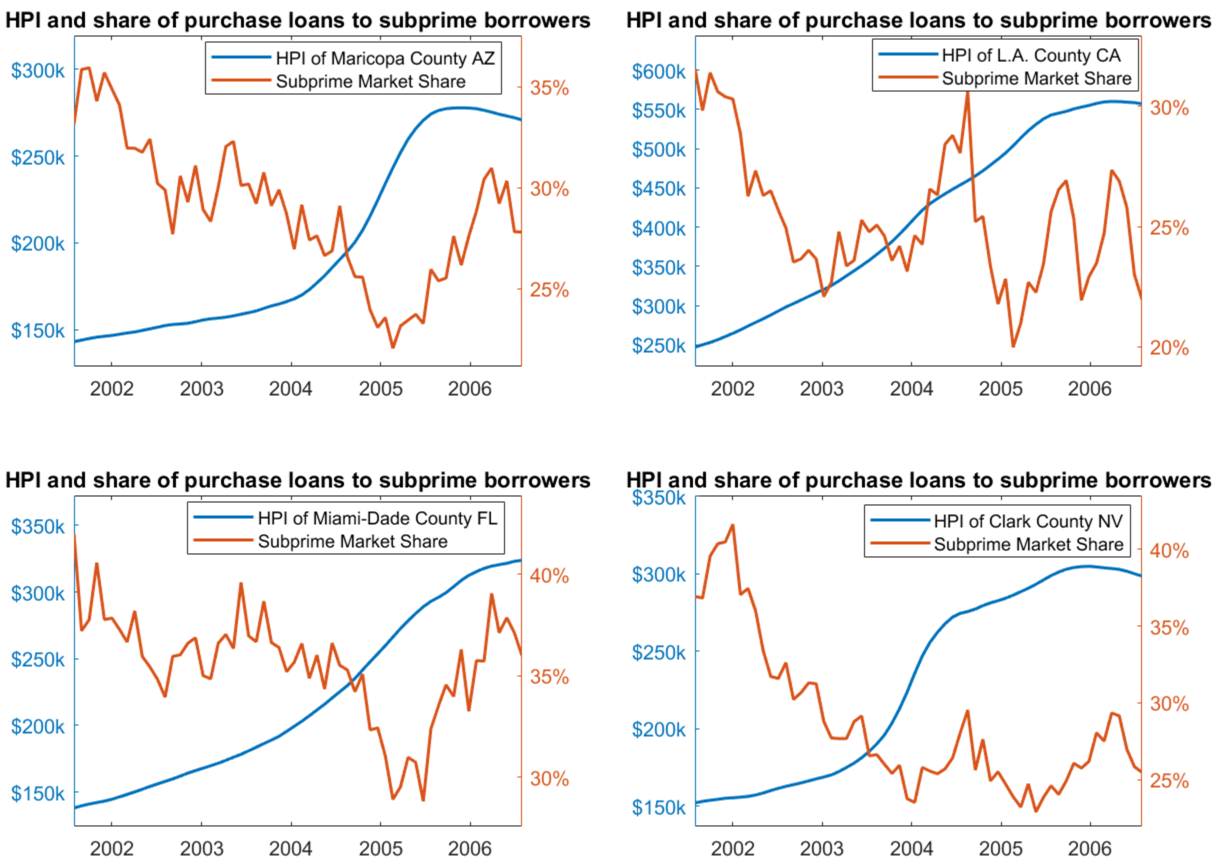
Source: McDash, ABSNet, authors' calculations. This figure plots the subprime shares of population and mortgages in different counties. Counties are sorted according to their cumulative home price appreciation between 2002 and 2006. Each panel of this figure represents counties in a specific HPA quartile.

Figure 3 . Share of Purchase Originations



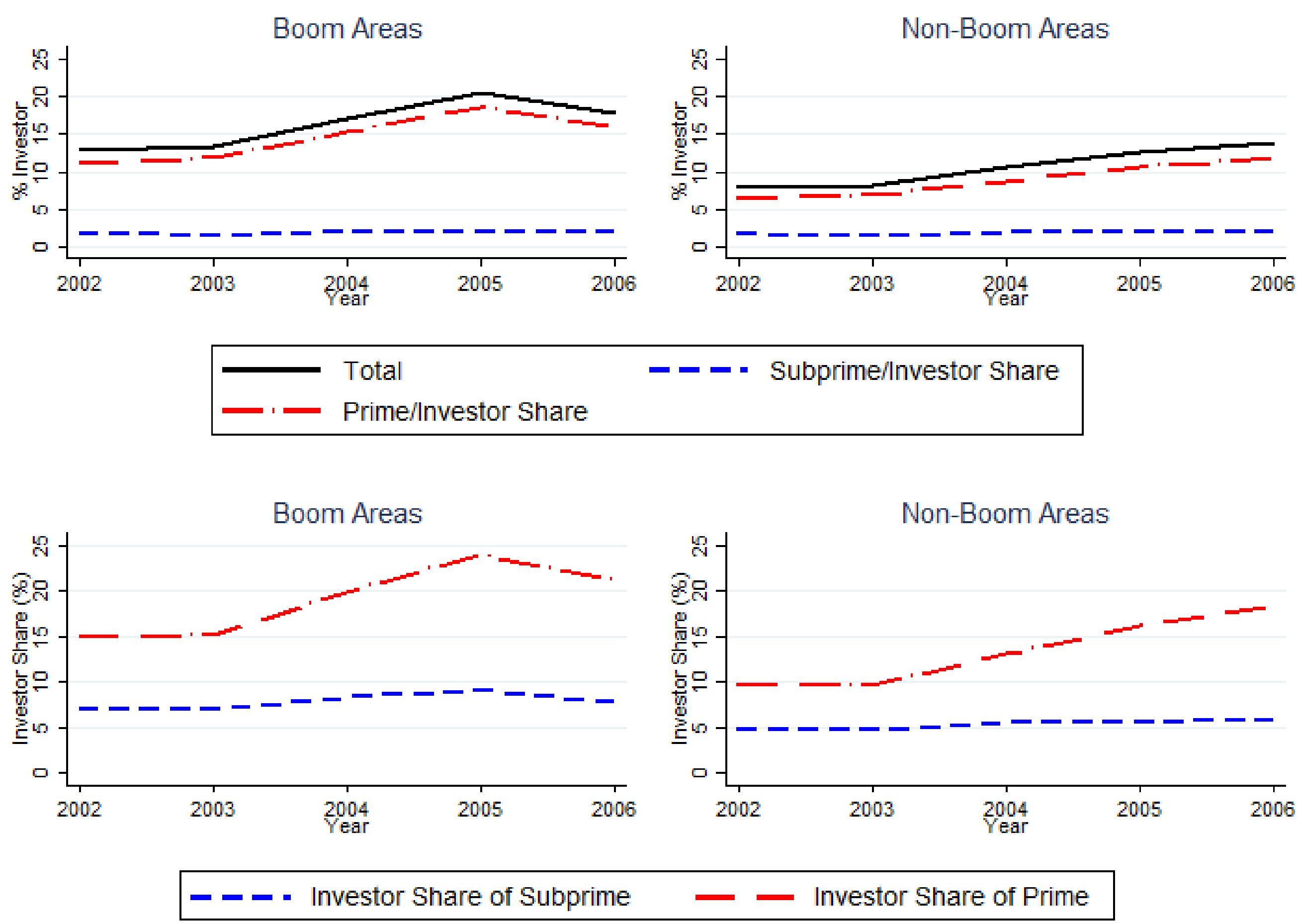
Source: McDash, ABSNet, authors' calculations. This figure shows the fraction of loans split by FICO scores. The loan sample is a merged sample of first-lien purchase mortgages between McDash and ABSNet, excluding all duplicates.

Figure 4 . Median Home Prices and Share of Purchase Mortgages to Subprime Borrowers in Four Counties



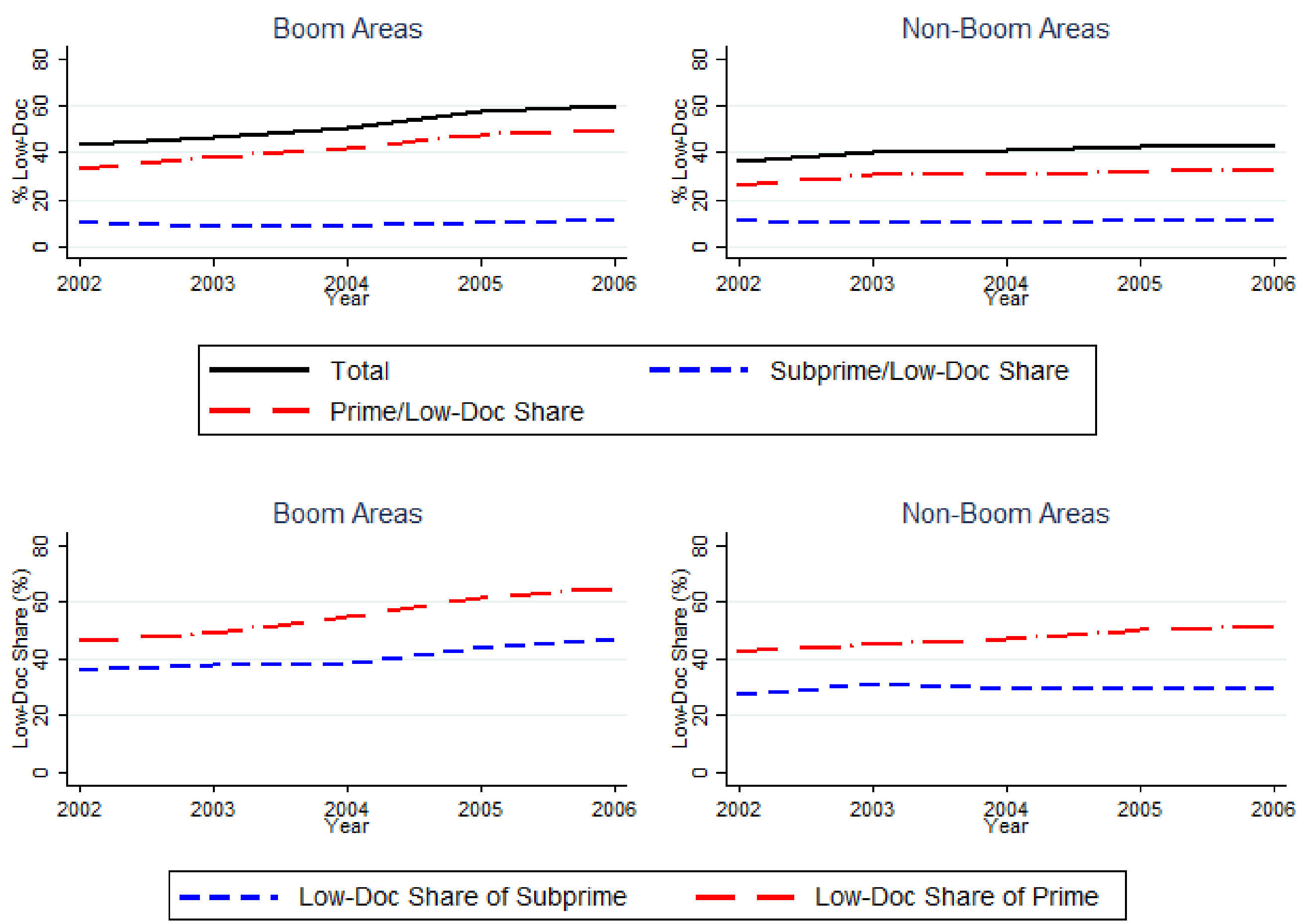
Source: Zillow, McDash, ABSNet, authors' calculations. This figure plots the subprime share of purchases and house prices for the most populous county in each sand state. The loan sample is a merged sample of first-lien purchase mortgages between McDash and ABSNet, excluding all duplicates. Subprime expansion is defined as the percentage changes in share of loans to subprime borrowers, defined as borrowers with FICO scores under 660. Home price index is from Zillow.

Figure 5 . Investor Share of Purchase Mortgages to Prime and Subprime Borrowers in Housing Boom and Non-Boom Counties: 2002–2006



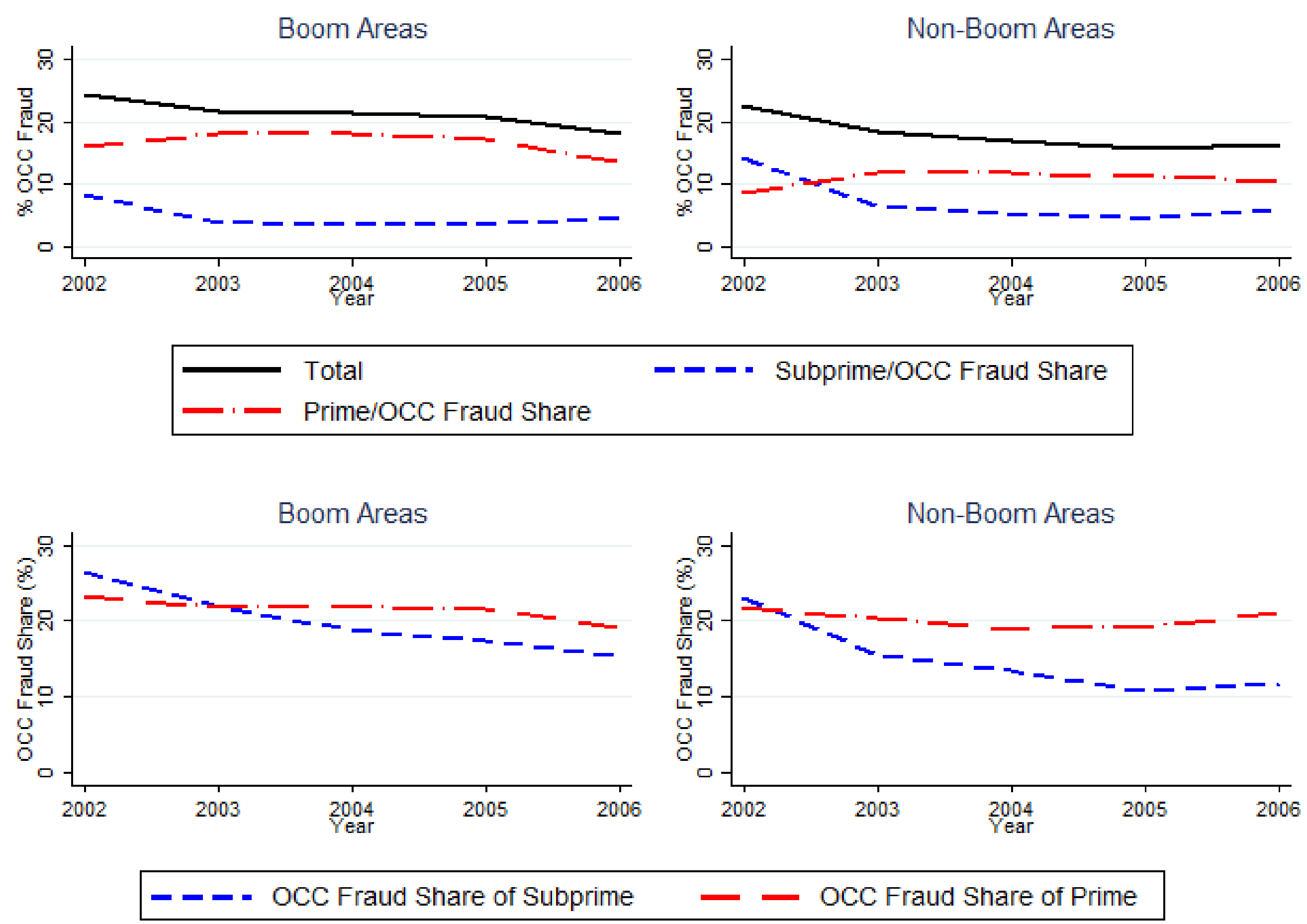
Source: McDash, ABSNet, FHFA, authors' calculations. The top two panels plot the total investor share (solid black line) of mortgage originations (the ratio of total investor-financed loans to total mortgage originations), and the investor share broken down into contributions from subprime and prime borrowers (dotted blue and red lines). The total investor share is the sum of the prime and subprime and subprime contributions (i.e. the subprime contribution is the number of subprime investor loans divided by the total number of originations). The bottom two panels plot the investor share of subprime and prime mortgage originations, respectively (i.e. the investor share of subprime is the number of subprime investor loans divided by the total number of subprime loans). Boom areas are defined as counties that experienced at least 20% price growth from 2002–2006. Other areas include counties that experienced less than a 20% increase in house prices from 2002–2006.

Figure 6 . Low Documentation Share of Purchase Mortgages to Prime and Subprime Borrowers in Housing Boom and Non-Boom Counties: 2002–2006



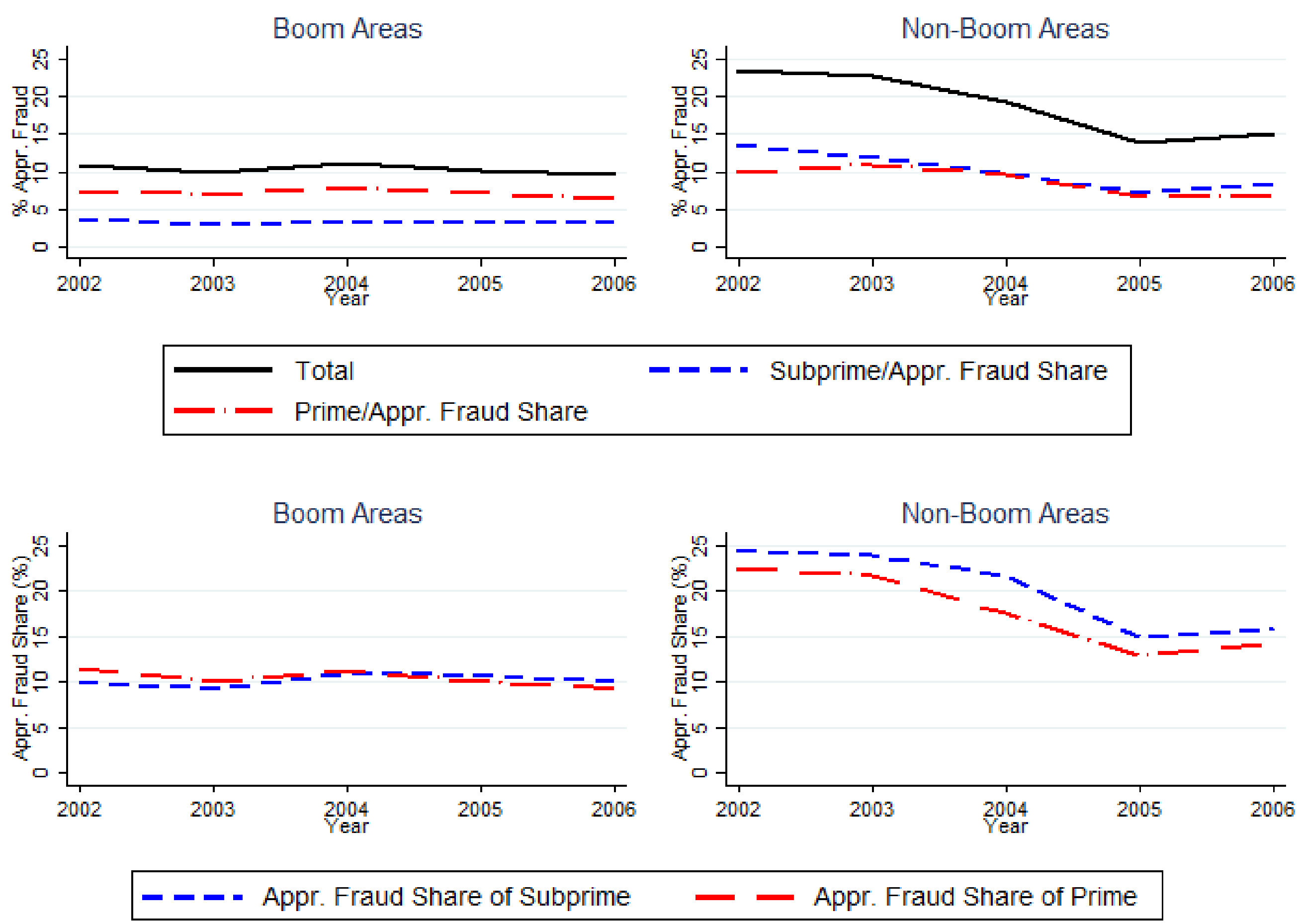
Source: McDash, ABSNet, FHFA, authors' calculations. The top two panels plot the total low-doc share (solid black line) of mortgage originations (the ratio of total low-doc loans to total mortgage originations), and the low-doc share broken down into contributions from subprime and prime borrowers (dotted blue and red lines). The total low-doc share is the sum of the prime and subprime and subprime contributions (i.e. the subprime contribution is the number of subprime low-doc loans divided by the total number of originations). The bottom two panels plot the low-doc share of subprime and prime mortgage originations, respectively (i.e. the low-doc share of subprime is the number of subprime low-doc loans divided by the total number of subprime loans). Boom areas are defined as counties that experienced at least 20% price growth from 2002–2006. Other areas include counties that experienced less than a 20% increase in house prices from 2002–2006.

Figure 7 . Estimated Incidence of Occupancy Fraud to Prime and Subprime Borrowers in Housing Boom and Non-Boom Counties: 2002–2006



Source: ABSNet, FHFA, authors' calculations. The top two panels plot the total occupancy fraud share (solid black line) of mortgage originations (the ratio of loans characterized by occupancy fraud to total mortgage originations), and the occupancy fraud share broken down into contributions from subprime and prime borrowers (dotted blue and red lines). The total occupancy fraud share is the sum of the prime and subprime and subprime contributions (i.e. the subprime contribution is the number of subprime loans characterized by occupancy fraud divided by the total number of originations). The bottom two panels plot the occupancy fraud share of subprime and prime mortgage originations, respectively (i.e. the occupancy fraud share of subprime is the number of subprime loans characterize by occupancy fraud divided by the total number of subprime loans). Boom areas are defined as counties that experienced at least 20% price growth from 2002–2006. Other areas include counties that experienced less than a 20% increase in house prices from 2002–2006.

Figure 8 . Estimated Incidence of Appraisal Fraud to Prime and Subprime Borrowers in Housing Boom and Non-Boom Counties: 2002–2006



Source: ABSNet, FHFA, authors' calculations. The top two panels plot the total appraisal fraud share (solid black line) of mortgage originations (the ratio of loans characterized by appraisal fraud to total mortgage originations), and the appraisal fraud share broken down into contributions from subprime and prime borrowers (dotted blue and red lines). The total appraisal fraud share is the sum of the prime and subprime and subprime contributions (i.e. the subprime contribution is the number of subprime loans characterized by appraisal fraud divided by the total number of originations). The bottom two panels plot the appraisal fraud share of subprime and prime mortgage originations, respectively (i.e. the appraisal fraud share of subprime is the number of subprime loans characterize by appraisal fraud divided by the total number of subprime loans). Boom areas are defined as counties that experienced at least 20% price growth from 2002–2006. Other areas include counties that experienced less than a 20% increase in house prices from 2002–2006.

Table 1 . Summary Statistics

	Full sample		Counties sorted by cumulative HPA during the period 2002–2006			
	weighted average	Median	HPA $\geq$ 70%	40% $\leq$ HPA $<$ 70%	20% $\leq$ HPA $<$ 40%	HPA $<$ 20%
			(1)	(2)	(3)	(4)
Δ log(home price index)	0.352	0.185	0.625	0.438	0.254	0.116
	(0.207)		(0.057)	(0.056)	(0.043)	(0.039)
Δ log(subprime share of purchase loans)	-0.028	0.092	-0.112	-0.067	0.001	0.056
	(0.179)		(0.125)	(0.155)	(0.209)	(0.175)
Δ log(subprime share of refinance loans)	0.462	0.406	0.424	0.500	0.257	0.419
	(0.210)		(0.194)	(0.236)	(0.172)	(0.205)
Δ log(subprime share of population)	-0.023	-0.03	-0.022	-0.040	-0.024	-0.009
	(0.049)		(0.053)	(0.049)	(0.047)	(0.041)
Δ log(average wage)	0.106	0.099	0.131	0.116	0.101	0.081
	(0.037)		(0.024)	(0.028)	(0.330)	(0.038)
Δ unemployment rate	-1.26%	-0.7%	-1.84%	-1.40%	-1.15%	-0.71%
	(1.02%)		(0.72%)	(1.00%)	(1.05%)	(0.93%)
Subprime share of purchase loans in 2002	0.292	0.330	0.278	0.234	0.303	0.342
	(0.087)		(0.065)	(0.082)	(0.087)	(0.078)
# Purchase loans in 2002	9,235	136	18,302	5,812	5,842	6,269
	(12,023)		(16,741)	(5,353)	(8,812)	(8,193)
Subprime share of refinance loans in 2002	0.256	0.310	0.247	0.212	0.266	0.310
	(0.085)		(0.060)	(0.076)	(0.081)	(0.093)
# Refinance loans in 2002	19,330	210	38,302	14,137	12,942	6,269
	(30,646)		(45,944)	(15,470)	(19,003)	(6,582)
Subprime share of population in 2002	0.308	0.308	0.311	0.28	0.301	0.328
	(0.058)		(0.042)	(0.055)	(0.071)	(0.056)
Average wage in 2002	$39,031	$29,051	$35,744	$42,797	$38,131	$39,462
	($9,648)		($7,410)	($11,582)	($9,470)	($8,716)
Unemployment rate in 2002	5.73%	5.5%	6.01%	5.67%	5.59%	5.63%
	(1.40%)		(1.58%)	(1.46%)	(1.38%)	(1.14%)
# Counties	2,384		136	353	722	1,173
# Subprime purchase loans 2002–2006 (million)	16.6		4.6	4.0	3.1	4.9
# Subprime refinance loans 2002–2006 (million)	20.8		6.3	5.7	3.8	5.0

Notes: This table displays averages and standard deviations (in parenthesis) for all variables included in the regressions in Section 3. The underlying data come from McDash Analytics and ABSNet. Statistics on purchase mortgages except the medians are averages weighted by the number of purchase mortgages between 2002 and 2006. Statistics on refinance mortgages except the medians are averages weighted by the number of refinance mortgages between 2002 and 2006.

Table 2 . Growth of U.S. County Level House Prices and the Share of Purchase Mortgages to Subprime Borrowers

Dependent Variable:	Δ log(HPI) 2002-2006					
	Weighted			Unweighted		
	(1)	(2)	(3)	(4)	(5)	(6)
Δ log(Subprime Share)	-0.444***	-0.215***	-0.135***	-0.122***	-0.154***	-0.073***
	(0.113)	(0.075)	(0.026)	(0.027)	(0.020)	(0.012)
Covariates	N	Y	Y	N	Y	Y
State FE	N	N	Y	N	N	Y
Observations	2,384	2,384	2,384	2,384	2,384	2,384
Adjusted R^2	0.15	0.55	0.90	0.07	0.41	0.76

Notes: This table reports estimates from a regression of change in the log FHFA county house price index from 2002 to 2006 on the contemporaneous change in log county subprime share of purchases. Observations are at the county level and regressions in columns (1)–(3) are weighted by the total number of loans in the county across both years, 2002 and 2006. Regressions in columns (4)–(6) are unweighted. Subprime share is calculated based on the fraction of first-lien purchase mortgage originations in a county with a FICO score less than 660. The other covariates include both level and change variables. The level variables (measured in 2002) are the percentage of the county population that is subprime, the county subprime share of mortgages, the number of loans originated, county average wages (from IRS), and the unemployment rate. The change variables include: change in log wages, change in log(% of subprime population), and the change in unemployment. Robust standard errors are in parentheses and are clustered at the state level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table 3 . House price growth and county subprime share growth over different periods

Dependent Variable:	Δ log(HPI)					
	2002-2006	2002-2005	2001-2006	2001-2005	2003-2005	2003-2006
	(1)	(2)	(3)	(4)	(5)	(6)
Δ log(Subprime Share)	-0.135***	-0.146***	-0.146***	-0.154***	-0.120***	-0.129***
	(0.026)	(0.031)	(0.030)	(0.030)	(0.032)	(0.022)
Covariates	Y	Y	Y	Y	Y	Y
State FE	Y	Y	Y	Y	Y	Y
Observations	2,384	2,384	2,384	2,384	2,384	2,384
Adjusted R^2	0.90	0.90	0.90	0.91	0.90	0.90

Notes: This table reports estimates from a regression of change in the log FHFA county house price index over different time periods on the contemporaneous change in the log county subprime share of purchases. Observations are at the county level and are weighted by the total number of loans in the county across both years, 2002 and 2006. Subprime share is calculated based on the fraction of first-lien purchase mortgage originations in a county with a FICO score less than 660. The other covariates include both level and change variables. The level variables (measured in the initial year) are the percentage of the county population that is subprime, the county subprime share of mortgages, the number of loans originated, county average wages (from IRS), and the unemployment rate. The change variables include: change in log wages, change in log(% of subprime population), and the change in unemployment. Robust standard errors are in parenthesis and are clustered at the state level. The underlying data come from McDash Analytics and ABSNet. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table 4 . House price growth and county subprime share growth for different FICO thresholds

Dependent Variable:	Δ log(HPI) 2002-2006					
Subprime Definition:	FICO $\leq$ 620			FICO $\leq$ 580		
	(1)	(2)	(3)	(4)	(5)	(6)
Δ log(Subprime Share)	-0.403***	-0.274***	-0.138***	-0.270***	-0.188***	-0.091***
	(0.050)	(0.039)	(0.020)	(0.034)	(0.026)	(0.013)
Covariates	N	Y	Y	N	Y	Y
State FE	N	N	Y	N	N	Y
Observations	2,304	2,304	2,304	2,037	2,037	2,037
Adjusted R^2	0.41	0.67	0.91	0.38	0.68	0.91

Note: This table reports estimates from a regression of change in the log FHFA county house price index from 2002 to 2006 on the contemporaneous change in the log county subprime share of purchases. Observations are at the county level and weighted by the total number of loans in the county across both years, 2002 and 2006. Subprime share is calculated based on the fraction of first-lien purchase mortgage originations in a county with a FICO score less than 620 (columns (1)–(3)) and 580 (columns (4)–(6)). The other covariates include both level and change variables. The level variables (measured in 2002) are the percentage of the county population that is subprime, the county subprime share of mortgages, the number of loans originated, county average wages (from IRS), and the unemployment rate. The change variables include: change in log wages, change in log(% of subprime population), and the change in unemployment. Robust standard errors are in parenthesis and are clustered at the state level. The underlying data come from McDash Analytics and ABSNet. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table 5 . House price growth and alternative measures of subprime purchase activity

Dependent Variable:	Δ log(HPI) 2002-2006					
	(1)	(2)	(3)	(4)	(5)	(6)
Δ (Subprime Share)	-1.603***	-0.915***	-0.465***			
	(0.387)	(0.262)	(0.095)			
Δ log(# Subprime Purchase Loans)				-0.101	-0.056	-0.018
				(0.107)	(0.061)	(0.013)
Covariates	N	Y	Y	N	Y	Y
State FE	N	N	Y	N	N	Y
Observations	2,398	2,398	2,398	2,384	2,384	2,384
Adjusted R^2	0.10	0.56	0.90	0.01	0.51	0.90

Note: This table reports estimates from a regression of change in the log FHFA county house price index from 2002 to 2006 on the contemporaneous change in the subprime purchase share at the county level (columns (1)-(3)) and the change in the log of the number of subprime purchases mortgages at the county level (columns (4)-(6)). Observations are at the county level and weighted by the total number of loans in the county. The other covariates include both level and change variables. The level variables (measured in 2002) are the percentage of the county population that is subprime, the county subprime share of mortgages, the number of loans originated, county average wages (from IRS), and the unemployment rate. The change variables include: change in log wages, change in log(% of subprime population), and the change in unemployment. Robust standard errors are in parenthesis and are clustered at the state level. The underlying data come from McDash Analytics and ABSNet. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table 6 . Growth of U.S. County Level House Prices and the Share of Refinance Mortgages to Subprime Borrowers

Dependent Variable:	Δ log(HPI) 2002-2006					
	Weighted			Unweighted		
	(1)	(2)	(3)	(4)	(5)	(6)
Δ log(Subprime Share)	-0.040	-0.004	0.094	0.105***	-0.042	-0.040**
	(0.183)	(0.093)	(0.101)	(0.034)	(0.038)	(0.017)
Covariates	N	Y	Y	N	Y	Y
State FE	N	N	Y	N	N	Y
Observations	2,392	2,392	2,392	2,392	2,392	2,392
Adjusted R^2	0.15	0.55	0.90	0.04	0.31	0.75

Note: This table reports estimates from a regression of change in the log FHFA county house price index from 2002 to 2006 on the contemporaneous change in log county subprime share of refinance loans. Observations are at the county level and regressions in columns (1)–(3) are weighted by the total number of loans in the county. Subprime share is calculated based on the fraction of first-lien refinance mortgage originations in a county with a FICO score less than 660. The other covariates include both level and change variables. The level variables (measured in 2002) are the percentage of the county population that is subprime, the county subprime share of mortgages, the number of loans originated, county average wages (from IRS), and the unemployment rate. The change variables include: change in log wages, change in log(% of subprime population), and the change in unemployment. Robust standard errors are in parentheses and are clustered at the state level. The underlying data come from McDash Analytics and ABSNet. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table 7 . Growth of U.S. County Level House Prices and the Share of Purchase Mortgages to Low Income Borrowers

Dependent Variable:	Δ log(HPI) 2002-2006		
	(1)	(2)	(3)
Δ log(Low Income Share)	-0.160***		
	(0.016)		
Δ log(Low Wage Share)		-0.146***	
		(0.014)	
Δ log(Share Income $<$ 50K)			-0.156***
			(0.015)
Covariates	Y	Y	Y
State FEs	Y	Y	Y
Observations	2598	2591	2599
Adjusted R^2	0.92	0.94	0.93

Note: This table reports estimates from a regression of change in the log FHFA county house price index from 2002 to 2006 on the contemporaneous change in log county low income share of purchase loans from HMDA. Observations are at the county level and regressions in columns (1)–(3) are weighted by the total number of loans in the county. In column (1), low income share is calculated based on the fraction of first-lien purchase mortgage originations in a county where applicant(s) income is less than the county household median income from GeoFred data. In column (2), low wage share is calculated based on the fraction of first-lien purchase mortgage originations in a county where applicant(s) income is less than the average county wage from IRS data. In column (3), Share Income $<$ 50K is calculated based on the fraction of first-lien purchase mortgage originations in a county where applicant(s) income is less than \$50,000. The other covariates include both level and change variables. The level variables (measured in 2002) are the county share of purchase mortgages to low income borrowers, the number of loans originated, county average wages (from IRS), and the unemployment rate. The change variables include: change in log wages and the change in unemployment. Robust standard errors are in parentheses and are clustered at the state level. The underlying data come from HMDA. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

References

ADELINO, M., A. SCHOAR, AND F. SEVERINO (2013): "Credit supply and house prices: evidence from mortgage market segmentation," *NBER Working Paper # 17832*.

——— (2015): "Loan originations and defaults in the mortgage crisis: Further evidence," *NBER Working Paper # 21320*.

——— (2016): "Loan originations and defaults in the mortgage crisis: The role of the middle class," *The Review of Financial Studies*, 29(7), 1635–1670.

AGARWAL, S., B. W. AMBROSE, AND V. W. YAO (2014): "The limits of regulation: Appraisal bias in the mortgage market," *Penn State University Working Paper*.

AGARWAL, S., I. BEN-DAVID, AND V. YAO (2015): "Collateral valuation and borrower financial constraints: Evidence from the residential real estate market," *Management Science*, 61(9), 2220–2240.

ALBANESI, S., G. DE GIORGI, AND J. NOSAL (2017): "Credit growth and the financial crisis: A new narrative," *NBER Working Paper # 23740*.

AMBROSE, B. W., J. CONKLIN, AND J. YOSHIDA (2016): "Credit rationing, income exaggeration, and adverse selection in the mortgage market," *The Journal of Finance*, 71(6), 2637–2686.

AVERY, R., K. BREVOORT, AND G. CANNER (2007): "Opportunities and issues in using HMDA data," *Journal of Real Estate Research*, 29(4), 351–380.

BEN-DAVID, I. (2011): "Financial constraints and inflated home prices during the real estate boom," *American Economic Journal: Applied Economics*, 3(3), 55–87.

BERKOVEC, J., Y. CHANG, AND D. A. MCMANUS (2012): "Alternative lending channels and the crisis in US housing markets," *Real Estate Economics*, 40(S1), S8 – S31.

BHUTTA, N. (2015): "The ins and outs of mortgage debt during the housing boom and bust," *Journal of Monetary Economics*, 76, 284–298.

BHUTTA, N., AND B. J. KEYS (2016): "Interest rates and equity extraction during the housing boom," *The American Economic Review*, 106(7), 1742–1774.

BLACKBURN, M. L., AND T. VERMILYEA (2012): "The prevalence and impact of misstated incomes on mortgage loan applications," *Journal of Housing Economics*, 21(2), 151–168.

BRUECKNER, J. K., P. S. CALEM, AND L. I. NAKAMURA (2012): "Subprime mortgages and the housing bubble," *Journal of Urban Economics*, 71(2), 230–243.

BRUNNERMEIER, M. K. (2009): "Deciphering the liquidity and credit crunch 2007-2008," *Journal of Economic Perspectives*, 23(1), 77–100.

CALEM, P. S., L. LAMBIE-HANSON, AND L. I. NAKAMURA (2017): "Appraising Home Purchase Appraisals," *FRB of Philadelphia Working Paper No. 17-23*.

CHINCO, A., AND C. MAYER (2015): "Misinformed speculators and mispricing in the housing market," *The Review of Financial Studies*, 29(2), 486–522.

COLEMAN IV, M., M. LACOUR-LITTLE, AND K. D. VANDELL (2008): "Subprime lending and the housing bubble: Tail wags dog?," *Journal of Housing Economics*, 17(4), 272–290.

CONKLIN, J., E. COULSON, M. DIOP, AND T. LE (2017): "Competition and Appraisal Inflation," *Working Paper*.

DEMIROGLU, C., AND C. JAMES (2016): "Indicators of Collateral Misreporting," *Management Science*, 64(4), 1747–1760.

DI MAGGIO, M., AND A. KERMANI (2017): "Credit-induced boom and bust," *The Review of Financial Studies*, 30(11), 3711–3758.

DING, L., AND L. NAKAMURA (2016): "The impact of the home valuation code of conduct on appraisal and mortgage outcomes," *Real Estate Economics*, 44(3), 658–690.

DWYER, G. P., AND P. TKAC (2009): "The financial crisis of 2008 in fixed-income markets," *Journal of International Money and Finance*, 28(8), 1293–1316.

FAVARA, G., AND J. IMBS (2015): "Credit supply and the price of housing," *American Economic Review*, 105(3), 958–92.

FAVILUKIS, J., S. C. LUDVIGSON, AND S. VAN NIEUWERBURGH (2017): "The macroeconomic effects of housing wealth, housing finance, and limited risk sharing in general equilibrium," *Journal of Political Economy*, 125(1), 140–223.

FERREIRA, F., AND J. GYOURKO (2011): "Anatomy of the beginning of the housing boom: US neighborhoods and metropolitan areas, 1993-2009," *NBER Working Paper # 17374*.

——— (2015): "A new look at the US foreclosure crisis: Panel data evidence of prime and subprime borrowers from 1997 to 2012," *NBER Working Paper # 21261*.

FOOTE, C. L., L. LOEWENSTEIN, AND P. S. WILLEN (2016): "Cross-sectional patterns of mortgage debt during the housing boom: Evidence and implications," *NBER Working Paper # 22985*.

FUSTER, A., AND J. VICKERY (2014): "Securitization and the fixed-rate mortgage," *Review of Financial Studies*, 28(1), 176–211.

GERARDI, K., A. LEHNERT, S. M. SHERLUND, AND P. WILLEN (2008): "Making Sense of the Subprime Crisis," *Brookings Papers on Economic Activity*, 39(2 (Fall)), 69–159.

GRIFFIN, J. M., AND G. MATURANA (2016a): "Did dubious mortgage origination practices distort house prices?," *The Review of Financial Studies*, 29(7), 1671–1708.

——— (2016b): "Who facilitated misreporting in securitized loans?," *The Review of Financial Studies*, 29(2), 384–419.

HAUGHWOUT, A., D. LEE, J. S. TRACY, AND W. VAN DER KLAAUW (2011): "Real estate investors, the leverage cycle, and the housing market crisis," *FRB of New York Staff Report No. 514*.

JIANG, W., A. A. NELSON, AND E. VYTLACIL (2014): "Liar's loan? Effects of origination channel and information falsification on mortgage delinquency," *Review of Economics and Statistics*, 96(1), 1–18.

KAPLAN, G., K. MITMAN, AND G. VIOLANTE (2015): "Consumption and house prices in the Great Recession: Model meets evidence," *New York University Working Paper*.

KRUGER, S. A., AND G. MATURANA (2017): "Collateral Misreporting in the RMBS Market," *Working Paper*.

LAEVEN, L., AND A. POPOV (2017): "Waking up from the american dream: on the experience of young americans during the housing boom of the 2000s," *Journal of Money, Credit and Banking*, 49(5), 861–895.

MAYER, C., K. PENCE, AND S. M. SHERLUND (2009): "The rise in mortgage defaults," *The Journal of Economic Perspectives*, 23(1), 27–50.

MIAN, A., AND A. SUFI (2009): "The consequences of mortgage credit expansion: Evidence from the US mortgage default crisis," *The Quarterly Journal of Economics*, 124(4), 1449–1496.

——— (2017a): "Fraudulent income overstatement on mortgage applications during the credit expansion of 2002 to 2005," *The Review of Financial Studies*, 30(6), 1832–1864.

——— (2017b): "Household Debt and Defaults from 2000 to 2010: The Credit Supply View," in *Evidence and Innovation in Housing Law and Policy*, ed. by L. A. Fennell, and B. J. Keys, pp. 257 – 288. Cambridge University Press, Cambridge.

PISKORSKI, T., A. SERU, AND J. WITKIN (2015): "Asset quality misrepresentation by financial intermediaries: Evidence from the RMBS market," *The Journal of Finance*, 70(6), 2635–2678.

SHI, L., AND Y. ZHANG (2015): "Appraisal inflation: Evidence from the 2009 GSE HVCC intervention," *Journal of Housing Economics*, 27, 71–90.

SOMMER, K., P. SULLIVAN, AND R. VERBRUGGE (2013): "The equilibrium effect of fundamentals on house prices and rents," *Journal of Monetary Economics*, 60(7), 854 – 870.