

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 12, 2019

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. By contrast, this paper documents a robust, negative correlation between the growth in the share of purchase mortgages to subprime borrowers and house price appreciation at the county-level during this time. Using two different instrumental variables approaches, we also establish causal evidence that house price appreciation lowered the share of purchase loans to subprime borrowers. Further analysis using micro-level credit bureau data shows that higher house price appreciation lowered the transition rate into first-time homeownership for subprime individuals. Finally, the paper documents that subprime borrowers did not play a significant role in the increased speculative activity and underwriting fraud that the literature has linked directly to the housing boom. Taken together, these results are more consistent with subprime borrowers being priced out of housing boom markets rather than inflating prices in those markets.

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

*We thank Manuel Adelino, Brent Ambrose, Stefan Nagel, Wayne Archer (discussant), Zahi Ben-David, Shaun Bond, John Clapp, Daniel Cooper, Matteo Crosignani (discussant), Richard Crump, Fernando Duarte, David Echeverry, Andreas Fuster, Pedro Gete, John Griffin, Deeksha Gupta (discussant), Andrew Haughwout, David Humphrey, Michael Lee, Zhenguo Lin, David Lucca, Ronan Lyons, Gonzalo Maturana (discussant), Brian Melzer, Nuno Mota (discussant), Christopher Palmer, Amiyatosh Purnanandam, Michael Reher, Eva Steiner, Randal Verbrugge, James Vickery, Nancy Wallace, Paul Willen, and seminar participants at the Federal Reserve Bank of Dallas, Federal Reserve Bank of New York, the 2018 WFA Conference, the 2018 Notre Dame Roundtable on Housing and Mortgage Markets, the 2018 Cambridge-UF-NUS Real Estate Finance Investment Symposium, the 2019 AREUEA annual meeting, and the 2019 FIRS conference 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 Dallas, scott.frame@dal.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. What role the expansion of mortgage lending to subprime borrowers played in driving the housing boom is a key question that continues to be hotly debated in the academic literature.¹ 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.

The predominant narrative linking expanded mortgage lending to subprime borrowers and home price growth during the 2000s is based on the empirical findings presented in Mian and Sufi (2009). This narrative, which Mian and Sufi (2017b) term the *credit supply* view of the 2000s housing cycle, argues that there was a large expansion of mortgage debt to previously rationed marginal borrowers (defined by credit score or income) that coincided with an increase in private securitization activity riddled with incentive conflicts. This mechanism drove home prices to unsustainable levels and ultimately a crash. The principal alternative view, labeled the *passive credit view* in Mian and Sufi (2017b), is that the expansion in mortgage credit was not directly responsible for inflating prices, but was instead a response to the high house price appreciation that was fueling extremely optimistic expectations of future price growth. This alternative view is detailed in Foote, Gerardi, and Willen (2012) and Adelino, Schoar, and Severino (2016).

A straightforward cross-sectional implication of the *credit supply view* of the housing boom is that we should see higher house price growth in areas that experienced high growth in subprime

¹There 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). Depending on the question at-hand and data availability, most researchers have identified subprime mortgages as: (i) loans to borrowers with low credit scores; or (ii) loans backing privately issued mortgage-backed securities and marketed to investors as being 'subprime'.

purchase mortgage activity. However, the data presented in Figure 1 actually suggests the opposite. 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 first-lien 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. Multivariate 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. 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 indices, and credit score thresholds.

One natural explanation for this negative correlation is reverse causality. That is, high house price appreciation made property increasingly unaffordable for subprime borrowers. We explore this “pricing out” effect in two ways. First, we identify the causal effect of house price appreciation on subprime purchase mortgage lending at the county level using two instrumental variable strategies. Our first strategy uses a recent county-level land unavailability index constructed by Lutz and Sand (2017) as an instrument for cross-sectional variation in house price growth over the 2002–2006 period. This index is similar in spirit to the MSA-level land unavailability measure of Saiz (2010), which has been used as an instrument for house price growth by numerous studies in the literature. Our second strategy follows the methodology of Guren, McKay, Nakamura, and Steinsson (2018), who instrument for local house price appreciation using its historical sensitivity to regional house price appreciation (holding constant some local, regional, and national economic factors). Both instrumental variable strategies indicate that house price growth has an economically meaningful, negative, causal effect on the growth in the share of subprime purchase mortgage

lending at the county level.

Second, we estimate the effect of house price appreciation on the likelihood of a subprime individual transitioning into homeownership for the first time between 2002 and 2006.² Using the Federal Reserve Bank of New York Consumer Credit Panel, we find that higher house price growth lowered the likelihood of a subprime borrower becoming a homeowner. These two articles of evidence are consistent with the pricing out effect generating a negative, causal relationship between house price growth and the growth in the share of subprime purchase lending during the mid-2000s housing boom.

We complement our evidence on the relationship between house price appreciation and subprime purchase share growth with new evidence on the role played by subprime mortgage lending in two activities that have previously been linked to the housing boom and bust. Recent research suggests that speculative behavior by real estate investors played a significant role in driving house price growth in many areas of the country during the 2000s U.S. housing boom (Haughwout, Lee, Tracy, and Van der Klaauw, 2011; Chincó 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 further 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

²While we are not aware of a similarly direct test of the effect of house prices on the transition into homeownership in the literature, previous studies have found results consistent with a pricing out effect. For example, Bhutta (2015) and Foote, Loewenstein, and Willen (2016) document that first-time home buying dropped disproportionately for low credit score borrowers during the mid-2000s housing boom.

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 credit score thresholds. By using credit scores to define subprime borrowers, our intent is to focus on marginal borrowers in terms of credit risk. Furthermore, using income to identify marginal borrowers is problematic since income misrepresentation was prevalent during the U.S. housing boom (Jiang, Nelson, and Vytlačil, 2014; Adelino, Schoar, and Severino, 2015; Ambrose, Conklin, and Yoshida, 2016; Mian and Sufi, 2017a). Nonetheless, we conduct robustness checks using income to define marginal borrowers and find that growth in the share of purchase mortgages to low-income borrowers is also negatively correlated with house price growth.

Our main analysis focuses on mortgage originations used to finance home purchases, as house prices are determined by marginal buyers in the market. 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 house price growth and the growth in the share of refinance mortgages to subprime borrowers at the county level. We find no evidence of a positive correlation.

The paper is related to a recent debate in the literature about the nature of the expansion in mortgage credit during the mid-2000s. In their seminal study, Mian and Sufi (2009) argued that credit growth was concentrated principally among subprime borrowers and that this growth was decoupled from income growth. However, recent research has shown that credit growth occurred in a uniform manner across the entire credit score and income 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.

The paper is also related to recent empirical evidence showing that an expansion of mortgage credit can cause 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. Dokko, Keys, and Relihan (2019) find that county-level increases in the use of non-fixed rate mortgages preceded local house price booms during the 2000s.³ We do not view our results as being inconsistent with the results of these empirical studies, which indicate that increases in credit supply can increase housing demand and result in higher home 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.

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 explores the causal effect of house price growth on subprime purchase lending. Section 5 shows that several types of speculative mortgage products were not biased towards subprime borrowers. Section 6 concludes.

³In a related paper, Berkovec, Chang, and McManus (2012) document that the ZIP code-level correlation between house price growth and growth in the share of interest-only and negative amortization mortgages was stronger than the correlation between price growth and growth in alternative lending channels.

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 duplicate 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 our 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 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 (2002-2006). 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.⁴ 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.

⁴Appendix Table B.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.

housing boom across metropolitan areas (e.g., 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.⁵ We also observe that, on average, the share of purchase mortgages to subprime 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 housing boom. 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. Counties with the strongest house price growth also experienced the highest 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

⁵The 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.

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 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 purchase mortgage growth

This section more formally considers the relationship between house price growth and growth in the share of purchase mortgage originations to subprime borrowers at the county level. 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.

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:

$$\Delta p_i^{2002-2006} = \beta \Delta s_i^{2002-2006} + \rho X_i + \pi_{state} + \epsilon_i \quad (3.1)$$

where $\Delta p_i^{2002-2006}$ is the growth in the FHFA house price index in county i between 2002 and 2006 and $\Delta s_i^{2002-2006}$ is the growth over the same period in the county share of first lien purchase mortgages to borrowers with a FICO score less than 660.⁶

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.

⁶The growth variables in equation 3.1 are measured as the natural logarithm of the ratio of the variables over the two time periods. For example, $\Delta p_i^{2002-2006}$ is calculated as $\ln\left(\frac{HPI_{2006}}{HPI_{2002}}\right)$. The FHFA county price indices have much broader coverage than the CoreLogic or Zillow indices, thus, we use the FHFA indices for most of our analysis. Our main results are materially unchanged when using CoreLogic or Zillow indices.

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 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. Hence, 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.⁷ 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 2006.⁸ 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

⁷Although 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.

⁸Our results are robust to using alternative house price indices. Appendix Tables C.3 and C.4 show similar results using CoreLogic and Zillow house price indices, respectively. Additionally, our results remain unchanged if we perform the analysis using panel data. Regressions using annual county level growth rates, rather than overall growth from 2002–2006, are reported in Appendix Table C.5 and are consistent with our main results.

and Gyourko (2011) and Bhutta and Keys (2016), there was significant geographic variation in the timing of the boom-bust period. 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 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, as noted above, the empirical relationship between house price growth and that in the share of subprime purchase mortgages may depend on the initial share of subprime purchase mortgages. While our measure treats them the same, a doubling of the subprime purchase share from 3% to 6% may have a very different effect on house prices than an increase from 15% to 30%. We attempted to address this issue by including county level subprime purchase shares in 2002 in the covariate set, but this may not completely resolve the issue. In Table 5 we present regression results based on equation 1 where we have split the sample into counties above and below the median initial subprime purchase share. Here we see that the initial subprime purchase share is negatively related to house price growth, although this relationship is statistically insignificant for counties with below median initial subprime purchase shares. This implies that counties with high levels of subprime purchase mortgage activity in 2002 were not the areas those experiencing rapid house price appreciation during the housing boom. More importantly, the negative coefficient on

subprime share growth is of similar magnitude across all three columns. Regardless of the initial level of subprime purchase share, growth in this share is associated with lower house price growth.

Fourth, we examine alternative measures of subprime purchase mortgage activity during the housing boom in Table 6 : (i) the change (rather than growth) in subprime purchase shares (columns (1)–(3)), and (ii) the growth in the number (rather than growth in the share) of subprime purchase mortgages (columns (4)–(6)). Consistent with our main result, both alternative 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.⁹

Fifth, in Table 7 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.¹⁰ The results in Table 7 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.

Sixth, we use an alternative method of defining marginal borrowers. By 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 8 , we investigate the relationship between house price appreciation and growth in the low income share of purchase originations. Since borrower income is not available in our mortgage servicing data, we turn to

⁹Note that the growth in purchase mortgages to prime borrowers is implicitly contained in the dependent variable in all regressions that involve subprime purchase share. However, in columns (4) – (6) we do not control for the growth in prime purchase lending. If we do so, the coefficients on $\Delta \log(\# \text{ Subprime Purchase Loans})$ are negative and statistically significant.

¹⁰For example, the large increase in mortgage equity withdrawal via cash-out refinances during the 2000s could have led to an increase in consumption and an increase in local economic activity, which could have then put upward pressure on house prices.

loan level information disclosed by financial institutions under the Home Mortgage Disclosure Act (HMDA). The HMDA data provides the most comprehensive coverage of home lending activity in the United States and includes a measure of borrower income for each loan.¹¹ However, the drawback of the HMDA data is that it includes little information about loan level risk characteristics (e.g., loan-to-value ratios or credit scores), so we are unable to use it in our main analysis. Table 8 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, county level 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 have been a first order driver of the housing boom.

Finally, since several related influential papers use ZIP codes as the unit of analysis, we also provide results at the ZIP code level. While our county-level analysis covers approximately 80% of those jurisdictions, the dearth of zip code level house price information reduces our coverage to 13% - 35% of these areas depending on the index used (Corelogic and FHFA). As described in Appendix A, the structure of this analysis includes some subtle differences. At this level of granularity we find consistent results – a negative relationship between the growth in subprime purchase shares in a zip code and house price appreciation.

Taken together, the analysis presented in this section indicates a robust negative conditional correlation between the growth in home purchase financing to subprime borrowers and the growth in house prices during the U.S. housing boom. This finding is inconsistent with the credit supply view of the 2000s housing cycle.

¹¹As noted by others, income is likely measured with error in the HMDA data (Avery, Brevoort, and Canner (2007), Mian and Sufi (2017a), and Adelino, Schoar, and Severino (2016)). For example, lenders do not verify the accuracy of income on low-documentation loans; and a borrower need not document all sources of income on the mortgage application.

4 Causal effect of house prices on subprime purchase mortgage lending

One natural explanation for the negative relationship between county-level house price appreciation and county-level growth in purchase mortgage lending to subprime borrowers during the 2002-2006 U.S. housing boom is reverse causality. Rapidly increasing home prices in some counties made homeownership unaffordable for many prospective subprime borrowers, who tend to have lower income and wealth on average. Previous research has found results consistent with such a “pricing out” hypothesis, but to our knowledge this has not been formally tested.¹² We begin by investigating whether county-level house price growth had a causal effect on home purchase lending to subprime borrowers by leveraging two separate instrumental variables (IVs) for local home prices recently developed by Lutz and Sand (2017) and Guren, McKay, Nakamura, and Steinsson (2018). Using both IV strategies, we find strong evidence that house price growth has a negative effect on growth in the share of home purchases by subprime borrowers at the county-level. Next, we use the Federal Reserve Bank of New York Consumer Credit Panel (CCP) data to examine the relationship between house price growth and the likelihood of an individual transitioning into first time homeownership. We find that high house price growth lowered the relative likelihood of subprime individuals transitioning into homeownership. Both of the findings presented in this section are consistent with the “pricing out” hypothesis generating a negative, causal relationship between house price growth and growth in subprime purchase mortgage shares.

¹²For example, Laeven and Popov (2017) show that young households’ conversion to homeownership slowed during the U.S. housing boom, 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. These latter two papers also find that the share of home purchases by subprime borrowers declined during the boom. We extend this finding by showing that this decline was largest in those areas with higher house price appreciation.

4.1 Land unavailability as an instrument for house price growth

To explore the causal effect of county-level house price appreciation on county-level growth in the share of lending to subprime home purchase borrowers during the 2002–2006 U.S. housing boom we start with the following empirical model:

$$\Delta s_i^{2002-2006} = \alpha \Delta p_i^{2002-2006} + \kappa X_i + \pi_{state} + \epsilon_i, \quad (4.2)$$

where $\Delta s_i^{2002-2006}$ is the change in the log subprime purchase share from 2002 to 2006 in county i and $\Delta p_i^{2002-2006}$ is contemporaneous county house price appreciation.¹³ All other variables are as defined above in equation (3.1).

We are primarily interested in α , but the naive regression in equation (4.2) will likely produce a biased estimate of the causal effect of house price appreciation on subprime purchase share growth due to endogeneity. One source of bias is reverse causation, as explored above in equation (3.1). There may also be omitted variable bias insofar as there could be variables that are correlated with both house price growth and growth in home purchase lending to subprime borrowers. For example, unobserved labor demand shocks, such as an increase in demand for highly skilled workers, may affect both subprime mortgage credit demand and house prices.

In order to overcome these sources of bias, we start by using cross-sectional variation in the share of undevelopable land to instrument for house price growth. The amount of developable land is a key component of local housing supply elasticities, which determine the extent to which prices respond to housing demand shocks. In an area with relatively elastic housing supply, a shift in housing demand will result in modest house price growth because the shock will largely be absorbed through a housing supply response. However, in an area characterized by inelastic housing supply, demand shocks will primarily be reflected in house price fluctuations. The most

¹³To be consistent with prior literature, $\Delta s_i^{2002-2006}$ and $\Delta p_i^{2002-2006}$ are represented by $g_i^{02-06}(SubShare)$ and $g_i^{02-06}(HPI)$ in equation (3.1). We change the notation in this section to ease interpretation of the identification strategy, particularly in Section 4.2.

commonly used proxy for supply elasticities was developed in Saiz (2010).¹⁴ He argued that the share of land that is undevelopable at the MSA level due to water (e.g., ocean, lake) or steep slopes (e.g., mountains), is largely determined by geologic forces, and thus can be considered plausibly orthogonal to other demand factors that may also be correlated with the dependent variable of interest.¹⁵

Saiz’s measure has an important limitation in the context of our study. Our analysis is performed at the county level, while his measure is only available at the MSA level. For this reason, we use an alternative land unavailability measure created by Lutz and Sand (2017) that is available at more disaggregated geographic levels, including counties. While the measure does have some subtle differences from the Saiz index, Lutz and Sand (2017) shows that the two measures are highly correlated.¹⁶

We employ the following instrumental variables (IV) specification:

$$\Delta p_i^{2002-2006} = \delta Land\ Unavailable_i + \theta X_i + \eta_{state} + v_i, \quad (4.3)$$

$$\Delta s_i^{2002-2006} = \alpha \widehat{\Delta p_i}^{2002-2006} + \kappa X_i + \pi_{state} + \epsilon_i, \quad (4.4)$$

where the instrument, $Land\ Unavailable_i$, is the share of undevelopable land at the county level from Lutz and Sand (2017), and $\widehat{\Delta p_i}^{2002-2006}$ is predicted county house price growth from the first stage regression in equation (4.3). All other variables are as described above. Standard errors are clustered at the state level.

¹⁴See, for example, Stroebel and Vavra (2019), Mian, Rao, and Sufi (2013), and Mian and Sufi (2011).

¹⁵By contrast, other components of housing supply elasticities such as land-use regulations, are determined by policymakers and thus would violate the exclusion restriction in many contexts.

¹⁶There is also a second reason for using the measure constructed by Lutz and Sand (2017) rather than the Saiz index. Davidoff (2016) argues that land unavailability at the MSA level is correlated with other demand factors that are likely to affect house prices and outcomes of interest, and hence the Saiz index violates the exclusion restriction necessary for a valid instrument. Lutz and Sand (2017) provide evidence that their more disaggregated measure is not plagued by the Davidoff (2016) critique, and thus serves as a valid instrument for house price growth.

Table 9 presents coefficient estimates from the first stage regression. Consistent with previous studies using the more aggregated (MSA) Saiz measure of land unavailability, column (1) shows that $\hat{\delta}$ is positive and statistically significant, indicating that land unavailability predicts house price growth. Land unavailability alone explains 30% of the variation in house price growth in the 2002–2006 period, and thus is a fairly strong instrument.¹⁷ In column (2) we add covariates and again see that land unavailability strongly predicts county level house price growth. After adding state fixed effects in column (3), the coefficient on land unavailability is still positive, albeit not significantly different from zero. The F-stat drops to 1.8, which suggests that the share of undevelopable land is not a strong predictor of within-state variation in house price growth. This is likely due to two factors. First, panel (a) in Figure 4 shows that there isn't very much within-state variation in the share of undevelopable land. Second, as we discussed above, most of the county-level variation in house price appreciation is across states, rather than within them. Thus it is not surprising that the unavailability measure is a weak instrument for within-state house price appreciation.

Table 10 presents the instrumental variable estimates. Columns (1)–(3) report OLS estimates as a benchmark, while columns (4)–(6) present second stage estimates corresponding to the first stage regressions in Table 9. The OLS estimates show a strong negative correlation between house price growth and subprime purchase share growth, consistent with our main results in Section 3. In column (3), a one standard deviation increase in house price appreciation is associated with a 10% decrease in subprime purchase share growth. The instrumental variable estimates in columns (4)–(6) are also all negative and slightly larger in magnitude than their OLS counterparts. The lack of statistical significance in column (6) is likely due to the fact that the instrument is not a strong predictor of within state variation in house price growth. However, taken together, the results in columns (4)–(6) suggest that house price appreciation had a causal (negative) effect on subprime purchase mortgage lending, consistent with the pricing out hypothesis described above. To our

¹⁷Since standard errors are clustered (at the state level), there is no widely accepted procedure to test the null hypothesis that the instrument is weak. But, the combination of a statistically significant coefficient, a high R^2 , and a Kleibergen-Paap Wald F statistic of approximately 42 suggests that the Lutz and Sand instrument is not weak

knowledge, this is the first evidence of a causal effect of house price growth on subprime purchase lending growth.

4.2 An Alternative instrument for house price growth

In this section, we explore an alternative instrumental variables strategy to identify the effect of house price appreciation on subprime purchase share growth during the 2002–2006 U.S. housing boom. We start with the panel version of the cross sectional model specified by equation (4.2):

$$\Delta s_{i,r,t} = \alpha \Delta p_{i,r,t} + \kappa X_{i,r,t} + \pi_i + \phi_{r,t} + \epsilon_{i,r,t}. \quad (4.5)$$

where $\Delta s_{i,r,t}$ is the quarterly change in log subprime purchase share in county i at time t , $\Delta p_{i,r,t}$ is contemporaneous county house price appreciation, $X_{i,r,t}$ is a vector of controls, π_i are county fixed effects, $\phi_{r,t}$ are region-time fixed effects, and $\epsilon_{i,r,t}$ is an error term.

An estimate of α from the naive regression of equation (4.5) will be subject to the same biases described in Section 4.1. In order to overcome these sources of bias, we take an approach recently developed by Guren, McKay, Nakamura, and Steinsson (2018), hereafter referred to as GMNS, which develops an instrument for local house price appreciation.¹⁸

The GMNS identification strategy relies on the fact that variation in local housing supply elasticities will lead to heterogeneous responses of local house prices to regional economic shocks. The authors estimate the historical sensitivity of city-level house price growth (core-based statistical areas, or CBSAs, to be precise) to changes in more aggregated house prices (Census Region), and then use the estimated historical sensitivities interacted with aggregate house price growth as an instrument for local house price appreciation. To provide the intuition of their strategy, GMNS use two cities located in the same Census Region (Providence, RI and Rochester, NY). They show that house prices in Providence historically have been much more sensitive to regional house price

¹⁸We are grateful to Adam Guren for providing his code.

shocks. Moreover, the differences in sensitivity across these two cities are fairly stable over time. From these facts, it can be inferred that housing supply is relatively inelastic in Providence (as compared to Rochester). Although Providence and Rochester are used as an illustrative example, similar patterns exist across many cities within different regions.

Based on this intuition, the identification strategy is straightforward. First, using quarterly data from 1975 to 2000, we estimate the county-level sensitivity of house prices to regional house price movements with the following regression:

$$\Delta p_{i,r,t} = \gamma_i \Delta P_{r,t} + \Psi Z_{i,r,t} + \delta_i \Delta y_{i,r,t} + \mu_i \Delta Y_{r,t} + \varphi_i + \zeta_{r,t} + v_{i,r,t}, \quad (4.6)$$

where i indexes county, r indexes Census Division, and t indicates the quarter-year. $\Delta P_{r,t}$ is the change in regional (Census Division) house prices in quarter-year t , and $Z_{i,r,t}$ is a vector of controls.¹⁹ Throughout this section we use CoreLogic house price indices because they are reported on a more frequent basis (quarterly) than the FHFA county indices (annually), which allows for more power in the estimation of equation (4.6). The drawback of the CoreLogic data – and the reason we do not use it as the main source of house price information in our research – is that it has much less county-level coverage than the FHFA data. We follow GMNS in terms of the control variables ($Z_{i,r,t}$) and by conditioning on changes in local and regional retail employment per capita ($y_{i,r,t}$ and $Y_{r,t}$, respectively).

The main goal of estimating the regression specified by equation (4.6) is to recover the γ_i estimates ($\hat{\gamma}_i$), which are proxies for county level (inverse) housing supply elasticities. It is important to stress that the controls in equation (4.6) – $Z_{i,r,t}$, $y_{i,r,t}$ and $Y_{r,t}$ – are crucial to the identification strategy. As GMNS point out, if local house prices react differently to regional price shocks through channels other than variation in housing supply elasticities, then identification may not

¹⁹For each of the parameters that are subscripted with an i , we estimate a county-specific coefficient. For example, $\gamma_i \Delta P_{r,t}$ is actually $\sum \gamma_i \Delta P_{r,t} I_i$, where I_i is a dummy variable for county i . This simplified notation is used throughout this section.

hold. An example used by GMNS to illustrate this point is variation in local industrial structure that causes differences in the sensitivity of local employment to aggregate business cycle shocks, which have nothing to do with housing elasticities. This could cause bias in $\hat{\gamma}_i$ due to reverse causality in equation (4.6). The inclusion of the controls should account for these sources of bias as it means that the $\hat{\gamma}_i$'s are estimated using only the variation in county-level house prices that is orthogonal to local and regional changes in retail employment and time-varying differences in county-level industry shares.

After obtaining the $\hat{\gamma}_i$'s, we then estimate the following first-stage regression:

$$\Delta p_{i,r,t} = \lambda(\hat{\gamma}_i \Delta P_{r,t}) + \tau X_{i,r,t} + \theta_i + \sigma_{r,t} + \eta_{i,r,t}. \quad (4.7)$$

The current regional shock to house prices, multiplied by the historical proxy for county supply elasticity, serves as an instrument for quarterly county-level house price appreciation from 2002–2006. $X_{i,r,t}$ includes the same control variables as equation (3.1).²⁰ We then estimate the following second stage regression:

$$\Delta s_{i,r,t} = \alpha \widehat{\Delta p}_{i,r,t} + \kappa X_{i,r,t} + \pi_i + \phi_{r,t} + \epsilon_{i,r,t}. \quad (4.8)$$

It is important to note that this approach is similar in spirit to the identification strategies that use Saiz's (2010) measure of housing supply elasticities as an instrument for house price changes (e.g., Mian, Rao, and Sufi (2013) and Mian and Sufi (2014)).²¹ However, there is a key difference in the GMNS approach. Whereas Saiz's (2010) elasticities are constructed based exclusively on measures of land unavailability, GMNS is more flexible in that it allows for a more comprehensive set of housing supply determinants. For example, the GMNS approach allows for differential

²⁰Note that the GMNS IV strategy requires the inclusion of county-by-time and division-by-time fixed effects for identification, which precludes using both the Lutz and Sand (2017) land unavailability measure and the GMNS inverse elasticity estimate as instruments in the same first stage regression.

²¹It is also similar to the strategy used by Palmer (2015) which involves instrumenting for house prices in the Great Recession using the historical variance of a city's house price index interacted with the national change in house prices.

expectations about future housing supply elasticities, due perhaps to differences in expected population growth, to affect house price dynamics. Because of this difference, GMNS argue that their measure provides a better instrument for house price growth.

Panel (b) of Figure 4 plots the $\hat{\gamma}_i$'s – proxies for county (inverse) supply elasticities – estimated from equation (4.6) using house price data from 1975–2000. These historical elasticity estimates are then interacted with quarterly regional house price growth from 2002–2006 to create the instrument.

Table 11 presents estimates from the first stage regression. Note that there is only one column in the table because, as mentioned above, the GMNS strategy requires the inclusion of county and division-by-time fixed effects. We see that the GMNS instrument is significantly positively related to house price growth. As in Section 4.1, the combination of a statistically significant coefficient, a high R^2 , and a Kleibergen-Paap Wald F statistic of almost 50 suggests that the GMNS instrument is not weak.

Coefficient estimates from the regression specified in equation (4.8) are presented in Table 12. For comparison purposes, column (1) displays results using actual county house price appreciation, while column (2) displays the second stage results using our instrument for house price growth in the first stage. Although column (1) shows a negative relationship between house price growth and subprime share growth, we note that this is a conditional correlation. The IV estimate in column (2), by contrast, suggests that house price growth *caused* a relative decline in purchase lending to subprime borrowers – consistent with our results using the land unavailability instrument in Section 4.1.

4.3 House prices, subprime lending, and first time homebuyers

We now turn to the mechanism generating the negative causal relationship between house price growth and subprime purchase share growth by exploring the relationship between house price

growth and the likelihood of an individual transitioning into homeownership. Specifically, we test the hypothesis that higher house price appreciation during the 2002–2006 housing boom decreased the rate at which subprime borrowers transitioned from renting to owning. To do this, we use the Federal Reserve Bank of New York Consumer Credit Panel (CCP) data, which is based on consumer credit records from Equifax. The data is a nationally representative sample of individuals with a Social Security number and a credit record, and tracks information reported to credit bureaus over time for individuals at a quarterly frequency. Important for our analysis, the CCP data includes an individual’s location of residence, individual debts, and credit score at the end of each quarter. For consistency with our previous analysis we collect CCP data for 2001–2006.²²

To examine the relationship between county-level house price growth and homeownership transitions for prime and subprime individuals we estimate a conditional linear probability model of the following form:²³

$$Pr(FTHB_{i,c,a,t} = 1) = \delta_1 Subprime_{i,t} + \delta_2 \Delta p_{c,t} + \delta_3 Subprime_{i,t} \cdot \Delta p_{c,t} + \Psi X_{c,t} + Age_a + Year_t + County_c + \epsilon_{i,a,c,t}, \quad (4.9)$$

where i indexes individual, c indexes county, a indexes age, and t indexes the year-quarter. The dependent variable is an indicator that takes a value of one if the individual buys a home in quarter t , conditional on not owning a home at any time prior to t .²⁴ As is common in the literature, we use first-time mortgage borrowing as a proxy for first-time home purchase. *Subprime* is an indicator variable that takes a value of one if the individual has a credit score below 660 in the current

²²For a detailed discussion of the data, see Lee and Van der Klaauw (2010) and Bhutta (2015).

²³Note that we focus on the transition for first-time homebuyers as opposed to the transition from renting to owning more broadly due, in part, to data restrictions. It is not possible to directly identify homeowners versus renters in the CCP. It is only possible to identify individuals with and without mortgage debt. An individual with no mortgage debt is either a renter, or is a homeowner with no outstanding mortgage. The CCP does provide a separate field that tells us if an individual has ever had mortgage debt in their credit profile, which we use along with a change from zero to positive mortgage debt to identify a first-time homebuyer.

²⁴An individual that becomes an owner at time t is not included in the sample in subsequent periods.

quarter, and zero otherwise.²⁵ Δp is county-level house price appreciation over the previous year. To control for changes in county-level economic conditions, X includes wage and unemployment rate changes. Since the probability of transitioning into homeownership likely varies over the life-cycle, we account for borrower age with a full set of indicator variables. To control for time-varying macroeconomic conditions, we include year fixed effects. We also include county fixed effects to control for time-invariant differences in the likelihood of transitioning into ownership across counties. Since we observe individuals over time, and treatment (house price growth) is at the county level, standard errors are adjusted for clustering at both the individual and county levels.

Regression results are reported in Table 13 . In column (1) we begin with a simple measure of county house price growth – an indicator for a “boom area” defined as a county that experienced at least 20% house price appreciation from 2002–2006. (The 20% threshold corresponds approximately to the median county house price growth in our sample.) We find that subprime individuals are less likely than prime individuals to transition into ownership. In non-boom areas, subprime individuals are 0.46 percentage points less likely than prime individuals to transition into ownership; and the gap is larger in boom areas, where subprime individuals are 0.59 percentage points less likely to transition into ownership. The mean (unconditional) transition rate is 0.80%, so the economic magnitudes of these differences are meaningful.

Although differences in the propensity to transition into ownership across prime and subprime individuals is interesting in its own right, our primary goal in this section is to examine how house price growth affects this transition and whether the impact varies across prime and subprime individuals. Column (1) shows that for prime individuals, high house price appreciation does not significantly affect the likelihood of becoming a homeowner. However, subprime individuals in boom areas are 0.13 percentage points less likely to transition into ownership than subprime borrowers in non-boom areas. This translates into a 16% decrease relative to the mean transition

²⁵Credit scores in the CCP data are Equifax Risk Scores rather than FICO scores. Both scoring systems have approximately the same range of credit scores and there is a strong correlation between the two measures (Lee and Van der Klaauw (2010), Jambulapati and Stavins (2014), and Laufer and Paciorek (2016)).

rate.

In column (2) we use recent county house price appreciation, rather than a boom indicator, as our measure of house price growth. The results are similar to column (1). Subprime individuals are less likely than prime individuals to transition into homeownership, regardless of where they are located, and recent house price growth reduces the likelihood that subprime individuals make the transition. In column (3) we rely on within-county variation in house prices to identify the effect of house price growth on transition into ownership. Recent house price appreciation (within county) is positively related to transition into ownership, but the effect is significantly smaller for subprime borrowers. In other words, house price growth is associated with a relative decline in the likelihood of transitioning into ownership for subprime individuals.

Recall that county and Census Division-by-time fixed effects are necessary to identify the causal impact of house price growth in the instrumental variables estimation using the GMNS methodology. Thus, column (4) presents OLS results including county and division-by-time fixed effects, and serves as a benchmark for the IV estimation reported in column (5).²⁶ In both columns (4) and (5) recent house price growth increases the likelihood of transition into ownership, but the magnitude of the effect is reduced considerably for subprime individuals, which suggests that recent house price appreciation causes a relative decline in the likelihood of transitioning into ownership for subprime individuals.

To summarize, using individual level credit data we document that subprime individuals were less likely to transition into homeownership during the U.S. housing boom – a finding consistent with previous studies conducted at aggregated levels (Bhutta, 2015; Foote, Loewenstein, and Willen, 2016). We also expand on the existing literature by showing that house price growth causes a decline in the likelihood of transitioning into ownership for subprime individuals (relative to prime individuals). This is consistent with subprime individuals being priced out of the high house price growth markets, rather than being responsible for the 2002–2006 U.S. housing boom.

²⁶First stage results for column (5) are reported in Appendix Table C.6 .

5 Subprime borrowers, speculative activity, and mortgage fraud

The evidence presented in Sections 3 and 4 suggests that growth in subprime mortgage lending was not a principal driver of the 2000s U.S. house price boom. But it may still have played an indirect role by facilitating activities that have been directly linked to the boom. Recent evidence in the literature has focused on two such activities: (i) speculation by real estate investors; and (ii) mortgage fraud in the form of misrepresentation of income, misrepresentation of occupancy status, and appraisal inflation.

5.1 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 they comprised roughly 50% of purchase mortgage originations in four states that experienced the most extreme housing cycle during the 2000s (AZ, CA, FL, and NV). 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 had 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 housing bust 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, such as 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 5.3 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 property purchased 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 purchase originations to 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 housing boom and was greater in areas that experienced higher house price appreciation. 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 counties.²⁷ 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 subprime borrowers fueled the housing boom.

²⁷The key difference between the top and bottom panels is the denominator. In the top left (right) panel, the denominator is the total number of privately securitized purchase mortgage 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.

5.2 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 U.S. 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 mortgage fraud related to the misrepresentation of borrower income was a common occurrence during the mid-2000s (Blackburn and Vermilyea, 2012; Jiang, Nelson, and Vytlačil, 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 was 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 mortgage 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 purchase originations to prime and subprime borrowers separately. Both series increased over time in both boom and non-boom counties. 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.

5.3 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 ($HPA < 20\%$) counties. 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 (private mortgage-backed securities) data are matched to public records data for property sales. The occupancy status reported in ABSNet is compared to the occupancy status reported in the public records.²⁸ 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, which potentially creates some survival bias.²⁹

²⁸Although 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.

²⁹While evidence exists suggesting that subprime and prime loans end in foreclosure at similar rates once current

Figure 7 presents the rate of estimated occupancy fraud across boom and non-boom counties using the sample of privately securitized mortgages from ABSNet. 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 reported by Griffin and Maturana (2016b).³⁰ 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.

5.4 Appraisal inflation

Although appraisals are supposed to be unbiased estimates of market value, an overwhelming amount of evidence from the U.S. housing cycle of the mid-2000s 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, 2019; 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).

loan-to-value 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.

³⁰As 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).

In this section we ask whether appraisal inflation for purchase mortgages 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.³¹ Because both the appraisal 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 privately securitized 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 purchase 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 purchase loans tracks very closely. This finding suggests that appraisal fraud was not concentrated in loans to subprime borrowers.

³¹Griffin 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.

6 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 purchase mortgage lending to subprime borrowers.

One explanation for the negative spatial correlation between house price growth and growth in purchase mortgages to subprime borrowers at the county level over the housing boom is reverse causality. That is, that rapidly increasing home prices made homeownership unattainable for many prospective subprime borrowers, which tend to have lower income and wealth on average. Using two different instrumental variables approaches, we establish a negative causal relationship running from house price growth to growth in subprime purchase lending. Using data from the Federal Reserve Bank of New York Consumer Credit Panel, we also illustrate that county house price growth had a negative effect on subprime individuals becoming first-time home buyers.

We complement the aforementioned findings with new evidence showing that subprime borrowers did not play a prominent role in activities that have been tied to the U.S. house price boom of the 2000s. 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 illustrate 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.

The findings in this paper contribute to the “new narrative” that rapid U.S. house price appreciation during the 2000s was mainly driven by prime borrowers. Hence, policy prescriptions

intended to limit access to credit for marginal borrowers are unlikely to prevent a future housing boom.

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.
- 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.

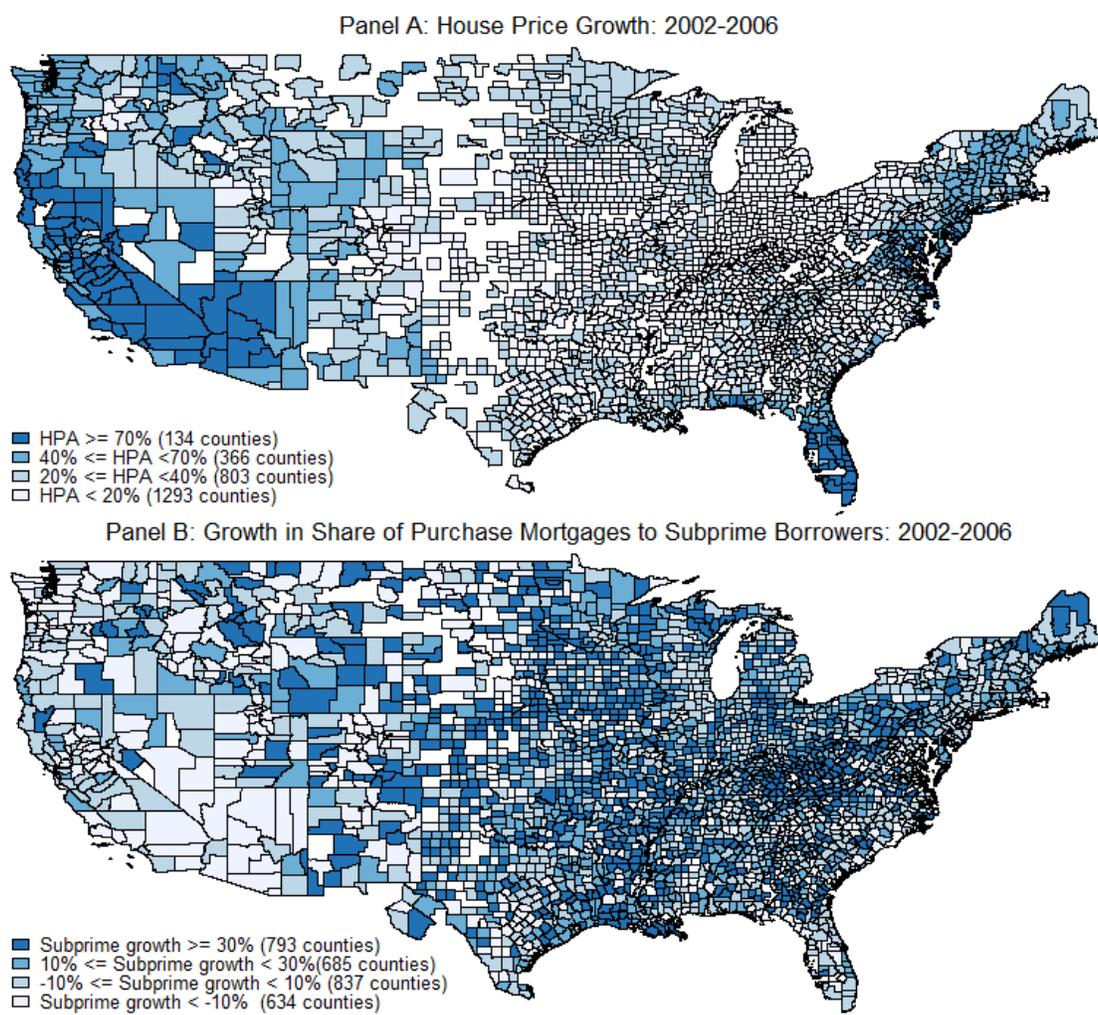
- CONKLIN, J., E. COULSON, M. DIOP, AND T. LE (2019): “Competition and Appraisal Inflation,” *The Journal of Real Estate Finance and Economics*, Forthcoming.
- DAVIDOFF, T. (2016): “Supply constraints are not valid instrumental variables for home prices because they are correlated with many demand factors,” *Critical Finance Review*, 5(2), 177–206.
- 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.
- DOKKO, J., B. J. KEYS, AND L. RELIHAN (2019): “Affordability, Financial Innovation, and the Start of the Housing Boom,” .
- FAVARA, G., AND J. IMBS (2015): “Credit supply and the price of housing,” *American Economic Review*, 105(3), 958–92.
- 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., K. S. GERARDI, AND P. S. WILLEN (2012): “Why did so many people make so many ex post bad decisions? The causes of the foreclosure crisis,” Discussion paper, National Bureau of Economic Research.
- 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.

- GUREN, A. M., A. MCKAY, E. NAKAMURA, AND J. STEINSSON (2018): “Housing wealth effects: The long view,” Discussion paper, National Bureau of Economic Research.
- 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*.
- JAMBULAPATI, V., AND J. STAVINS (2014): “Credit CARD Act of 2009: What did banks do?,” *Journal of Banking & Finance*, 46, 21–30.
- 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.
- 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.
- LAUFER, S., AND A. PACIOREK (2016): “The effects of mortgage credit availability: Evidence from minimum credit score lending rules,” .
- LEE, D., AND W. VAN DER KLAAUW (2010): “An introduction to the FRBNY consumer credit panel,” *FRB of New York Staff Report*, (479).
- LUTZ, C., AND B. SAND (2017): “Highly Disaggregated Topological Land Unavailability,” .
- 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., K. RAO, AND A. SUFI (2013): “Household balance sheets, consumption, and the economic slump,” *The Quarterly Journal of Economics*, 128(4), 1687–1726.
- 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.
- (2011): “House prices, home equity-based borrowing, and the US household leverage crisis,” *American Economic Review*, 101(5), 2132–56.
- (2014): “What explains the 2007–2009 drop in employment?,” *Econometrica*, 82(6), 2197–2223.
- (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.
- PALMER, C. (2015): “Why did so many subprime borrowers default during the crisis: Loose credit or plummeting prices?,” .
- 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.
- SAIZ, A. (2010): “The geographic determinants of housing supply,” *The Quarterly Journal of Economics*, 125(3), 1253–1296.
- SHI, L., AND Y. ZHANG (2015): “Appraisal inflation: Evidence from the 2009 GSE HVCC intervention,” *Journal of Housing Economics*, 27, 71–90.
- STROEBEL, J., AND J. VAVRA (2019): “House Prices, Local Demand, and Retail Prices,” *Journal of Political Economy*, 127(3), 1391–1436.

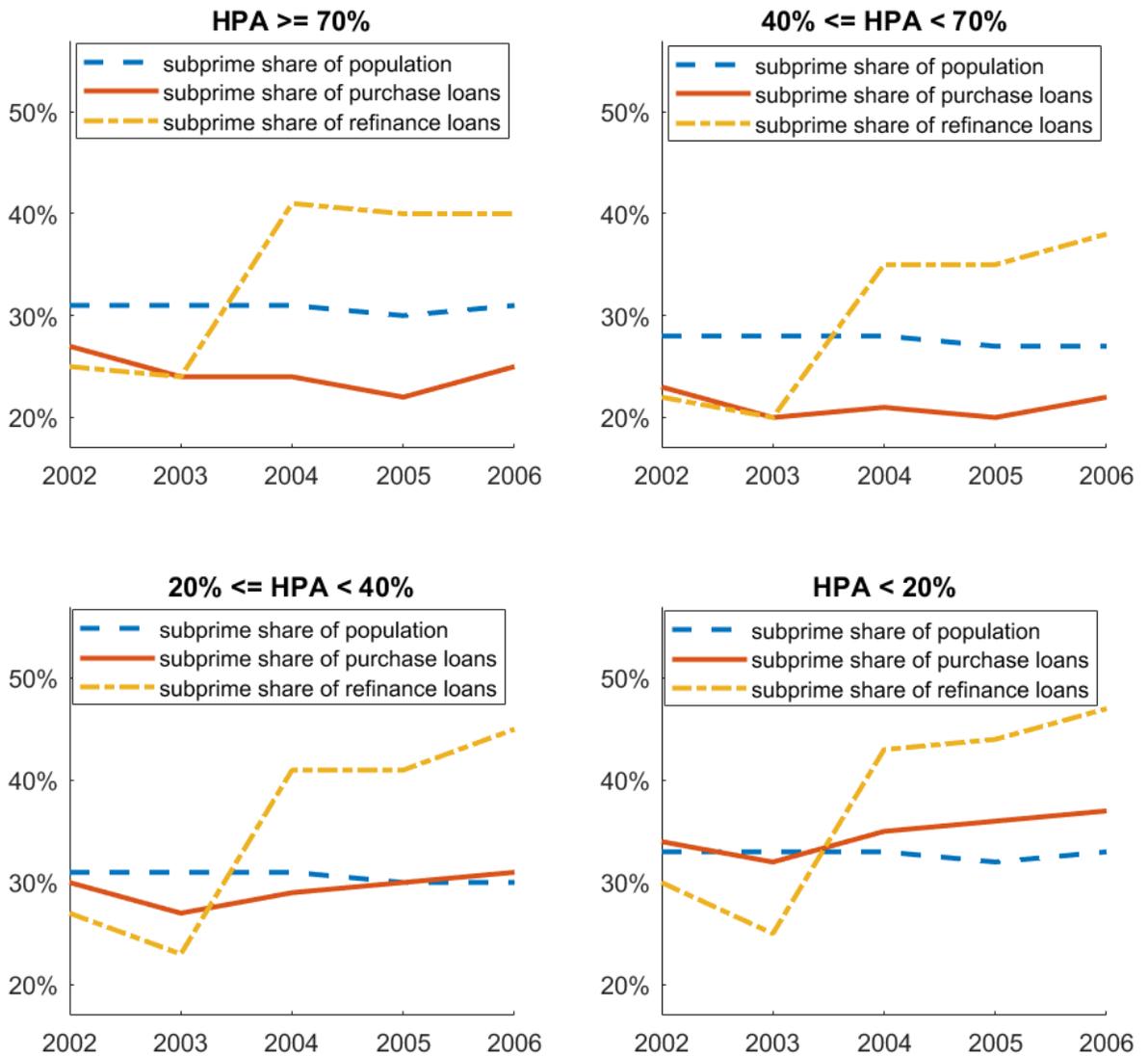
Figures and Tables

Figure 1 . Growth in U.S. County House Prices and Share of Purchase Mortgages to Subprime Borrowers between 2002 and 2006



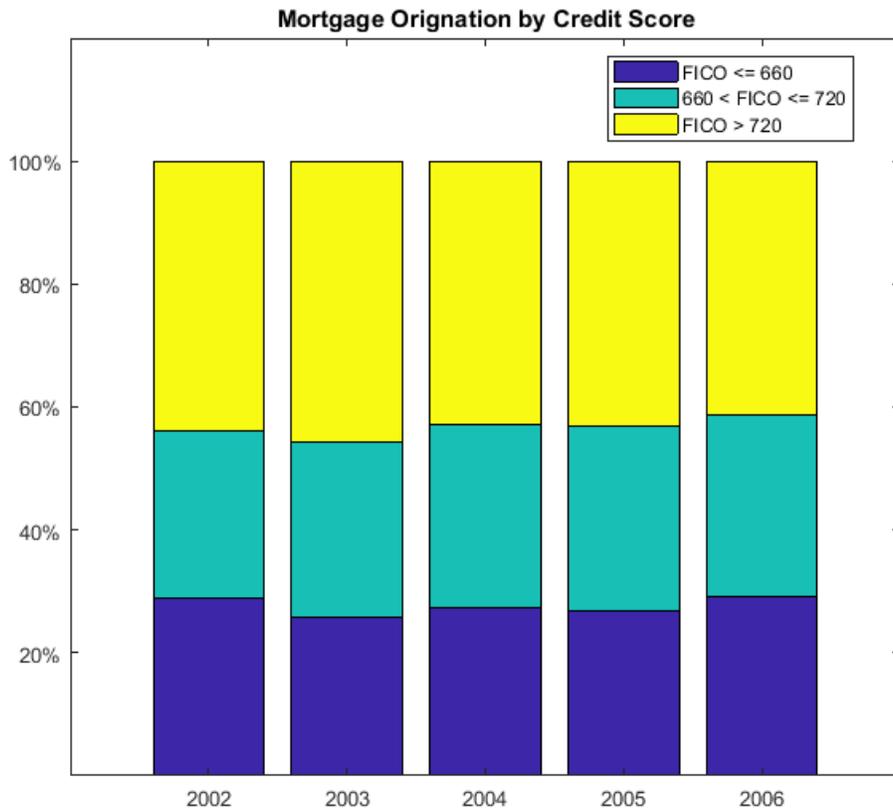
Sources: FHFA, McDash, ABSNet, and authors' calculations. Panel A presents U.S. county house price appreciation (HPA) between 2002 and 2006 from the FHFA. Color indicates HPA ranging from light blue (low HPA) to dark blue (high HPA). Panel B illustrates the change in the share of first-lien purchase mortgages to subprime borrowers (FICO \geq 660) using merged servicing data from McDash and ABSNet (excluding duplicates). Color indicates subprime growth ranging from light blue (subprime contraction) to dark blue (subprime expansion).

Figure 2 . Average Annual Share of U.S. County Subprime Population and Mortgage Originations Grouped by County House Price Growth between 2002 and 2006



Sources: GeoFred, McDash, ABSNet, and authors' calculations. This figure plots average shares of U.S. county subprime population and first-lien mortgages (delineated by purchases and refinances) grouped by house price appreciation (HPA) annually between 2002 and 2006. First lien mortgages to subprime borrowers (FICO < 660) are identified in merged servicing data from McDash and ABSNet (excluding duplicates). Each panel of this figure represents county averages by HPA quartile (HPA \geq 70%, 40% \leq HPA < 70%, 20% \leq HPA < 40%, and HPA < 20%).

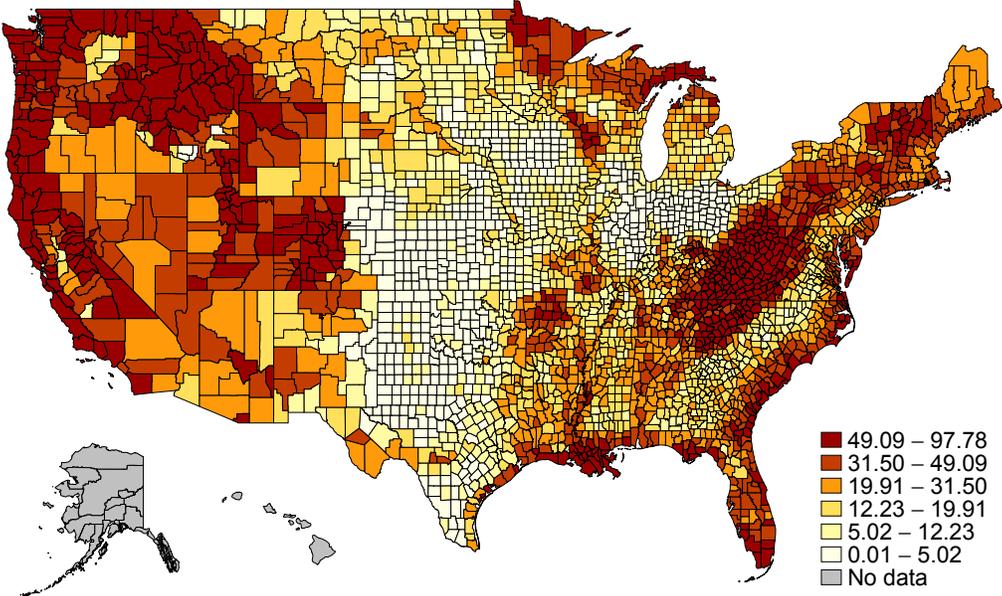
Figure 3 . Annual Share of Home Purchase Mortgage Originations by FICO Score: 2002-2006



Sources: McDash, ABSNet, and authors' calculations. This figure presents the share of first-lien home purchase mortgage originations grouped by FICO credit scores (FICO > 720, 660 < FICO ≤ 720, and FICO ≤ 660) annually between 2002 and 2006. First lien mortgages are identified in merged servicing data from McDash and ABSNet (excluding duplicates).

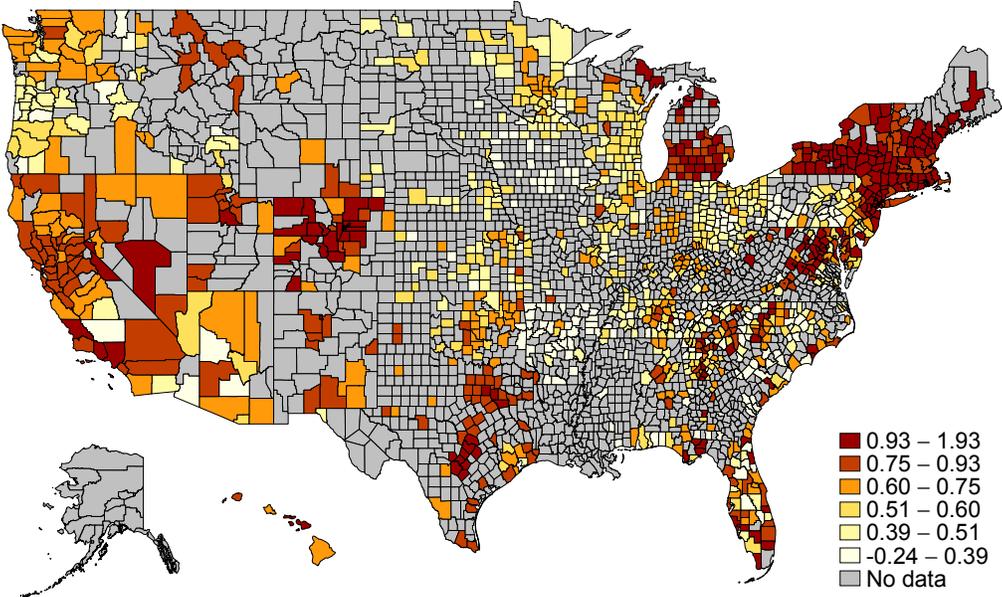
Figure 4 . Instruments for County House Price Growth

(a) Share of Undevelopable Land



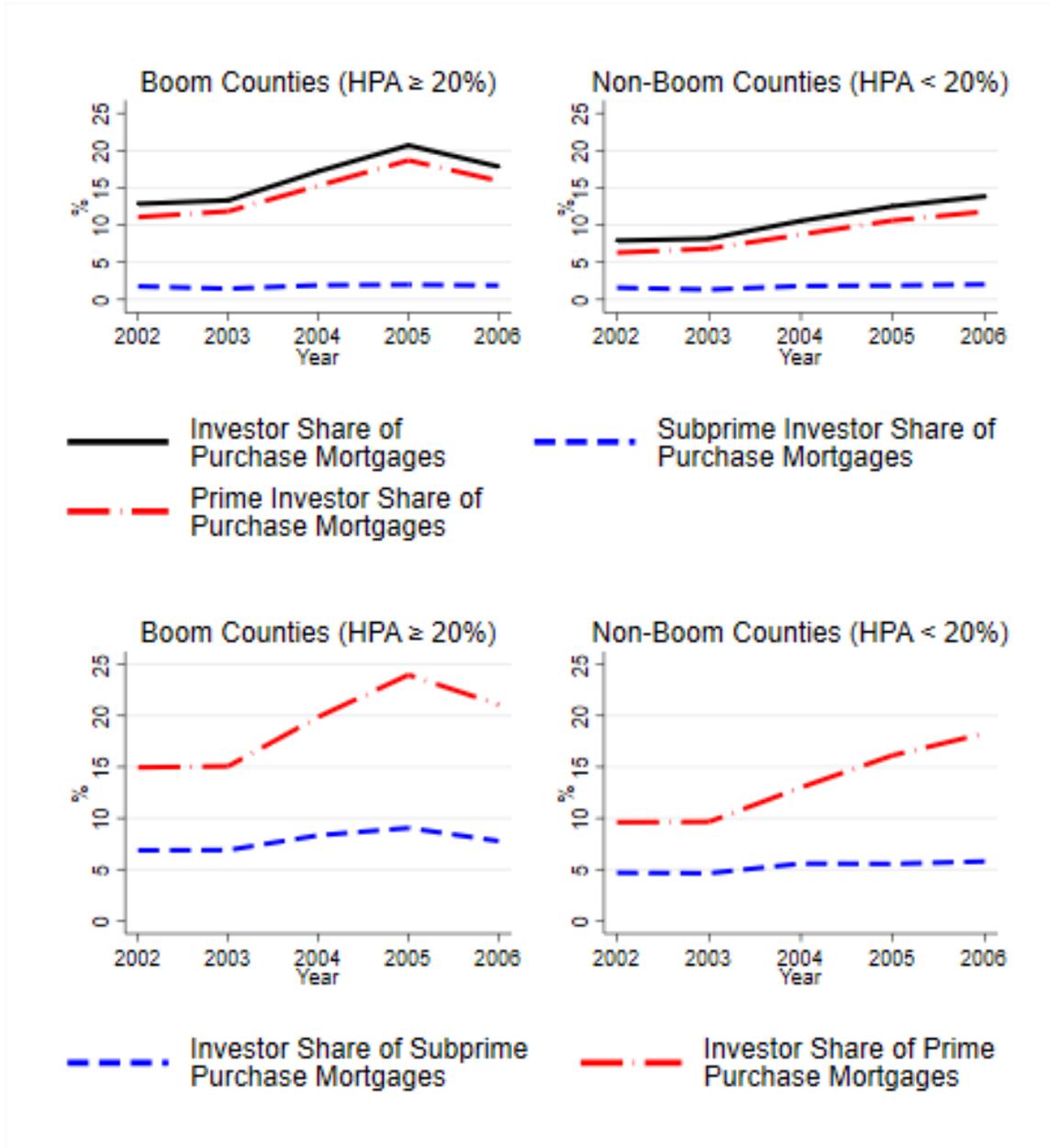
Source: Lutz and Sand (2017). This Figure presents the share of undevelopable land at the county level. Darker colors correspond to larger shares of undevelopable land.

(b) County-level GMNS γ 's



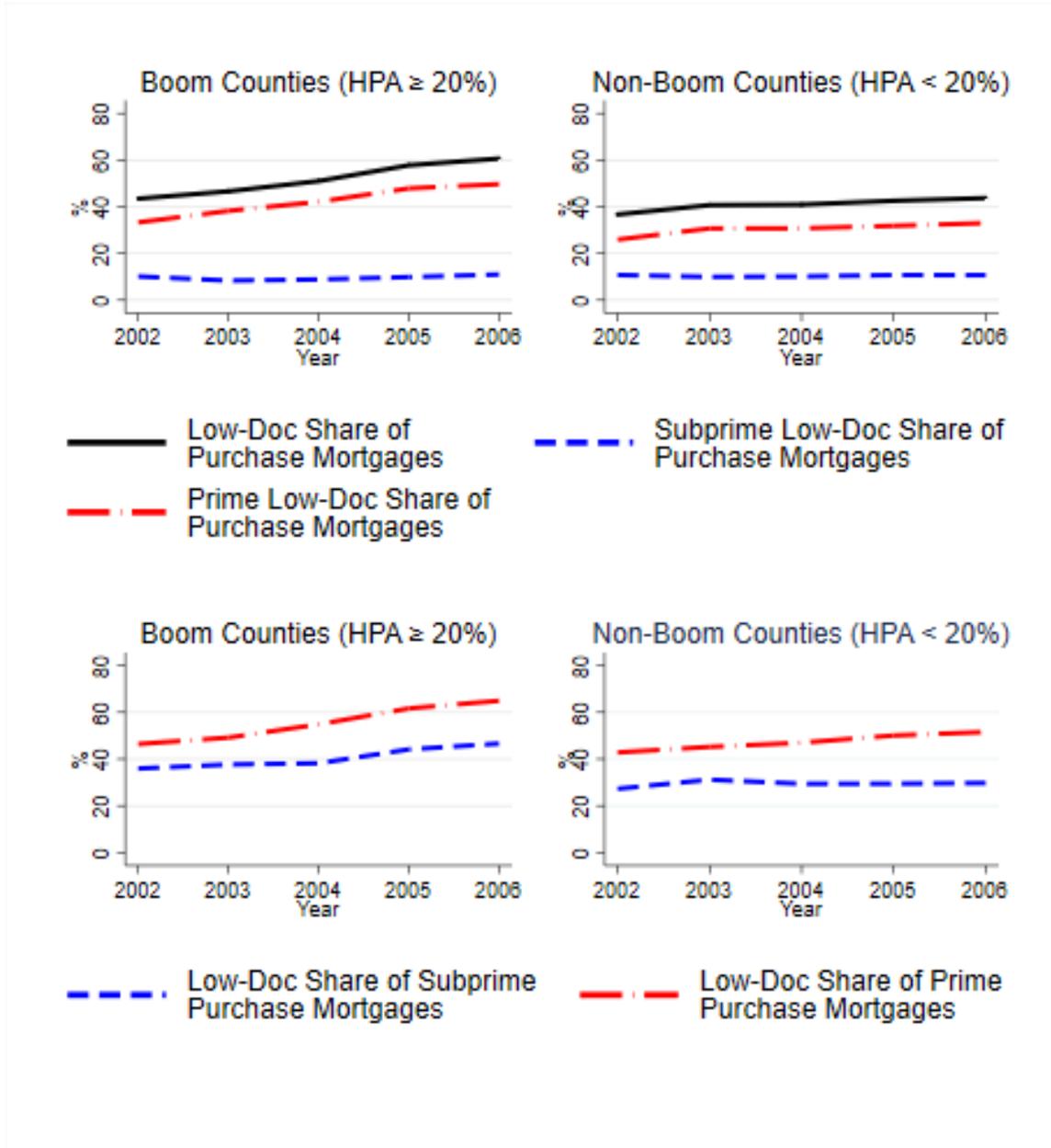
Source: CoreLogic and authors' calculations. This figure plots estimated inverse housing supply elasticities based on the method presented in Guren, et al. (2018) leveraging CoreLogic quarterly county house price indices from 1975-2000.

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



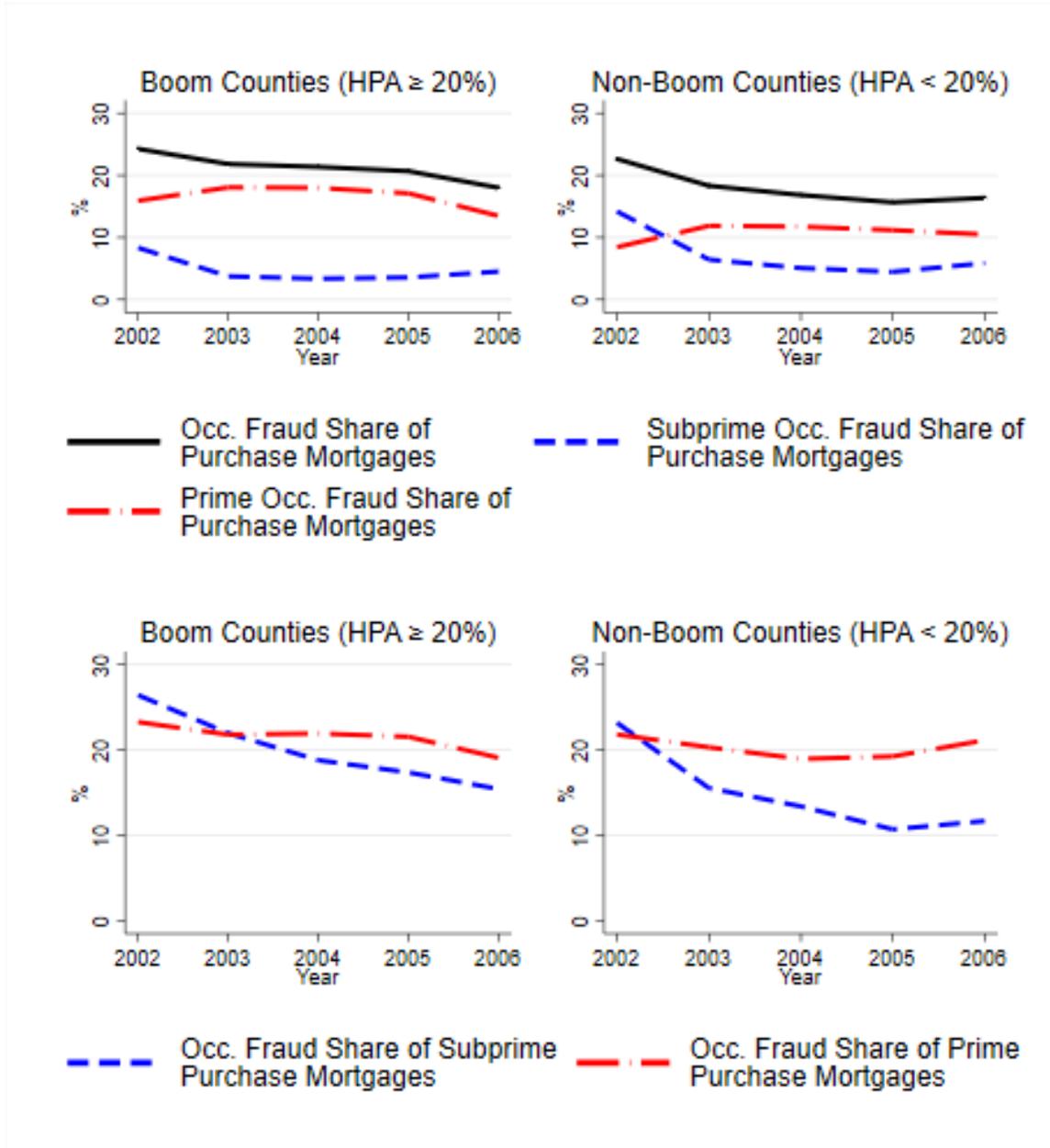
Source: McDash, ABSNet, FHFA, and authors' calculations. The top two panels plot the annual investor share of total purchase mortgage originations (solid black line) and the investor share broken down into contributions from subprime (FICO < 660) and prime (FICO ≥ 660) borrowers (dotted blue and red lines respectively). The bottom two panels plot the investor share of subprime and prime purchase mortgage originations, respectively (i.e. the investor share of subprime mortgages is the number of subprime purchase loans to investors divided by the total number of subprime purchase loans). For both the top and bottom panels, the series are delineated for boom counties (HPA $\geq 20\%$) and non-boom counties (HPA $< 20\%$).

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



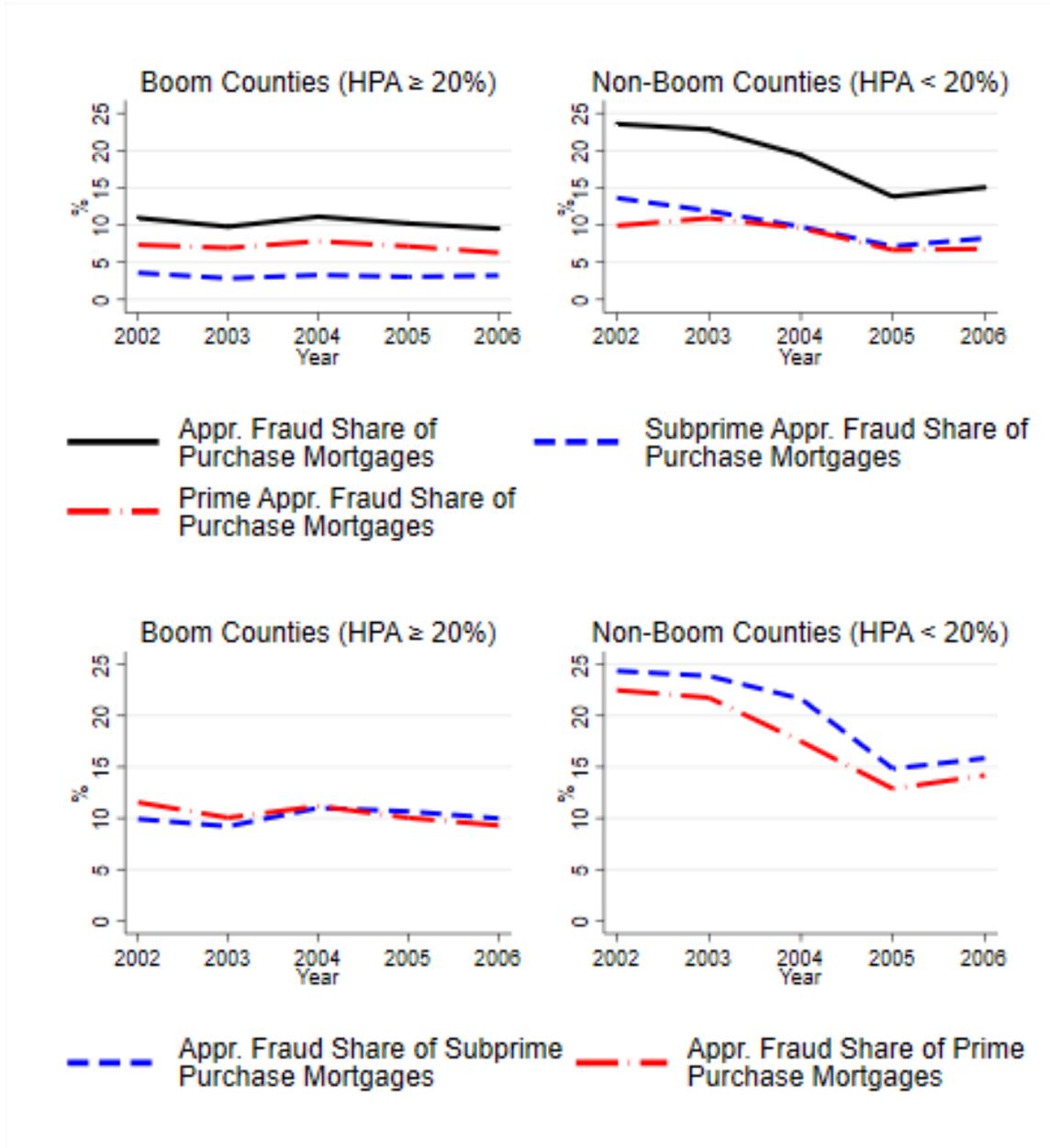
Sources: McDash, ABSNet, FHFA, and authors' calculations. The top two panels plot the annual low documentation (low-doc) share of total purchase mortgage originations (solid black line) and the low-doc share broken down into contributions from subprime (FICO < 660) and prime (FICO ≥ 660) borrowers (dotted blue and red lines, respectively). The bottom two panels plot the low-doc share of subprime and prime purchase mortgage originations, respectively (i.e. the low-doc share of subprime is the number of subprime low-doc purchase loans divided by the total number of subprime purchase loans). For both the top and bottom panels, the series are delineated for boom counties (HPA ≥ 20%) and non-boom counties (HPA < 20%).

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



Sources: McDash, ABSNet, FHFA, and authors' calculations. The top two panels plot the annual occupancy fraud share of purchase mortgage originations (solid black line) and the occupancy fraud share broken down into contributions from subprime (FICO < 660) and prime (FICO ≥ 660) borrowers (dotted blue and red lines, respectively). The bottom two panels plot the occupancy fraud share of subprime and prime purchase mortgage originations, respectively (i.e. the occupancy fraud share of subprime mortgages is the number of subprime purchase loans characterized by occupancy fraud divided by the total number of subprime purchase loans). For both the top and bottom panels, the series are delineated for boom counties (HPA ≥ 20%) and non-boom counties (HPA < 20%).

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



Sources: ABSNet, FHFA, and authors' calculations. The top two panels plot the annual appraisal fraud share of total purchase mortgage originations (solid black line) and the appraisal fraud share broken down into contributions from subprime (FICO < 660) and prime (FICO ≥ 660) borrowers (dotted blue and red lines, respectively). The bottom two panels plot the appraisal fraud share of subprime and prime purchase mortgage originations, respectively (i.e. the appraisal fraud share of subprime mortgages is the number of subprime purchase loans characterized by appraisal fraud divided by the total number of subprime loans). For both the top and bottom panels, the series are delineated for boom counties (HPA ≥ 20%) and non-boom counties (HPA < 20%).

Table 1 . Summary Statistics

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

Sources: ABSNet, BLS, FHFA, GeoFred, IRS, and McDash. This table displays means, standard deviations (in parenthesis), and medians for all variables used in the regression analysis 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 in U.S. County House Prices and Share of Purchase Mortgages to Subprime Borrowers between 2002 and 2006

Dependent Variable:	$\Delta \log(\text{HPI})$ 2002–2006					
	Weighted			Unweighted		
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta \log(\text{Subprime Share})$	-0.444*** (0.113)	-0.215*** (0.075)	-0.135*** (0.026)	-0.122*** (0.027)	-0.154*** (0.020)	-0.073*** (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 the change in the log of the FHFA county house price index from 2002 to 2006 on the contemporaneous change in log of county subprime share of purchase mortgage originations. Regressions in columns (1) - (3) are weighted by the total number of loans in the county across both years (2002 and 2006), while 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 share of the county population that is subprime, the county subprime share of purchase mortgages, the number of mortgages originated, county average wages, and county unemployment rate. The change variables include: the change in the log of the subprime share of the population, the change in log of average wages, and the change in the unemployment rate. Robust standard errors are in parentheses and are clustered at the state level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table 3 . Growth in U.S. County House Prices and Share of Purchase Mortgages to Subprime Borrowers over Different Periods

Dependent Variable:	$\Delta \log(\text{HPI})$					
	2002–2006	2002–2005	2001–2006	2001–2005	2003–2005	2003–2006
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta \log(\text{Subprime Share})$	-0.135*** (0.026)	-0.146*** (0.031)	-0.146*** (0.030)	-0.154*** (0.030)	-0.120*** (0.032)	-0.129*** (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 the change in the log of the FHFA county house price index over different time periods on the contemporaneous change in the log of county subprime share of purchase mortgage originations. Regressions are weighted by the total number of loans in the county across starting and ending years. 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 purchase mortgages, the number of loans originated, county average wages, and county unemployment rate. The change variables include: the change in the log of the subprime share of the population, the change in log of average wages, and the change in the unemployment rate. Robust standard errors are in parenthesis and are clustered at the state level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table 4 . Growth in U.S. County House Prices and Share of Purchase Mortgages to Subprime Borrowers for Different FICO Thresholds

Dependent Variable:	$\Delta \log(\text{HPI})$ 2002–2006					
Subprime Definition:	FICO \leq 620			FICO \leq 580		
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta \log(\text{Subprime Share})$	-0.403*** (0.050)	-0.274*** (0.039)	-0.138*** (0.020)	-0.270*** (0.034)	-0.188*** (0.026)	-0.091*** (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 the change in the log of the FHFA county house price index from 2002 to 2006 on the contemporaneous change in the log county subprime share of purchase mortgage originations using different FICO thresholds to define subprime. Regressions 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 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 purchase mortgages, the number of loans originated, county average wages, and county unemployment rate. The change variables include: the change in the log of the subprime share of the population, the change in log of average wages, and the change in the unemployment rate. Robust standard errors are in parenthesis and are clustered at the state level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table 5 . Growth of U.S. County House Prices and Share of Purchase Mortgages to Subprime Borrowers for Areas with High and Low Initial Subprime Shares

Dependent Variable:	$\Delta \log(\text{HPI})$ 2002–2006		
	All Counties	Low Initial Share	High Initial Share
	(1)	(2)	(3)
$\Delta \log(\text{Subprime Share})$ 2002-2006	-0.135*** (0.026)	-0.129*** (0.033)	-0.143*** (0.023)
$\log(\text{Subprime Share})$ 2002	-0.231* (0.133)	-0.155 (0.164)	-0.271*** (0.092)
Covariates	Y	Y	Y
State FE	Y	Y	Y
# Observations	2,384	1,192	1,192
Adjusted R^2	0.90	0.91	0.94

Note: This table reports estimates from a regression of the change in the log FHFA county house price index from 2002 to 2006 on the contemporaneous change in the log county subprime share of purchase mortgage originations separately for areas with low and high initial subprime purchase shares. Regressions are weighted by the total number of purchase loans in the county across both years (2002 and 2006). Column (1) includes all counties, column (2) includes only counties with 2002 subprime purchase share less than or equal to the sample median, and column (3) includes counties with 2002 subprime purchase share greater than the sample median. 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 purchase mortgages, the number of loans originated, county average wages, and county unemployment rate. The change variables include: the change in the log of the subprime share of the population, the change in log of average wages, and the change in the unemployment rate. Robust standard errors are in parenthesis and are clustered at the state level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table 6 . Growth of U.S. County House Prices and Share of Purchase Mortgages to Subprime Borrowers Using Alternative Measures of Subprime Purchase Activity

Dependent Variable:	$\Delta \log(\text{HPI})$ 2002–2006					
	(1)	(2)	(3)	(4)	(5)	(6)
Δ (Subprime Share)	-1.603*** (0.387)	-0.915*** (0.262)	-0.465*** (0.095)			
$\Delta \log(\#$ Subprime Purchase Loans)				-0.101 (0.107)	-0.056 (0.061)	-0.018 (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 the change in the log FHFA county house price index from 2002 to 2006 on the contemporaneous change in the subprime purchase share (columns (1) - (3)) and the change in the log of the number of subprime purchase mortgages (columns (4) - (6)). Regressions are weighted by the total number of loans in the county across both years (2002 and 2006). Subprime shares are identified as first-lien purchase originations 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 purchase mortgages, the number of loans originated, county average wages, and county unemployment rate. The change variables include: the change in the log of the subprime share of the population, the change in log of average wages, and the change in the unemployment rate. Robust standard errors are in parenthesis and are clustered at the state level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table 7 . Growth of U.S. County House Prices and Share of Refinance Mortgages to Subprime Borrowers

Dependent Variable:	$\Delta \log(\text{HPI})$ 2002–2006					
	Weighted			Unweighted		
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta \log(\text{Subprime Share})$	-0.040 (0.183)	-0.004 (0.093)	0.094 (0.101)	0.105*** (0.034)	-0.042 (0.038)	-0.040** (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 the 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. Regressions in columns (1) - (3) are weighted by the total number of loans in the county across both years (2002 and 2006), while regressions in columns (4) - (6) are unweighted. 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 purchase mortgages, the number of loans originated, county average wages, and county unemployment rate. The change variables include: the change in the log of the subprime share of the population, the change in log of average wages, and the change in the unemployment rate. Robust standard errors are in parentheses and are clustered at the state level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table 8 . Growth of U.S. County House Prices and the Share of Purchase Mortgages to Low Income Borrowers

Dependent Variable:	$\Delta \log(\text{HPI})$ 2002–2006		
	(1)	(2)	(3)
$\Delta \log(\text{Low Income Share})$	-0.160*** (0.016)		
$\Delta \log(\text{Low Wage Share})$		-0.146*** (0.014)	
$\Delta \log(\text{Share Income} < 50\text{K})$			-0.156*** (0.015)
Covariates	Y	Y	Y
State FE	Y	Y	Y
# Observations	2,598	2,591	2,599
Adjusted R^2	0.92	0.94	0.93

Note: This table reports estimates from a regression of the change in the log FHFA county house price index from 2002 to 2006 on the contemporaneous change in log county low income share of home purchase loans from HMDA. Regressions are weighted by the total number of loans in the county across both years (2002 and 2006). In column (1), the 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. In column (2), the 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. In column (3), Share Income < 50K is calculated as 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 percentage of the county population that is subprime, the county subprime share of purchase mortgages, the number of loans originated, county average wages, and county unemployment rate. The change variables include: the change in the log of the subprime share of the population, the change in log of average wages, and the change in the unemployment rate. Robust standard errors are in parentheses and are clustered at the state level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table 9 . 1st Stage Estimates of U.S. County House Price Growth Using Lutz and Sand (2017) Land Unavailability Measure

Dependent Variable:	$\Delta \log(\text{HPI})$ 2002–2006		
	(1)	(2)	(3)
Land Unavailable	0.006*** (0.001)	0.003*** (0.001)	0.001 (0.000)
Covariates	N	Y	Y
State FE	N	N	Y
# Observations	2,384	2,384	2,384
Adjusted R^2	0.31	0.57	0.90
1st Stage Wald F-Stat	42.0	21.7	1.8

Notes: This table reports estimates from a regression of the change in the log FHFA county house price index from 2002 to 2006 on the share of undevelopable land from Lutz and Sand (2017). Regressions are weighted by the total number of purchase loans in the county across both years (2002 and 2006). 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 purchase mortgages, the number of loans originated, county average wages, and county unemployment rate. The change variables include: the change in the log of the subprime share of the population, the change in log of average wages, and the change in the unemployment rate. The Kleibergen-Paap Wald F statistic is reported in all columns. Robust standard errors are in parentheses and are clustered at the state level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table 10 . Growth in the County Share of Purchase Mortgages to Subprime Borrowers and House Price Growth (Instrumented using Lutz-Sand Land Unavailability)

Dependent Variable	$\Delta \log(\text{Subprime Share})$ 2002–2006					
	OLS			2SLS		
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta \log(\text{HPI})$ 2002-2006	-0.333*** (0.083)	-0.267*** (0.085)	-0.504*** (0.121)			
$\Delta \log(\text{HPI})$ 2002-2006 (Instrumented)				-0.381** (0.155)	-0.562* (0.299)	-0.857 (0.888)
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	0.24	0.49			
1st Stage Wald F-Stat				42.0	21.7	1.8

Notes: This table reports estimates from a regression of the change in the log county subprime share of purchase mortgages on the contemporaneous change in the log FHFA county house price index from 2002 to 2006. Columns (1) - (3) are baseline OLS estimates using the change in the log FHFA county house price index, while columns (4) - (6) use the instrumented version based on the share of undevelopable land from Lutz and Sand (2017). Subprime shares are calculated based on the fraction of first-lien purchase mortgage originations in a county with a FICO score less than 660. Regressions are weighted by the total number of purchase loans in the county across both years (2002 and 2006). 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 purchase mortgages, the number of loans originated, county average wages, and county unemployment rate. The change variables include: the change in the log of the subprime share of the population, the change in log of average wages, and the change in the unemployment rate. The Kleibergen-Paap Wald F statistic is reported in all columns (4) - (6). Robust standard errors are in parentheses and are clustered at the state level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table 11 . 1st Stage Estimates of U.S. County House Price Growth Using GMNS Instrument

Dependent Variable:	$\Delta \log(\text{HPI})$ Quarterly
	(1)
GMNS Instrument	0.087*** (0.123)
Covariates	Y
County FE	Y
Division-by-Time FE	Y
# Observations	25,252
Adjusted R^2	0.73
1st Stage Wald F-Stat	49.6

Notes: This table reports estimates from a regression of the quarterly change in the log CoreLogic county house price index on the product of the change in the log CoreLogic Census Region house price index and estimated inverse housing supply elasticity using the Guren et al. (2018) methodology between 2002 and 2006. Regressions are weighted by the total number of purchase loans in the county during the quarter. 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 purchase mortgages, the number of loans originated, county average wages, and county unemployment rate. The change variables include: the change in the log of the subprime share of the population, the change in log of average wages, and the change in the unemployment rate. The Kleibergen-Paap Wald F statistic is also reported. Robust standard errors are in parentheses and are clustered at the state level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table 12 . Growth in the County Share of Purchase Mortgages to Subprime Borrowers and House Price Growth (Instrumented using GNMS)

Dependent Variable:	$\Delta \log(\text{Subprime Share})$ Quarterly	
	OLS	2SLS
	(1)	(2)
$\Delta \log(\text{HPI})$ Quarterly	-0.323*** (0.044)	
$\Delta \log(\text{HPI})$ Quarterly (Instrumented)		-0.475** (0.202)
Covariates	Y	Y
County FE	Y	Y
Division-by-Time FE	Y	Y
# Observations	25,252	25,252
Adjusted R^2	0.52	
1st Stage Wald F-Stat		49.6

Notes: This table reports estimates from a regression of the quarterly change in log county subprime share of purchase mortgages on the contemporaneous change in the log CoreLogic county house price index (actual and instrumented using the GNMS methodology) between 2002 and 2006. Regressions are weighted by the total number of purchase loans in the county during the quarter. Column (1) presents baseline OLS estimates using actual CoreLogic county house price growth as the primary independent variable of interest, while column (2) uses the instrumented version of the change in the log CoreLogic county house price index. Subprime share is calculated based on the fraction of first-lien purchase mortgage originations in a county with a FICO score less than 660. 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 purchase mortgages, the number of loans originated, county average wages, and county unemployment rate. The change variables include: the change in the log of the subprime share of the population, the change in log of average wages, and the change in the unemployment rate. The Kleibergen-Paap Wald F statistic from the 1st stage regression is reported in column (2). Robust standard errors are in parentheses and are clustered at the state level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table 13 . Individual First-Time Homeownership Transitions and U.S. County House Price Growth

Dependent Variable:	Pr(FTHB=1)				
	OLS				2SLS
	(1)	(2)	(3)	(4)	(5)
Subprime	-0.0046*** (0.0002)	-0.0052*** (0.0002)	-0.0051*** (0.0002)	-0.0050*** (0.0002)	-0.0047*** (0.0002)
Boom Area	0.0003 (0.0002)				
Subprime × Boom Area	-0.0013*** (0.0002)				
$\Delta \log(\text{HPI})$		0.0025 (0.0021)	0.0082*** (0.0008)	0.0068*** (0.0009)	
Subprime × $\Delta \log(\text{HPI})$		-0.0027* (0.0015)	-0.0027** (0.0013)	-0.0029** (0.0013)	
$\Delta \log(\text{HPI})$ (Instrumented)					0.0107*** (0.0034)
Subprime × $\Delta \log(\text{HPI})$ (Instrumented)					-0.0077*** (0.0024)
County Controls	Y	Y	Y	Y	Y
Age FE	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y
County FE	N	N	Y	Y	Y
Division-by-Time FE	N	N	N	Y	Y
# Observations	22,760,616	22,760,616	22,760,616	22,760,616	22,760,616
Adjusted R^2	0.00	0.00	0.00	0.01	
1 st Stage Wald F-Stat					11.3

Notes: This table reports estimates from a linear probability model of the quarterly likelihood of an individual becoming a first-time homebuyer (FTHB) on measures of contemporaneous county house price growth (actual and instrumented using the GNMS methodology) between 2002 and 2006. The dependent variable is based on data in the Federal Reserve Bank of New York Consumer Credit Panel and takes a value of one the first quarter that a mortgage appears on the individuals credit report. Subprime is an indicator that takes a value of one if the borrowers credit score is below 620 in the current quarter and zero otherwise. House price growth measures vary across the columns. Column (1) includes a dummy variable Boom Area that equals one if the individual is located in a county that experienced at least 20% price appreciation from 2002 - 2006. Columns (2) and (3) include the change in the log CoreLogic county house price index over the previous year. Columns (4) and (5) include the instrumented change in the log CoreLogic county house price index over the previous year based on the GNMS methodology. First stage estimates for column (5) are reported in Appendix Table C.6 . Covariates include the annual changes in county log wages and the change in the unemployment rate. All regressions include fixed effects for an individuals age and the year of observation. Robust standard errors are in parentheses and are clustered at the individual and county levels. *** p < 0.01, ** p < 0.05, * p < 0.10

Online Appendix

Appendix A: Procedure for Merging ABSNet with McDash

The following is the procedure used to construct a combined dataset from ABSNet and McDash and removing all duplicates.

- 1) Identify and remove duplicates within the McDash data by searching for loans with the identical closing month, loan purpose (refinance or purchase), loan amount, credit score, and ZIP Code. The McDash data used in this research includes 5-digit ZIP Codes for non-GSE loans and 3-digit ZIP Codes for GSE loans.
- 2) Identify and remove duplicates within the ABSNet data in similar fashion. The matching criteria are again the closing month, loan purpose, loan amount, credit score, and ZIP Code.
- 3) Identify and remove duplicates between ABSNet and McDash.
 - 3-A) Find loans with identical closing months, loan purposes, loan amounts, credit scores, and ZIP Codes.
 - 3-B) Note that the credit score match is complicated by the fact that McDash and ABSNet may report scores from different vendors (e.g., FICO v. Equifax). Dropping the matching credit score condition is likely to result in some duplicates entering the final sample. Thus, we replace the identical credit score condition with a requirements to match loan-to-value (LTV) and mortgage type (ARM/FRM). In other words, matching is based on loans having 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. Since there are multiple important dates in a real estate transaction, the closing date could 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 B: ZIP Code Level Analysis

Our main analysis is conducted at the county level for a few reasons. First, the vast majority of variation in house prices is across counties rather than within counties. Table B.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. By 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 U.S. counties, the limited availability of ZIP Code level house price information reduces our coverage to 13% - 35% of ZIP Codes depending on the index used.

This appendix provides results from ZIP Code level analysis, which has been the focus of some previous papers in the related literature (e.g., Mian and Sufi (2009), Adelino, Schoar, and Severino (2016), and Foote, Loewenstein, and Willen (2016)). Thus, it is important to make sure that our results are robust to this change. In addition, by estimating regressions at the ZIP Code level, 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 A. 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 5, which are not available in the CoreLogic dataset.

We display regression results at the ZIP Code level in Table B.2. We show results using two different 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.³² For each house price index, 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, the ZIP Code level analysis also suggests a negative relationship between the growth in subprime purchase shares and house price appreciation.

³²There 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.

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

	Zip Codes (1)	Counties (2)	Mean (3)	SD (4)	Between county (SD) (5)	Within county (SD) (6)
$\Delta \log(\text{home price index})$ [CoreLogic]	5,350	799	0.392	0.225	0.192	0.050
$\Delta \log(\text{home price index})$ [FHFA]	14,481	2,324	0.289	0.191	0.146	0.048
$\Delta \log(\text{subprime share of purchase loans})$	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.

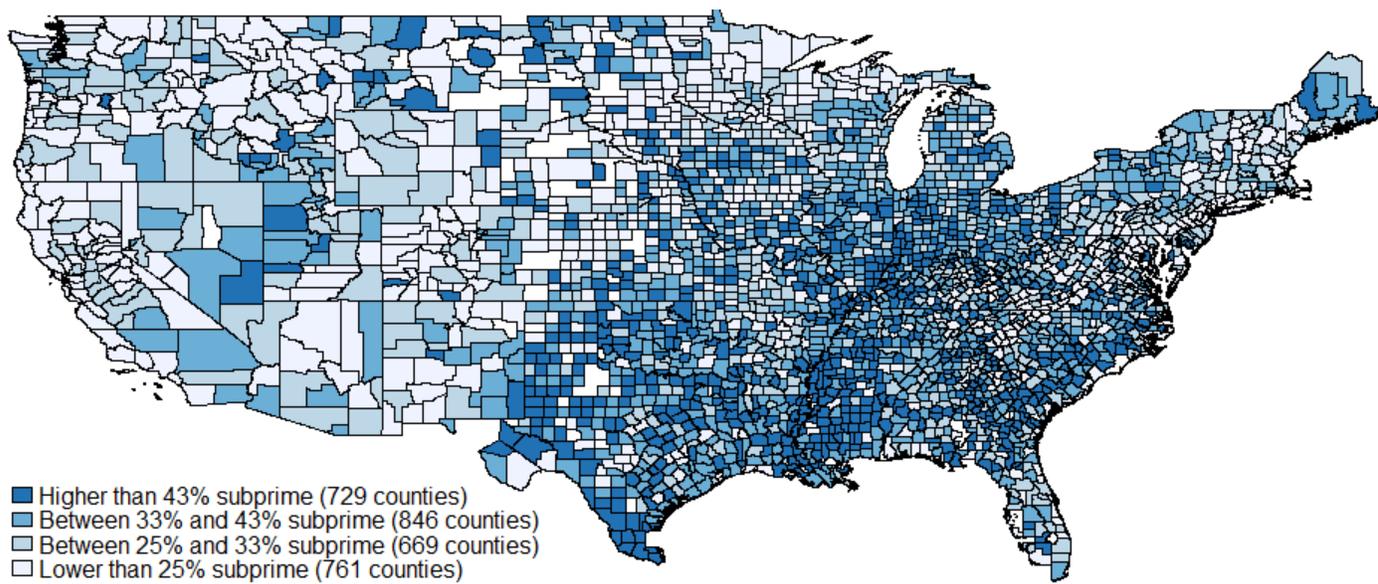
Table B.2 . Growth of U.S. ZIP Code Level House Prices and the Share of Purchase Mortgages to Subprime Borrowers

Dependent Variable:	$\Delta \log(\text{HPI})$ 2002-2006							
	CoreLogic				FHFA			
HPI Source:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\Delta \log(\text{Subprime Share})$	-0.33*** (0.093)	-0.18* (0.095)	-0.21** (0.101)	-0.02*** (0.005)	-0.21*** (0.043)	-0.12*** (0.049)	-0.03*** (0.010)	-0.01 (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 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.

Table C.3 . Growth of U.S. County Level House Prices (CoreLogic) and the Share of Purchase Mortgages to Subprime Borrowers

Dependent Variable:	$\Delta \log(\text{HPI})$ 2002-2006					
	Weighted			Unweighted		
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta \log(\text{Subprime Share})$	-0.519*** (0.124)	-0.248*** (0.086)	-0.185*** (0.035)	-0.259*** (0.052)	-0.211*** (0.032)	-0.094*** (0.023)
Covariates	N	Y	Y	N	Y	Y
State FE	N	N	Y	N	N	Y
Observations	1,323	1,323	1,323	1,323	1,323	1,323
R-squared	0.154	0.524	0.882	0.108	0.409	0.775

Notes: This table reports estimates from a regression of change in the log CoreLogic 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 C.4 . Growth of U.S. County Level House Prices (Zillow) and the Share of Purchase Mortgages to Subprime Borrowers

Dependent Variable:	$\Delta \log(\text{HPI})$ 2002-2006					
	Weighted			Unweighted		
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta \log(\text{Subprime Share})$	-0.488*** (0.166)	-0.144 (0.119)	-0.171*** (0.043)	-0.280*** (0.068)	-0.188*** (0.043)	-0.098*** (0.034)
Covariates	N	Y	Y	N	Y	Y
State FE	N	N	Y	N	N	Y
Observations	930	930	930	930	930	930
R-squared	0.112	0.507	0.891	0.080	0.385	0.750

Notes: This table reports estimates from a regression of change in the log Zillow 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 C.5 . Annual Growth of U.S. County Level House Prices and the Share of Purchase Mortgages to Subprime Borrowers

Dependent Variable	$\Delta \log(\text{HPI})$				
	(1)	(2)	(3)	(4)	(5)
$\Delta \log(\text{Subprime Share})$	-0.048*** (0.012)	-0.0938*** (0.014)	-0.049*** (0.013)	-0.106*** (0.027)	-0.112*** (0.019)
Covariates	N	Y	Y	Y	Y
State FE	N	N	Y	N	N
County FE	N	N	N	Y	Y
Year FE	N	N	N	N	Y
Observations	11,920	11,920	11,920	11,920	11,920
Adjusted R^2	0.01	0.49	0.67	0.76	0.81

Notes: This table reports estimates from a regression of the annual 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-year level and regressions in columns (1)–(5) are weighted by the total number of loans in the county-year. 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 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 C.6 . 1st Stage Estimates of County House Price Growth for Transition into Ownership Regressions

Dependent Variable	$\Delta \log(\text{HPI})$ Past Year	Subprime \times $\Delta \log(\text{HPI})$ Past Year
	(1)	(2)
Subprime	0.000532 (0.00033)	0.016100*** (0.00279)
GMNS Instrument	0.61600*** (0.12800)	-0.17800*** (0.04880)
Subprime \times GMNS Instrument	-0.01010 (0.00713)	1.00400*** (0.05230)
County Controls	Y	Y
Age FE	Y	Y
Year FE	Y	Y
County FE	Y	Y
Division-Time FE	Y	Y
Observations	22,760,616	22,760,616
R-squared	0.69	0.69

Notes: This table reports first stage estimates for column (5) of Table 13 . The GMNS instrument for house price changes is created following the methodology of Guren, McKay, Nakamura, and Steinsson (2018). County controls include the change in log wages and the change in unemployment. An individual's age enters all models through age fixed effects. Robust standard errors are in parentheses and are clustered at the individual and county levels. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$