

# Loan Originations and Defaults in the Mortgage Crisis: The Role of the Middle Class\*

Manuel Adelino, Duke

Antoinette Schoar, MIT and NBER

Felipe Severino, Dartmouth

Current version: June, 2015  
First version: November, 2014

## Abstract

We provide new facts on the debt dynamics leading up to the financial crisis of 2007. Earlier research suggests that distortions in the supply of mortgage credit, evidenced by a decoupling of credit flow from income growth, may have caused the rise in house prices and the subsequent housing market collapse. This paper shows that the increase in mortgage originations was shared across the whole distribution of borrowers, and that middle- and high-income borrowers made up the majority of originations even at the peak of the boom. Compared to prior years, middle- and high-income borrowers (not the poor), as well as those with medium and high credit scores, made up a much larger share of delinquencies in the crisis relative to earlier years. We show that the relation between individual mortgage size and income growth during the housing boom was always strongly positive, also in line with previous periods (and independent of how income is measured). These results are most consistent with an expectations based view of the financial crisis in which both homebuyers and lenders were buying into increasing housing values and defaulted once prices dropped.

Keywords: Mortgage credit, mortgage origination, financial crisis, income

JEL codes: R30, D30, G21

---

\* We thank Nittai Bergman, Markus Brunnermeier, Ing-Haw Cheng, Matthieu Gomez, Raj Iyer, Amir Kermani, Andrew Lo, Sendhil Mullainathan, Debarshi Nanda, Jonathan Parker, Adriano Rampini, David Robinson, Steve Ross, Amit Seru, Andrei Shleifer, Jeremy Stein, Jialan Wang and seminar participants at Berkeley, Cass, Brandeis, CFPB Research Forum, Dartmouth, Duke, Harvard Business School, Maryland, McGill, MIT, NBER Corporate Finance, Princeton, UNC and Imperial College London for thoughtful comments. This paper was previously circulated under the title “Changes in Buyer Composition and the Expansion of Credit during the Boom.”

## 1. Introduction

Understanding the origins of the housing crisis of 2007/08 has been an enduring challenge for financial economists and policy makers alike. One of the predominant narratives that has emerged is that fundamental changes in the origination technology and incentives of the financial sector significantly contributed to unsustainable levels of lending and borrowing. The central assertion is that unprecedented lending to low-income and poor credit quality borrowers caused the acceleration of house prices and the subsequent crash in the housing market. This interpretation builds on a key finding by Mian and Sufi (2009) that growth in mortgage credit for home purchase at the zip code level became negatively correlated with per capita income growth in the run-up to the financial crisis, suggesting that lending became decoupled from income, especially in areas with strong house price growth. As a result, there has been a significant emphasis on understanding the role of the financial industry in providing credit to low-income borrowers in particular.<sup>1</sup>

We provide new analysis on the debt dynamics leading up to the financial crisis, which runs counter to the previous interpretation that large distortions in the supply of credit to low-income borrowers set off a housing bubble and was primarily responsible for the crash. We show that between 2002 and 2006 mortgage origination increased proportionally across the whole income distribution, not just for low-income borrowers. While mortgage credit origination to all income groups expanded significantly between 2002 and 2006, the majority of purchase mortgages by value were originated for middle class and high-income segments of the population even at the peak of the boom, in line with prior years. Similarly, we show that the share of originations to low- versus high-credit-score borrowers remained largely stable across the pre-crisis period. While the pace of origination rose in low-income zip codes, this increase translates into small changes in the overall distribution of credit, given that it starts from a low base (as borrowers in these zip codes obtain fewer and significantly smaller mortgages on average).<sup>2</sup> These results are inconsistent with the hypothesis that large distortions in the supply of credit to poor or marginal borrowers drove the house price increases and the subsequent crash in the mortgage market.

Delinquency patterns also shed doubt on the interpretation that credit to low-income borrowers caused the crisis. Using delinquency information from Lender Processing Service (LPS), we show that the share of mortgage dollars in delinquency stemming from the lowest income groups actually *decreased* during the financial crisis. In contrast, middle- and high-income borrowers constituted a much larger share of mortgage dollars in delinquency than in any prior years. The magnitudes are large: If we look at purchase mortgages originated in 2003 to borrowers in zip codes in the top quintile of the income distribution, we find that they make up only 13% of mortgage dollars in delinquency three years later, even though they represent 33% of the mortgage dollars that were originated in that year. This is what one would expect: borrowers in high-income zip codes usually have better credit and receive larger mortgages than lower-income groups. In 2006, however, this picture is reversed: mortgages originated to borrowers in zip codes in the top income quintile make up 23% of the delinquencies three years out. This is almost a doubling of their share of defaults. In contrast, over the same time period, the contribution of delinquencies from the zip codes in the lowest 20% of the income distribution fell from 22% to only 11%.<sup>3</sup>

---

<sup>1</sup> “The expansion of mortgage credit to *marginal borrowers* kicked off an explosion in household debt in the U.S. between 2000 and 2007.” (Mian and Sufi, 2009, emphasis in original text).

<sup>2</sup> The average purchase mortgage for the borrowers in the lowest income quartile of zip codes is about \$97K as of 2002, and approximately 2 mortgages are originated per 100 residents per year. Borrowers in the top quartile of zip codes obtain average mortgages of over \$246K, and about 3 mortgages per 100 residents.

<sup>3</sup> In this article we look at where leverage and mortgage default occurred, but we are not making welfare statements. While lower income borrowers did not contribute to delinquency at a higher volume than other income groups, it is likely that lower income households (and zip codes) suffered more from defaults than higher income ones. This might be driven by more limited non-housing wealth, worse income shocks, or lack of other funding possibilities.

We find a similar pattern when we look at credit scores: borrowers with high credit scores *increased* their share of defaults on purchase mortgages significantly during the crisis. The fraction of mortgage dollars in delinquency from high-credit-score borrowers (those with a FICO score above 720) that took a loan for home purchase in 2003 was very low (9%), even though these borrowers account for 55% of the credit being originated. However, high-credit-score borrowers made up 23% of delinquencies in the 2006 cohort of purchase loans. The inverse pattern emerges for borrowers below a credit score of 660 (a common cutoff for subprime borrowers, see, e.g., Mian and Sufi 2009), where the share of delinquencies goes from 71% in the 2003 mortgage cohort to 39% for the 2006 cohort. In other words, what is different about the crisis is not that defaults became particularly concentrated in low-income or low-credit-score borrowers but rather that they became less so, as middle- and high-income households (and middle- and high-credit-score borrowers) significantly increased their share in the total mortgages in default.

Earlier research proposes that a decoupling between mortgage growth and income growth constitutes evidence in favor of supply-side distortions, which in turn caused the house price run-up (Mian and Sufi 2009). This evidence relies on a regression of the growth in total dollar value of originations by zip code on the growth in average household income from the Internal Revenue Service (IRS). The growth in zip-code-level originations, however, combines increases at the intensive margin (changes in average mortgage size) with the extensive margin (growth in the number of mortgages originated in a zip code). Given that households take on mortgages, not zip codes, it is only the relation between individual mortgage size and income that is informative about potential supply-side distortions.

We show that the negative correlation of income and purchase mortgage credit is driven entirely by a change in the velocity of mortgage origination (the number of mortgages), not a decoupling of the growth in average mortgage size from income growth. In fact, growth in individual mortgage size is strongly positively related to the growth in IRS household income throughout the pre-crisis period. The apparent decoupling of zip-code-level credit growth and per capita income growth is due solely to the negative relation between the *number* of new originations and per capita income growth.<sup>4</sup> That means that the pace of mortgage origination increased inversely with per capita income growth, but at stable leverage of the average household. This negative correlation is concentrated in high-income zip codes that saw fast per capita income growth and moderate growth in the number of mortgages during this time period. For the bottom 75% of zip codes, the relation between growth in dollar volume of originations and per capita income growth is always positive.

An additional conceptual problem arises when using zip code level aggregates as in Mian and Sufi (2009), since growth in zip code income (in this case household income data from the IRS) confounds the income of new homebuyers with the stock of the average income of the existing residents in an area. Buyers, in fact, have very different income levels from the average household in a zip code—approximately double that of the average resident in an area—over this period and historically. We find that total mortgage credit and average mortgage size track growth in buyer income. In addition, when we look at a longer period, between 1996 and 2007, we confirm that there was neither a reversal of the sign nor a change in the slope between credit flows and income growth using individual borrower income.

---

<sup>4</sup> Our tests also show that the coefficient of aggregate mortgage growth on zip code household income is negative only if we include county fixed effects (as proposed in Mian and Sufi 2009). Without the county fixed effects, the coefficient on average household income is positive even for the regressions using total mortgage credit. If the aim is to test whether credit is increasingly allocated to zip codes with slower income growth, county fixed effects should not be included. The analysis with county fixed effects only tests whether, within a given county, zip codes that are growing more quickly than the county average are disproportionately receiving more credit. However, differences in growth rates in both credit and income between counties are also large (as shown in the summary statistics in Mian and Sufi, 2009), which is lost by the within-county analysis. We show results with and without fixed effects throughout the paper.

One concern in interpreting the evidence with regard to buyer income is that it might be driven by aggressive overstatement of reported income. A number of recent studies have shown that misreporting of borrower characteristics increased during the pre-crisis period (see, for example, Jiang, Nelson, and Vytlačil 2014, or Ambrose, Conklin, and Yoshida 2015 for a careful analysis of this phenomenon). To rule out that this is driving the results, we repeat this analysis for zip codes with more and fewer agency loans (that is, those purchased by one of the government-sponsored enterprises, or GSEs) and show that they are unchanged. Because GSE loans adhere to much stricter underwriting standards, even during the boom period, overstatement is a smaller concern for this sample. The same is true when we break out the data by prime and subprime lenders, which again proxies for the propensity to misreport income across different areas.<sup>5</sup> We are of course not arguing that income misreporting did not happen during the run-up to the crisis, simply that it does not explain the patterns we show in this paper. We provide an even more detailed discussion of why income overstatement does not drive our results in Adelino, Schoar, and Severino (2015).

An additional concern could be that by focusing on mortgage debt for home purchases, we are missing an important part of the distortions in housing leverage, such as cash-out refinancing or home equity lines of credit. We show that the origination of cash-out refinances and second-lien loans are also concentrated in middle-class and upper-middle-class borrowers during this time period. The results on the relation between growth in purchase mortgages and growth in income are also largely unchanged when we consider only refinancing transactions from the Home Mortgage Disclosure Act (HMDA) mortgage data set, as well as data from LPS, which includes cash-out refinances and second liens.

These results provide a new picture of the mortgage expansion prior to 2007. We show that in the aggregate mortgage originations expanded across the entire income distribution and not disproportionately for poorer households, and high and middle income borrowers accounted for a much larger fraction of delinquencies in the crisis relative to before. Similarly, at the individual level, there was no decoupling of mortgage growth from income growth. The results suggest that expectations about house price appreciation played a central role in both borrowers' and lenders' decisions during the credit expansion (including refinances and second-lien mortgages) and subsequent defaults. A number of prior papers show that credit rose significantly more in areas that saw high rates of house price appreciation from 2002 to 2006. Our results confirm these findings but demonstrate that a decoupling of household debt from income for purchase mortgages did not cause this increase in debt. Instead, second liens and cash-out refinancing were particularly common in areas with high house price growth (consistent with Hurst and Stafford 2004, Lehnert 2004, Campbell and Cocco 2007, Bostic, Gabriel, and Painter 2009, Mian and Sufi 2011, and Brown, Stein, and Zafar 2013). In addition, we show that there was a significant increase in the flipping of properties—that is, an increase in the velocity with which properties turned over—in high house price growth areas. As a result of the increased number of transactions, a larger fraction of households held larger mortgages, which had recently been originated. Although both effects increase household leverage, they build on the availability of credit after house prices have appreciated; they are not the *source* of price increases. In line with this expectations-based view of the crisis, we also show that the increase in defaults by middle-class and high-income borrowers (and those for middle- and high-credit-score borrowers) happened predominantly in zip codes in which house prices increased substantially in the pre-crisis period and subsequently dropped (consistent with the role of house prices in driving defaults shown in Foote, Gerardi, and Willen 2008,

---

<sup>5</sup> In addition, the magnitudes of overstatement documented in the literature are too small to explain our results. Jiang, Nelson, and Vytlačil (2014) show that income was overstated by 20% to 25% for low-documentation or no-documentation loans, themselves a small fraction of loans originated in this period (about 30%). Ambrose, Conklin, and Yoshida (2015) estimate an 11% mean overstatement in the sample of borrowers most likely to exaggerate income. However, the difference in buyer income and zip code household income is more than 75% even at the beginning of the boom.

Haughwout, Peach, and Tracy 2008, Mayer, Pence, and Sherlund 2009, Palmer 2014, and Ferreira and Gyourko, 2015).

It is beyond the scope of this paper to analyze what drove house price expectations. Our results are consistent with a “savings glut hypothesis”, whereby low interest rates lead to more lending and borrowing broadly across the population, as proposed for example by Bernanke (2007). As Rajan (2010) argues, the cumulative effect of low interest rates over the decade leading up to the housing boom may have increased house prices through lowering user costs and increased demand for credit (consistent with the discussion also in Himmelberg, Mayer and Sinai 2005). At the same time irrationally exuberant or extrapolative expectations may have played a role in driving up house prices. Among many others, Foote, Gerardi and Willen (2012), Cheng, Raina and Xiong (2014), Shiller (2014), and Glaser and Nathanson (2015) argue that buyers as well as investors in the mortgage market held highly optimistic beliefs about house price growth. Similarly, Chincio and Mayer (2014) document an increased inflow of out-of-state buyers who seem to have been buying for speculative purposes. Coleman, LaCour-Little, and Vandell (2008) argue that subprime lending may have been a joint product, rather than the cause of the increase in house prices.<sup>6</sup> Several papers on the consequences of mortgage securitization focus on the expansion of credit to riskier or more marginal borrowers (Nadauld and Sherlund 2009, Loutskina and Strahan 2009, Keys, Mukherjee, Seru, and Vig 2010, Demyanyk and Van Hemert 2011, Dell’Ariccia, Igan, and Laeven 2012, Agarwal, Amromin, Ben-David, Chomsisengphet, and Evanoff, 2014, and Landvoigt, Piazzesi, and Schneider 2014). Our focus complements this literature, since we analyze both how credit expanded along the whole distribution of borrowers and who contributed most significantly to aggregate defaults.

It is of central importance to get the origins of the credit crisis right: only a proper diagnosis will allow for a meaningful response to prevent future similar events. An analysis that focuses on supply-side distortions in lending to the poor prescribes tight micro-prudential regulation on the debt-to-income (DTI) of bank lending standards, especially when lending to low-income borrowers. This is what many current policy efforts are focused on. In contrast, our results point to a more complex picture, suggesting that financial institutions and borrowers alike bought into the house price appreciation, which might have led to a buildup of systemic risk across the entire housing market. Once prices fell, defaults were not more concentrated among low income or low credit score borrowers, but in fact less so. These dynamics point toward a need for macro-prudential regulation to ensure that there is sufficient slack in the financial system to guard against such systemic shocks. It also points towards a central role of the financial sector: if the build up of systemic risk can have widespread economic impact, macro-prudential regulation ultimately has to trade-off how much to restrict lending upfront to minimize potential losses from the household sector versus how to assign who bears the losses in case of a crisis.

## 2. Data description

The analysis in this paper uses data from three primary sources: the Home Mortgage Disclosure Act (HMDA) mortgage dataset, income data from the Internal Revenue Service (IRS) at the zip code level, and a 5% random sample of all loans in the Lender Processing Services data (LPS, formerly known as McDash). The HMDA dataset contains the universe of US mortgage applications in each year. The variables of interest for our purposes are the loan amount, the applicant’s income, the purpose of the loan (purchase, refinance, or remodel), the action type (granted or denied), the lender identifier, the location of the borrower (state, county, and census tract), and the year of origination. We match census tracts from HMDA to zip codes using the Missouri Census Data Center bridge.

---

<sup>6</sup> Also Glaeser, Gottlieb and Gyourko (2010) argue that “easier” access to credit cannot explain the increase in house prices during the “boom”. On the other hand, Corbae and Quintin (2014), Kermani (2012), and Di Maggio and Kermani (2014) argue that looser credit standards helped feed the boom in housing prices and led to the subsequent bust.

This is a many-to-many match, and we rely on population weights to assign tracts to zip codes.<sup>7</sup> We drop zip codes for which census tracts in HMDA cover less than 80% of a zip code's total population.<sup>8</sup> With this restriction, we arrive at 23,385 individual zip codes in the data.

IRS income data is obtained directly from the IRS, and we use the adjusted gross income of households that filed their taxes in a particular year in that zip code. Besides total income and per capita income, we use the number of tax filings in a zip code to construct an estimate of the population in a zip code in each year.<sup>9</sup> We obtain house price indices from Zillow.<sup>10</sup> The zip-code-level house prices are estimated using the median house price for all homes in a zip code as of June of each year. Zillow house prices are available for only 8,619 zip codes in the HMDA sample for this period, representing approximately 70% of the total mortgage origination volume.

To identify subprime loans, we rely on the subprime and manufactured home lender list constructed by the Department of Housing and Urban Development (HUD) for the years between 1993 and 2005. This list includes lenders that specialize in such loans, and are identified by a combination of features like the average origination rate, the proportion of loans for refinancing, and the share of loans sold to Fannie Mae or Freddie Mac, among others.<sup>11</sup> The data contain lender names, agency codes, and lender identification numbers, and we use these identifiers to match this list to HMDA.

We also use a loan-level data set from LPS that covers approximately 60% of the US mortgage market and contains detailed information on the loan and borrower characteristics for both purchase mortgages and mortgages used to refinance existing debt. This data set is provided by the mortgage servicers, and we use a 5% sample of the data. The LPS data include not only loan characteristics at origination but also the performance of loans after origination, which allows us to look at ex-post delinquency and defaults. One constraint of using the LPS data is that coverage improves over time, so we start the analysis in 2003 when we use this data set.

In two robustness tables of the Online Appendix we also use data from Blackbox Logic and Freddie Mac (a random sample of 50,000 loans per year from the single-family loan data set) on mortgage originations and delinquencies. The Blackbox Logic data include approximately 90% of privately securitized loans in the 2002–2006 period, so they include almost the whole population of subprime loans that were privately securitized. The public Freddie Mac data, on the other hand, include higher-quality loans that were included in Freddie Mac securities and had to conform to that agency's guidelines.

Last, we use household income and debt data from the 2001, 2004, and 2007 waves of the Federal Reserve Board Survey of Consumer Finances. We use these data to construct a debt-to-income measure that includes all mortgage-related debt and to ask where along the income distribution we observe an increase in DTI levels.

---

<sup>7</sup> In other words, zip codes often have more than one census tract associated with them, and census tracts can overlap with more than one zip code. The Missouri Census Data Center bridges of tracts to zip codes using population weights are obtained from <http://mcdc.missouri.edu/websas/geocorr90.shtml> for the 1990 definitions of tracts (used in the HMDA data up to 2002) and <http://mcdc2.missouri.edu/websas/geocorr2k.html> (for 2000 tract definitions, used in HMDA starting in 2003).

<sup>8</sup> This restriction drops 180 zip codes out of 23,565.

<sup>9</sup> IRS zip code information is available at [http://www.irs.gov/uac/SOI-Tax-Stats-Individual-Income-Tax-Statistics-ZIP-Code-Data-\(SOI\)](http://www.irs.gov/uac/SOI-Tax-Stats-Individual-Income-Tax-Statistics-ZIP-Code-Data-(SOI)). Data are available on the website for 1998, 2001, and 2004 onward, and we obtained the 2002 data on a CD from the IRS directly. The zip code population is approximated by multiplying the number of exemptions by a factor of 0.9 (this factor is obtained based on 2008 population estimates constructed by adding the number of returns, the number of returns filing jointly, and the number of dependents).

<sup>10</sup> Zillow house prices are available at <http://www.zillow.com/research/data/>.

<sup>11</sup> The complete list, as well as the detailed criteria for inclusion of lenders in the list, is available at <http://www.huduser.org/portal/datasets/manu.html>. Mayer and Pence (2009) discuss on advantages and disadvantages of using this list to identify subprime loans.

### 3. Descriptive statistics

Table 1 presents the descriptive statistics for the main variables in our sample. The first column reports the average and standard deviation for the full sample, while the next three columns break out the averages for the top quartile, the middle two quartiles, and the bottom quartile of zip codes by household income from the IRS as of 2002. In the last three columns (Columns (5)– (7)), we break out the data by the level of house price growth in a zip code. We report summary statistics for the highest and lowest quartiles and combine the two middle quartiles into one number (“Middle”). The sample is based on the 8,619 zip codes that have nonmissing house price data at the zip code level from Zillow.

The first row shows the zip code average household income based on IRS adjusted gross income data as of 2002. The average household income is \$51k for our sample of zip codes. The average income in the highest quartile is \$84k versus about \$31k for the zip codes in the lowest quartile. Interestingly, when we look at the income reported for the home buyers in HMDA as of 2002 (that is, at the beginning of the boom period), the individuals in those zip codes who actually took out a loan to buy a house (a purchase mortgage) have an average income that is much higher than the average for their zip code, at \$92k. This figure is about \$143k for the zip codes with the highest household income (about 1.7 times the average income) and \$63k for the zip codes in the lowest income group (more than twice the average income in these zip codes). The average original mortgage balance (as of 2002) is also strongly increasing in average zip code income. The average original mortgage balance is \$155k for the pooled sample, but it is \$246k in the highest income quartile and \$97k in the lowest. In addition, the number of mortgages varies across income bins. There are 3.1 mortgages originated yearly per 100 residents for zip codes in the highest income quartile, compared with only 2.1 mortgages originated per 100 residents for those in the lowest quartile.

Panel A of Table 1 also reports purchase mortgage DTI calculated from HMDA as the mortgage amount over the reported income. The DTI for the zip codes in the highest income quartile is 2.26, while the ratio for the lowest quartile is 1.97. As we show in Figures A1 and A2 of the Online Appendix, however, the change in DTI between 2002 and 2006 is small and not significantly different across income groups.

The last three columns of Panel A of Table 1 show that the zip codes that experienced the biggest house price run-ups between 2002 and 2006 had similar household income levels to the other zip codes in the sample. However, buyer income as of 2002 was already higher than the buyer income in any of the other quartiles, and these zip codes already had relatively high average mortgage sizes as of 2002, especially compared to zip codes with small house price increases during this period. Also, in 2002 there were already more mortgages originated per 100 residents in the zip codes that later experienced large house price increases (3.4 compared to 2.4 for the other quartiles). There are no large differences in terms of DTI between zip codes in different quartiles of house price growth.

Panel A also reports the growth rates in mortgages and income. First, we document the (annualized) nominal growth rate of IRS household income between 2002 and 2006. The growth rate of household income is about 4.6% on average, with 6.4% for the highest income zip codes and 3.5% for the lowest income ones. When we consider the annualized growth rate of the income reported for the group of home buyers in HMDA, we see that they are relatively similar across household income quartiles (all around 6% to 7%). There are, however, large differences in both household and buyer income growth depending on whether zip codes experienced large or small increases in house prices during this period. In both cases, zip codes with larger house price run-ups have bigger contemporaneous income increases (although more so for growth in buyer income).

We also show the annualized growth rate in the total mortgage credit originated for home purchases by zip code between 2002 and 2006. This growth rate includes the growth in the average mortgage balance at origination, as well as the increase in the number of mortgages originated in an area. The growth rate is about 8% in the zip codes in the highest income quartile and double this amount (16%) for the lowest quartile. When we focus on the change in the average mortgage size, we see that the growth is about 7.5% in the highest quartile and about 6.9% in the lowest quartile, a relatively uniform shift across income levels. In the next row we see a much larger difference across areas when we consider the increase in the number of mortgages. The areas in the highest income quartile see an annual increase in the number of purchase mortgages originated of only about 1%, whereas the lowest quartile has an increase of almost 10% annually. This suggests that the bulk of the increase in the number of mortgages originated in lower-income areas is driven by a steep increase in the number of transactions. In other words, there was a larger impact on the extensive margin than on the intensive margin.<sup>12</sup> A similar picture emerges from Figures A1 and A2 of the Online Appendix, where zip codes along the whole distribution show small increases in DTI.<sup>13</sup>

When we consider neighborhoods with different levels of house price appreciation, changes in the average size of mortgages account for the bulk of the increase in mortgage credit for zip codes that experience large house price increases (12.4% annual increase in average mortgage size). Average mortgage size increases by much less (2.1%) for zip codes with small house price increases, as we would expect.<sup>14</sup> The growth in the number of purchase mortgages originated is relatively similar across zip codes with high and low house price appreciation.

Panel B of Table 1 shows descriptive statistics for the 5% sample of the LPS dataset. The first line shows the average mortgage balance at origination for the 2003 mortgage cohort (\$188.7k), slightly above the number for the whole HMDA dataset of \$155k for the year of 2002. If we take into account an average growth rate of about 7% over 2002 to 2003, the discrepancy is about \$23k between the two datasets. The average credit score in the data is 711, and average scores are increasing in zip code household income, as one would expect. Average delinquency rate in the 2003 mortgage cohort is 3.7%, with a rate of 1.5% in the top quartile of zip codes by income, and 7% in the bottom quartile. A mortgage is defined as being delinquent if payments become more than 90 days past due (i.e., 90 days, 120 days or more, in foreclosure or REO) at any point during the three years after origination. The delinquency rates are significantly higher for the 2006 cohort, at 18%, and they are again monotonically decreasing in zip code household income. Importantly, the proportional increase in default rates is much more dramatic for the top quartile of zip codes by income than at the bottom, which then implies a larger fraction of overall delinquencies for the zip codes in the high income bucket that we describe below.

## 4 Aggregate Results

### 4.1 Aggregation by income

The first set of questions we address is whether the overall shares of mortgages originated changed significantly between 2002 and 2006 along the income distribution. If, indeed, credit decoupled from income and started flowing disproportionately to poorer households, we would expect to see an increase in the share of credit originated to low-income homebuyers. For this purpose, we use

---

<sup>12</sup> This increase in the number of purchase mortgages originated can be the result of new homeowners moving into these areas (as in Guerrieri, Hartley, and Hurst 2013) or of more transactions by existing residents (home “flipping”).

<sup>13</sup> In Panel B of Figure A1 of the Online Appendix, DTI is calculated by the lender, and it is obtained as the sum of mortgage payments, insurance, and taxes divided by the monthly borrower income. Debt service to income figures show a similar modest rise in Jaffee (2009).

<sup>14</sup> Ferreira and Gyourko (2011) show that income growth has important explanatory power for local housing booms. The relation between business cycles and real estate prices is an issue of some debate (see, e.g., Leamer 2007 and Ghent and Owyang 2010). See Ghysels, Plazzi, Torous, and Valkanov (2012) for a survey on predictability of real estate prices.



individual transaction-level data from HMDA and use income data from both the IRS (average zip-code-level household income) and HMDA (buyer income). We restrict our attention to zip codes with nonmissing Zillow house price data.

In Panel A of Figure 1 we break out the total dollar volume of mortgages originated for home purchase in each year by the quintile that each borrower falls into based on applicant income. We sum the mortgage amounts originated to all the households within an income quintile and divide this number by the total amount of mortgage debt originated in the United States in a given year. As of 2002, the buyer income cutoff for the bottom quintile is \$41k, the second quintile corresponds to \$58k, the third quintile corresponds to \$78k, and the fourth quintile corresponds to \$112k. This picture highlights that credit was flowing predominantly to middle-class and richer borrowers: the proportion of mortgages originated is strongly monotonically increasing in income, and it is very stable over time between 2002 and 2006. The highest income borrowers (those in the top quintile) account for about 36% of the mortgage credit and the next income quintile of borrowers for about 22% of credit, whereas the bottom quintile accounts for about 11%.

The picture using IRS household income to form quintiles (shown in Panel B) shows a similar pattern. For this panel and all other figures using quintiles of IRS income, the cutoffs are as follows: the bottom quintile cutoff corresponds to an average household income in the zip code as of 2002 of \$34k, for Q2 it is \$40k, for Q3 it is \$48k, and for Q4 it is \$61k. We see that the top quintile shows a reduction in its share of total dollar value of originations from 35% of the total in 2002 to 30% in 2006, and this reduction is accounted for by increases of 1 to 2 percentage points in the quintiles below. The bottom quintile accounts for about 10% of mortgage dollars originated in 2002 and rises to 12% in 2006. The conclusion from the picture is that purchase mortgage credit was allocated very similarly at the peak of the boom to what it was earlier across income levels.

In Panel A of Figure A1 of the Online Appendix we calculate the average DTI (calculated as the mortgage balance divided by applicant income) for mortgage holders in each income quintile across time. We compute DTI as the ratio of the mortgage balance at origination divided by the applicant income. Not surprisingly, when we look across income quintiles, we see that poorer households are significantly more leveraged than richer ones even in 2002. The average household in the lowest decile (1) has a first mortgage DTI of about 3, whereas the average mortgage holder in the top income quintile has a DTI of only about 1.5. We see that DTI levels measured in HMDA change little over from 2002 to 2006, and more important, DTI does not become differentially higher for low-income borrowers than for high-income ones. In Panel B of Figure A1 we show the DTI measure typically used in the industry, namely a measure of recurring mortgage payments divided by monthly borrower income. This includes payments on interest, second liens, insurance, and taxes. The increase in DTI using this measure is relatively modest, and, again, borrowers at all income levels move in lockstep. As we discuss in Section 6, this is consistent with the patterns for all mortgage-related debt in the Survey of Consumer Finances.

## 4.2 Delinquency by income groups

In this section we consider how the distribution of mortgage credit compares with that of mortgage delinquency. Much of the prior literature focuses on the fact that delinquency *rates* are higher for lower-quality and lower-income borrowers, but here we show a breakdown of the dollar volume of credit that is past due by income level and cohort of loans. Rather than focusing on the number of mortgages that become delinquent, we focus on the volume of delinquencies broken out by the income level of borrowers. With this approach, we are able to take into account not only the likelihood of default for a given income group, but also its share of overall mortgage credit at origination.

In Figure 2 we start by repeating the analysis from the prior section using a 5% random sample of LPS data. These data allow us to follow borrowers and observe loan performance after origination. We start our analysis in 2003 because the LPS data are not representative before then. In Panel A of Figure 2 we break out loan origination by income quintiles using IRS household income at the zip code level as of 2002 to show that the origination patterns are in line with those obtained with the HMDA data. The cutoffs for each quintile are given in the note to Figure 1. We confirm the results from the previous section: purchase mortgage origination did not become disproportionately skewed toward the lowest income quintiles, and the shares across quintiles are very close to those we obtain using all of the HMDA data in Figure 1.<sup>15</sup>

In Panel B of Figure 2 we use the same decomposition of zip codes to look at the fraction of mortgages in delinquency at some point during the first three years after origination for each cohort of loans. So, for a mortgage originated in 2003, we look at its performance up to 2006 and ask whether it became seriously delinquent (90 days or more past due), was in foreclosure, or was real estate owned (REO) at any point during that period. This measure follows a common definition of default used elsewhere in the literature (see, for example, Demyanyk and Van Hemert 2011). We find a surprising reversal in the pattern of the share of defaults across income groups: in 2003, as one might have expected, there was a negative relation between income quintiles and the fraction of mortgage dollars in delinquency. For example, the top income quintile rises from 12% in 2003 to 18% in 2006, and the second highest income quintile rises from 21% to 24%. In contrast, we see that the lowest income quintiles constitute a smaller share of the dollars in delinquency than before: the lowest quintile drops from 22% to 19% and the second lowest from 23% to 19%.

In Panels C and D of Figure 2 we repeat this analysis but use buyer income from the HMDA data set to sort zip codes into quintiles. Because LPS does not report applicant income, we use the average applicant income at the zip code level from HMDA as of the beginning of the sample and merge it to LPS.<sup>16</sup> In 2003 only a small fraction of purchase mortgage volume (9% of the total) originated in the lowest quintile of zip codes by buyer income, whereas 33% and 25% of mortgages (by dollar volume) were originated in the highest and second highest income quintiles, respectively. As before, we see that this pattern of mortgage originations does not change meaningfully from 2003 to 2006. When we look at the distribution of delinquencies across quintiles we see an even more pronounced increase in the share of mortgage dollars in default for the highest income zip codes. For example, for the 2003 cohort, only 13% of the mortgage values in delinquency come from borrowers in the top income quintile (even though these zip codes make up 33% of originations), 19% come from the second highest income quintile, and 22%–23% come for each of the three lowest income quintiles. This is not surprising, since poorer borrowers have higher default rates in the cross section. However, from 2003 to 2006 the middle and even the highest income quintiles become much more important in default. In fact, for the 2006 purchase mortgage cohort, three years out (that is, delinquencies up to 2009), we find that 23% of the value of delinquencies comes from the top income quintile, 26% from the second highest quintile, and only 18% and 11% from the lowest two income quintiles.

The results show that although the crisis was associated with a large increase in the overall volume of delinquencies, this was due primarily not to a concentration of defaults in low-income zip codes or borrowers but rather to an increase in the share of delinquencies by borrowers in higher income groups, where delinquencies are usually much less common.

### 4.3 Credit Scores

---

<sup>15</sup> This is an important check that the 5% LPS sample is representative of the overall origination data.

<sup>16</sup> Note that by fixing applicant income as of the beginning of the period this analysis cannot be contaminated by concerns of income misreporting during the mortgage boom. Also, below we show all of these patterns using income data from the IRS.

In Figure 3 we analyze origination and delinquency patterns by borrower credit score. We use FICO scores as our measure of credit scores because these are the most widely used measure in the industry and other academic research (see, among many others, Keys, Mukherjee, Seru, and Vig 2009, and Mayer, Pence, and Sherlund 2009). Credit scores give us another dimension to determine whether marginal and low-quality borrowers were primarily responsible for driving up delinquencies in the crisis. We focus on a cutoff of 660 because this is a common FICO cutoff for subprime borrowers used by, specifically, the Federal Reserve Board (FRB), Office of the Comptroller of the Currency (OCC), Federal Deposit Insurance Corporation (FDIC), and Office of the Thrift Supervision (OTS), as well as several researchers, including Mian and Sufi (2009) and Demyanyk and Van Hemert (2011).<sup>17</sup> We also consider a cutoff of 720, which is approximately the median credit score in the LPS sample (the median is 721 in 2003, 716 in 2004, 718 in 2005, and 715 in 2006). Panel A of Figure 3 shows that originations across credit scores remained largely stable. In the 2003 cohort of purchase mortgages, 55% of the mortgage dollars originated from people with credit scores above 720. About 28% of mortgage dollars originated from people with credit scores between 660 and 720, and only 17% of mortgages went to people with credit scores below 660. Interestingly, this pattern stays basically unchanged from 2003 to 2006. In other words, the origination of purchase mortgages also did not grow disproportionately for households with very low credit scores. As we see in the summary statistics, overall origination of purchase mortgages expanded significantly in the pre-crisis period, but it expanded roughly proportionately for households at all credit score levels.

However, when we look at delinquency patterns for the purchase mortgage cohorts between 2003 and 2006, we see a reversal in the share of delinquencies across credit score groups, just as we did when we considered income levels. In 2003 only 9% of purchase mortgage dollars in delinquency originated from borrowers with credit scores above 720, 20% came from borrowers between 720 and 660, and the majority (71%) of mortgage dollars in delinquency in the three years after origination occurred in borrowers with credit scores below 660. However, these fractions reverse for the 2006 cohort. Now 23% of dollars in delinquency come from the borrowers with credit scores above 720, 38% from the middle bin (borrowers between 660 and 720), and 39% from the lowest credit score bin (those below 660). The picture is virtually identical if we restrict the analysis to mortgages in each cohort that reach foreclosure (Panel C). Figure A2 of the Online Appendix shows the absolute value of the purchase mortgages in delinquency in each cohort. As expected, it shows a significant increase in the volume of delinquencies, but borrowers with a FICO score below 660 represent less than half of the total volume for the later cohorts (just like Figure 3, by construction because it is based on the same data).<sup>18</sup>

In Figure 4 we sort zip codes by IRS household income as of 2002, as well as by those that fall above and below the FICO threshold of 660, and ask where the increase in the share of delinquencies for middle- and high-income zip codes comes from. In 2003, borrowers with a credit score above 660 in the highest income zip codes in the country made up only 4% of all delinquencies measured by dollar value. By 2006, they represented 13% of delinquencies, a tripling of their share of the total. Similarly, those in the fourth quartile by household income and a credit score above 660 make up 7% of delinquencies by value in the 2003 cohort, and this increases to 15% for the 2006 cohort. Interestingly, even in low income zip codes, the main feature of the data is a dramatic increase in the share of delinquencies by borrowers with a credit score above 660, and a

---

<sup>17</sup> See <https://www.fdic.gov/news/news/press/2001/pr0901a.html> for a press release describing the guidelines for the FRB, the OCC, the FDIC, and the OTS.

<sup>18</sup> Following our paper, Mian and Sufi (2015) replicate our analysis using credit bureau data and confirm these results: the increase in mortgage debt (purchase mortgages, as well as all other mortgage-related debt) was broadly shared among all borrowers up to the 80<sup>th</sup> percentile in credit scores. They also show a significant reduction in the share of delinquencies coming from low-credit-score borrowers in the crisis relative to the earlier period. Please see Adelino, Schoar and Severino (2015) for a detailed discussion and a comparison between the results in both papers.

simultaneous reduction in the share coming from low credit score borrowers. This tabulation again reinforces the fact that what changed was not that the crisis was concentrated in low income, low credit score individuals, but rather that it was distributed across the whole distribution along both dimensions, with a large increase of delinquencies at the top of both distributions.

One potential concern with the LPS data are that they may underrepresent certain market segments, in particular low-credit-score borrowers, and that this might be responsible for the patterns we observe. We repeat the same exercise shown in Figure 3 using data on securitized mortgages in the subprime, Alt-A, and jumbo private-label market. The data come from Blackbox Logic, and as we point out in Section 2, above, the data set has excellent coverage (about 90%) of this market. We show the splits by the same credit score cutoffs in Figure A3 of the Online Appendix. The figure confirms all the conclusions from Figure 3, naturally with somewhat different shares for each group of borrowers (given that the data come from just a section of the market). The evolution of the shares is essentially the same as what we observe for the market as a whole, with the borrowers in the bottom bin accounting for a much smaller share of delinquencies in the 2006 cohort than they did in the 2003 cohort. In Panels C and D of Figure A3 we show the same splits using data from the publicly available Freddie Mac data set. These data, unlike those of Blackbox Logic, focus on the “prime” segment of the market, because mortgages originated for the Freddie Mac securitized mortgage pools must conform to stricter underwriting standards than those in the private-label market. We again see very stable shares of mortgage origination for the three groups of borrowers and a steep increase in the share of defaults from the borrowers above 660.

Last, Figure A6 shows the evolution of the shares of delinquencies as a function of outstanding mortgages as of the last quarter of each year (instead of by cohort, as all other figures). The message from this figure is the same as before, namely that the proportion of delinquent mortgages coming from low-credit-score borrowers drops, while there is an increase over time of both the 660–720 FICO score borrowers and those above 720.

#### 4.4 House prices

The fact that middle- and high-income and middle- and high-credit-score borrowers are the ones that significantly increase their share of defaults in the crisis period points to a systematic shift in borrower behavior and drivers of default. A key driver of default during this period was the behavior of house prices (Foote, Gerardi, and Willen 2008, Haughwout, Peach, and Tracy 2008, Mayer, Pence, and Sherlund 2009, Palmer 2014, and Ferreira and Gyourko 2015, who emphasize the importance of negative equity in particular). We start by confirming that house prices have an important impact in both originations and delinquency patterns, as discussed in the previous literature. In Figure A4 we analyze origination and delinquency patterns by sorting borrowers into quartiles of zip codes with the highest to lowest house price growth from 2002 to 2006. The picture that emerges supports the idea that areas with high house price growth played a larger role in originations during the boom. Panel A shows that the share of purchase mortgage originations increased from 29% to 35% for the quartile of zip codes with the highest house price growth. In comparison, it dropped proportionally for all the other quartiles; for example, for the zip codes with the lowest house price growth, it decreased from 19% to 16%. In Panel B, consistent with previous work, we find that house prices are strongly related to the importance of zip codes in delinquencies. In particular, we find that the areas with the largest house price increases represent 23% of delinquencies in the 2003 cohort but make up 56% of all delinquencies in 2006. The places with the slowest house price growth in the boom represent a much smaller share of the total delinquencies in the later cohorts relative to the earlier ones (dropping from 34% to just 10% of the total).

We are interested, though, in which borrowers make up this significant increase in the share of delinquencies in the zip codes that experience rapid house price increases in the boom (and drops in prices in the bust). To address this issue, we do a double sort, and break out delinquency patterns by

credit score and the house price growth experienced between 2002 and 2006. We show these splits of delinquent mortgages in Figure 5, first for the 2003 purchase mortgage cohort (Panel A) and then for the 2006 cohort (2006). For the 2003 cohort, the default patterns have the expected shape: 71% of defaults come from low-credit-score borrowers, and they are roughly proportionally distributed across neighborhoods with different house price growth patterns. Another 26% come from the lowest house price quartile, 15% from the second lowest, 13% from the third lowest, and about 17% from the quartile of zip codes with the highest house price growth. However, this pattern fundamentally changes when we examine the post-crisis period. If we look at the cohort of borrowers who took a loan in 2006 and follow them three years out, 62% of defaults come from borrowers above the subprime threshold of 660 (the result from Figure 3), and these defaults are concentrated in the two quartiles of zip codes with the highest house price growth in the previous period: 37% of defaults come from borrowers above 660 in the highest quartile of house price growth, and 14% from those in the second highest.

The result that borrowers above the subprime cutoff began to default in particular after house prices went down is consistent with house price dynamics playing a first-order effect in the crisis.

Figure 6 repeats the same splits as Figure 5, but zip codes are binned according to the proportion of lending by subprime borrowers as defined by the HUD subprime lender list. Something interesting jumps out from this chart—even in the 2003 cohort, areas with more subprime lending make up a large fraction of overall delinquencies. The top quartile of zip codes makes up 40% of all delinquent dollars. Out of these 40%, 29 percentage points come from lending to borrowers with a credit score below 660. In the 2006 cohort, these same zip codes are responsible for 59% of all delinquent mortgage dollars, a result that is consistent with many results shown previously in the literature. In this case, however, borrowers below 660 account for 22 percentage points, and the other 37 percentage points come from borrowers with a credit score above 660. This result emphasizes the importance of distinguishing zip codes and individual borrowers—even in areas where risky lenders were highly active, the growth in delinquencies is dominated by borrowers that are above the 660 cutoff typically associated with subprime.

The fact that low-credit-score borrowers represent a smaller fraction of overall delinquencies suggests that strategic default by borrowers with FICO scores above 660 may have been an important driver of delinquencies during this period. As suggestive evidence that this may have been the case, Figure A5 of the Online Appendix splits the sample into recourse and non-recourse states. The idea is that strategic default is more likely in states where borrowers are not personally liable for their secured (in this case, mortgage-related) debt. Indeed, the figure shows that the share of delinquencies coming from borrowers with credit scores in the 660 to 720 range, as well as that coming from borrowers above 720, is significantly higher and increases by more in non-recourse states. Of course, some of these states also experienced a large boom and bust in house prices (for example, Arizona and California), so we cannot definitively point to strategic default as a major driver of the patterns we uncover.

## **5. Mortgage credit and income**

### **5.1 Intensive and extensive margins**

The discussion in Section 4 emphasizes the stability of purchase mortgage origination along income and credit scores between 2002 and 2006, as well as the role that middle- and high-income and credit-score borrowers played in the rise in defaults once house prices dropped. In this section we revisit evidence presented by Mian and Sufi (2009) on the relation between purchase mortgage credit growth and income growth during this same period. It is worth emphasizing that income growth and income levels are strongly positively correlated during this period, so that the observation that

credit grew more in areas with slow income growth is closely related to the observation that high-income zip codes saw a relative reduction in their overall share in originations. As we saw above, however, the reduction in the share of the top quintile was accompanied by an increase in all quintiles below (that is, it does not just go to the bottom income quintile), and this change in allocations is small in the aggregate (at about 4–5 percentage points in total).

In Panel A of Table 2 we show the results of cross-sectional regressions of purchase mortgage origination growth and income growth between 2002 and 2006. Importantly, we also break down the growth in total mortgage origination into growth in the average loan size at the zip code level (the intensive margin) and growth in the total number of loans given out in those zip codes (the extensive margin). As discussed above, this distinction is crucial to differentiate an increase in individual leverage from higher volume of transactions in the housing market. The starting point is the same regression used in Mian and Sufi (2009):

$$g_{02-06} (Mtg)_i = \alpha_0 + \alpha_1 * g_{02-06} (PerCapitalInc_i) + \eta_{county} + \varepsilon_i$$

In Column (1) we regress the annualized growth in the total dollar value of mortgage credit originated for home purchase at the zip code level from 2002 to 2006 on the growth in average household income from the IRS. The sample includes all zip codes with nonmissing house price data from Zillow, and all growth rates are annualized. In the first column we include county fixed effects to follow Mian and Sufi (2009). However, if the aim is to test whether credit is increasingly allocated to zip codes with declining incomes, county fixed effects should not be included. By focusing on growth rates, the dependent variable and the right-hand side variable already take out zip code specific fixed effects (in levels). The analysis with county fixed effects then tests whether *within a given county*, zip codes that are growing more quickly than the county average are disproportionately receiving more credit. By construction, this test misses changes in credit allocation that happen between counties. In other words counties with homogenous zip codes are heavily underweighted in this regression. Column (2) does not include county fixed effects, since it is likely that cross-county variations in both mortgage growth and income are large and important (in fact, consistent with the descriptive statistics in Mian and Sufi 2009). We show results with and without fixed effects throughout the paper, but we highlight that the sign on the coefficient of interest is often reversed when we do not include the county fixed effects.

In Columns (3)–(6) we decompose the aggregate mortgage growth into growth in the average mortgage size and growth in the number of mortgages generated in a zip code.

Column (1) of Panel A replicates the main result on mortgage credit growth and income growth from Mian and Sufi (2009), namely that zip code aggregate credit growth and income growth were negatively correlated during the 2002–2006 housing boom. We find a negative and significant coefficient of 0.182, which means that the total value of mortgage originations at the zip code level dropped by 0.182 percent for every 1 percent increase in income relative to county averages. Column (2) of Panel B shows that the coefficient of aggregate mortgage growth on zip code income is strongly positive and statistically significant when we do not include county fixed effects. This means that when we use all of the within and across county variation in mortgage growth and income, there is no decoupling of total purchase mortgage growth and income growth.

In Columns (3)–(6), we decompose the growth in the total dollar value of mortgage originations into the growth in the size of the average mortgage and the growth in the total number of mortgages originated in a zip code. The first is a measure of the intensive margin and allows us to evaluate whether leverage increased for the people who bought a home. This is the relevant specification for understanding whether credit decisions became distorted on average during this period. The second measure captures the extensive margin and asks whether more mortgages were originated in a given neighborhood as income grew.

The result in Column (3) of Panel A shows that the relationship between growth in average mortgage size and per capita income is strongly positive. The point estimate is 0.239 with county fixed effects. This means the average mortgage size grows by about 0.27% for every percentage point relative increase in per capita income within a county. So the growth in average mortgages sizes is closely correlated with IRS income growth from 2002 to 2006. Column (4) confirms this result without county fixed effects.

In the fifth column of Panel A we now use as the dependent variable the growth in the number of purchase mortgages originated in a given zip code (extensive margin). We see that the relation between growth in the number of mortgages and IRS income is strongly negative. So, the source of the negative correlation we observe in Column (1) stems from the fact that the pace of mortgage originations (and possibly home buying) increased relatively more in zip codes where per capita income was growing less quickly. This result also holds without county fixed effects (Column (6)), although the point estimate shrinks in magnitude.

### 5.1.1 Panel specification

Panel B of Table 2 implements a similar regression as Panel A, but it makes use of yearly data, rather than the growth rate between 2002 and 2006. This specification allows us to assess whether the slope of the relation between income and mortgage growth changed over time. Specifically, we use the following specification:

$$\text{Ln}(Mtg_{it}) = \alpha_0 + \sum_j \alpha_j [\text{Ln}(\text{ZipInc})_{it} * Y_t] + FE_t + FE_i + \varepsilon_{it}$$

The independent variables are the logarithm of the average IRS income of households in the zip code interacted with a full set of dummies for all years in the sample (denoted  $Y_t$ );  $FE_t$  is a year fixed effect, and  $FE_i$  is a zip code fixed effect. Including zip code fixed effects and interactions of the variables of interest with year dummies allows us to test how changes in the sensitivity of mortgage levels to income levels changed over time within zip codes.

In Columns (1), (3), and (5) we simply include the logarithm of the income variable (and not the full set of interactions with the year dummies) to compare the results in this panel setup to the results in Panel A. The dependent variable in Column (1) is the aggregate mortgage origination in a zip code in a given year. The coefficient on the IRS income is positive and significant, and very similar in magnitude to our prior results. As before, we break out total mortgage origination into the average mortgage size by zip code and year (Column (3)) and the number of mortgages in a given zip code and year (Column (5)). The results confirm that average loan size is strongly positively related to the IRS income of existing buyers in a zip code. And the effects are close to zero and insignificant when the number of loans per zip code and year are used as the dependent variable.

Column (2) shows an interesting pattern about the evolution of the relation between income and mortgage credit. Although the direct effect of mortgage growth and existing IRS income is positive and significant, the interaction terms with the year dummies are negative and significant in all years. This means that the relation between the growth in mortgage origination and the growth in average household income from the IRS became flatter over time. Column (4) shows that this is not through a flattening of the relation between the size of mortgages and income, as there was no flattening in the slope in the relation between the average mortgage size and IRS income. Instead, we see in Column (6) that the number of new mortgages in an area becomes progressively less correlated with household income over the run-up to the crisis. This result points to the velocity of house buying and selling increasing in areas with lower income growth (and by symmetry slowing down in higher

income growth areas). But again is important to remember that the average change in loan size did not decouple from income growth at the same time.

## 5.2 Cross-sectional heterogeneity by zip code income

In this subsection we explore how mortgage and income are related within low-, middle-, and high-income zip codes in the sample. In Table 3 we break out the data into quartiles based on the average IRS household income in a zip code as of 2002. The analysis follows exactly the specifications of Table 2. Panel A of Table 3 includes the regressions with county fixed effects, and Panel B excludes county fixed effects. Columns (1)–(3) of Panel A, Table 3, show that the relation is not the same across the different zip code income quartiles. Only the top quartile by income (Column (1)) shows a negative but not significant coefficient on the measure of average IRS income growth ( $\square$  0.191). For the lower three income quartiles in Columns (2) and (3), we find a positive (but not always significant) relation between mortgage and household income growth (0.218 for the middle group and 0.144 for the lowest quartile).

In Columns (4)–(6) we repeat the analysis for the average mortgage size. We find that the coefficient on IRS household income is strongly positive and significant, and the magnitude of the coefficient is extremely stable across all zip codes (it varies between 0.222 and 0.258). In contrast, the negative correlation of the growth in the number of mortgages and income is prominent only in the highest income zip codes (the coefficient is  $\square$  0.406 and highly statistically significant). For the other three quartiles of zip codes we do not find a correlation between the number of mortgages and zip code income growth. The estimated coefficients are almost zero and not significant.

We repeat these regressions in Panel B without county fixed effects and find that these patterns are consistent and even stronger. The first three columns show that the coefficient from the regressions of growth in total mortgage value is positive across all income quartiles. For the average mortgage size (Columns (4)–(6)), we find a strongly positive coefficient for all income groups, and the coefficient is largest for the zip codes in the lowest income quartile (0.869), and only half the size for the highest income quartile (0.432). Last, when we look at Columns (7)–(9), we find that the coefficient between the growth in the number of mortgages and the growth in IRS income is negative only in the top income quartile.

Taken together, we do not find evidence that poorer zip codes were changing their leverage levels disproportionately relative to income growth. In fact, the relation between credit and borrower income is strongest for lower income zip codes, which runs against the idea that credit flowed disproportionately to poorer and marginal borrowers. The relation between average household income and the number of mortgages originated is negative only for the zip codes with the highest income.

## 5.3 Buyer income vs. household income from the IRS

In Table 4 we analyze the relation between mortgage growth and the growth in applicant income from the HMDA data set. As in Table 2, the sample includes the zip codes that have nonmissing zip code level house prices from Zillow. As we discuss above, the income of home buyers differs substantially from that of the average resident in a zip code, and this has been the case historically. A report by the Census Bureau shows that in 2002–2003, about 7.4% of homeowners moved.<sup>19</sup> Of these 7.4%, about 40% moved across counties (which mechanically means that growth in their income cannot be accounted for in either the old or the new county using IRS data). Given that zip codes are much smaller geographic units than counties, we posit that an even larger proportion of

---

<sup>19</sup> Jason P. Schachter, Geographical mobility: 2002 to 2003, Census Bureau, Current Population Reports, issued March 2004.



movers move across zip codes. Given these facts, we argue that another important measure of borrower fundamentals is the income of the people who actually buy a property in the zip code during a given year, as opposed to the average household in a zip code. We use individual-level transaction data from HMDA to measure the income growth of the individual buyers and aggregate up to the zip code level by taking the average for each zip code. We follow the exact specification as in Table 2 and show results with and without county fixed effects.

Column (1) of Panel A shows that between 2002 and 2006 there is a positive relation between the growth in total credit originated for home purchase in a zip code and the growth in buyer income. The coefficient on the income growth of the individual buyers is positive at 0.369 and highly significant at the 1% level. The point estimate is higher, as well as highly significant, when we do not include county fixed effects (Column (2)). In Column (3) we regress the growth in the average size of mortgages (the intensive margin) on the income growth of borrowers. The coefficient on buyer income growth is positive at 0.282 and statistically very significant. Again, without county fixed effects the result is even stronger. This result confirms that mortgage size grew proportionally with income throughout the pre-crisis period and was not decoupled from income growth.

In Column (5) we use the annualized growth rate of the number of mortgages originated in a given zip code as the dependent variable (the extensive margin). We find a positive relation between the growth in the number of mortgages and the growth in buyer income (the estimated coefficient is 0.117 and significant at the 1% level). The coefficient is zero without county fixed effects and not statistically significant.

### 5.3.1 Robustness to income misreporting

One important consideration in the run-up to the financial crisis is that lenders and borrowers began to misreport income levels in order to justify higher leverage than these borrowers could normally afford. It is important for our study to rule out that changes in the reporting of income itself could be the source of the strong relation between buyer income and total mortgage growth we find in all specifications.<sup>20</sup> This is not a test of whether there was significant income misreporting (which has been well established in earlier literature, for example, Jiang, Nelson, and Vytlačil 2014, and Ambrose, Conklin, and Yoshida 2015) but rather a test of whether this misreporting is responsible for the findings in Panel A of Table 4.

We use several approaches to analyze whether this is a first-order concern for our findings. First, in Panel B of Table 4 we break out our sample into different quartiles based on the fraction of mortgages originated and sold to Fannie Mae and Freddie Mac (the government-sponsored enterprises, or GSEs) in the zip code, as well as the fraction of loans that were originated by subprime lenders based on the subprime lender list constructed by the US Department of Housing and Urban Development (HUD, see Section 2 for details). Loans that were sold to (and then guaranteed by) GSEs had to conform to higher origination standards than those sold to other entities and thus were less likely to have unverified applicant income.<sup>21</sup> The idea in these tests is to see whether zip codes with a lower fraction of loans sold to GSEs exhibit a stronger relation between mortgage growth and buyer income. Similarly, loans originated by subprime lenders were much more likely to have low or no documentation status, and, if the correlations shown above were driven by misreporting, we would expect the splits based on this fraction to generate meaningful variation in the estimated coefficients.

---

<sup>20</sup> There is also evidence of other forms of misreporting during this time, including the value of transactions (Ben-David 2011) and mortgage quality in contractual disclosures in the secondary market (Piskorski, Seru, and Witkin 2013, and Griffin and Maturana 2014).

<sup>21</sup> Previous work, including Pinto (2010), has noted that origination standards for GSEs dropped between 2002 and 2006, but we find similar results when we split the sample directly by the fraction of loans originated by subprime lenders.

For both measures of quality of origination we do not find that coefficients on buyer income vary significantly. In fact, the coefficient on buyer income growth is very similar in magnitude and significance levels across all quartiles of both the GSE origination fraction (Columns (1)–(3) of Panel B) and the fraction originated by subprime lenders (Columns (4)–(6)).

In a next step we repeat our regressions of credit growth on buyer income growth for different periods (Panel C). We consider 4 subperiods: 1996–1998, 1998–2002, 2002–2006, and 2007–2011. The coefficient from the regression of growth in total mortgage origination on buyer income growth is positive and significant for all periods, and importantly, it does not become flatter in the pre-crisis years. The coefficient rises from 0.260 in 1998–2001 to 0.368 in 2002–2006, which means that, if anything, the sensitivity of mortgage growth to income growth increased before the crisis. We also look at the relation between average mortgage size growth and income growth in Panel C (Columns (4)–(6)) and again find a strongly positive and largely stable coefficient throughout all periods.

Taken together, the evidence in Panels B and C suggests that the boom period does not represent a “special” period in how mortgage credit growth tracked buyer income growth, nor is there evidence that the income misreporting that other work has found during this period contaminates the findings with regard to the basic relation we uncover.

#### 5.4 Individual level mortgage origination regressions

The zip-code-level regressions show that the negative correlation between zip code income growth and mortgage growth between 2002 and 2006 is driven by the extensive margin, that is, the relative increase in the number of buyers in places where income grew less (relative to the county average). This regression also shows that there was no decoupling between average mortgage amount and income—the intensive margin. In this section we consider the intensive margin using individual transactions, which allows for even finer geographic controls than before. To this end, in Table 5 we use the following specification:

$$\text{Ln}(Mtg_{it}) = \alpha_0 + \alpha_1 \text{Ln}(Census\ Tract\ Inc)_{it} + FE_t + FE_{census\ tract} + \varepsilon_{it}$$

where  $i$  indicates an individual borrower.  $FE_t$  is a year fixed effect, and  $FE_{census\ tract}$  is a census tract fixed effect, the finest geographic breakdown available in the HMDA data set. The independent variable of interest is the logarithm of the average income of households in that tract. Because we do not have data on the average household income by tract, we use the same zip-code-to-tract population-weighted bridge as before (from the University of Missouri Census Data Center) to impute average tract income based on zip code household income. Including census tract fixed effects allows us to test how the sensitivity of mortgage levels to income levels changed within census tracts over time.

Panel A of Table 5 shows the results for this loan-level specification. Consistent with the zip-code-level regressions, the coefficients on census tract income are positive and significant, and the result is unchanged when we replace county fixed effects with census tract fixed effects (Column (3)). Columns (2) and (4) show that the sensitivity of mortgage size to buyer income increases over time during our sample period (2002–2006), whereas the sensitivity to average household income either decreases or does not change. Overall these results reaffirm the conclusion of the zip-code-level analysis and suggest that credit supply did not decouple from income during the boom period.

#### 5.5 Refinancing mortgages and income

All of the regressions in this section focus on purchase mortgage credit, and it is possible that, although purchase mortgages grew in line with income, refinancing credit flowed more strongly to

areas with declining income. In this subsection we analyze the relation between refinancing mortgage credit and income. In Panel A of Table 6 we repeat the regressions in Panel A of Table 2, but we focus on the growth in refinancing transactions in the HMDA data set. In this panel we consider all refinancing transactions, as we are not able to distinguish between cash-out and rate refinancing transactions using the HMDA data. The picture that emerges is similar to what we find for purchase mortgages. In Column (1) we consider the growth in aggregate refinancing credit and find a negative and significant coefficient on the change in zip-code-level income ( $\square 0.579$ ). The coefficient is smaller in magnitude, but still negative, without county fixed effects. In Columns (3)–(6) we decompose the aggregate effect into the average mortgage size and the number of mortgages as before. The estimated coefficient on IRS income growth is zero with county fixed effects, but again positive and significant without fixed effects. The results for the number of transactions are very similar to those we obtain for purchase mortgages: we find negative coefficients on IRS income growth and positive but smaller coefficients on the borrower income growth.

Panel B of Table 6 implements a zip-code-level panel regression similar to the one in Panel B of Table 2. The coefficients show that the relation between refinancing mortgage growth and IRS income growth becomes progressively flatter over time. As in Table 2, almost all of the variation in the slope of these relations comes from the extensive margin (the number of transactions) rather than from changes in the average mortgage balance.

Because the HMDA data set does not distinguish between rate refinancing transactions and cash-out refinancing, we turn to the LPS data and run loan-level regressions in Panel B of Table 5.<sup>22</sup> The first message from Panel B of Table 5 is that the results using LPS data are very similar to those using HMDA shown in Panel A for purchase mortgages. The second message is that the results are slightly smaller in magnitude, but also positive and highly statistically significant for cash-out refinancing transactions, confirming that there is no visible decoupling between individual mortgage size for refinancing transactions and income.

## 6. Increase in household leverage

The results so far (with the exception of Section 5.5) have focused largely on purchase mortgages at origination, which specifically means first mortgages originated when a house is purchased. We show above that there is no decoupling of credit flows from income growth at origination (Section 5) and that the distribution of credit flows across the income distribution stays very stable over the boom (Section 4). This calls into question an interpretation that emphasizes the role of distortions in the supply of credit in *causing* the house price run-up and crash. However, we know that household leverage increased in the aggregate in the run-up to the financial crisis (see, among many examples, the quarterly Federal Reserve Bank of New York’s Household Debt and Credit Report).<sup>23</sup> There are two predominant margins by which credit could have increased during this time period, namely (1) through a quicker pace of house purchase transactions, which causes mortgage debt to reach a higher level, since at the time of home purchase households often lever up to their maximum debt capacity (and then pay down over time); or (2) through equity extraction after origination through second liens or the refinancing of mortgage debt. Cash-out refinancing loans in particular allow homeowners to take out equity that has accrued over time. The second channel is the consequence of house price growth, not its cause, but it is important for us to investigate whether these margins increased particularly for low-income or marginal borrowers or follow similar patterns to what we observe for purchase mortgages.

Below we analyze these different channels through which household leverage might have risen and

---

<sup>22</sup> We have access to only a 5% sample of the LPS data, which makes the data unsuitable for zip-code-level analysis.

<sup>23</sup> The quarterly Federal Reserve Bank of New York’s Household Debt and Credit Report provides data on household trends in borrowing and indebtedness and is available at <http://www.newyorkfed.org/microeconomics/hhdc.html>.

then show the combined impact of these different channels using data from the Survey of Consumer Finances.

## 6.1 Mortgage-related debt: Cash-out and rate refinances, second liens

In Figure 7 we use LPS data in 2006 to report aggregate splits of the loan values originated for borrowers across income quintiles, but we compare different loan products, namely purchase mortgages, cash-out refinance loans, rate refinance loans and second liens. We focus on 2006 to ensure good coverage of all products in LPS, and we split zip codes by average household income from the IRS as of 2002. We also report the total mortgage debt originated in 2006 that includes unclassified mortgages in LPS. The purpose of this figure is to ask whether these loan products show very different distributions from what we found for purchase mortgages. Panel A shows that the dollar distribution of all three product types are generally similar when we break zip codes out by quintiles based on borrower income. The top quintile shows 3 and 8 percentage points lower weight, respectively, for cash-out refinances and second liens relative to purchase mortgages, and this can be accounted for by 1 to 3 percentage points higher weights in the quintiles below. The distribution of rate-refinancing mortgages is very close to that of purchase mortgages. Given that the majority of mortgages originated are purchase mortgages or rate-refinancing mortgages (total origination is shown above the bars for each product in the figure), the total distribution of debt is very close to that of purchase mortgages we have focused on for all other figures.

When we consider the distribution of delinquent mortgages (measured in dollars) for all mortgage-related products (Panel B of Figure 7), the distribution again looks similar for all products, with a higher weight at the top quintiles for purchase mortgages and rate-refinancing mortgages, and a somewhat lower share in particular for the highest income quintile for second liens. The picture for the total amount of debt in delinquency for the 2006 cohort again mirrors that of purchase mortgages. Figure A7 of the Online Appendix also shows a similar message using credit score cutoffs rather than IRS household income. This suggests that the conclusion of the previous tables that high- and middle-income borrowers accounted for a large proportion of origination and delinquency also applies to mortgage debt other than purchase mortgages.

## 6.2 Velocity of transactions

The second channel by which leverage can increase even if DTI at origination is not dramatically distorted is if the velocity of home buying and selling increases in the boom. Typically, homeowners are at their highest DTI at the moment of purchase, and then, over time, their DTI goes down as they repay their mortgage.<sup>24</sup> However, if the velocity of home buying (and selling) increases, a larger fraction of homeowners will be at their maximum DTI.

To offer suggestive evidence of the role of the increased pace of transactions, and of house “flipping” in particular, we use data from Zillow on the fraction of transactions in a zip code that also transacted in the preceding twelve months (shown in Figure 8). In the time period between 1998 and 2006 the fraction of properties that sold twice within a year increased steeply, from about 4% annually in 1998 to almost 8% in 2006. We also see that this trend is prevalent across all zip codes, but it is at a higher pace in zip codes that saw higher house price increases during that time.<sup>25</sup> This suggests that an increase in the velocity of home purchases contributed to the aggregate increase in household leverage. Of course, the magnitude of the contribution to leverage cannot be inferred from this turnover statistic, since in these data we do not know the leverage position of the

---

<sup>24</sup> Notable exceptions to this rule are negatively amortizing mortgages.

<sup>25</sup> We also replicate this result using credit registry data from Dataquik, where we consider the fraction of properties that were sold within recent periods (last two, last three, and last four years).

people buying the homes, nor are we able to track all homes in a region (only those transacted one year earlier). For that purpose we consider the aggregate increase in leverage at the household level in Section 6.3.

### 6.3 Aggregate leverage

Finally, we use data from the Survey of Consumer Finances (SCF) to look at the total mortgage leverage at the household level.<sup>26</sup> The SCF is a household survey that asks consumers for detailed information about their finances and savings behavior and is conducted every 3 years as a repeated cross section. We focus on the survey waves in 2001, 2004 and 2007. In Figure 9 we report the average DTI for households with non-zero mortgage debt sorted by income quintiles (Panel A), as well as the median (Panel B). The measure of debt includes all mortgage-related debt including home equity loans, and we divide the total debt by household annual income to obtain the DTI.<sup>27</sup> Using SCF data we confirm that credit expanded across the income distribution and it again looks like it expanded more for middle class borrowers. Looking at the 2001 wave, we find a similar pattern to what we found with purchase mortgages: lower income groups have higher DTIs than high income groups. For example, households in the lowest income quintile and positive debt have a DTI of about 2.7 in 2001 while the top 20% have a DTI of around 1.

The question of interest, again, is whether debt ratios decoupled from income over the boom period before 2007, and went up disproportionately for the lower part of the distribution. When we look at the same distribution of households by income quintiles in 2007 and compare to 2001, we find that all income quartiles experienced on average a 50% increase in DTI.<sup>28</sup> This figure provides a more complete view of household leverage across the income distribution, since it includes the net stock of mortgage debt by household, and the picture again confirms that debt to income in the country overall did not become disproportionately skewed toward poor or marginal households. A similar picture emerges when we consider the medians in each income bin.

## 7. Conclusion

This paper shows that credit increased proportionally across all income levels, so that the distribution of mortgage credit across the income distribution was largely stable leading up to the housing crisis. This means that even in 2006 high- and middle-income borrowers accounted for the majority of credit originated in the mortgage market. At the same time, there was no decoupling of mortgage credit growth and income growth at the individual level during the period before the financial crisis. Instead, analysis at the individual level demonstrates that mortgage and income moved in the same direction. Similarly, we see that high- and middle-income borrowers, as well as borrowers with a credit score above 660, accounted for a much larger fraction of dollars of credit in delinquency once the crisis hit relative to earlier periods, especially in areas where house prices dropped more. Because these borrowers held much larger mortgages, a small increase in their default rate has a large impact on the amount of dollars in delinquency. These results paint a picture of the run-up to the crisis in which lenders and borrowers bought into high house-price expectations and ignored the potential equilibrium effects that would result from a large fraction of borrowers being close to their maximum leverage capacity.

Obtaining a clear view of the patterns of origination and delinquency during the house price boom and the crisis is crucial because it affects the policy recommendations that are proposed and implemented. A supply-side view in which the price run-up and subsequent defaults were driven by

---

<sup>26</sup> We thank Matthieu Gomez for the suggestion to replicate our DTI analysis using SCF data.

<sup>27</sup> Mortgage related debt includes SCF items *MRTHEL* (Mortgage and Home Equity Loan, Primary Residence) and *RESDBT* (Other residential debt)

<sup>28</sup> And again it is important to note that this increase in overall household DTI was fueled by a re-leveraging cycle via increased velocity of house buying and cash-out refinancing.

unsustainable lending to low-income borrowers argues for a policy response of tight micro-prudential regulation on bank lending standards, especially when lending to low-income borrowers. Following on this, some scholars argue that the response to the crisis should have focused more aggressively on principal debt forgiveness, since it would have funneled dollars only to those marginal households with a high marginal propensity to consume. For example, Mian and Sufi call the lack of a widespread principal reduction program “the biggest policy mistake of the Great Recession” (2014, p. 141). Of course, given the costs involved in principal forgiveness, this solution would have been viable only if a small fraction of homeowners, in particular the poor, were responsible for the problem in the housing market. However, given that credit went up across the income distribution and that defaults went up significantly at the middle and high end of the income distribution (and for high FICO borrowers), such a solution becomes very hard to implement. The dollar amounts needed would have been unrealistically high (as discussed in Eberly and Krishnamurthy 2014, who compare the costs of such programs with those aimed at providing liquidity to households, and the effects on consumption of both types of approaches), and the ensuing moral hazard problems might have plagued mortgage markets for a long time. These dynamics point, instead, toward a need for macro-prudential regulation that ensures that systemic risk does not build up in the economy because lenders and borrowers are chasing asset price increases.

## REFERENCES

- Adelino, M., K. Gerardi, and P. S. Willen. 2013. Why don't lenders renegotiate more home mortgages? Redefaults, self-cures and securitization. *Journal of Monetary Economics* 60(7), 835–853.
- Adelino, Manuel, Antoinette Schoar, and Felipe Severino. 2015. Loan Originations and Defaults in the Mortgage Crisis: Further Evidence. Working Paper.
- Agarwal, Sumit, Gene Amromin, Itzhak Ben-David, Souphala Chomsisengphet, and Douglas D. Evanoff. 2014. Predatory lending and the subprime crisis. *Journal of Financial Economics* 113(1), 29–52.
- Ambrose, Brent W., James Conklin, and Jiro Yoshida. 2015. Reputation and exaggeration: Adverse selection and private information in the mortgage market. Working paper.
- Ben-David, Itzhak. 2011. Financial constraints and inflated home prices during the real-estate boom. *American Economic Journal: Applied Economics* 3(3), 55–78.
- Bernanke, Ben S. 2007. Global Imbalances: Recent Developments and Prospects. Bundesbank Lecture. Berlin, Germany.
- Bostic, Raphael, Stuart Gabriel, and Gary Painter. 2009. Housing wealth, financial wealth, and consumption: New evidence from micro data. *Regional Science and Urban Economics* 39.1: 79-89.
- Brown, Meta, Sarah Stein, and Basit Zafar. 2015. The impact of housing markets on consumer debt: credit report evidence from 1999 to 2012. *Journal of Money, Credit and Banking* 47.S1: 175-213.
- Campbell, John Y., and Joao F. Cocco. 2007. How do house prices affect consumption? Evidence from micro data. *Journal of Monetary Economics* 54.3: 591-621.
- Cheng, I., Shail Raina, and Wei Xiong. 2014. Wall Street and the housing bubble. *American Economic Review* 104(9).
- Chinco, A., and C. Mayer. 2014. Misinformed speculators and mispricing in the housing market.
- Coleman IV, Major, Michael LaCour-Little, and Kerry D. Vandell. 2008. Subprime lending and the housing bubble: Tail wags dog? *Journal of Housing Economics* 17(4), 272–290.
- Corbae, Dean, and Erwan Quintin. 2014. Leverage and the foreclosure crisis. *Journal of Political Economy*, forthcoming.
- Dell’Ariccia, G., D. Igan, and L. Laeven. 2012. Credit booms and lending standards: Evidence from the subprime mortgage market. *Journal of Money, Credit and Banking* 44(2–3), 367–384.
- Demyanyk, Y., and O. Van Hemert. 2011. Understanding the subprime mortgage crisis. *Review of Financial Studies* 24(6), 1848–1880.
- Di Maggio, M., and Amir Kermani. 2014. Credit-induced boom and burst. Working paper.
- Eberly, Janice and Arvind Krishnamurthy. 2014. Efficient credit policies in a housing debt crisis. *Brookings Papers on Economic Activity*, Fall.

- Ferreira, Fernando, and Joseph Gyourko. 2011. Anatomy of the beginning of the housing boom: US neighborhoods and metropolitan areas, 1993–2009. National Bureau of Economic Research working paper No. 17374.
- Ferreira, Fernando, and Joseph Gyourko. 2015. A New Look at the U.S. Foreclosure Crisis: Panel Data Evidence of Prime and Subprime Borrowers from 1997 to 2012. National Bureau of Economic Research working paper No. 21261.
- Foote, Christopher L., Kristopher S. Gerardi, and Paul S. Willen. 2008. Negative equity and foreclosure: Theory and evidence. *Journal of Urban Economics* 64(2), 234–245.
- Foote, Christopher L., Kristopher S. Gerardi, and Paul S. Willen. 2012. Why did so many people make so many ex post bad decisions: The causes of the foreclosure crisis. FRB Boston Public Policy Discussion Paper Series, paper no. 12-2.
- Ghent, Andra C., and Michael T. Owyang. 2010. Is housing the business cycle? Evidence from US cities. *Journal of Urban Economics* 67(3), 336–351.
- Ghysels, Eric, Alberto Plazzi, Walter N. Torous, and Rossen Valkanov. 2012. Forecasting real estate prices. *Handbook of Economic Forecasting* 2.
- Glaeser, Edward L., Joshua D. Gottlieb, and Joseph Gyourko. 2013. Can cheap credit explain the housing boom? *Housing and the Financial Crisis* (2013), 301.
- Glaser, Edward, and Charles G. Nathanson. 2015. An extrapolative model of house price dynamics. National Bureau of Economic Research working paper No. 21037.
- Griffin, John, and Gonzalo Maturana. 2014. Who facilitated misreporting in securitized loans? *Journal of Finance*, forthcoming.
- Guerrieri, Veronica, Daniel Hartley, and Erik Hurst. 2013. Endogenous gentrification and housing price dynamics. *Journal of Public Economics* 100 (2013), 45–60.
- Haughwout, Andrew, Richard Peach, and Joseph Tracy. 2008. Juvenile delinquent mortgages: Bad credit or bad economy? *Journal of Urban Economics* 64(2), 246–257.
- Himmelberg C, Chris Mayer, and Todd Sinai. 2005. Assessing high house prices: bubbles, fundamentals, and misperceptions. *Journal of Economic Perspectives* 19(4): 67–92.
- Hurst, Erik, and Stafford, Frank. 2004. Home is where the equity is: Mortgage refinancing and household consumption. *Journal of Money, Credit and Banking*: 985-1014.
- Jaffee, Dwight M. 2009. The US subprime mortgage crisis: Issues raised and lessons learned. *Urbanization and Growth* (2009), 197.
- Jiang, W., A. A. Nelson, and E. Vytlačil. 2014. Liar's loan? Effects of origination channel and information falsification on mortgage delinquency. *Review of Economics and Statistics*, forthcoming.
- Kermani, A. 2012. Cheap credit, collateral and the boom-bust cycle.
- Keys, B. J., T. Mukherjee, A. Seru, and V. Vig. 2010. Did securitization lead to lax screening? Evidence from subprime loans. *Quarterly Journal of Economics*, 125(1), 307–362.

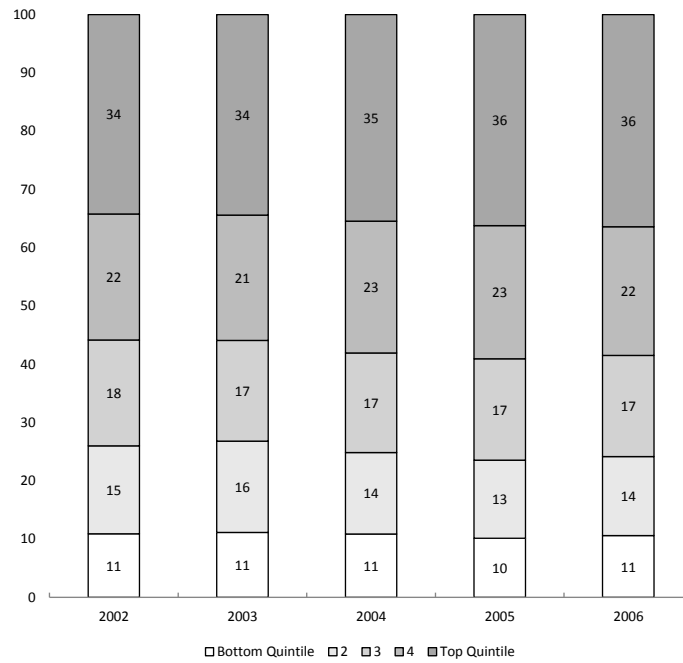


- Landvoigt, Tim, Monika Piazzesi, and Martin Schneider. 2014. The housing market (s) of San Diego. *American Economic Review*, forthcoming.
- Leamer, Edward E. 2007. Housing is the business cycle. *Proceedings - Economic Policy Symposium - Jackson Hole*, Federal Reserve Bank of Kansas City, 149–233.
- Lehnert, Andreas. 2004. Housing, consumption, and credit constraints. Working Paper.
- Loutskina, E., and P. E. Strahan. 2009. Securitization and the declining impact of bank finance on loan supply: Evidence from mortgage originations. *Journal of Finance* 64(2), 861–889.
- Mayer, Chris, and Karen Pence. 2009. Subprime mortgages: What, where, and to whom? In Edward L. Glaeser and John M. Quigley, eds., *Housing Markets and the Economy: Risk, Regulation, and Policy*. Cambridge, MA: Lincoln Institute of Land Policy.
- Mayer, Christopher, Karen Pence, and Shane M. Sherlund. 2009. The rise in mortgage defaults. *Journal of Economic Perspectives* 23(1), 27–50.
- Mian, Atif, and Amir Sufi. 2009. The consequences of mortgage credit expansion: Evidence from the US mortgage default crisis. *Quarterly Journal of Economics* 124(4), 1449–1496.
- Mian, Atif, and Amir Sufi. 2011. House prices, home equity–based borrowing, and the US household leverage crisis. *American Economic Review* 101(August), 2132–2156.
- Mian, Atif, and Amir Sufi. 2014. *House of Debt*. Chicago: University of Chicago Press.
- Mian, Atif, and Amir Sufi. 2015. Household debt and defaults from 2000 to 2010: Facts from Credit Bureau Data. Working paper.
- Nadauld, T. D., and S. M. Sherlund. 2009. The role of the securitization process in the expansion of subprime credit. Divisions of Research & Statistics and Monetary Affairs, Federal Reserve Board.
- Palmer, C. 2014. Why did so many subprime borrowers default during the crisis: Loose credit or plummeting prices?
- Pinto, Edward J. 2010. Government housing policies in the lead-up to the financial crisis: A forensic study. Discussion draft, American Enterprise Institute, Washington, DC.
- Piskorski, Tomasz, Amit Seru, and James Witkin. 2013. Asset quality misrepresentation by financial intermediaries: Evidence from RMBS market. National Bureau of Economic Research working paper w18843.
- Rajan, Raghuram. 2010. *Fault Lines: How Hidden Fractures Still Threaten the World Economy*. Princeton University Press.
- Shiller, Robert J. 2014. Speculative asset prices. *American Economic Review* 104(6), 1486–1517.

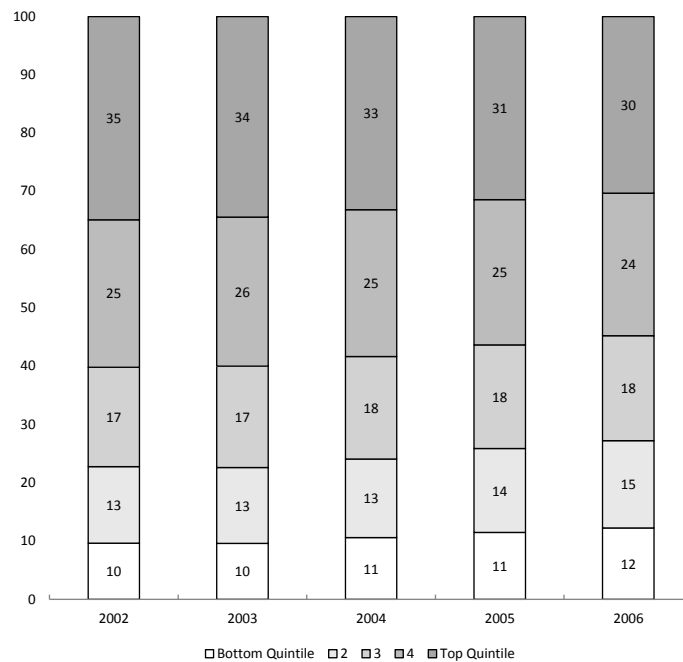
## Figure 1: Mortgage origination by income

This figure shows the fraction of total dollar volume of purchase mortgages in the HMDA dataset originated by income quintile. In Panel A we form quintiles based on the income of each individual buyer (as of 2002 the buyer income cutoff for the bottom quintile is \$41k, the second quintile corresponds to \$58k, the third quintile corresponds to \$78k, and the fourth quintile corresponds to \$112k.). In Panel B we use household income from the IRS as of 2002 (i.e., the zip codes in each bin are fixed over time). The cutoff for the bottom quintile corresponds to an average household income in the zip code as of 2002 of \$34k, for Q2 it is \$40k, for Q3 it is \$48k, and for Q4 it is \$61k. Sample includes zip codes with nonmissing house price data from Zillow.

### Panel A. Income quintiles based on buyer income from HMDA



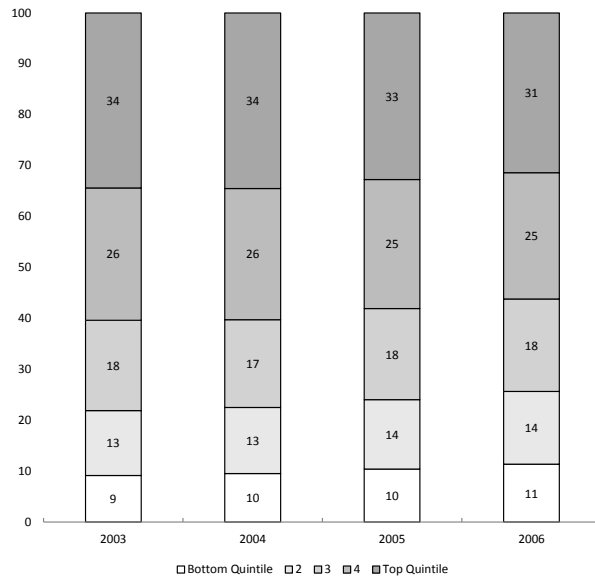
### Panel B. Income quintiles based on IRS household income as of 2002



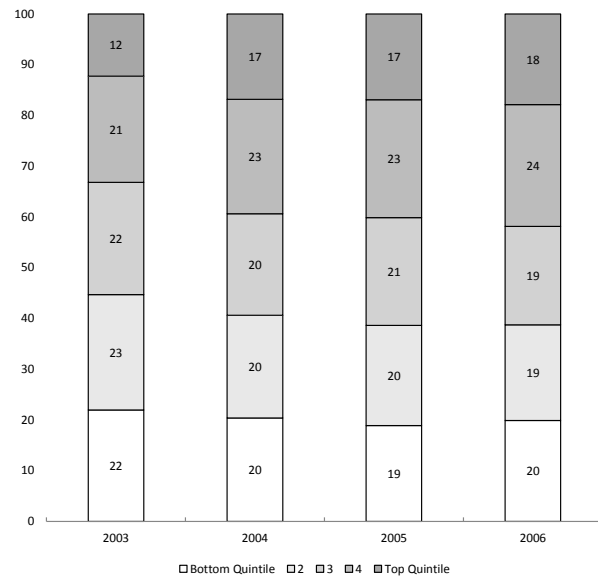
## Figure 2: Mortgage origination and delinquency by income

This figure shows the fraction of total dollar volume of purchase mortgages originated, as well as the total dollar volume of delinquent mortgages by cohort, split by income quintile. A mortgage is defined as being delinquent if payments become more than 90 days past due (i.e., 90 days, 120 days or more, in foreclosure or REO) at any point during the three years after origination. Data are from the 5% sample of the LPS dataset and the sample includes zip codes with nonmissing Zillow house price data. In Panels A and B we form quintiles based on household income from the IRS as of 2002 (i.e., in all panels zip codes are fixed as of 2002, cutoffs are the same as those given in Figure 1), and in Panels C and D we form quintiles based on average buyer income from HMDA in the zip code as of 2002 (as of 2002 the zip code average buyer income cutoff for the bottom quintile is \$59k, the second quintile corresponds to \$69k, the third quintile corresponds to \$83k, and the fourth quintile corresponds to \$109k).

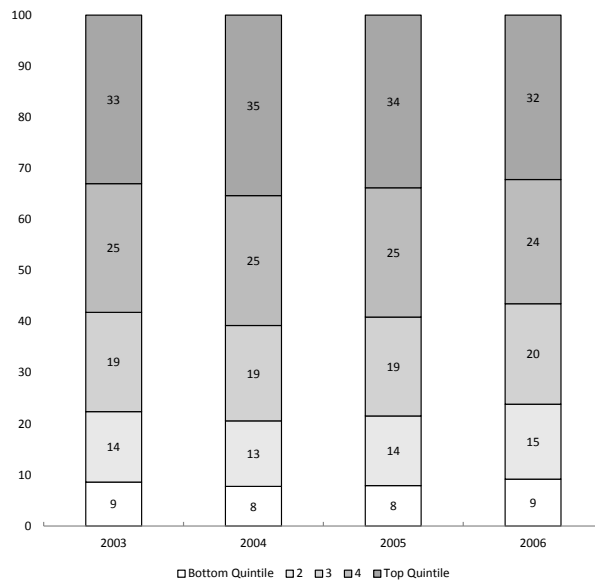
**Panel A. Origination (IRS household income quintiles as of 2002)**



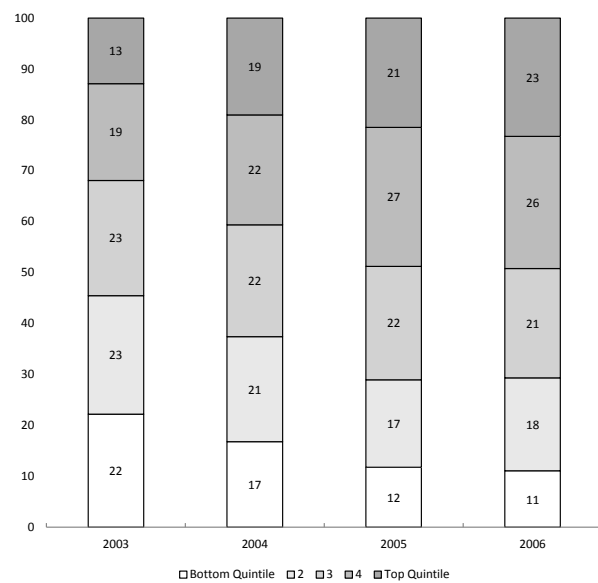
**Panel B. Delinquency (IRS household income quintiles as of 2002)**



**Panel C. Origination (HMDA buyer income quintiles as of 2002)**



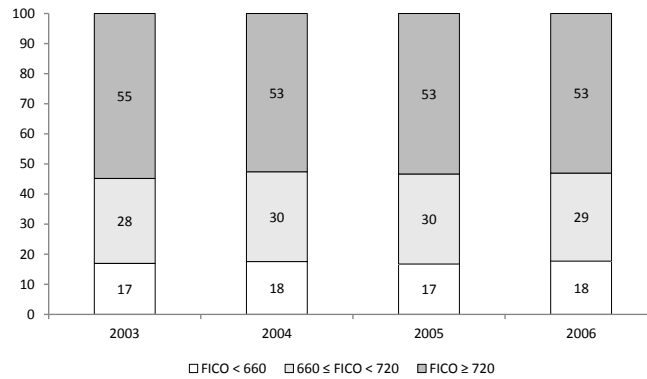
**Panel D. Delinquency (HMDA buyer income quintiles as of 2002)**



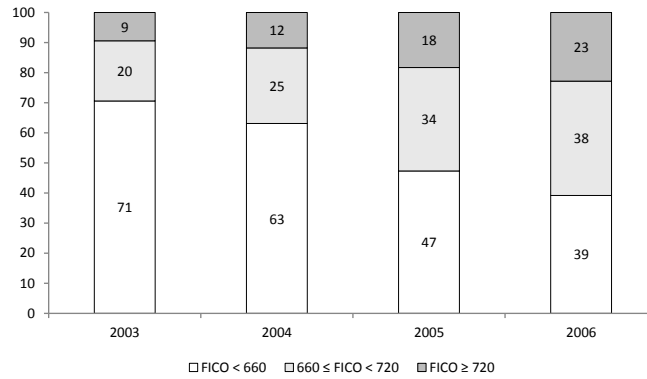
### Figure 3: Origination and delinquency by credit score

This figure shows the fraction of total dollar volume of purchase mortgages originated, as well as the total dollar volume of delinquent mortgages by cohort, split by credit score (FICO). A mortgage is defined as being delinquent if payments become more than 90 days past due (i.e., 90 days, 120 days or more, in foreclosure or REO) at any point during the three years after origination. Foreclosure is defined similarly, but includes only mortgages in foreclosure or REO. Data are from the 5% sample of the LPS dataset and the sample includes zip codes with nonmissing Zillow house price data. A FICO score of 660 corresponds to a widely used cutoff for subprime borrowers and 720 is close to the median FICO score of borrowers in the data (the median is 721 in 2003, 716 in 2004, 718 in 2005 and 715 in 2006).

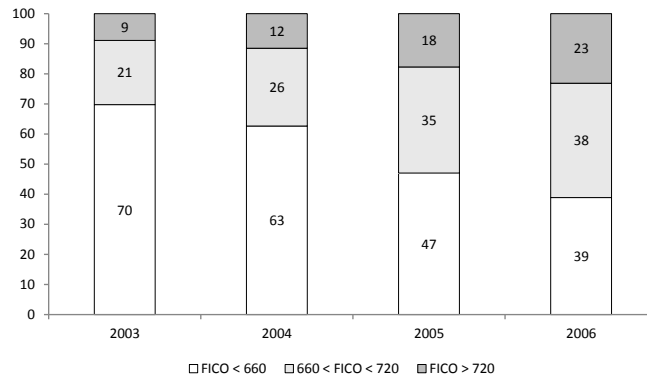
#### Panel A. Origination



#### Panel B. Delinquency



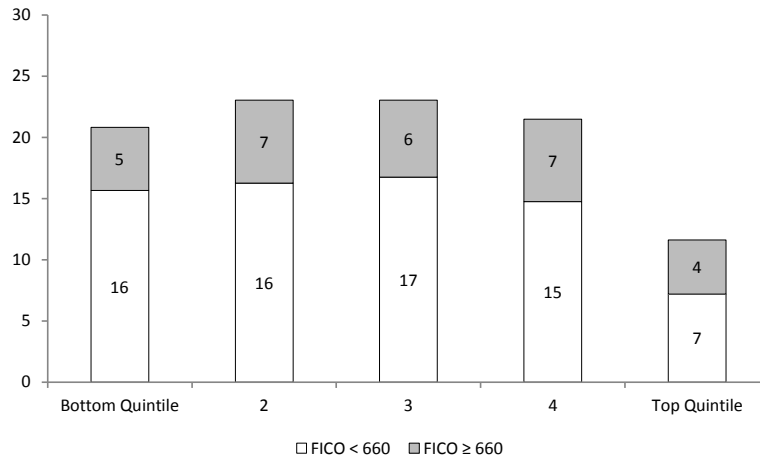
#### Panel C. Foreclosure



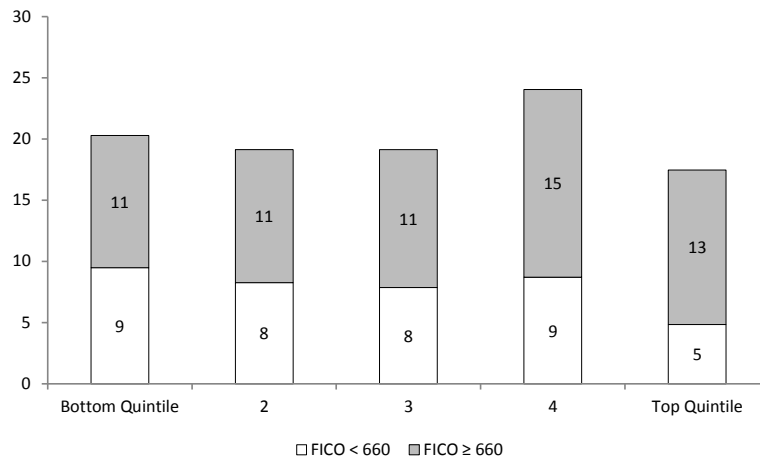
### Figure 4: Origination and delinquency by income and credit score

This figure shows the fraction of the dollar volume of purchase mortgages more than 90 days delinquent at any point during the three years after origination for the 2003 and 2006 origination cohorts. Panels show splits by quartiles of IRS household income as of 2002 (cutoffs are the same as those given in Figure 1), as well as by whether the borrower is above or below a credit score of 660 (a common FICO cutoff for subprime borrowers). In each panel fractions sum to 100 (the total amount of delinquent mortgages for each cohort), up to rounding error. Sample includes zip codes with nonmissing Zillow house price data. Data are from the 5% sample of the LPS dataset and the sample includes zip codes with nonmissing Zillow house price data.

#### Panel A. 2003 mortgage cohort



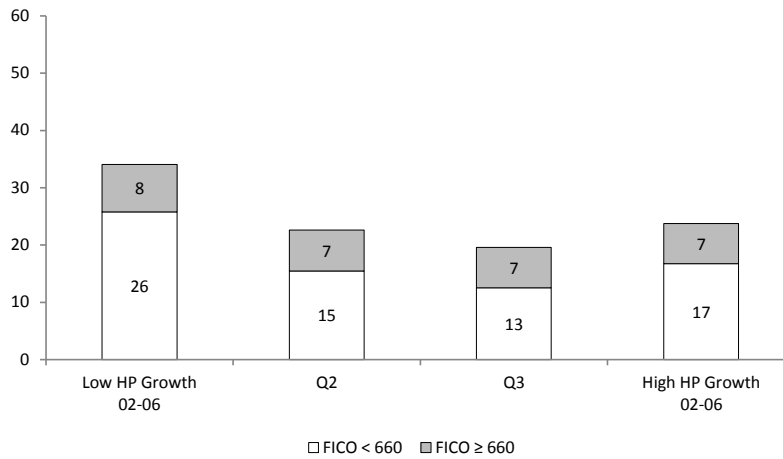
#### Panel B. 2006 mortgage cohort



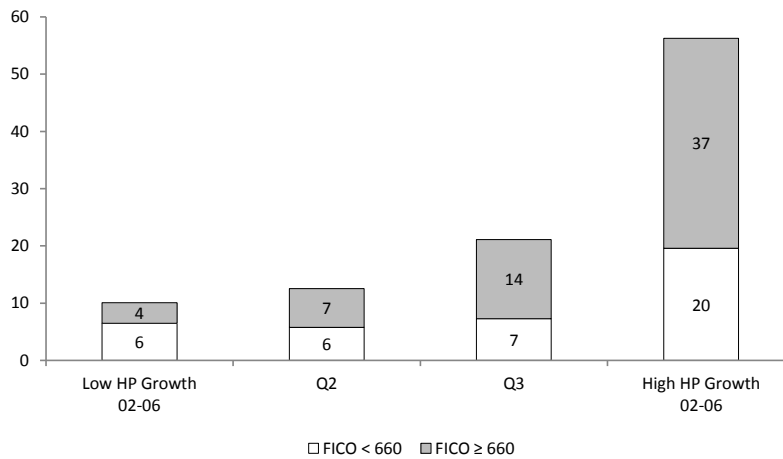
### Figure 5: Delinquency by house price growth and credit score

This figure shows the fraction of the dollar volume of purchase mortgages more than 90 days delinquent at any point during the three years after origination for the 2003 and 2006 origination cohorts. Panels show splits by quartiles of house price appreciation that the zip code experienced during the 2002-2006 period, as well as by whether the borrower is above or below a credit score of 660 (a common FICO cutoff for subprime borrowers). In each panel fractions sum to 100 (the total amount of delinquent mortgages for each cohort), up to rounding error. Sample includes zip codes with nonmissing Zillow house price data. Data are from the 5% sample of the LPS dataset and the sample includes zip codes with nonmissing Zillow house price data.

#### Panel A. 2003 mortgage cohort



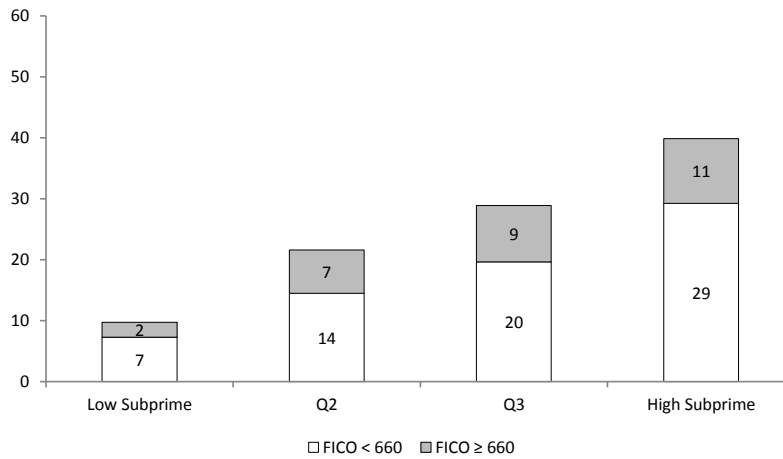
#### Panel B. 2006 mortgage cohort



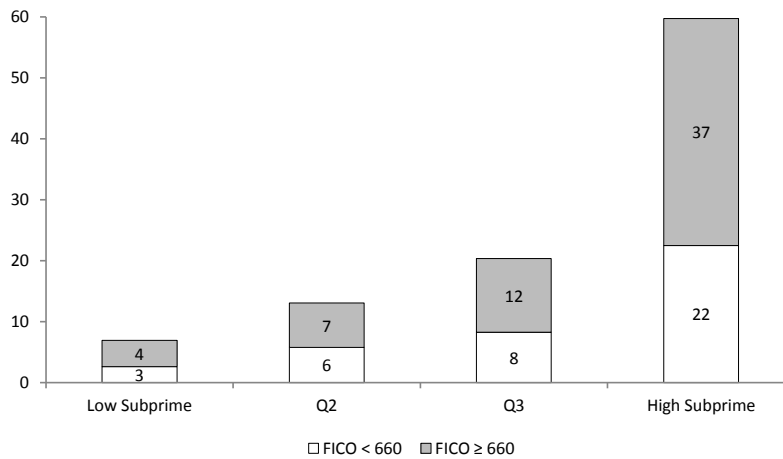
### Figure 6: Delinquency by proportion of subprime originations and credit score

This figure shows the fraction of the dollar volume of purchase mortgages more than 90 days delinquent at any point during the three years after origination for the 2003 and 2006 origination cohorts. Panels show splits by quartiles of subprime origination (defined as the percentage of loans made by lenders on the HUD subprime lender list), as well as by whether the borrower is above or below a credit score of 660 (a common FICO cutoff for subprime borrowers). In each panel fractions sum to 100 (the total amount of delinquent mortgages for each cohort), up to rounding error. Sample includes zip codes with nonmissing Zillow house price data. Data are from the 5% sample of the LPS dataset and the sample includes zip codes with nonmissing Zillow house price data.

#### Panel A. 2003 Mortgage cohort



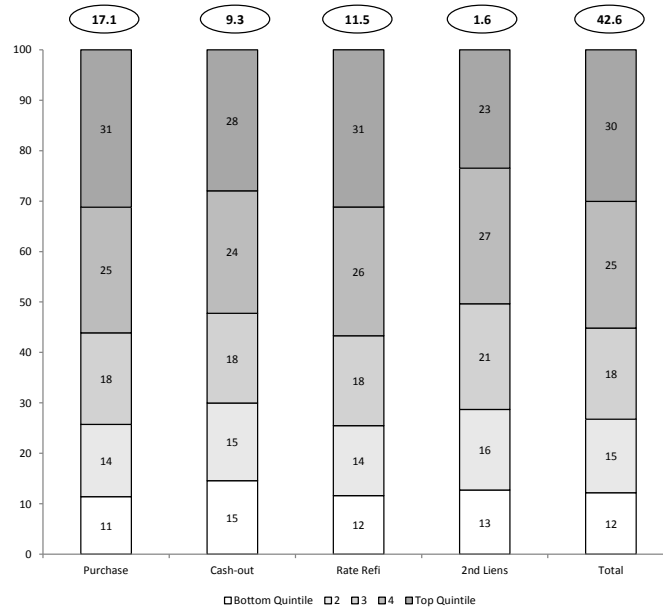
#### Panel B. 2006 Mortgage cohort



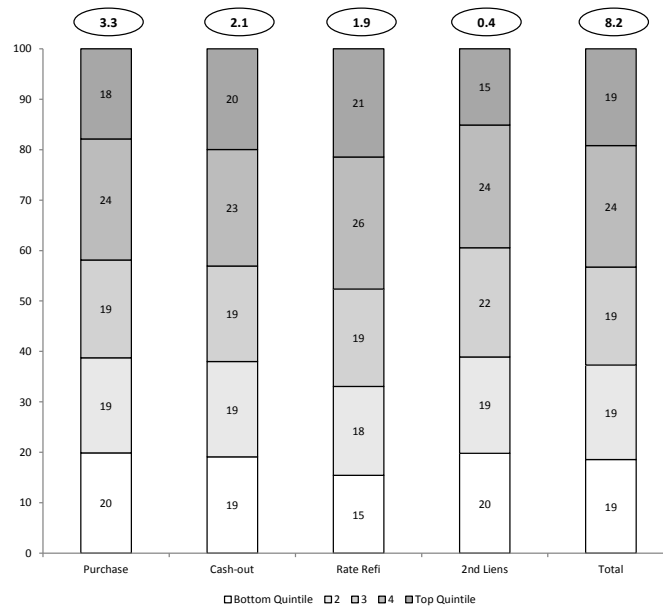
## Figure 7: Mortgage origination and delinquency by product type and income (2006)

This figure shows the fraction of the total dollar volume of purchase, cash-out refinance, rate refinance, and 2<sup>nd</sup> lien mortgages, as well as the total across all categories in the LPS dataset originated in 2006, as well as the dollar value of mortgages that became 90 days delinquent or more over the subsequent 3 years. The “Total” category includes mortgages that are unclassified in the dataset. Total origination in billions of dollars in the LPS sample is shown above each bar. Sample includes zip codes with nonmissing Zillow house price data. Quintiles are based on household income from the IRS as of 2002 and the cutoffs for each quintile are given in the notes to Figure 1.

### Panel A. Origination, 2006 cohort (IRS household income quintiles as of 2002)



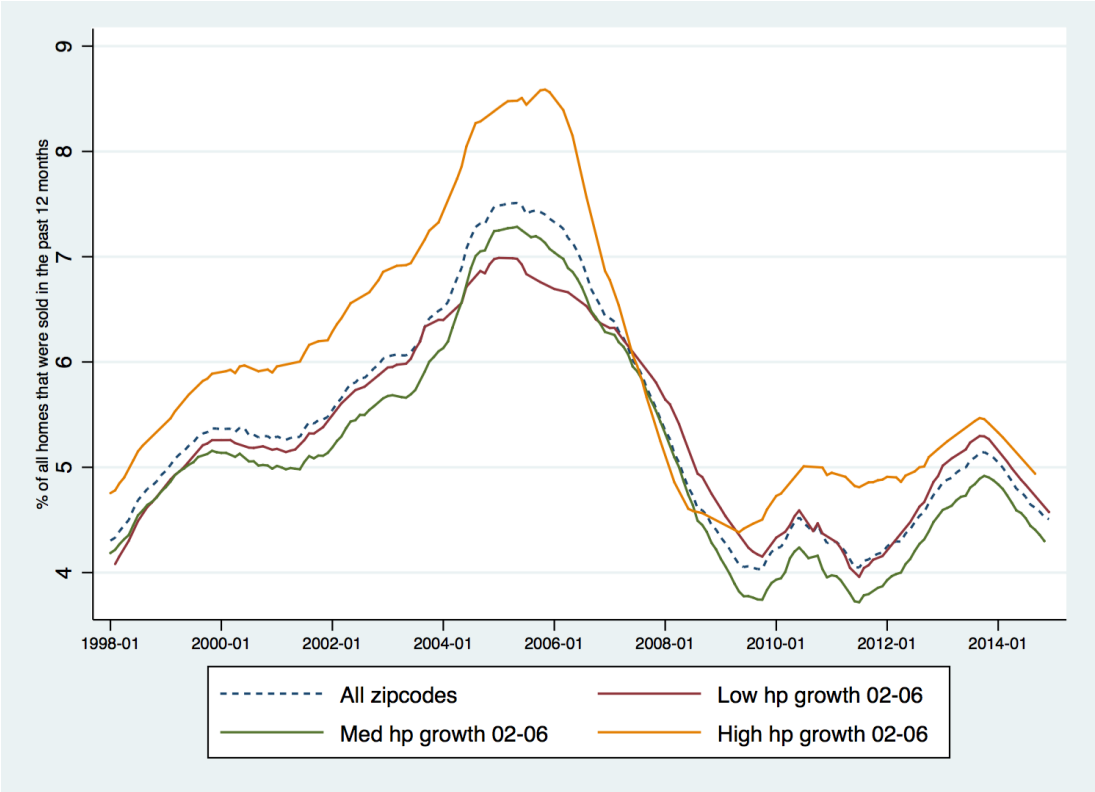
### Panel B. Delinquency, 2006 cohort (IRS household income quintiles as of 2002)





**Figure 8: Percentage of homes sold in past 12 months**

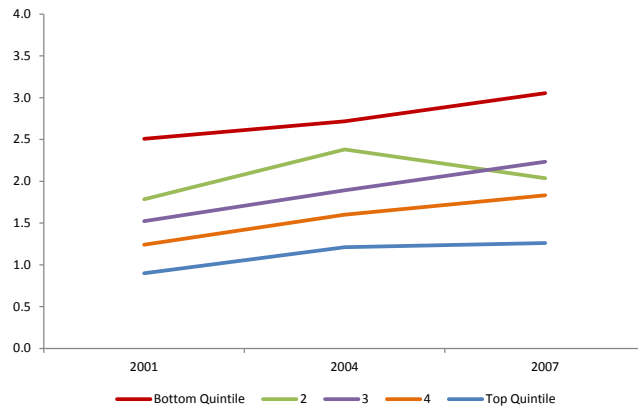
This figure shows the percentage of all transactions in each month for homes that also sold in the last 12 months (a measure of “flipping”). Data is provided by Zillow, and zip codes are broken down by house price growth between 2002 and 2006.



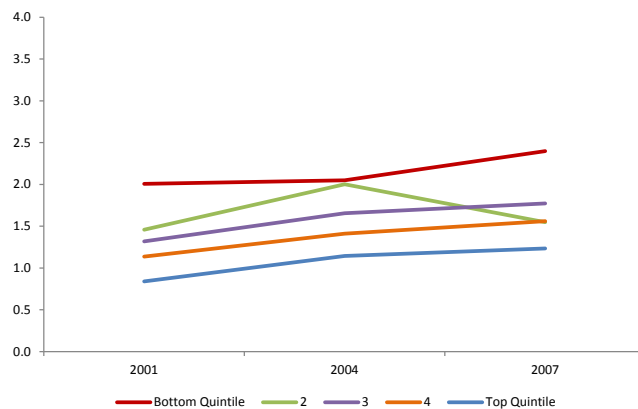
## Figure 9: Mortgage-related DTI by income

The figure shows the average and median DTI of households in the Survey of Consumer Finances. DTI is defined as the ratio of all mortgage-related debt over annual household income. Panel A shows value-weighted means within bin, Panel B shows medians within each bin. The sample includes households with positive mortgage debt. As of 2004, the cutoff for the bottom quintile corresponds to an annual household income of \$25.3k, the second quintile corresponds to \$44.3k, the third quintile corresponds to \$69.7k, and the fourth quintile corresponds to \$112.7k. Mortgage related debt includes SCF items *MRTHL* (Mortgage and Home Equity Loan, Primary Residence) and *RESDBT* (Other residential debt).

### Panel A. Average



### Panel B. Median



**Table 1. Summary statistics**

Panel A reports summary statistics for all zip codes in the HMDA sample with nonmissing house price data from Zillow. Column 1 shows the pooled summary statistics. Columns 2 to 4 show the summary statistics by household income as of 2002 divided into the highest quartile (column 2), the middle two quartiles (column 3), and the lowest quartile (column 4). Columns 5 to 7 do a similar split by house price growth in the zip code between 2002 and 2006. For each variable we show the average and standard deviation (in parenthesis). *IRS Household Income* is the average adjusted gross household income by zip code from the IRS. *HMDA Buyer Income* is the average applicant income by zip code from HMDA. *Average Purchase Mortgage Size* is the average balance at origination of purchase mortgages by zip code. *Number of mortgages originated per 100 residents* is the average number of purchase mortgages originated per 100 residents by zip code. *Debt to income* is the average ratio of the mortgage balance at the time of origination divided by the buyer income from HMDA. Panel B reports summary statistics for the 5% random sample from the LPS dataset.

**Panel A. HMDA data**

	Whole Sample	IRS Household Income, 2002			Zip Code House Price Growth, 2002-2006		
		High N=2088	Middle N=4346	Low N=2185	High N=2020	Middle N=4407	Low N=2192
IRS Household Income, 2002, '000s	50.93 (28.24)	84.81 (39.42)	44.75 (5.92)	30.85 (3.92)	47.40 (25.45)	54.44 (30.41)	47.13 (25.08)
HMDA Buyer Income, 2002, '000s	92.18 (67.26)	143.75 (98.40)	82.27 (46.87)	62.62 (24.85)	99.83 (70.94)	95.11 (70.58)	79.24 (53.87)
Average Purchase Mortgage Size, 2002, '000s	154.93 (86.70)	246.37 (113.33)	139.95 (46.49)	97.33 (36.46)	160.97 (76.74)	166.79 (95.63)	125.50 (67.57)
Mortgages Originated per 100 Residents, 2002	2.60 (2.16)	3.09 (3.12)	2.64 (1.78)	2.07 (1.52)	3.38 (3.42)	2.36 (1.53)	2.37 (1.47)
Debt to Income, 2002	2.13 (0.38)	2.26 (0.35)	2.16 (0.35)	1.97 (0.41)	2.18 (0.36)	2.17 (0.39)	2.03 (0.36)
Growth of IRS Household Income, 2002-2006, Annualized	0.046 (0.028)	0.064 (0.035)	0.042 (0.022)	0.035 (0.021)	0.053 (0.029)	0.047 (0.027)	0.036 (0.025)
Growth of HMDA Buyer Income, 2002-2006, Annualized	0.065 (0.061)	0.068 (0.063)	0.062 (0.058)	0.068 (0.064)	0.108 (0.066)	0.062 (0.050)	0.032 (0.052)
Growth in Total Purchase Mortgage Origination, 2002-2006, Annualized	0.121 (0.148)	0.078 (0.141)	0.119 (0.143)	0.168 (0.151)	0.170 (0.165)	0.123 (0.138)	0.074 (0.136)
Growth in Average Purchase Mortgage Size, 2002-2006, Annualized	0.067 (0.054)	0.075 (0.052)	0.062 (0.051)	0.069 (0.059)	0.124 (0.042)	0.063 (0.040)	0.021 (0.038)
Growth in Number of Purchase Mortgages, 2002-2006, Annualized	0.055 (0.129)	0.007 (0.131)	0.057 (0.124)	0.096 (0.121)	0.046 (0.144)	0.059 (0.126)	0.054 (0.119)

**Panel B. LPS data, 2002-2006 purchase mortgage cohorts (N = 272,077 for all cohorts)**

	Whole Sample	IRS Household Income, 2002			Zip Code House Price Growth, 2002-2006		
		High	Middle	Low	High	Middle	Low
Balance at Origination, 2003 cohort N = 51,947	188.69 (140.31)	276.80 (196.66)	167.36 (93.23)	118.35 (66.94)	197.94 (151.53)	204.49 (148.17)	148.47 (97.60)
Credit Score (FICO), 2003 cohort N = 44,750	711.8 (62.7)	729.0 (53.8)	710.1 (62.8)	691.7 (67.5)	711.1 (61.5)	716.0 (61.1)	704.7 (66.2)
3-Yr Delinquency Rate, 2003 cohort	0.037	0.015	0.038	0.070	0.027	0.033	0.057
3-Yr Delinquency Rate, 2006 cohort	0.183	0.115	0.177	0.271	0.301	0.131	0.148



**Table 3. Purchase mortgage origination and income by income level as of 2002**

This table shows OLS regressions of growth in total purchase mortgage credit, the average purchase mortgage size and the number of purchase mortgages originated at the zip code level on the growth rate of household income (from the IRS). Growth rates are annualized and computed between 2002 and 2006. Zip codes are separated into quartiles based on the household income as of 2002. The “High” column includes the top quartile, the “Medium” column includes the second and third quartiles, and “Low” includes the lowest quartile. In Panel A we partial out county fixed effects estimated over the whole sample. Standard errors are clustered by county (shown in parenthesis). \*, \*\*, \*\*\* indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

**Panel A. With county fixed effects**

	Growth in Total Purchase Mortgage Origination			Growth in average mortgage size			Growth in number of mortgages originated		
	High	Medium	Low	High	Medium	Low	High	Medium	Low
Growth of IRS Household Income	-0.191 (0.126)	0.218* (0.118)	0.144 (0.226)	0.227*** (0.034)	0.258*** (0.033)	0.222*** (0.061)	-0.406*** (0.110)	-0.037 (0.107)	-0.125 (0.193)
County FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Number of observations	2,088	4,346	2,185	2,088	4,346	2,185	2,088	4,346	2,185
R2	0.00	0.00	0.00	0.04	0.03	0.01	0.01	0.00	0.00

**Panel B. Without county fixed effects**

	Growth in Total Purchase Mortgage Origination			Growth in average mortgage size			Growth in number of mortgages originated		
	High	Medium	Low	High	Medium	Low	High	Medium	Low
Growth of IRS Household Income	0.222* (0.131)	1.398*** (0.139)	1.734*** (0.210)	0.432*** (0.056)	0.766*** (0.065)	0.869*** (0.086)	-0.197* (0.119)	0.589*** (0.117)	0.666*** (0.157)
County FE	N	N	N	N	N	N	N	N	N
Number of observations	2,088	4,346	2,185	2,088	4,346	2,185	2,088	4,346	2,185
R2	0.00	0.05	0.06	0.08	0.11	0.10	0.00	0.01	0.01

**Table 4. Mortgage origination and growth in buyer income**

Panel A of this table shows OLS regressions of growth in total mortgage credit, the average mortgage size and the number of mortgages originated at the zip code level on the growth rate of average buyer income in the zip code (obtained from HMDA). Panel B shows OLS regressions of annualized growth in total mortgage credit at the zip code level on the annualized growth rate of average buyer income in the zip code (from HMDA). Results are split by the proportion of loans sold to Fannie Mae and Freddie Mac (the GSEs) as of 2006, and by the proportion of loans originated by subprime lenders as of 2006 (subprime lenders are defined by the HUD subprime lender list). Panel C shows the same regressions as in Panel A for total mortgage origination and the average mortgage size for alternative time periods. Growth rates are all annualized and computed between 2002 and 2006. Sample includes zip codes with house price data from Zillow Standard errors are clustered by county (shown in parenthesis). \*, \*\*, \*\*\* indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

**Panel A. Mortgage growth measures between 2002 and 2006**

	Total Purchase Mortgage Origination		Average Mortgage Size		Number of Mortgages	
Growth of Buyer Income (HMDA)	0.369*** (0.047)	0.524*** (0.047)	0.282*** (0.015)	0.539*** (0.033)	0.117*** (0.040)	0.002 (0.052)
County FE	Y	N	Y	N	Y	N
Number of observations	8,619	8,619	8,619	8,619	8,619	8,619
R2	0.35	0.05	0.72	0.37	0.31	0.00

**Panel B. Heterogeneity by propensity for income misreporting**

	Growth in Total Purchase Mortgage Origination					
	High GSE	Med GSE	Low GSE	High Subprime	Med Subprime	Low Subprime
	Fraction	Fraction	Fraction	Fraction	Fraction	Fraction
Growth of Buyer Income (HMDA)	0.335*** (0.077)	0.387*** (0.054)	0.348*** (0.098)	0.470*** (0.090)	0.313*** (0.059)	0.375*** (0.080)
County FE	Y	Y	Y	Y	Y	Y
Number of observations	2,203	4,355	2,062	2,120	4,326	2,174
R2	0.01	0.02	0.02	0.03	0.01	0.02

**Panel C. Alternative time periods**

	Growth in Total Purchase Mortgage Origination				Growth in Average Mortgage Size			
	1996-1998	1998-2002	2002-2006	2007-2011	1996-1998	1998-2002	2002-2006	2007-2011
Growth of Buyer Income (HMDA)	0.260*** (0.033)	0.258*** (0.024)	0.368*** (0.047)	0.341*** (0.029)	0.261*** (0.015)	0.179*** (0.015)	0.282*** (0.015)	0.307*** (0.015)
County FE	Y	Y	Y	Y	Y	Y	Y	Y
Number of observations	8,597	8,609	8,620	8,550	8,597	8,609	8,620	8,550
R2	0.57	0.44	0.35	0.48	0.46	0.57	0.72	0.64

**Table 5. Mortgage origination and income at the transaction level**

Panel A shows regressions of the logarithm of purchase mortgage size at the individual level on the logarithm of average household income in the census tract (inferred using zip code household income from the IRS). The unit of observation is an individual loan in HMDA in zip codes with nonmissing Zillow house price data. IRS data is available for 2002, 2004, 2005, and 2006. In columns 2 and 4 the income variable is interacted with a linear trend for the years in the sample. Panel B shows regressions of the logarithm of mortgage size at the individual level on the logarithm of average household income in the zip code (inferred using zip code household income from the IRS). The unit of observation is an individual loan in the LPS originated between 2003 and 2006 in zip codes with nonmissing Zillow house price data. The first four columns include only purchase mortgages, and the last four include only cash-out refinancing mortgages. Standard errors are clustered by county (shown in parenthesis). \*, \*\*, \*\*\* indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

**Panel A. Purchase mortgages (HMDA)**

	(1)	(2)	(3)	(4)
Ln(IRS Household Income)	0.600*** (0.014)	0.626*** (0.017)	0.397*** (0.029)	0.396*** (0.038)
Ln(IRS Household Income) x Linear trend		-0.010** (0.005)		0.000 (0.004)
Year FE and county FE	Y	Y	N	N
Year FE and census tract FE	N	N	Y	Y
Number of observations	17,220,064	17,220,064	17,220,064	17,220,064
R2	0.23	0.23	0.29	0.29

**Panel B. Purchase and cash-out refinancing mortgages (LPS)**

	Purchase				Cash-out Refinancing			
Ln(IRS Household Income)	0.623*** (0.014)	0.688*** (0.021)	0.454*** (0.047)	0.619*** (0.067)	0.550*** (0.016)	0.587*** (0.027)	0.284*** (0.060)	0.302*** (0.090)
Ln(IRS Household Income) x Linear trend		-0.018*** (0.004)		-0.018*** (0.005)		-0.009 (0.006)		-0.002 (0.009)
Year FE and county FE	Y	Y	N	N	Y	Y	N	N
Year FE and zip code FE	N	N	Y	Y	N	N	Y	Y
Number of observations	272,077	272,077	272,077	272,077	108,097	108,097	108,097	108,097
R2	0.58	0.58	0.63	0.63	0.50	0.50	0.57	0.57

**Table 6. Mortgage refinancing and income**

Panel A shows OLS regressions of growth in total origination of mortgages for refinancing, the average refinancing mortgage size and the number of refinancing mortgages originated at the zip code level on the growth rate of household income (from the IRS). Growth rates are annualized and computed between 2002 and 2006. Panel B shows fixed effects regressions of the logarithm of total refinancing mortgage credit at the zip code level, the logarithm of average refinancing mortgage size, and the logarithm of the total number of refinancing mortgages on the logarithm of household income. IRS data is available for 2002, 2004, 2005, and 2006. In columns 2, 4 and 6 the income variable is interacted with indicator variables for each year in the sample. Sample includes zip codes with house price data from Zillow. Standard errors are clustered by county (shown in parenthesis). \*, \*\*, \*\*\* indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

**Panel A. Mortgage growth measures between 2002 and 2006**

	Total Refinancing Mortgage		Average Mortgage Size		Number of Mortgages	
	Origination					
Growth of IRS Household Income	-1.200*** (0.129)	-0.332** (0.149)	-0.005 (0.030)	0.459*** (0.041)	-1.100*** (0.096)	-0.651*** (0.115)
County FE	Y	N	Y	N	Y	N
Number of observations	8,622	8,622	8,622	8,622	8,622	8,622
R2	0.55	0.00	0.75	0.05	0.48	0.02

**Panel B. Panel specification**

	Total Refinancing Mortgage		Average Mortgage Size		Number of Mortgages	
	Origination					
Ln(IRS Household Income)	-0.123 (0.129)	1.300*** (0.129)	0.322*** (0.031)	0.314*** (0.050)	-0.440*** (0.115)	1.001*** (0.098)
Ln(IRS Household Income) x Year 2004		-0.502*** (0.026)		-0.001 (0.013)		-0.506*** (0.020)
Ln(IRS Household Income) x Year 2005		-0.604*** (0.039)		0.015 (0.018)		-0.623*** (0.027)
Ln(IRS Household Income) x Year 2006		-0.717*** (0.044)		-0.002 (0.019)		-0.719*** (0.031)
Zip Code FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Number of observations	36,265	36,265	36,265	36,265	36,265	36,265
R2	0.96	0.97	0.96	0.96	0.96	0.97

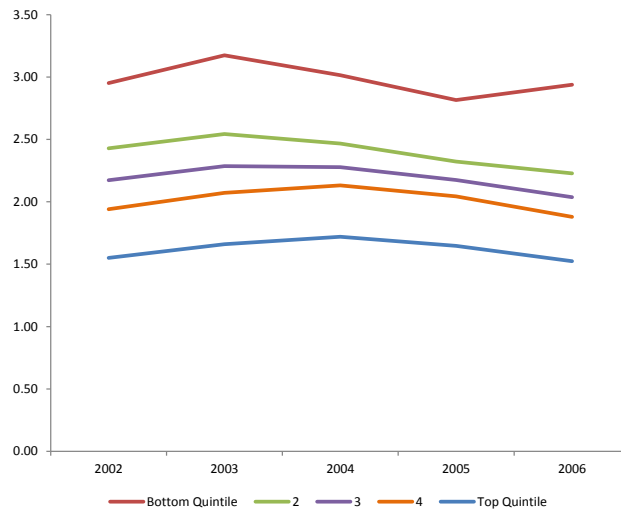


## ONLINE APPENDIX (NOT FOR PUBLICATION)

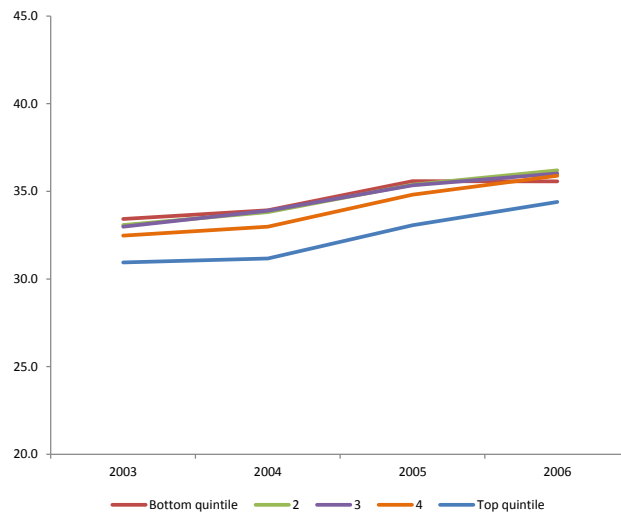
### Figure A1. Purchase mortgage DTI

Panel A shows average debt-to-income for individuals in each HMDA buyer income quintile. Debt-to-income is defined as the ratio of the mortgage balance at origination divided by the income of the applicant. Data are from HMDA and sample includes zip codes with nonmissing house price data from Zillow. Panel B shows debt-to-income for purchase mortgages according to a standard industry definition that calculates the ratio of recurring mortgage debt payments over monthly income. Income quintiles based on IRS household income as of 2002. Data are from the 5% sample of LPS and sample includes zip codes with nonmissing house price data from Zillow.

#### Panel A. DTI (individual buyer income, HMDA)

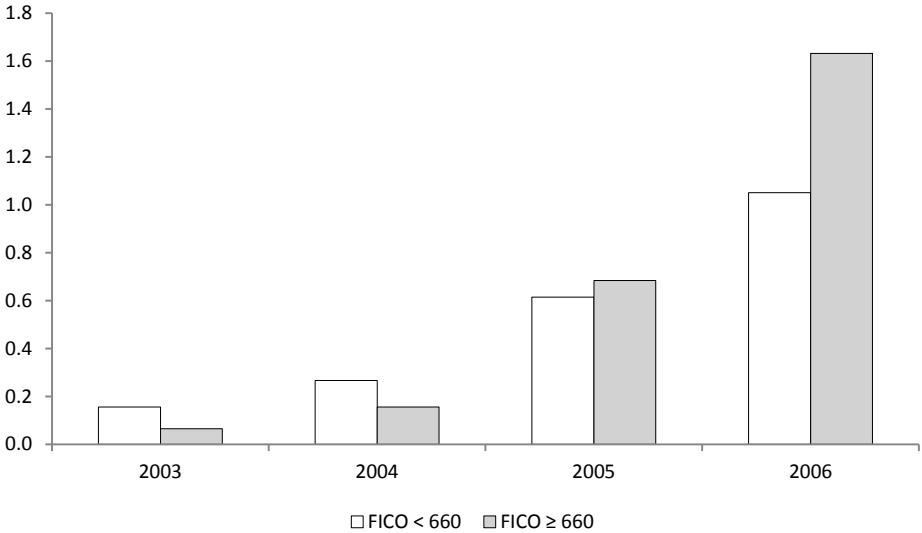


#### Panel B. Mortgage payments as a percentage of income (quintiles based on IRS income as of 2002)



### Figure A2. Delinquency by credit score – dollar amounts

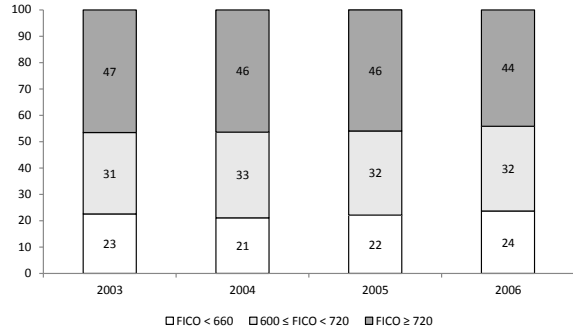
This figure shows the total dollar volume of delinquent purchase mortgages by cohort, split by credit score (FICO). A mortgage is defined as being delinquent if payments become more than 90 days past due (i.e., 90 days, 120 days or more, in foreclosure or REO) at any point during the three years after origination. Data are from the 5% sample of the LPS dataset and the sample includes zip codes with nonmissing Zillow house price data. A FICO score of 660 corresponds to a widely used cutoff for subprime borrowers.



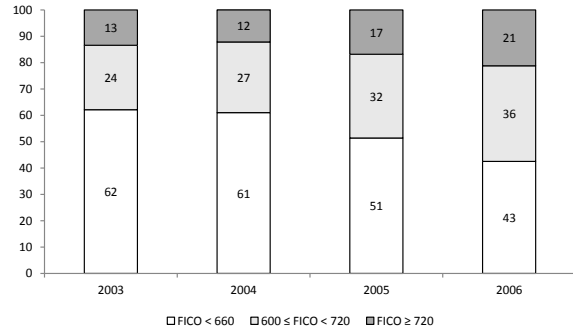
### Figure A3: Mortgage origination and delinquency by credit score, alternative datasets

This figure shows the fraction of total dollar volume of purchase mortgages originated, as well as the total dollar volume of delinquent mortgages by cohort, split by income quintile. Data in Panel A is from Blackbox Logic, a dataset of private-label securitized mortgages, and data in Panel B comes from the public Freddie Mac single-family home dataset. The sample includes zip codes with nonmissing Zillow house price data. A mortgage is defined as being delinquent if payments become more than 90 days past due (i.e., 90 days, 120 days or more, in foreclosure or REO) at any point during the three years after origination.

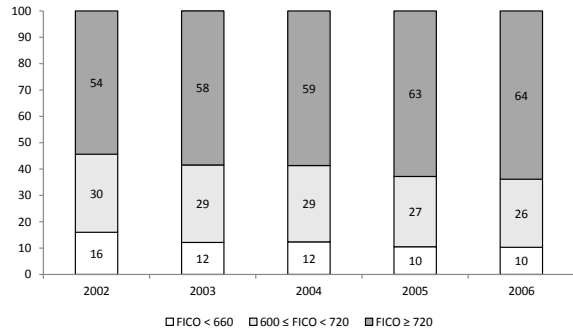
**Panel A. Origination (Blackbox Logic data)**



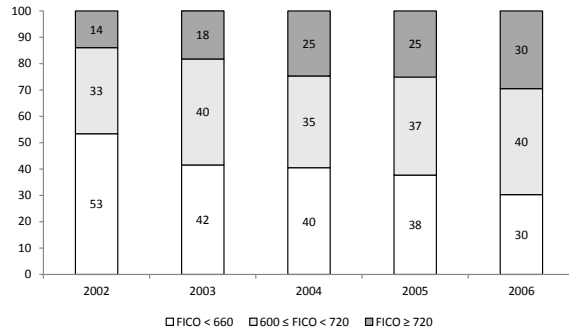
**Panel B. Delinquency (Blackbox Logic data)**



**Panel C. Origination (Freddie Mac data)**



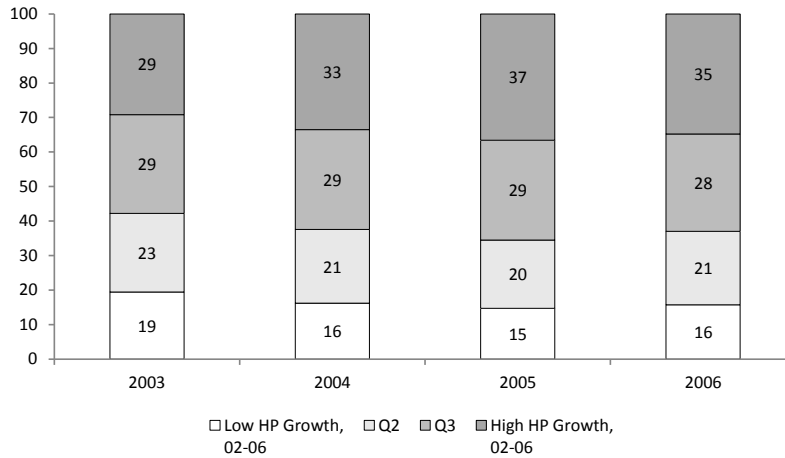
**Panel D. Delinquency (Freddie Mac data)**



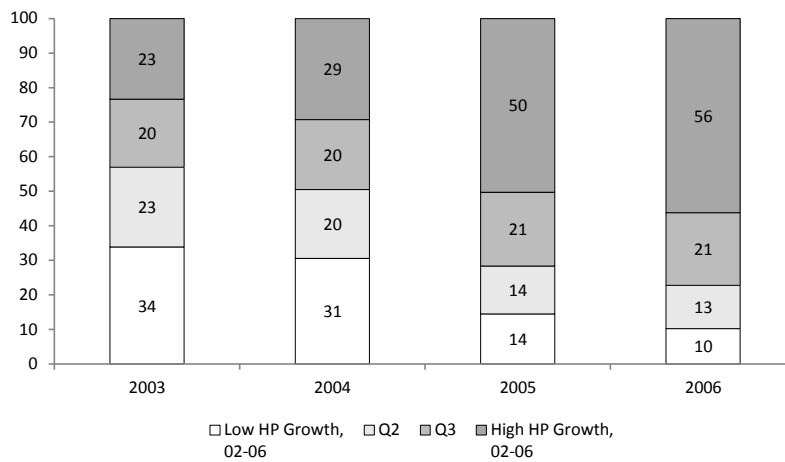
## Figure A4: Origination and Delinquency by House Price Growth

This figure shows the fraction of total dollar volume of purchase mortgages originated, as well as the total dollar volume of delinquent mortgages by cohort, split by quartiles of house price growth between 2002 and 2006. A mortgage is defined as being delinquent if payments become more than 90 days past due (i.e., 90 days, 120 days or more, in foreclosure or REO) at any point during the 3 years after origination. Data is from the 5% sample of the LPS dataset and the sample includes zip codes with non-missing Zillow house price data.

### Panel A. Origination



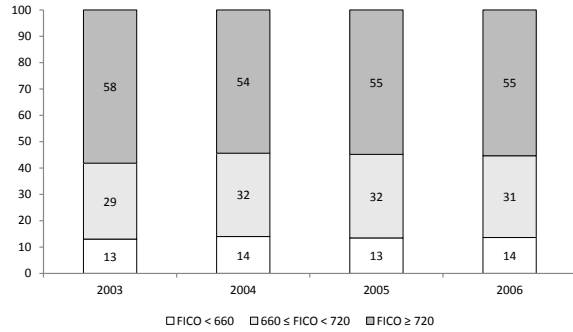
### Panel B. Delinquency



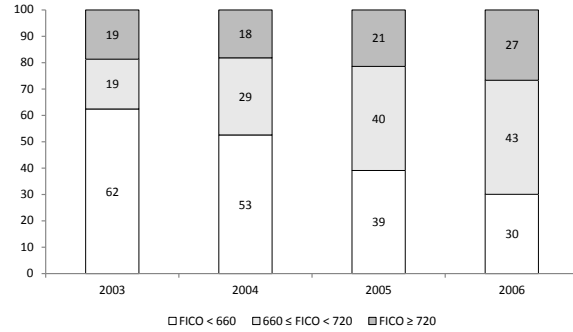
## Figure A5: Mortgage origination and delinquency in recourse and non-recourse states

This figure shows the fraction of total dollar volume of purchase mortgages originated, as well as the total dollar volume of delinquent mortgages by cohort, split by recourse and non-recourse states, as well as credit scores. Non-recourse states include AK, AZ, CA, HI, MN, MT, ND, OK, OR, and WA. A mortgage is defined as being delinquent if payments become more than 90 days past due (i.e., 90 days, 120 days or more, in foreclosure or REO) at any point during the three years after origination. A FICO score of 660 corresponds to a widely used cutoff for subprime borrowers and 720 is close to the median FICO score of borrowers in the data (please see Table 1 for additional summary statistics). Data are from the 5% sample of the LPS dataset and the sample includes zip codes with nonmissing Zillow house price data.

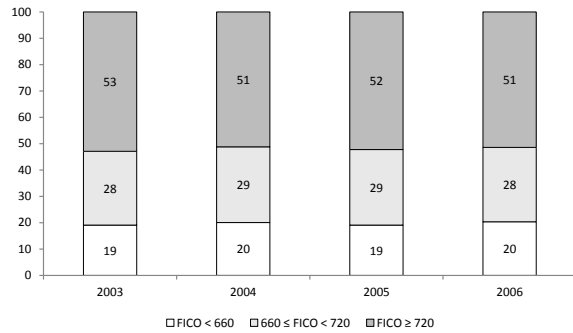
**Panel A. Origination, non-recourse states**



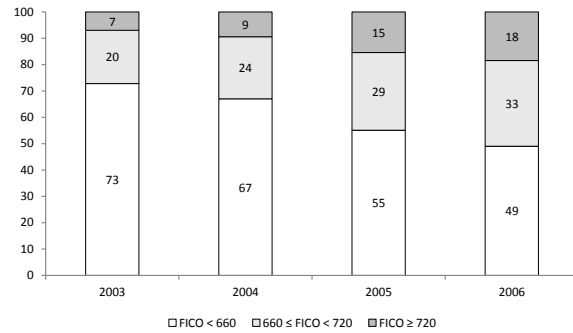
**Panel B. Delinquency, non-recourse states**



**Panel C. Origination, recourse states**



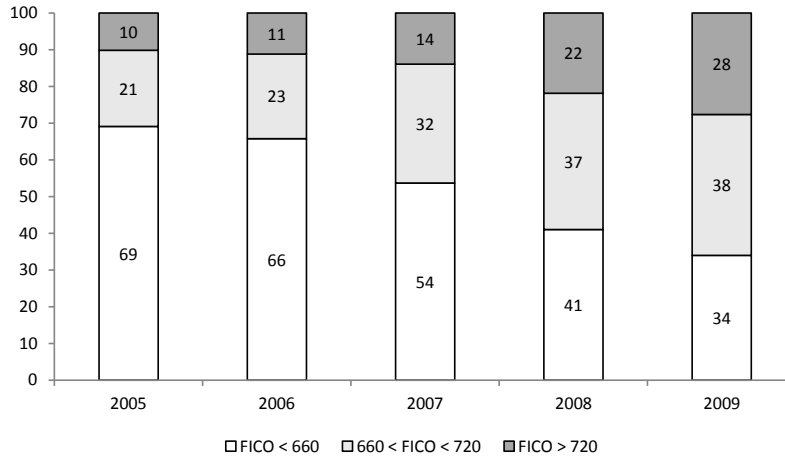
**Panel D. Delinquency, recourse states**



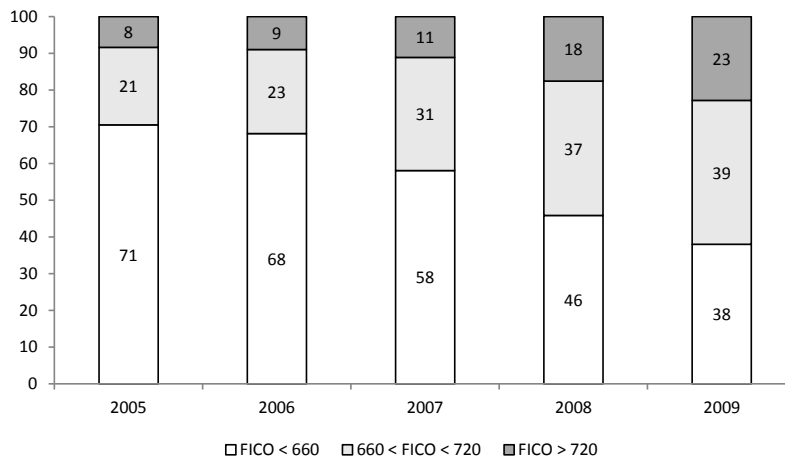
## Figure A6: Delinquency by year and credit score

This figure shows the fraction of total dollar volume of delinquent purchase mortgages, as well as of all types of mortgages available in the LPS data, split by credit score (FICO). Fractions are based on the total dollar amount of delinquent mortgages outstanding as of the last quarter of each year. A mortgage is defined as being delinquent if payments become more than 90 days past due (i.e., 90 days, 120 days or more, in foreclosure or REO) at any point during the three years after origination. Data are from the 5% sample of the LPS dataset and the sample includes zip codes with nonmissing Zillow house price data. A FICO score of 660 corresponds to a widely used cutoff for subprime borrowers and 720 is close to the median FICO score of borrowers in the data (please see Table 1 for additional summary statistics).

### Panel A. Purchase mortgages



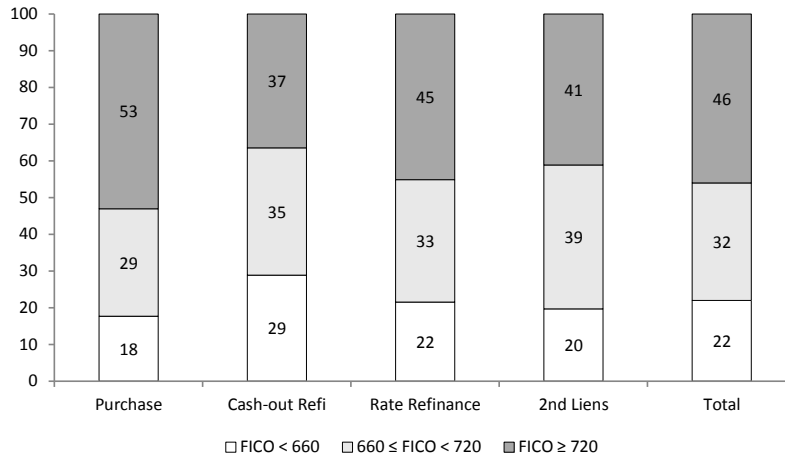
### Panel B. All mortgages



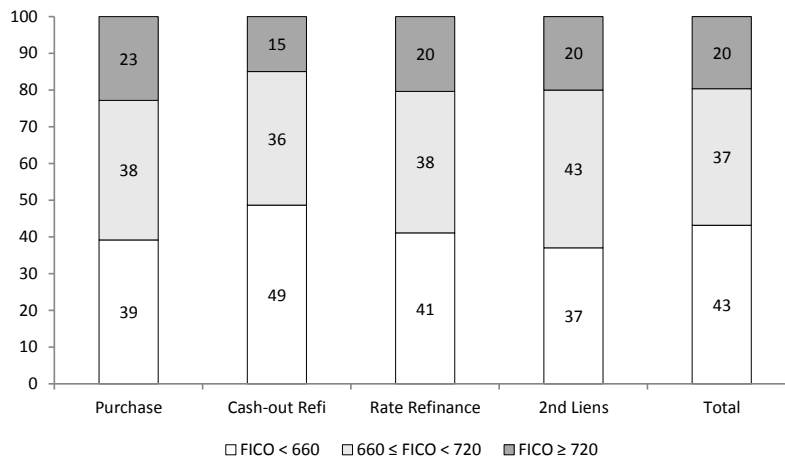
## Figure A7: Origination and Delinquency by Credit Score – All Mortgage Types

This figure shows the fraction of total dollar volume of originations, as well as the total dollar volume of delinquent mortgages, of all types available in the LPS data, split by credit score (FICO). A mortgage is defined as being delinquent if payments become more than 90 days past due (i.e., 90 days, 120 days or more, in foreclosure or REO) at any point during the 3 years after origination. Data is from the 5% sample of the LPS dataset and the sample includes zip codes with non-missing Zillow house price data. A FICO score of 660 corresponds to a widely used cutoff for subprime borrowers and 720 is close to the median FICO score of borrowers in the data (please see Table 1 for additional summary statistics).

### Panel A. Origination, 2006 cohort



### Panel B. Delinquency, 2006 cohort



**Table A1. Summary statistics for full HMDA sample**

This table reports summary statistics for all zip codes in the HMDA sample. Column 1 shows the pooled summary statistics. Columns 2 to 4 show the summary statistics by household income as of 2002 divided into the highest quartile (column 2), the middle two quartiles (column 3), and the lowest quartile (column 4). For each variable we show the average and standard deviation (in parenthesis). *IRS Household Income* is the average adjusted gross household income by zip code from the IRS. *HMDA Buyer Income* is the average applicant income by zip code from HMDA. *Average Purchase Mortgage Size* is the average size of purchase mortgages originated by zip code. *Number of mortgages originated per 100 residents* is the average number of purchase mortgages originated per 100 residents by zip code. *Debt to income* is the average ratio of the mortgage balance at the time of origination divided by the buyer income from HMDA.

	Whole sample	Zip code household income, 2002		
	N =27385	High N =6936	Middle N =14126	Low N =6323
IRS Household Income, 2002, '000s	39.41 (24.68)	63.57 (39.10)	34.16 (3.65)	24.63 (3.41)
HMDA Buyer Income, 2002, '000s	74.21 (56.34)	111.06 (84.89)	64.11 (34.21)	56.33 (34.54)
Average Purchase Mortgage Size, 2002, '000s	110.47 (70.44)	182.26 (93.37)	93.10 (36.33)	70.53 (32.51)
Number of mortgages originated per 100 residents, 2002, purchase mortgages only	1.91 (6.08)	3.06 (8.49)	1.69 (5.73)	1.14 (2.39)
Debt to Income, 2002	1.83 (0.48)	2.16 (0.38)	1.79 (0.42)	1.55 (0.48)
Growth of IRS Household Income, 2002-2006, Annualized	0.043 (0.031)	0.051 (0.035)	0.039 (0.025)	0.045 (0.036)
Growth of HMDA Buyer Income, 2002-2006, Annualized	0.058 (0.072)	0.062 (0.064)	0.055 (0.065)	0.059 (0.090)
Growth in Total Purchase Mortgage Origination, 2002-2006, Annualized	0.164 (0.210)	0.110 (0.166)	0.172 (0.197)	0.204 (0.261)
Growth in Average Purchase Mortgage Size, 2002-2006, Annualized	0.066 (0.063)	0.065 (0.054)	0.063 (0.058)	0.076 (0.081)
Growth in Number of Purchase Mortgages, 2002-2006, Annualized	0.092 (0.180)	0.046 (0.150)	0.104 (0.169)	0.117 (0.220)



**Table A2. Mortgage origination and income, IRS and buyer income**

The Table shows OLS regressions of annualized growth in total mortgage credit, the average mortgage size and the number of mortgages originated at the zip code level on the annualized growth rate of average household income (from the IRS) and the annualized growth rate of average buyer income in the zip code (obtained from HMDA). The data only includes mortgages for home purchase. Columns 3, 6 and 9 also include zip code house price growth from Zillow as a control. Sample includes zip codes with house price data from Zillow. Standard errors are clustered by county (shown in parenthesis). \*, \*\*, \*\*\* indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

**Panel A. With county fixed effects**

	Growth in total mortgage origination			Growth in average mortgage size			Growth in number of mortgages originated		
Growth of Buyer Income (HMDA)	0.369*** (0.047)	0.376*** (0.047)	0.349*** (0.047)	0.282*** (0.015)	0.276*** (0.015)	0.266*** (0.015)	0.117*** (0.040)	0.130*** (0.040)	0.116*** (0.040)
Growth of IRS Household Income		-0.224** (0.088)	-0.214*** (0.079)		0.208*** (0.023)	0.212*** (0.021)		-0.417*** (0.075)	-0.411*** (0.071)
Zip code house price growth			0.559*** (0.139)			0.198*** (0.023)			0.281** (0.122)
County FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Number of observations	8,619	8,619	8,619	8,619	8,619	8,619	8,619	8,619	8,619
R2	0.35	0.35	0.35	0.72	0.73	0.74	0.31	0.32	0.32

**Panel B. Without county fixed effects**

	Growth in total mortgage origination			Growth in average mortgage size			Growth in number of mortgages originated		
Growth of Buyer Income (HMDA)	0.524*** (0.047)	0.511*** (0.045)	0.292*** (0.048)	0.539*** (0.033)	0.506*** (0.034)	0.285*** (0.016)	0.002 (0.052)	0.023 (0.050)	0.041 (0.045)
Growth of IRS Household Income		0.150 (0.101)	0.020 (0.089)		0.372*** (0.029)	0.239*** (0.022)		-0.227*** (0.084)	-0.217*** (0.083)
Zip code house price growth			0.466*** (0.057)			0.473*** (0.017)			-0.038 (0.053)
County FE	N	N	N	N	N	N	N	N	N
Number of observations	8,619	8,619	8,619	8,619	8,619	8,619	8,619	8,619	8,619
R2	0.05	0.05	0.08	0.37	0.41	0.63	0.00	0.00	0.00

**Table A3. Mortgage origination and income for alternative time periods**

The Table shows OLS regressions of growth in total purchase mortgage credit at the zip code level, and the growth in average mortgage size on the growth rate of household income from the IRS. Growth rates are annualized and computed between 2002 and 2006. Sample includes zip codes with house price data from Zillow. Standard errors are clustered by county (shown in parenthesis). \*, \*\*, \*\*\* indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Growth in Total Mortgage Origination			Growth in Average Mortgage Size		
	1998-2002	2002-2006	2007-2011	1998-2002	2002-2006	2007-2011
Growth of IRS Household Income	0.616*** (0.065)	-0.182** (0.090)	-0.512*** (0.065)	0.170*** (0.023)	0.239*** (0.026)	-0.056* (0.033)
County FE	Y	Y	Y	Y	Y	Y
Number of observations	8,605	8,619	8,529	8,605	8,619	8,529
R2	0.44	0.33	0.46	0.53	0.68	0.58