CHANGES IN BUYER COMPOSITION AND THE EXPANSION OF CREDIT DURING THE BOOM[1]

Manuel Adelino, Duke

Antoinette Schoar, MIT and NBER

Felipe Severino, Dartmouth

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

Abstract

We provide a novel interpretation of the debt dynamics leading up to the financial crisis of 2007. Earlier research suggests that distortions in the supply of mortgage credit, in particular a decoupling of credit flow from income growth, may have been responsible for the rise in house prices and the subsequent collapse of the housing market. By focusing on individual transactions rather than zip codes, we show that the relationship between individual mortgage size and income growth during the housing boom never decoupled and was strongly positive, in line with previous periods (and independent of how income is measured). Zip codes with large house price increases experienced significant changes in the composition of buyers, i.e. home buyers had increasingly higher income than the average residents in an area. Instead, there was a significant expansion of credit along the *extensive margin* due to the increase in the pace of home buying. The fact that a larger fraction of people were holding recent mortgages and were thus close to their maximum debt capacity likely contributed to the build-up of systemic risk prior to the crisis. Compared to prior years, middle and high income borrowers (not the poor), as well as those with high fico scores, made up a much larger share of delinquencies in the crisis. These results are most consistent with an expectations based view of the financial crisis where both home buyers and lenders were buying into an unfolding housing bubble.

[1] We thank Nittai Bergman, Ing-Haw Cheng, Andrew Lo, Sendhil Mullainathan, Debarshi Nanda, David Robinson, Steve Ross, Andrei Shleifer, Jeremy Stein, and seminar participants at Berkeley, Brandeis, Dartmouth, Duke, Harvard Business School, UNC and MIT for thoughtful comments.

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 in the literature is that fundamental changes in the origination technology of lenders significantly contributed to unsustainable levels of borrowing and ultimately caused an acceleration of house prices. This interpretation builds on a key finding by Mian and Sufi (2009) that growth in mortgage credit at the zip code level became negatively correlated with per capita income growth in the run-up to the financial crisis, suggesting that lending was 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, which is often referred to as the credit supply side view of the housing crisis.[2]

We provide a novel interpretation of the debt dynamics leading up to the financial crisis. We show that between 2002 and 2006 mortgage origination increased across the whole income distribution and not just the poor. At the same time credit growth at the individual level did not decouple from income growth. In other words debt to income levels (DTI) did not disproportionately go up or for the poorer part of the income distribution. Instead, there was a significant expansion of credit along the *extensive margin* due to the increase in the pace of home buying. This meant that a larger fraction of people were holding recent mortgages when the crisis hit. The fact that a larger fraction of homeowners (in particular middle and high income home buyers) were close to their debt capacity might have led to a build-up of systemic risk that ultimately contributed to the downfall of the housing market. In line with this interpretation, we show that the share of mortgage dollars in delinquency after 2007 went up for borrowers with above average income levels, and also for borrowers with credit scores above 660, a typical subprime cutoff. These results are most consistent with an expectations based view of the financial crisis where both home buyers and lenders were riding the increase in house prices and defaulted when prices dropped.

We show that using zip codes as the unit of observation for analyzing the growth in lending and per capita income, as proposed in Mian and Sufi (2009), confounds changes in individual leverage and growth due to more mortgages being originated. By focusing on individual borrower transactions, we can decompose the growth in total mortgage origination due to the intensive margin (change in the average size of individual loans) from the extensive margin (the number of new loans that are originated in a zip code). An additional conceptual problem arises when using zip code level data, since growth in zip code income (in this case IRS is income), confounds the income of new homebuyers with the stock of the average income of the residents in an area.

We provide three main new findings about the relationship between credit origination and income in the run-up to the financial crisis. First, when we relate individual mortgage size to per capita income, measured either using borrower income from mortgage applications or average household income from the IRS, we see that the growth in individual mortgage size is strongly positively related to income growth throughout the pre-crisis period. This means that there was never a decoupling of mortgage growth and income growth at the *individual* level, the relevant measure of leverage at loan origination. The apparent decoupling of zip code level credit growth and per capita income is driven entirely by a negative relationship between the *number* of new originations and per capita income growth. That means that the pace of mortgage origination increased but not the leverage of the average household (measured in terms of debt-to-income, or DTI). Importantly, this negative correlation is concentrated in high income zip codes that saw fast per capita income growth and

[2] 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 et al. 2010, Demyanyk and Van Hemert 2011, Dell'Ariccia, Igan and Laeven, 2012 Agarwal, Amromin, Ben-David, Chomsisengphet, and Evanoff (2014) or Landvoigt, Piazzesi and Schneider, 2014). The focus of this paper is, instead, on understanding the relationship between credit and income along the whole distribution of borrowers.

moderate mortgage growth. For the bottom 75 percent of zip codes, the relationship between growth in dollar volume of originations and per capita income growth is positive.

Second, we show that total mortgage credit tracked growth in buyer income, which is a more appropriate measure of income for the lending decision. We show that buyers have very different income levels from the average residents in a zip code, and that adjusting for this difference overturns the negative correlation between total credit growth and income. This happens because new home buyers had increasingly higher income levels than the average household living in these areas. In particular, neighborhoods that experienced strong house price growth see a rise in average mortgage size, but again at DTI levels close to previous periods, since the average income of these buyers also went up significantly.[3]

Third, we document how these correlations translate into changes in the aggregate mortgage origination in the U.S. across borrower income levels. The large majority of mortgage dollars originated between 2002 and 2006 are obtained by middle income and high income borrowers (not the poor). While there was a rapid expansion in overall mortgage origination during this time period, the fraction of new mortgage dollars going to each income group was stable. In other words, the poor did not represent a higher fraction of the mortgage loans originated over the period. In addition, borrowers in the middle and top of the income distribution, as well as those with credit scores above 660, are the ones that contributed most significantly to the increase in the dollar value of mortgages in default after 2007. Taken together, the evidence in the paper suggests that there was no decoupling of mortgage growth from income growth where unsustainable credit was flowing disproportionally to poor borrowers.

Using data on individual mortgage applications from the Home Mortgage Disclosure Act (HMDA) between 2002 and 2006, this paper decomposes the relationship between income and mortgage growth into the effects at the intensive margin (changes in average mortgage size) versus the extensive margin (growth in the number of mortgages originated) within a zip code. We show that changes in average mortgage size within a zip are strongly positively correlated with per capita income growth during the 2002-2006 period. We re-run this analysis using individual mortgage transactions (rather than zip code averages) and confirm that these results hold very strongly also at the individual level. In contrast, we find that it is the *number* of mortgages originated within a zip code that was negatively correlated with average household income growth in the zip code. This result is driven mainly by the top quartile of zip codes based on income per capita, where rapid income growth was not accompanied by similarly fast mortgage growth.

We then also distinguish the growth in average household income at the zip code level (from the IRS) and the income of individual borrowers (from HMDA). We find a strong positive relationship between total credit growth and the individual borrower income growth, but a negative relationship with the growth in average household income at the zip code level (the result highlighted in Mian and Sufi, 2009). This suggests that there was no decoupling of total credit growth and *borrower* fundamentals during this time period. In addition, when we look at a longer time 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.[4]

---

[3] These results are consistent with those in Gyourko, Mayer and Sinai (2013), who also show a trend of higher income in cities experiencing large house price increases.

[4] Our tests also show that the coefficient of aggregate mortgage growth on zip code income is only negative if we control for county fixed effects (as proposed in Mian and Sufi (2009)). Without the county fixed effects the coefficient on average household income is positive throughout. However, if the aim is to test if credit is increasingly allocated to zip codes with declining incomes, one should not include a county fixed effect. The analysis with county fixed effects only tests if within a given county, zip codes that are growing quicker than the county average are disproportionately receiving more credit. However, all the first order change in credit allocation might happen between counties, which would be lost by this analysis. We show results with and without fixed effects throughout the paper.

To relate the results on relative growth in originations to the aggregate housing market, we also look at how mortgage credit origination was distributed across different levels of the income distribution during this time period. We find that the dollar value of mortgage origination is disproportionally concentrated in the top income deciles throughout the period, since higher income individuals typically obtain larger mortgages. Similarly, the majority of credit goes to high and middle income neighborhoods across the country, and the shares of mortgages by income level were very stable throughout the boom. These results again suggest that there was not a significant reversal in the flow of credit to lower income households. Using data from Lender Processing Services (LPS), we also analyze the contribution of borrowers at different levels of the income distribution to total mortgage dollars in default. We find that zip codes at the top of the buyer income distribution contributed disproportionally to the total dollar amount of mortgages in delinquency in the crisis. Additionally, in the peak of the financial crisis there is a large increase in the share of defaults by borrowers with a credit score above 660. While in previous years most defaults came from borrowers with low credit scores, during the financial crisis we document a dramatic increase in the proportion of mortgages in default from borrowers with high credit scores. In fact, we see that a major fraction of these defaults are for borrowers with high income and high credit scores (above 660).

A central concern in interpreting our results could be that they might be driven by aggressive overstatement of reported income. A number of recent studies have shown that misreporting of borrower characteristics increased significantly during the pre-crisis period (see, for example Jiang, Nelson and Vytlacil, 2014 for a careful analysis of this phenomenon). To rule out that this effect is driving our results, we conduct a number of tests: First, we show that there is an equally strong positive relationship between mortgage growth and IRS income when using the correct measure of individual leverage rather than zip code averages. Second, we repeat the analysis using new buyer income and show that the result is equally strong for agency and non-agency loans (i.e., those that were not purchased by one of the government sponsored enterprises, the GSEs). Since loans purchased by the GSEs adhered to much stricter underwriting standards even during the boom period, overstatement is less of a concern for this sample. The same is true when we break out the data by prime and subprime lenders. We also find that the magnitudes of overstatement that have been documented in the literature are too small to explain our results. The best estimates of the overstatement (Jiang et al, 2014) are around 20% to 25% for low documentation or no documentation loans, themselves a small fraction of all loans originated in this period (about 30%).[5] However, the relevant difference in new buyer income and zip code average income in our analysis is 75% and above. In addition, we document that the difference between buyer income and IRS income is most significant in lower income zip codes that make up less than 6% of the market. While we are, of course, not arguing that income misreporting did not increase during the run up to the crisis, this evidence shows that it is not a first order factor to explain any of the patterns we focus on in this paper, including the manner in which leverage and subsequent defaults built up in the economy.

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 re-run our tests using only refinancing transactions from HMDA, as well as data from LPS, which includes cash-out refinances and second liens, and confirm that borrower income growth and these types of credit were positively correlated and this correlation did not change significantly throughout the run-up to the crisis. Also, cash-out refinances and second lien loans were concentrated in middle class and upper middle class borrowers, just like purchase mortgages. This again suggests that even home equity loan growth and income growth did not become significantly decoupled over the pre-crisis period.

[5] See, e.g., Adelino, Gerardi and Willen (2013) Table 4.

These results provide a new picture of the mortgage boom. The aggregate increase in debt across zip codes (and nationally) was accompanied by an increase in individual borrower income levels, and was the result of an expansion of credit along both the intensive and the extensive margin: This suggests that home buyers increased the pace of home buying and therefore were holding more recent mortgages.[6] As a result of the increased churning, zip codes as a whole became more levered, since a larger fraction of households held mortgages which had recently been originated. The problem was not that levels of individual DTI at origination were grossly out of line with prior periods, but that a larger fraction of homeowners were close to their maximum debt capacity. These results suggest that house prices played a central role in both borrowers' and lenders' decision making.[7] For example, Cheng, Raina and Xiong (2014), Foote, Gerardi and Willen (2012), Shiller (2014), among many others, argue that buyers as well as investors in the mortgage market had overoptimistic beliefs about house price growth. Similarly, Chinco 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.[8] Our results are also related and consistent with Haughwout, Peach and Tracy (2008), Foote, Gerardi and Willen (2008), Mayer, Pence and Sherlund (2009), and Palmer (2014), who suggest that declining house prices were key for explaining increased defaults.

2. Data description

The analysis in this paper primarily uses data from three sources: the Home Mortgage Disclosure Act (HMDA) individual mortgage dataset, income data from the IRS at the zip code level, and a 5% random sample of all loans in the Lender Processing Services (LPS) data. The HMDA dataset contains the universe of mortgages applications in the US in each year. The variables of interest for our purposes are the loan amount, the applicant income, the purpose of the loan (purchase, refinance or remodeling), 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 tract from HMDA to zip codes using the Missouri Census Data Center bridge. This is a many-to-many match, and we rely on population weights to assign tracts to zip codes.[9] We drop zip codes for which census tracts in HMDA cover less than 80% of a zip code's total population.[10] With this restriction, we end up with 23,385 individual zip codes in the data.

IRS zip code income is obtained directly from IRS and represents the adjusted gross income of households that filed their taxes in a particular year in that zip code. Besides the 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.[11]

---

[6] In previous research, we show that new mortgages in the years before 2007 tended to be originated at a loan to value ratio of eighty percent or more, see Adelino et al (2013). In addition home ownership levels in the US did not increase during the 2002 to 2006 period.

[7] There is possibly also a complementary channel, namely that banks were becoming more likely to lend to people with more volatile income, even though the DTI and LTV ratios of these mortgages were reasonable at origination. Again, this would suggest that expectations about home prices were affecting lending decisions.

[8] 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.

[9] In other words, many zip codes will have more than census tract associated to them, and census tract could potentially overlap with more than one zip code. Missouri census tract to zip code bridge by population are obtained from http://mcdc.missouri.edu/websas/geocorr90.shtml and http://mcdc2.missouri.edu/websas/geocorr2k.html

[10] This drops only 180 zip codes out of 23,565.

[11] IRS zip code information is available at http://www.irs.gov/uac/SOI-Tax-Stats-Individual-Income-Tax-Statistics-ZIP-Code-Data-(SOI). 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).

The house price indices used in the paper are obtained from Zillow.[12] The zip code level house prices are estimated using the median house price for all homes in a zip code in June of each year. Zillow house prices are only available for 8,619 zip codes in the HMDA sample, representing approximately 70% of the total mortgage volume in the US during our sample period.

In order to identify subprime loans, we rely on the subprime and manufactured home lender list constructed by the U.S. Department of Housing and Urban Development (HUD) for the years between 1993 and 2005. This list includes lenders that specialize in these types of loans, and they are identified by a combination of features that include the average origination rate of loans by these lenders, the proportion of loans for refinancing, the share of loans sold to Fannie Mae or Freddie Mac, among others. [13] The data contains lender names, their agency codes, and lender identification numbers and we use these identifiers to match this list to HMDA and identify loans that were originated by subprime lenders.

We also use a dataset provided by Lender Processing Services (LPS, formerly known as the McDash dataset). This is a loan-level dataset that covers approximately 60 percent of the U.S. mortgage market and contains detailed information on the loan and borrower characteristics for both purchase mortgages and mortgages used to refinance existing debt. This dataset is provided by the mortgage servicers, and we use a 5% sample of the data. The LPS data includes 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.

Finally, we use the Saiz (2010) supply elasticity measure. This measure is constructed using geographical and local regulatory constraints to new construction and it correlates strongly with house price growth in the period of 2002 to 2007. This measure is available for 269 metropolitan statistical areas that we match to a total of 776 counties using the correspondence between MSAs and counties for the year 1999 provided by the Census Bureau.[14]

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 code income per capita. In the last three columns (columns 5 through 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 (labeled "Middle"). The sample is based on the 8619 zip codes that are part of the Zillow house price panel.

The first row shows the zip code average household income based on IRS adjusted gross income reporting as of 2002. The average household income is $50K 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 levels reported for the home buyers in HMDA (row 2), i.e. the individuals in those zip codes who actually took out a loan for buying a house (a purchase mortgage), we see that the average income of buyers is much higher than the average for their zip code, at $92K. This figure is about $143K for the highest quartile and $63K for the lowest quartile, which is almost twice the average income for the total population in these groups. The

[12] Zillow house prices are available at http://www.zillow.com/research/data/

[13] The whole 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) provide a detailed discussion on the advantages and disadvantages of the use of this list to identify subprime loans.

[14] This correspondence is available at http://www.census.gov/population/estimates/metro-city/a99mfips.txt and also http://www.census.gov/population/estimates/metro-city/a99nfips.txt for the New England Metropolitan Area Components used by Saiz (2010).

average original balance of mortgages (as of 2002) is also strongly increasing in the average zip code income. The average original mortgage balance is $155K, but it is $246K in the highest quartile and $97K in the lowest. In addition, the number of mortgages varies across income bins. There are 3.1 mortgages per 100 residents for zip codes in the high income quartile, while there are only 2.1 mortgages per 100 residents for those in the lowest quartile.

Table 1 also reports debt-to-income (DTI) calculated from HMDA as the mortgage amount over the reported income, and loan-to-value (LTV) across zip codes calculated as the average LTV from LPS, in 2003[15], since HMDA does not provide information about the value of the property that is being purchased. The DTI for the highest quartile is 2.26 while the ratio for the lowest quartile it is 1.97. Crucially, the change in DTI (shown in the second-to-last row of the table) between 2002 and 2006 is indistinguishable across income quartiles, which already suggests that there was no differential increase in individual leverage across rich and poor neighborhoods. LTV ratio in the highest quartile it is 0.73 and in the lowest 0.86. The next row shows the fraction of low documentation loans in 2002 across the different income quartiles. Interestingly, we see that buyers in richer neighborhoods (high household income) have a 20% likelihood of obtaining a low documentation loan in 2002, slightly higher than the remaining three (lower) income bins, with about 18% of all loans classified as being low (or no) documentation loans. In the last row on this table we show the change in the fraction of low documentation loans leading up to the crisis. Interestingly, the change in the fraction of low documentation loans was largest in the high income quartile of zip codes (a 21% increase), whereas there was only a 16% increase for the other three quartiles. We also report the Saiz elasticity measure for each of the subgroups. The average elasticity measure for the full sample is 1.7 but it is slightly lower for the highest income quartile at 1.4 and 1.9 for the lowest income quartile. As it has been shown in a number of other papers, the elasticity measure is strongly correlated with house price growth.

The last three columns 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, already at the beginning of the period there were 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 LTV, DTI or the fraction of low documentation loans between zip codes in different quartiles of house price growth.

In Table 1 we also report the growth rates of the main variables of interest, i.e. 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 high income zip codes and only 3.5% for the lowest income ones. However, 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. They all hover 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.

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 size, as well as the growth in the number of mortgages originated in an area. The growth rate is about 8% in the zip codes in the highest income quartile, while it is double this amount (16%) for

[15] The value is presented in 2003, given that the coverage in 2002 is not very comprehensive.

the lowest quartile. When we focus only on the change in the average mortgage size, we see that the growth in the highest quartile is about 7.5% while in the lowest quartile it is about 6.9%. This means that the average loan size increased during the 2002-2006 period, but the differential growth rate across higher and lower income areas was relatively limited. In the next row we see a much larger difference across areas when we consider the growth in the number of mortgages. The areas in the highest income quartile only see an annual increase in transactions of about 1%, while the lowest quartile has an increase of almost 10% annually. This suggests that the bulk of the increase in the total amount of mortgages originated in lower income areas is driven by the fact that these saw 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. [16] As we see in the second-to-last row of the table, this increase in the number of mortgages was not accompanied by a large increase in DTI for the low income zip codes. A similar picture emerges from Appendix Figure 2, where zip codes along the whole distribution show small increases DTI, where DTI comes from LPS and is computed by the lender (rather than by us). It is obtained as the sum of mortgage payments, insurance, and taxes divided by the monthly borrower income. A similar picture is shown in Jaffee (2009), where debt service to income figures show a similar modest rise.

When we consider neighborhoods with different levels of house price appreciation, changes in the average size of mortgages accounts 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 increased by much less (2.1%) for zip codes with small house price increases, as we would expect.[17] The growth in the number of mortgages is relatively similar across zip codes with high and low house price appreciation.

4. Mortgage credit and income

4.1 Intensive and extensive margins

In Table 2 we 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). The starting point is the same regression used in Mian and Sufi (2009), showed below:

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

In column (1) we repeat the regression in Mian and Sufi (2009) using the sample of zip codes with nonmissing house price data from Zillow. In particular, in column (1) the left-hand side of the regression is the annualized growth in the total dollar value of mortgage credit used for home purchase at the zip code level from 2002 to 2006. The right-hand side of all regressions in this table is the growth in the average household income obtained from the IRS in that zip code (which we also refer to as per capita income growth).

Column 1 of Table 2 replicates the main result on mortgage credit growth and income growth from Mian and Sufi (2009), namely that credit and income growth were negatively correlated. We find a negative and significant coefficient of 0.182, which means that the total value of mortgage originations on average dropped by 0.182 percent for every 1 percent increase in income relative to

[16] This increase in the number of mortgages 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.

[17] Ferreira and Gyourko (2011) show that income growth has important explanatory power for local housing booms. The relationship between business cycles and real estate prices is an issue of some debate (see, for example, Leamer (2007) and Ghent and Owyang (2010). See Ghysels, Plazzi, Torous and Valkanov for a survey on predictability of real estate prices.

county averages (because of the county fixed effects). Our tests also show that the coefficient of aggregate mortgage growth on zip code income is only negative if we control for county fixed effects (as proposed in Mian and Sufi (2009)). Without the county fixed effects (Panel B of Table 2), the coefficient on per capita income growth from the IRS is positive. The remaining patterns follow those in Panel A. The relationship between average mortgage size and income is positive, and the number of loans has a negative and significant relationship with income growth.

In columns 2 and 3 of Panels A, B and C 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 asks whether leverage increased for the people who buy a home. This is the relevant specification for understanding whether the average buyer's mortgage size changed with per capita income. The second measure captures the extensive margin and asks whether more mortgages were originated in a given neighborhood.

The second column of both Panels A and B show that the relationship between growth in the average mortgage size and per capita income growth is strongly positive, both with and without county fixed effects. The point estimate is 0.239 with county fixed effects (meaning that average mortgage size grows by about .27 percent for every percentage point increase in per capita income. The estimate is significantly larger at 0.587 without county fixed effects. In the third column of both panels we show that the relationship with growth in the number of mortgages is, however, strongly negative. This is, then, the source of the negative correlation we observe when in column (1) of Panel A. We show below that this negative relationship is concentrated in the top quartile of the income distribution.

4.2 Buyer income vs. per capita income from IRS

In Table 3 we analyze the relationship between mortgage growth and applicant income from HMDA. As in Table 2, the sample includes the zip codes that have non-missing zip code level house prices from Zillow. As we discuss above, the relevant measure of borrower fundamentals is the income of the people who actually bought 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. Panel A includes county fixed effects to follow the approach in Mian and Sufi (2009). However, if the aim is to test if credit is increasingly allocated to zip codes with declining incomes, one should not include county fixed effects. The analysis with county fixed effects only tests if, within a given county, zip codes that are growing quicker than the county average are disproportionately receiving more credit. If all the first order changes in credit allocation happen between counties, this would be lost by this analysis. Panel B does not include county fixed effects, since by first differencing the dependent variable and the right-hand side variables we are already taking out zip code specific fixed effects. So, the relationship is estimated comparing changes in income and mortgage growth rates across all zip codes in the U.S. We show results with and without fixed effects throughout the paper, but we highlight that the sign on the coefficient often is reversed when we do not include the county fixed effects.

The base specification for these tests is:

$$\mathrm{g}_{02-06}\,(Mtg)_i = \alpha_0 + \alpha_1 * \mathrm{g}_{02-06}(\mathrm{BuyerInc_i}) + \alpha_2 * \mathrm{g}_{02-06}(PerCapitaInc_i) + \eta_{county} + \varepsilon_i$$

Where the left-hand side variable is, in turn, the annualized growth in total mortgage origination in zip code *i*, the annualized growth in the average mortgage size in zip code *i*, or the annualized growth in the total number of mortgages originated in that zip code. Similarly, the two right-had side variables of interest are the annualized growth in average buyer income (from HMDA) and the annualized growth in average household income from the IRS. We also add county fixed effects as discussed.

Column (1) of Table 3 shows that between 2002 and 2006 there is a positive relationship between the growth in total amount of 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. In column (2), we repeat this regression and include per capita income growth from the IRS. The reported coefficient on the growth in IRS income is -0.224 (significant at the 5% level), which is similar to the magnitude estimated by Mian and Sufi (2009). But, again, the coefficient on the individual buyers' income is strongly positive (0.376) and significant at the 1% level. Finally, in column (3) we add a measure for the growth rate of house prices at the zip code level and the estimated coefficients on the variables of interest do not change.

The dependent variable used in the first three columns measures the change in the total mortgage debt originated yearly at the zip code level. This variable is a combination of the change in the average size of mortgages and the number of mortgages originated. In columns (4) to (6) of Table 3 we use the growth in average mortgage size, similarly to Table 2. In column (5) we regress the growth in the average size of mortgages (the intensive margin) on the change in the average IRS zip code income and the income growth of borrowers, parallel to the regression in Column (2). The coefficient on the IRS average income growth is 0.208 and significant at the 1% level. The coefficient on the income growth for the buyers is similar in magnitude (0.276) and again very significant. In column (6) we add, as before, a measure of house price growth and the results again are unchanged. These results confirm that mortgage sizes grew proportionally with income throughout the pre-crisis period and were not decoupled from income growth. When we remove county fixed effects (Panel B), we see that the estimated coefficients on both the average IRS income growth, as well as the income growth for the buyers are positive. The coefficient on the IRS average income growth is 0.372 and significant at the 1% level. The coefficient on the income growth for the buyers is similar in magnitude (0.506) and again very significant.

In Columns (7) and (9) we repeat the same regressions as before, but use as the dependent variable the annualized growth rate of the number of mortgages originated in a given zip code (the extensive margin). In column (7) we find a positive relationship 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 results are qualitatively unchanged if we add county fixed effects in column (8) or when we control for house price appreciation in Column (9). Without county fixed effects we see a positive but insignificant coefficient on the measure of income growth of home buyers.

Overall, the results of the decomposition of total mortgage growth into average mortgage size and the number of mortgages supports the idea that there was a change in the composition of buyers relative to the existing residents of a zip code in the period before the financial crisis. Since the prior literature focused on zip codes as the unit of analysis, it was not able to differentiate the characteristics of the stock of residents from those of the flow of home buyers. Our results show that even in the run-up to the financial crisis there was a positive and significant relationship between the growth in average mortgage size and income growth of the home buyers, independent of how we measure income. So, there was no reversal relative to previous periods with credit disproportionately flowing to people with (relative) declining income. Below we explicitly test for differences relative to other time periods. Instead, the negative coefficient between mortgage growth and income growth at the aggregate level that has been shown by Mian and Sufi (2009) is driven by

the extensive margin. But the debt-to-income levels did not change differentially for rich and poor borrowers.

4.3 Full sample of HMDA zip codes

In Appendix Table 1 we replicate the regressions in Table 3, but we use the universe of all zip codes in HMDA. This increases the sample of zip codes by a factor of three. Given the importance of the composition effects we document in the previous section, we want to verify whether the results hold in the larger sample of zip codes. It is important to keep in mind that the majority of the zip codes added in this sample are much less densely populated and have fewer transactions than the zip codes in Zillow. In total, these zip codes make up only 30% of the annual volume of mortgages originated in the US. We follow the same analysis as in Panel A. Odd columns do not include county fixed effects, while even columns do. In column (1) we see that the coefficient from the regression of the growth in aggregate zip code mortgage origination on the growth in average household income is positive and significant using the whole HMDA sample. The coefficient on the income growth of the home buyers is positive and significant as before. Adding county fixed effect in column (2) reduces the magnitude of the coefficient on household income growth from the IRS but does not otherwise significantly change the results. This suggests that, in the wider set of zip codes, the relationship between growth in total mortgage credit and income was positive, independent of how we measure the growth in average income.

As before, we also decompose the aggregate credit growth into the average size of individual mortgages and the total number of mortgages. In columns (3) and (4) the dependent variable is the growth in the average mortgage size in a zip code between 2002 and 2006. The coefficients on both the average household income and the income of the buyers are positive and significant. When we include county fixed effects the coefficient on average IRS income growth drops by about 50%, while the coefficient on buyer income remains unchanged. Finally, in columns (5) and (6) we repeat the same regressions using the growth in the number of mortgages in a zip code as the dependent variable. The coefficients on both the average IRS income and the income of the buyers are positive and significant. These results show no decoupling of mortgage growth and income growth even on the extensive margin.

4.4 Cross sectional variation in house price increases

To understand whether the compositional changes of home buyers across neighborhoods are related to the house price run-up, we sort zip codes into quartiles based on the amount of house price growth they experienced in the period from 2002 to 2006. We repeat the previous regressions for each subgroup of house price growth in Table 4. We report the results for the high, middle and low quartiles, where, as before, the middle group combines the second and third quartiles. Panel A column (1) to (3) focuses on the aggregate growth in mortgage origination at the zip code level, combining the intensive and extensive margins. Columns (1) and (2) show that the negative relationship between the growth in mortgage origination and the growth in household income is concentrated in the three quartiles that have the highest growth in house prices. The coefficient for the highest and the middle groups are negative 0.370 and negative 0.283, respectively. As before, the coefficient on the growth of the buyers' income is positive and strongly economically and statistically significant. In contrast, the zip codes in the lowest quartile of house price growth show a strong positive relationship between zip code level income growth and mortgage growth (0.370), but a weaker relationship between mortgage growth and the change in income of buyers (0.194).

These findings highlight that the changing composition in the income of all residents relative to that of home buyers within a zip code was prominent in all areas where house prices were going up quickly. However, in neighborhoods where house prices did not go up by much, we do not see the

same strong divergence between the stock of residents and the flow of buyers. Of course, from these results we cannot establish a direction of causality: buyers with higher income than the average resident might have pushed up house prices in these areas, or vice versa, the rapid rise in house prices within an area might have caused the change in the composition of buyers, since only people with higher income could now afford to buy properties.

We again break out the aggregate mortgage amounts into the average mortgage size (Panel B) and the number of mortgages in a zip code (Panel C). We find a positive coefficient on the relationship between the growth in average mortgage size and the growth in average IRS incomes at the zip code level across all quartiles of house price growth. In addition, when looking at the growth in the income of the buyers, we again find a positive relationship throughout. However, when we consider the number of mortgages originated within a zip code (the extensive margin), we find a strong negative relationship between the number of mortgages and household income growth at the zip code level for the three quartiles with the highest house price growth (Panel C). But, again, the sign of the coefficient flips and is positive and significant for the lowest quartile. The relationship between the number of mortgage transactions and the growth of the income of the buyers is positive and significant as before.

These results confirm that the divergent growth in the income of the average resident and of home buyers was strongest for areas with rapid house price appreciation, and that the effect runs entirely through the expansion of the number of borrowers (extensive margin). In contrast, the mortgage size for the average buyer within a zip code is still strongly positively and significantly related to the income of the household (irrespective of how we measure income). This is consistent with a changing composition of households within neighborhoods.

4.5 Cross sectional heterogeneity in zip code income

Many observers of the housing market have been concerned that lending to lower income borrowers changed most dramatically over this time period. In this subsection we explore this dimension of cross-sectional heterogeneity in the data. In Table 5 we break out the data into quartiles based on the average income per capita in a zip code as of 2002. The analysis follows exactly the specifications of Table 4. Panel A, columns (1) through (3) of Table 5 show that the relationship is not identical across the different zip code income quartiles. Only the top quartile by income (column 1) shows a negative and significant coefficient on the measure of average IRS income growth (-0.239). For the lower three income quartiles in columns (2) and (3) we always find a positive relationship between mortgage and household income growth. The coefficient of total mortgage origination growth on average IRS household income growth in the middle income group (second column) is 0.160 and very significant, and the coefficient for the lowest income group in column (3) is similar at 0.163. The relationship between total mortgage growth and borrower income growth is positive and significant throughout (0.309, 0.282 and 0.575 for the highest, middle and lowest quartiles, respectively). Taken together, we do not find evidence that poorer zip codes were changing their leverage levels disproportionally relative to income growth. In fact, the relationship between credit and borrower income is strongest for lower income zip codes, and the negative relationship between average household income and credit is only negative for the zip codes with the highest income.

Since zip codes with higher house price growth usually also have lower average income (as documented in the descriptive statistics), Appendix Table 3 shows a "horse race" regression between different cross sectional variables interacted with the measures of income growth. This allows us to assess whether income level or house price growth better captures the cross-sectional variation in the relationship between mortgage growth and income growth. We also include interactions with a number of other zip code level variables that have been discussed in the prior literature, such as the Saiz (2010) measure of the elasticity of housing supply, the fraction of low

documentation loans in a zip code and the fraction of loans sold to Fannie Mae and Freddie Mac (the government-sponsored enterprises, or GSEs).

In this table we run the standard specification of growth in aggregate mortgage origination at the zip code level on growth in average zip code income and home buyer income, and then interact each of the income variables with the cross sectional measures. Column (1) replicates the same regressions as in Table 4 for the sample of zip codes for which we have the relevant cross sectional information. We see that the results are virtually unchanged relative to the prior sample. When we include the interaction terms in column (2) we see that the direct effect of zip code income growth and buyer income growth are both positive and significantly related to aggregate mortgage growth. However, the interaction terms of the dummy for high income areas with both average household income growth and buyer income growth are negative and very large (and significant at the 1% level). This suggests that, consistent with the previous tables, it is in the high income areas that the relationship between mortgage growth and income growth is weakest. In what follows we will show that this is the result of increased relative income growth in these areas compared to lower income zip codes in the US. We also find a negative coefficient on the interaction of house price growth and average zip code income growth, but a positive and very significant relationship with the growth in buyer income for the intensive margin. This suggests, as before, that in neighborhoods with higher house price growth, mortgage growth became less sensitive to the average zip code income, but it became more sensitive to the income of buyers. In columns (3) to (6) of Appendix Table 3 we again confirm that the change in sensitivity was mainly driven by the extensive margin, i.e. the number of new loans that were originated within the zip codes.

4.6 Longer run correlation

The previous sections show that there is no negative relationship between mortgage growth and borrower income growth, but it is still possible that the *slope* of the relationship was weaker in the pre-crisis period (2002-2006) than in previous periods. As a result, leverage might have become less closely tied to income levels or growth in income, suggesting that credit expansion might have deviated somewhat from fundamentals relative to other periods. To address this question, in Table 6 we explore how the relationship between mortgage growth and income growth changed over time.

In a first step we repeat our main regression from Table 3 but break it out by different time periods. We follow the example of Mian and Sufi (2009) and consider 4 sub-periods: 1996-98, 1998-2002, 2002-2006 and after 2006. The coefficient from the regression of growth in total mortgage origination on growth in average zip code income, is positive and significant for most time periods before the financial crisis, and it turns negative in the 2002 to 2006 period. However, as we have argued before, this is not the relevant income measure for assessing borrower fundamentals. When we look at the coefficient of mortgage growth on the growth in the income of the buyers, the coefficient is positive and significant across all the time periods, and, importantly, it does not become flatter in the pre-crisis period. If anything, the coefficient goes from 0.240 in the 1998-2001 period to 0.376 in 2002-2006, which means that the sensitivity of mortgage growth to income growth actually increased prior to the crisis. In Panel B of Table 6 we repeat the same regressions for the full HMDA sample. When we consider all zip codes in the data, the coefficients on both the buyer income and the average zip code household income are remarkably stable over time, with somewhat lower magnitudes in 1998-2002, but otherwise very similar across time periods.

4.7 Panel data specification

We now go a step further and expand the previous regression set up to a full panel specification. So far, we have related mortgage growth over the 2002 to 2006 period to contemporaneous growth in income. Here we turn to the full annual panel of zip code level mortgage origination to look at the relationship between mortgage growth and annual measures of household and buyer income. This is an alternative way to assess whether the slope of this relationship changed over time.

In Table 7 we now use the following specification,

$$\mathrm{Ln}(Mtg_{it}) = \alpha_0 + \Sigma_i \alpha_i \left[\mathrm{Ln}(BuyerInc)_{it} * Y_t\right] + \Sigma_j \alpha_j [\mathrm{Ln}(ZipInc)_{it} * Y_t] + FE_t + FE_i + \varepsilon_{it}$$

The independent variables are the logarithm of the income of buyers interacted with a full set of dummies for all years in the sample (denoted Y_t), the logarithm of the average income of households in the zip code also interacted with a set of year dummies, FE_t is a year fixed effect, and FE_i 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 within zip codes over time.

In columns (1), (3) and (5) we simply include the logarithm of the income variables (and not the full set of interactions with the year dummies) to compare the results in this panel setup to the previous tables. The dependent variable in column (1) is the aggregate mortgage origination in a zip code in a given year. The results show that the coefficients on the income of buyers and on the IRS average 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 buyer income but also to the income of existing buyers in a zip code. The effects are slightly less significant but go into the same direction when using the number of loans per zip code and year.

Given the panel structure of our data we can now analyze the dynamics over the run-up to the financial crisis, i.e. whether the sensitivity between mortgage level and the income variables changed over the run-up to the financial crisis. The omitted year is 2002, which is the base year in our sample. Column (2) shows an interesting pattern. While 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 relationship between the growth in mortgage origination and the growth in average household income from the IRS became flatter over time. However, the relationship between aggregate mortgage origination and buyer income in a given zip code is strongly positive and the interactions with the year dummies are positive and significant. This means that the combined effect increases from 0.231 in the base year (2002) to about 0.434 in 2005 and back to 0.271 in 2006. This finding reinforces our earlier findings that the growth in zip code leverage became more closely tied to the change in the income of the buyers (and less to that of the average household in a neighborhood).

We again decompose this effect into the average mortgage size and the number of mortgages. Column (4) shows that the estimated relationships are very similar to the ones for the aggregate mortgage levels when we use the average mortgage size as the dependent variable. The relationship with buyer income becomes stronger over the time period and, in contrast, the relationship with zip code level household income becomes weaker over that time period. The effect on the extensive margin (the number of mortgages, shown in column 6) is much noisier and again we see that the

number of new mortgages in an area becomes more negatively correlated with household income over time.

5 Individual leverage

5.1 Aggregation by income deciles

The zip code-level analysis shows that credit did not disproportionately flow to *borrowers* with declining income, but rather that along the extensive margin *zip codes* where the average household income saw a relative decline there was an increase. Ultimately, however, it is not neighborhoods who obtain (and are responsible for) loans, but rather individual home buyers. We now consider how this compositional change we document across neighborhoods affects where credit built up in the US during the 2002-2006 period. For this purpose, we use individual transaction level data from HMDA (rather than the aggregates at the zip code level).

Panel A of Table 8 shows the average size of mortgages obtained by borrowers in each income decile based on applicant income. The table shows that mortgage size grew significantly over this time period for all income groups. At the bottom of the income distribution, we see that the average mortgage balance at origination went from around $74K in 2002 to $85K in 2006 for the lowest income decile. Similarly, in the middle income deciles average loan amounts grew during the same period from $146K to 183K (decile 6). For the top decile, the average size of purchase mortgages went from $351K to a remarkable level of over $442K. This breakdown of the data highlights an interesting fact that has not received enough attention, which is that the largest increases in the size of mortgages were in the middle income and, in particular, the high income borrowers, not for the lower income buyers.

In Panel B of Table 8 we calculate the average debt-to-income ratio (calculated as the mortgage balance divided by applicant income) for mortgage holders in each of these income deciles across time. We compute debt-to-income as the ratio of the mortgage balance at origination divided by the applicant income. Not surprisingly, when we look across income deciles, we see that poorer households are significantly more levered than richer ones even in 2002. The average household in the lowest decile (1) has a first mortgage DTI of 3.0 while the average mortgage holder in the top income decile (10) only has a DTI of 1.32. In line with our prior analysis, we see that DTI levels measured in HMDA do not change much over the time period from 2002 to 2006. In deciles (2) through (10) the DTI is virtually constant or even slightly declining, e.g. from 2.17 to 2.01 in income bin (5). In the lowest income group the DTI went slightly up, but even there it moved from from 3.07 to 3.24. Importantly, DTI did not become differentially higher for low income borrowers than for high income ones. The same data is plotted in Appendix Figure 1, where it is immediately obvious that DTI levels are remarkably flat between 2002 and 2006. In Appendix Figure 2 we show the debt-to-income 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. Consistent with the results in Table 8, the figure shows that the increase in DTI is relatively modest, and that borrowers at all income levels move in lockstep.

In Panel A of Figure 1 we break down the total dollar volume of mortgages originated for home purchase in each year (from HMDA) by the quintile that each borrower falls into based on their applicant income (the same data is shown by deciles in Panel C of Table 8). We sum the mortgage amounts of all the households within an income quintile and divide this number by the total amount of mortgage debt originated in the US in a given year. This picture dramatically highlights the idea that credit was flowing predominantly to richer households: 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 total mortgage credit, whereas the bottom quintile accounts for about 11%. The picture using IRS income

per capita to form quintiles shows a similar pattern. The top quintile shows a small reduction in its share of total dollar value of originations from 35% of the total to 30%, and this reduction is accounted for by increases of 1 to 2 percentage points by the quintiles below. This is consistent with the results from the correlation of growth in income per capita and mortgage growth, as it is the top of the distribution that exhibits a rapid increase in income and a (relative) reduction in the number of loans. This again clearly shows that there was no significant decoupling of the distribution of credit from applicant income throughout this period, and that the share of originations is very similar in 2006 to what it was in 2002 across income levels.

5.2 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, i.e. 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 9 we use the following specification:

$$\begin{aligned}\mathrm{Ln}(Mtg_{it}) &= \alpha_0 + \alpha_1\ \mathrm{Ln}(BuyerInc)_{it} + \alpha_2\,\mathrm{Ln}(Census\ Tract\ Inc)_{it} + FE_t \\ &+ FE_{census\ tract} + \varepsilon_{it}\end{aligned}$$

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 HMDA. The independent variables are, again, the logarithm of the buyer's reported income and 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 tract over time.

Table 9 shows the results for this loan-level specification. Consistent with the zip code level regressions, both the coefficients on buyer income and 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 drawn from the zip code level analysis, and are supportive of the idea that credit supply did not decouple from income during the boom period.

6 Robustness

6.1 Misreporting of applicant income

One important consideration in the run-up to the financial crisis was that lenders started to misreport income levels of prospective borrowers in order to justify higher leverage levels than these borrowers would normally be able to afford. It is important for the purposes of our study to rule out that changes in the reporting of income itself could be the source of the strong relationship between

buyer income and total mortgage growth we find in all specifications, as this might hide a de facto increase in household leverage levels.[18]

We use a few different approaches to analyze whether this is a first order concern for our findings. First, in Panel A of Table 10 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) the GSEs had to conform to higher origination standards than those sold to other entities, and thus were less likely to have unverified applicant income.[19] The idea in these tests is to see whether zip codes with a lower fraction of loans sold to the GSEs are the ones that exhibit a stronger relationship 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.

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 and the fraction originated by subprime lenders (if anything, the coefficient is smaller for the zip codes with a relatively low fraction of GSE loans).

In Panel B we repeat the individual-level regressions in Table 9 split also by loans that are later sold to the GSEs, and also by loans originated by subprime lenders versus those originated by non-subprime lender. Again, the coefficient on buyer income is very stable and statistically identical across all subsamples, consistent with the notion that buyer income misreporting is not driven the results.

6.2 Refinancing

While the focus of the previous results linking income and mortgage origination has been on loans used for home purchase, it is possible that refinancing transactions, and especially cash-out re-financing transactions, became more important over the pre-crisis period and could have altered the indebtedness of low income relative to high income households. Increasing house prices and a growing willingness of lenders to provide such refinancing arrangements made it easier to unlock equity that borrowers had in their homes. It is possible that this happened disproportionately to income over this period, so it is important for us to consider this margin. The downside of using HMDA data for this analysis is that we only have an identifier for whether a transaction is a refinancing transaction (as opposed to a purchase transaction) but we do not know if it was used to change the existing mortgage level on the property (i.e., a cash-out refinance) or just to reset the interest rate without any change in the loan size. Only a cash-out refinancing would increase the leverage level of the household, so we would ideally want to restrict the analysis to cash-out loans only.

[18] There is also evidence of other forms of misreporting during this time period, including the value of transactions (Ben-David, 2011), or mortgage quality in contractual disclosures in the secondary market (Piskorski, Seru and Witkin, 2013 and Griffin and Maturana, 2014). These forms of misrepresentation do not, however, influence the analysis in this paper.

[19] Previous work, including Pinto (2010) have noted that origination standards for the 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.

In Panel A of Table 11 we repeat the regressions in Panel A of Table 3, but we focus on the growth in the income of borrowers who are engaged in refinancing transactions. The picture that emerges is very similar to the new mortgage transactions. 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 (-0.579), but an economically and statistically large and positive effect on the change in buyer income (1.113). In columns (3) and (4) we decompose the aggregate effect into the average mortgage size and the number of mortgages as before. The estimated coefficient on IRS income and borrower income growth are positive and significant when we look at the change in the average mortgage balance without fixed effects, and negative but very small when county fixed effect are included. The results for the number of transactions is very similar to purchase mortgages: we find negative coefficients on IRS income growth, and positive but smaller coefficients on the borrower income growth.

Panel B of Table 11 implements a zip code level panel regression similar to the one in Table 6, but only for refinancing transactions. The coefficients show that the relationship between refinancing mortgage growth and borrower income becomes steeper (not flatter) for the later years (2004, 2005 and 2006), and that it becomes progressively flatter for the IRS income measure. Almost all of this variation in the slope of these relationships is coming from the extensive margin (the number of transactions) rather than from changes in the average mortgage balance.

Because HMDA does not distinguish between regular refinancing transactions and cash-out refinancing, we turn to data from LPS and run loan-level regressions similar to the ones in Table 9.[20] We measure zip code income using IRS data, and we use the average borrower income obtained from HMDA (LPS does not include borrower income information, so we merge borrower income at the zip code and year level). Table 12 shows these regressions for both purchase mortgages and cash-out refinancing transactions. The first message from the table is that the results using LPS data are very similar to those using HMDA shown in Table 9. Both borrower income and average household income are strongly positively correlated with mortgage size, and this relationship becomes weaker over the 2002-2006 period for IRS household income, but it becomes stronger for borrower income. The second message from the table is that the results are almost indistinguishable for cash-out refinancing transactions. Both income measures are strongly positively associated with the size of cash-out refinancing transactions, and that relationship evolves in a similar way as for purchase transactions over time.

7 Distribution of mortgage delinquency

In this final section we consider how the distribution of mortgage credit compares with that of mortgage delinquency. Much of the prior literature has focused on the fact that delinquency rates are higher for lower quality borrowers, but here we show simple summary statistics on the dollar volume of credit that is past due (as opposed to the rate of loans that are delinquent as a fraction of the loans that are originated).

Panel A of Figure 2 uses data from LPS to first show the distribution of mortgage origination by applicant income quintile (similarly to Figure 1 that uses data from HMDA). We again restrict our attention to zip codes with nonmissing Zillow house price data. Because LPS does not provide individual level income, we use the average of applicant income from HMDA by zip code to create these figures. As before, the distribution of credit stable, and the proportion by quintile closely matches the proportions shown in Figure 1. Panel B shows the contribution by applicant income quintile to the total dollar value of mortgages that are 90 days delinquent or more during the subsequent three years by origination year. The dollar value of delinquent mortgages is simply

[20] We only have access to a 5% sample of the LPS data, which makes the data unsuitable for zip code level analysis.

calculated as the sum of the origination amount for mortgages that become delinquent at some point during the first 36 months after origination. Contrary to popular belief, it is the middle and top quintiles of zip codes based on applicant income that substantially increase their weight in the pool of delinquent mortgages over time. Of course, the total dollar value of mortgages that are delinquent went up dramatically for mortgages originated in 2006 relative to those originated in 2002, but clearly this is not driven primarily by low income borrowers. In Panels C and D we repeat the same summary statistics by income per capita from the IRS. The shares of each quintile are qualitatively similar to those shown in Panels A and B, and reiterate the message that it is the top quintiles that increase their share of delinquencies, not the bottom ones. They also strongly reject the notion that the mortgage delinquency crisis was concentrated in the lowest zip codes by income.

Figure 3 focuses on 2006 and compares different types of mortgages: first mortgages taken out with the purpose of purchasing a home, second liens, and cash-out refinance loans. The purpose of this figure is to ask whether the message from Figure 1 and Figure 2 is likely to be very different for mortgages other than purchase mortgages. Panel A shows that the dollar distribution of all three product types are remarkably similar when we break zip codes out by quintiles based on borrower income. The top quintile shows 5 and 1 percentage points lower weight respectively for second liens and cash-out refinances relative to purchase mortgages, and this can be accounted for by 1 or 2 percentage points higher weights in the quintiles below. When we consider the distribution of delinquent mortgages (measured in dollars), the distribution again looks very stable for all products, again with a higher weight of the first and second quintiles of borrowers. 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 applied to mortgages other than purchase mortgages.

Figure 4 shows the same breakdown of mortgage origination and delinquency for borrowers that have a credit score above and below 660, a cutoff often associated with subprime borrowers (although, as Mayer and Pence (2009), the precise definition of subprime is less straightforward). Panels A and B show that the fraction of total credit going to borrowers with a credit score below 660 is fairly stable across all quintiles. In fact, borrowers with a score below 660 represented about 17% of all originations (across all quintiles), and this fraction grew to about 19% in 2006. The patterns for delinquencies are, however, much more dramatic. Borrowers *above* 660, despite making up 83% of all originations, only accounted for 29% of the dollar value of mortgages originated in 2003 that became delinquent, which is the reflection of very low default rates. For mortgages originated in 2006, however, they made up 61% of the total dollar value of delinquencies.
The patterns are even more striking when we focus on the highest income quintile of zip codes. In 2003, borrowers with a credit score above 660 in the highest income zip codes in the country made up 31% of originations but only 4% of all delinquencies (both measured by dollar value). By 2006, they represented 29% of all originations, but a full 13% of delinquencies, a tripling of their share of the total. Panels D and E show very similar patterns for mortgages in foreclosure. 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.

8 Conclusion

This paper shows that there was no decoupling of mortgage credit growth and income growth during the period before the financial crisis. Instead, mortgage credit and income move in the same direction when we focus on individual mortgage growth and income growth. Additionally, we document that total mortgage credit moved in line with borrower income throughout the pre-crisis period. However, borrower income was becoming increasingly higher than the zip code level average, and especially neighborhoods with increasing house prices also saw significant increases in borrower income. This suggests that changes in borrower composition are important in understanding the relationship between credit growth and borrower fundamentals during this time period. The fact that DTI did not change differentially across rich and poor borrowers suggests that there were no severe distortions towards poor or low-income people in the way banks allocated capital at loan origination.

Instead, our results show that the dramatic rise in credit before the financial crisis was mainly driven by increases along the extensive margin, and that credit increased proportionally across all income levels, so that the distribution of mortgage credit across income quintiles was stable. This means that even in 2006 high and middle income borrowers accounted for the overwhelming majority of credit that was originated in the mortgage market, since these middle-class borrowers take out much larger (and more) loans than the poor. The increased house buying activity (churning) led, in the aggregate, to a larger fraction of potential buyers to be levered up to their maximum level of DTI and possibly LTV. Once the crisis hit, we see that high and middle income borrowers accounted for the majority of dollars of credit in delinquency, especially in areas where house prices dropped. Since these borrowers have much larger mortgages, a small increase in their default rate has large impact on the amount or dollars in delinquency. These delinquencies were then exacerbated in zip codes with a large fraction of owners that are highly levered, since there is less financial slack to buy homes from people who have to sell. As a result, there might have been an asset fire sale type of externality that led to (downward) pressure on prices. In this sense, the main change in the financial market was not that low income individuals became unsustainably levered relative to higher income borrowers. But rather that lenders and borrowers bought into high house price expectations and ignored potential equilibrium effects from a large fraction of borrowers being levered to their maximum DTI. A significant fraction of these home buyers later sold or foreclosed on mortgages when the option value of higher house price appreciation was not realized.

REFERENCES

Adelino, M., Gerardi, K., and Willen, P. S. 2013. Why don't Lenders renegotiate more home mortgages? Redefaults, self-cures and securitization. Journal of Monetary Economics, 60(7), 835-853.

Agarwal, Sumit, Gene Amromin, Itzhak Ben-David, Souphala Chomsisengphet, and Douglas D. Evanoff. "Predatory lending and the subprime crisis." Journal of Financial Economics 113, no. 1 (2014): 29-52.

Ben-David, Itzhak, 2011, "Financial Constraints and Inflated Home Prices during the Real-Estate Boom." American Economic Journal: Applied Economics 3(3), 55–78.

Cheng, I., Shail Raina, and Wei Xiong. "Wall Street and the Housing Bubble" American Economic Review (2014): 104(9)

Chinco, A., and Mayer, C. 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 (2008): 272-290.

Corbae, Dean, and Erwan Quintin, 2013, "Leverage and the foreclosure crisis". Journal of Political Economy, forthcoming.

Dell'Ariccia, G., Igan, D., and Laeven, L. 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 Van Hemert, O. 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.

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.

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 – 45

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"

Glaeser, Edward L., Joshua D. Gottlieb, and Joseph Gyourko, 2013, "Can Cheap Credit Explain the Housing Boom?" Housing and the Financial Crisis (2013): 301.

Ghent, Andra C., and Michael T. Owyang. 2010. "Is housing the business cycle? Evidence from US cities." Journal of Urban Economics 67, no. 3 (2010): 336-351.

Ghysels, Eric, Alberto Plazzi, Walter N. Torous, and Rossen Valkanov. 2012. "Forecasting real estate prices." Handbook of Economic Forecasting 2.

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.

Gyourko, Joseph, Christopher Mayer and Todd Sinai, 2013. "Superstar Cities." American Economic Journal: Economic Policy, vol. 5(4), pages 167-199.

Haughwout, Andrew, Richard Peach, and Joseph Tracy, 2008, "Juvenile delinquent mortgages: Bad credit or bad economy?," Journal of Urban Economics, 2008, 64 (2), 246 – 257.

Jaffee, Dwight M., 2009, "The US subprime mortgage crisis: issues raised and lessons learned." Urbanization and Growth (2009): 197.

Jiang, W., Nelson, A. A., and E. Vytlacil, 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., Mukherjee, T., Seru, A., 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.

Loutskina, E., and Strahan, P. E. (2009). Securitization and the declining impact of bank finance on loan supply: Evidence from mortgage originations. The Journal of Finance, 64(2), 861-889.

Leamer, Edward E. 2007. "Housing is the business cycle." Proceedings - Economic Policy Symposium - Jackson Hole, Federal Reserve Bank of Kansas City, pages 149-233.

Mayer, Chris, and Karen Pence, 2009, "Subprime Mortgages: What, Where, and to Whom?" in Glaeser, Edward L., 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 Vol. 23 (1), pages 27–50

Mian, A., and Sufi, A. 2009. The consequences of mortgage credit expansion: Evidence from the US mortgage default crisis. The Quarterly Journal of Economics, 124(4), 1449-1496.

Nadauld, T. D., and Sherlund, S. M. 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 D.C.

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.

Saiz, A. 2010. The geographic determinants of housing supply. The Quarterly Journal of Economics, 125(3), 1253-1296

Shiller, Robert J. 2014. "Speculative asset prices." American Economic Review 104.6 (2014): 1486-1517.

Table 1. Summary Statistics

This table reports summary statistics for all counties in the sample that have nonmissing zip code level house prices from Zillow. Column 1 shows the summary statistics for the entire sample. Columns 2 to 4 show the summary statistics for the zip codes in the highest quartile of average income per capita in 2002 (High), second and third quartiles of average income per capita (Intermediate), and lowest quartile of average income per capita (Low). 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). *Zip code income* is the average adjusted gross income per capita by zip code from the IRS. *Buyer Income* is the average applicant income by zip code from HMDA. *Average mortgage size* is the total mortgage amount originated in a zip code used for home purchase in 2002. *Number of mortgages originated per 100 residents* is the average number of mortgages originated per 100 residents by zip code. Debt-to-income is the zip code average ratio of the mortgage balance at the time of origination over applicant income from HMDA. *Loan to value* is the average loan to value (LTV) in a zip code calculated from LPS. *Fraction of low documentation loans* is the fraction of low documentation loans originated in a zip code from LPS. *Population* is the average zip code-level population estimated from IRS returns. *Elasticity of housing supply* is the Saiz (2010) measure of housing supply elasticity at the MSA level.

Zip code income growth refers to the percentage change between 2002 and 2006 in household adjusted gross income by zip code from the IRS. *Buyer income growth* is the percentage change in average applicant income by zip code from HMDA. *Growth in total mortgage origination* refers to the percentage change in total mortgage credit originated in a zip code used for home purchase between 2002 and 2006 calculated using HMDA. *Growth in average mortgage size* refers to the percentage change in the average balance of individual mortgages in a zip code between 2002 and 2006 (also calculated using HMDA). *Growth in number of mortgages originated* is the percentage change in the number mortgages originated between 2002 and 2006. *Change in debt-to-income* and *Change in loan to value* is the change in Debt to Income and LTV by zip code between 2002 and 2006.

	Whole sample	Zip code household income, 2002			Zip code house price growth, 2002-2006		
		High	Middle	Low	High	Middle	Low
	N =8619	N =2088	N =4346	N =2185	N =2020	N =4407	N =2192
Zip code household income (USD thousands), IRS, 2002	50.93	84.81	44.75	30.85	47.40	54.44	47.13
	(28.24)	(39.42)	(5.92)	(3.92)	(25.45)	(30.41)	(25.08)
Buyer income (USD thousands), HMDA, 2002	92.18	143.75	82.27	62.62	99.83	95.11	79.24
	(67.26)	(98.40)	(46.87)	(24.85)	(70.94)	(70.58)	(53.87)
Average mortgage size (USD thousands), 2002, purchase mortgages only	154.93	246.37	139.95	97.33	160.97	166.79	125.50
	(86.70)	(113.33)	(46.49)	(36.46)	(76.74)	(95.63)	(67.57)
Number of mortgages originated per 100 residents, 2002, purchase mortgages only	2.60	3.09	2.64	2.07	3.38	2.36	2.37
	(2.16)	(3.12)	(1.78)	(1.52)	(3.42)	(1.53)	(1.47)
Debt to income, 2002	2.13	2.26	2.16	1.97	2.18	2.17	2.03
	(0.38)	(0.35)	(0.35)	(0.41)	(0.36)	(0.39)	(0.36)
Loan to value, LPS, 2003 (N=13,555)	0.80	0.73	0.82	0.86	0.80	0.79	0.83
	(0.11)	(0.10)	(0.09)	(0.09)	(0.10)	(0.11)	(0.10)
Fraction of low documentation loans, LPS, 2003 (N=13,555)	0.18	0.20	0.18	0.17	0.20	0.18	0.17
	(0.23)	(0.23)	(0.23)	(0.24)	(0.22)	(0.24)	(0.23)
Population (000s), IRS, 2002	142.13	152.90	141.77	132.55	166.16	132.44	139.45
	(113.67)	(110.27)	(114.96)	(113.44)	(128.36)	(109.75)	(103.37)
Elasticity of housing supply (N=11887), Saiz (2010)	1.72	1.41	1.78	1.93	1.24	1.65	2.22
	(0.88)	(0.71)	(0.88)	(0.93)	(0.51)	(0.84)	(0.91)
Zip code income growth (annualized), IRS, 2002-2006	0.046	0.064	0.042	0.035	0.053	0.047	0.036
	(0.028)	(0.035)	(0.022)	(0.021)	(0.029)	(0.027)	(0.025)
Buyer income growth (annualized), HMDA, 2002-2006	0.065	0.068	0.062	0.068	0.108	0.062	0.032
	(0.061)	(0.063)	(0.058)	(0.064)	(0.066)	(0.050)	(0.052)
Growth in total mortgage origination (annualized), 2002-2006, purchase only	0.121	0.078	0.119	0.168	0.170	0.123	0.074
	(0.148)	(0.141)	(0.143)	(0.151)	(0.165)	(0.138)	(0.136)
Growth in average mortgage size (annualized), 2002-2006, purchase only	0.067	0.075	0.062	0.069	0.124	0.063	0.021
	(0.054)	(0.052)	(0.051)	(0.059)	(0.042)	(0.040)	(0.038)
Growth in number of mortgages originated (annualized), 2002-2006, purchase only	0.055	0.007	0.057	0.096	0.046	0.059	0.054
	(0.129)	(0.131)	(0.124)	(0.121)	(0.144)	(0.126)	(0.119)
Change in debt to income, 2002-2006	-0.004	0.008	-0.011	-0.001	0.063	0.016	-0.107
	(0.286)	(0.275)	(0.280)	(0.306)	(0.293)	(0.274)	(0.276)
Change in loan to value, 2003-2006	-0.017	0.008	-0.022	-0.034	-0.036	-0.010	-0.013
	(0.100)	(0.100)	(0.099)	(0.096)	(0.095)	(0.104)	(0.093)

Table 2. Mortgage Origination and Growth in IRS 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 income per capita (from the IRS). Sample includes zip codes with house price data from Zillow. Standard errors are clustered by county. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A. With county fixed effects

	Mortgage growth measure		
	Total	Average	Number
Per capita income growth	-0.182**	0.239***	-0.402***
(IRS)	(0.090)	(0.026)	(0.075)
County FE	Y	Y	Y
Number of observations	8,619	8,619	8,619
R2	0.33	0.68	0.31

Panel B. Without county fixed effects

	Mortgage growth measure		
	Total	Average	Number
Per capita income growth	0.368***	0.587***	-0.218**
(IRS)	(0.109)	(0.038)	(0.091)
County FE	N	N	N
Number of observations	8,619	8,619	8,619
R2	0.00	0.09	0.00

Table 3. 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. *, **, *** 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		
Buyer income growth	0.369***	0.376***	0.349***	0.282***	0.276***	0.266***	0.117***	0.130***	0.116***
	(0.047)	(0.047)	(0.047)	(0.015)	(0.015)	(0.015)	(0.040)	(0.040)	(0.040)
Per capita income growth (IRS)		-0.224**	-0.214***		0.208***	0.212***		-0.417***	-0.411***
		(0.088)	(0.079)		(0.023)	(0.021)		(0.075)	(0.071)
Zip code house price growth			0.559***			0.198***			0.281**
			(0.139)			(0.023)			(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		
Buyer income growth	0.524***	0.511***	0.292***	0.539***	0.506***	0.285***	0.002	0.023	0.041
	(0.047)	(0.045)	(0.048)	(0.033)	(0.034)	(0.016)	(0.052)	(0.050)	(0.045)
Per capita income growth (IRS)		0.150	0.020		0.372***	0.239***		-0.227***	-0.217***
		(0.101)	(0.089)		(0.029)	(0.022)		(0.084)	(0.083)
Zip code house price growth			0.466***			0.473***			-0.038
			(0.057)			(0.017)			(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 4. Mortgage Origination and Income by House Price Growth

The Table shows OLS regressions of annualized growth in total mortgage credit, the average mortgage size and the number of mortgages at the zip code level for purchase mortgages only 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). Zip codes are separated into quartiles based on the growth in house prices between 2002 and 2006 in the zip code. The "high" column includes the top quartile, the "middle" column includes the second and third quartiles, and "low" includes the lowest quartile. Panel A, partial out county fixed effect estimated over the whole sample. Standard errors are clustered by county. *, **, *** 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		
	High	Med	Low	High	Med	Low	High	Med	Low
Per capita income growth	-0.370*	-0.339***	0.213	0.221***	0.198***	0.232***	-0.564***	-0.512***	-0.037
(IRS)	(0.193)	(0.106)	(0.161)	(0.035)	(0.030)	(0.048)	(0.165)	(0.097)	(0.143)
Buyer income growth	0.449***	0.422***	0.185**	0.249***	0.331***	0.210***	0.210***	0.127**	0.024
	(0.089)	(0.069)	(0.088)	(0.031)	(0.018)	(0.027)	(0.076)	(0.063)	(0.069)
County FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Number of observations	2,020	4,407	2,192	2,020	4,407	2,192	2,020	4,407	2,192
R2	0.03	0.03	0.01	0.20	0.24	0.13	0.02	0.02	0.00

Panel B. Without county fixed effects

	Growth in total mortgage origination			Growth in average mortgage size			Growth in number of mortgages originated		
	High	Med	Low	High	Med	Low	High	Med	Low
Per capita income growth	-0.121	-0.144	0.639***	0.235***	0.256***	0.301***	-0.342**	-0.393***	0.306*
(IRS)	(0.181)	(0.121)	(0.194)	(0.046)	(0.031)	(0.050)	(0.155)	(0.118)	(0.172)
Buyer income growth	0.330***	0.422***	0.263***	0.305***	0.408***	0.224***	0.038	0.049	0.072
	(0.070)	(0.062)	(0.085)	(0.037)	(0.018)	(0.024)	(0.069)	(0.058)	(0.070)
County FE	N	N	N	N	N	N	N	N	N
Number of observations	2,020	4,407	2,192	2,020	4,407	2,192	2,020	4,407	2,192
R2	0.02	0.02	0.03	0.28	0.31	0.14	0.00	0.01	0.01

Table 5. Mortgage Origination and Income by IRS Income in 2002

The Table shows OLS regressions of annualized growth in total mortgage credit, the average mortgage size and the number of mortgages at the zip code level for purchase mortgages only 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). Zip codes are separated into quartiles based on the average household income per capita in 2002. The "high" column includes the top quartile, the "middle" column includes the second and third quartiles, and "low" includes the lowest quartile. In Panel A we partial out county fixed effect estimated over the whole sample. Standard errors are clustered by county. *, **, *** 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		
	High	Med	Low	High	Med	Low	High	Med	Low
Per capita income growth	-0.239*	0.160	0.163	0.173***	0.206***	0.229***	-0.410***	-0.051	-0.114
(IRS)	(0.125)	(0.120)	(0.223)	(0.028)	(0.030)	(0.059)	(0.109)	(0.108)	(0.193)
Buyer income growth	0.309***	0.282***	0.575***	0.344***	0.253***	0.234***	0.024	0.068	0.324***
	(0.086)	(0.072)	(0.084)	(0.031)	(0.018)	(0.028)	(0.081)	(0.061)	(0.066)
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	0.02	0.01	0.05	0.30	0.19	0.12	0.01	0.00	0.02

Panel B. Without county fixed effects

	Growth in total mortgage origination			Growth in average mortgage size			Growth in number of mortgages originated		
	High	Med	Low	High	Med	Low	High	Med	Low
Per capita income growth	0.124	1.169***	1.375***	0.242***	0.450***	0.639***	-0.129	0.658***	0.549***
(IRS)	(0.128)	(0.135)	(0.193)	(0.041)	(0.050)	(0.074)	(0.116)	(0.115)	(0.154)
Buyer income growth	0.252***	0.370***	0.777***	0.493***	0.511***	0.496***	-0.175**	-0.112**	0.254***
	(0.081)	(0.057)	(0.069)	(0.048)	(0.029)	(0.051)	(0.083)	(0.052)	(0.082)
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	0.02	0.07	0.16	0.42	0.42	0.38	0.01	0.01	0.03

Table 6. Mortgage Origination and Income for Alternative Time Periods

The Table shows OLS regressions of annualized growth in total mortgage credit at the zip code level (purchase mortgages only) 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 Table shows the regressions for the five time periods shown in the first row. Panel A includes zip codes with house price data from Zillow, and Panel B includes all zip codes in the HMDA data. Standard errors are clustered by county. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Sample with house price data (Zillow sample)

	1996-1998	1998-2002	2002-2006	2007-2011
Per capita income growth		0.563***	-0.224**	
(IRS)		(0.063)	(0.088)	
Buyer income growth	0.260***	0.240***	0.376***	0.341***
	(0.033)	(0.023)	(0.047)	(0.029)
County FE	Y	Y	Y	Y
Number of observations	8,597	8,605	8,619	8,550
R2	0.57	0.45	0.35	0.48

Panel B. All zip codes in HMDA

	1996-1998	1998-2002	2002-2006	2007-2011
Per capita income growth		0.173***	0.227***	
(IRS)		(0.038)	(0.063)	
Buyer income growth	0.544***	0.371***	0.551***	0.297***
	(0.040)	(0.023)	(0.032)	(0.017)
County FE	Y	Y	Y	Y
Number of observations	28,306	28,019	27,385	30,998
R2	0.60	0.53	0.33	0.56

Table 7. Mortgage Origination and Income, Panel Specification, 2002-2006

The Table shows OLS regressions of the logarithm of total mortgage credit at the zip code level (purchase mortgages only), the logarithm of average mortgage size, and the logarithm of the total number of mortgages on the logarithm of average household income (from the IRS) and the logarithm of average buyer income in the zip code (obtained from HMDA). IRS data is available for all years between 2002 and 2006, but 2003. The Table shows the regressions for the average treatment effect in columns 1, 3 and 5. In columns 2, 4 and 6 the income variables are interacted with indicator variables for each year in the sample. Standard errors are clustered by county. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Ln(Total mortgage origination)		Ln(Average mortgage size)		Ln(Total number of mortgages)	
Ln(Buyer income)	0.313***	0.231***	0.322***	0.166***	0.018	0.089**
	(0.033)	(0.041)	(0.029)	(0.018)	(0.030)	(0.035)
Ln(Buyer income)		0.195***		0.192***		0.007
x Year 2004		(0.028)		(0.021)		(0.031)
Ln(Buyer income)		0.203***		0.270***		-0.063
x Year 2005		(0.036)		(0.026)		(0.044)
Ln(Buyer income)		0.040		0.259***		-0.216***
x Year 2006		(0.040)		(0.023)		(0.042)
Ln(Zip code income)	0.277***	1.004***	0.339***	0.345***	-0.074	0.651***
	(0.077)	(0.090)	(0.027)	(0.031)	(0.070)	(0.075)
Ln(Zip code income)		-0.267***		-0.118***		-0.155***
x Year 2004		(0.030)		(0.019)		(0.027)
Ln(Zip code income)		-0.380***		-0.160***		-0.224***
x Year 2005		(0.034)		(0.023)		(0.035)
Ln(Zip code income)		-0.399***		-0.189***		-0.214***
x Year 2006		(0.040)		(0.021)		(0.037)
Zip code and year FE	Y	Y	Y	Y	Y	Y
Number of observations	36,299	36,299	36,299	36,299	36,299	36,299
R2	0.97	0.97	0.98	0.98	0.97	0.97

Table 8. Summary statistics by buyer income decile and year, 2002, 2004-2006

The Table shows mean and standard deviation of a tabulation of loan size (Panel A), debt-to-income (Panel B) and the fraction of overall mortgage credit originated (Panel C) for purchase mortgages by decile of buyer income and year. Buyers are sorted yearly into deciles based on the applicant income from HMDA. Debt-to-income is defined as the mortgage amount divided by the applicant income. The fraction of mortgage credit originated is calculated as the sum of total mortgage amount in each decile divided by the total mortgage amount originated in each year shown in HMDA. The first row in each year shows averages and the second row shows standard deviations.

Panel A. Loan size (USD thousands)

	Applicant income decile									
Year	1	2	3	4	5	6	7	8	9	10
2002	70.5	91.6	105.3	117.6	129.9	143.7	158.9	178.6	209.6	316.8
	36.5	40.4	46.8	54.2	61.0	69.6	79.8	95.3	121.2	263.4
2004	78.1	99.9	114.4	129.4	144.8	162.2	184.3	211.0	250.7	377.2
	42.7	50.0	59.0	68.7	78.7	89.9	104.3	123.9	154.3	323.8
2005	79.8	101.4	115.8	132.2	149.0	169.0	193.2	223.5	269.1	405.5
	45.5	55.8	65.9	77.2	88.3	102.0	118.9	141.0	176.9	361.1
2006	82.4	104.7	119.5	132.8	149.5	168.8	190.6	220.9	265.8	412.3
	47.2	56.9	68.3	78.3	90.8	104.4	121.2	144.2	180.5	383.0

Panel B. Debt-to-income

	Applicant income decile									
Year	1	2	3	4	5	6	7	8	9	10
2002	3.07	2.58	2.43	2.29	2.17	2.06	1.95	1.83	1.67	1.32
	6.05	1.13	1.08	1.05	1.02	0.99	0.98	0.97	0.96	0.94
2004	3.19	2.62	2.45	2.32	2.23	2.17	2.10	2.01	1.85	1.46
	5.13	1.30	1.26	1.23	1.21	1.20	1.19	1.17	1.13	1.05
2005	2.95	2.47	2.30	2.19	2.12	2.07	2.01	1.93	1.78	1.38
	4.59	1.35	1.30	1.27	1.25	1.25	1.24	1.21	1.16	1.06
2006	3.24	2.42	2.24	2.10	2.01	1.94	1.86	1.78	1.64	1.30
	10.50	1.31	1.28	1.24	1.22	1.19	1.18	1.16	1.11	1.01

Panel C. Fraction of mortgages originated per year

	Applicant income decile									
Year	1	2	3	4	5	6	7	8	9	10
2002	0.051	0.057	0.071	0.078	0.084	0.100	0.099	0.118	0.135	0.207
2004	0.047	0.058	0.062	0.085	0.071	0.093	0.107	0.120	0.142	0.216
2005	0.049	0.049	0.069	0.073	0.076	0.093	0.105	0.121	0.150	0.215
2006	0.048	0.059	0.062	0.069	0.084	0.090	0.100	0.120	0.145	0.223

Table 9. Mortgage Origination and Income at the Transaction Level, 2002-2006.

The Table shows OLS regressions of the logarithm of 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) and the logarithm of buyer income (obtained from HMDA). The unit of observation is an individual loan in HMDA. IRS data is available for all years between 2002 and 2006, but 2003.The table shows the regressions for the average treatment effect in columns 1 and 2. Columns 2 and 4 shows the effect interacted with a linear trend. Standard errors are clustered by county. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Zillow Sample				Whole sample			
Ln(Buyer income)	0.403***	0.366***	0.340***	0.313***	0.393***	0.365***	0.333***	0.312***
	(0.008)	(0.008)	(0.006)	(0.007)	(0.007)	(0.008)	(0.006)	(0.007)
Ln(Buyer income)		0.015***		0.012***		0.012***		0.009***
x Linear trend		(0.002)		(0.002)		(0.002)		(0.002)
Ln(Census tract household income)	0.382***	0.409***	0.313***	0.302***	0.374***	0.403***	0.250***	0.251***
	(0.012)	(0.015)	(0.024)	(0.030)	(0.012)	(0.015)	(0.023)	(0.032)
Ln(Census tract household income)		-0.011***		-0.004		-0.011***		-0.005
x Linear trend		(0.004)		(0.004)		(0.004)		(0.004)
Year FE and county FE	Y	Y	N	N	Y	Y	N	N
Year FE and census tract FE	N	N	Y	Y	N	N	Y	Y
Number of observations	17,220,064	17,220,064	17,220,064	17,220,064	23,943,309	23,943,309	23,943,309	23,943,309
R2	0.30	0.30	0.33	0.33	0.32	0.32	0.35	0.35

Table 10. Misreporting of income – robustness tests

Panel A shows OLS regressions of annualized growth in total mortgage credit 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). 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 B shows OLS regressions of the logarithm of 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) and the logarithm of buyer income (obtained from HMDA) for tract within our Zillow sample. Results are split by subprime and prime lenders, as well as GSE or non-GSE loans. Standard errors are clustered by county. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Zip code-level

	High GSE Fraction	Med GSE Fraction	Low GSE Fraction	High Subp Fraction	Med Subp Fraction	Low Subp Fraction
Per capita income growth	-0.072	-0.046	-0.495***	-0.190	-0.109	-0.098
(IRS)	(0.160)	(0.112)	(0.170)	(0.179)	(0.138)	(0.123)
Buyer income growth	0.338***	0.389***	0.363***	0.477***	0.316***	0.379***
	(0.089)	(0.060)	(0.104)	(0.098)	(0.065)	(0.092)
County FE	Y	Y	Y	Y	Y	Y
Number of observations	2,203	4,355	2,061	2,119	4,326	2,174
R2	0.01	0.02	0.03	0.03	0.02	0.02

Panel B. Loan-level

	GSE loan	Non-GSE loan	Subprime lenders	Prime lenders
Ln(Buyer income)	0.305***	0.313***	0.342***	0.306***
	(0.005)	(0.008)	(0.011)	(0.007)
Ln(Buyer income)	-0.001	0.016***	0.004	0.012***
x Linear trend	(0.001)	(0.002)	(0.004)	(0.002)
Ln(Census tract household income)	0.213***	0.307***	0.307***	0.281***
	(0.026)	(0.034)	(0.043)	(0.030)
Ln(Census tract household income)	-0.016***	-0.008*	-0.012	0.000
x Linear trend	(0.003)	(0.004)	(0.007)	(0.003)
Year FE and county FE	N	N	N	N
Year FE and census tract FE	Y	Y	Y	Y
Number of observations	3,804,113	13,415,951	2,338,973	14,881,091
R2	0.53	0.33	0.29	0.35

Table 11. Mortgage Refinancing and Income

Panel A shows OLS regressions of annualized growth in total mortgage credit between 2002 and 2006, 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 used for refinancing. Panel B shows OLS regressions of the logarithm of total mortgage credit at the zip code level (refinancing mortgages only), the logarithm of average mortgage size, and the logarithm of the total number of mortgages on the logarithm of average household income (from the IRS) and the logarithm of average buyer income in the zip code (obtained from HMDA). The Table shows the regressions for the average treatment effect in columns 1, 3 and 5. In columns 2, 4 and 6 the income variables are interacted with indicator variables for each year in the sample. Standard errors are clustered by county. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Zip code cross-sectional regression, 2002-2006

	Growth in total mortgage origination		Growth in average mortgage size		Growth in number of mortgages originated	
Zip code income growth	-0.579***	-1.300***	0.283***	-0.013	-0.746***	-1.100***
	(0.145)	(0.123)	(0.033)	(0.024)	(0.110)	(0.094)
Refinancing borrower income growt	1.113***	0.705***	0.793***	0.424***	0.424***	0.370***
	(0.084)	(0.069)	(0.057)	(0.022)	(0.060)	(0.052)
County FE	N	Y	N	Y	N	Y
Number of observations	8,622	8,622	8,622	8,622	8,622	8,622
R2	0.09	0.57	0.37	0.81	0.04	0.49

Panel B. Zip code panel regression, 2002, 2004-2006

	Ln(Total mortgage origination)		Ln(Average mortgage size)		Ln(Total number of mortgages)	
Ln(Buyer income)	0.691***	0.621***	0.488***	0.260***	0.259***	0.423***
	(0.062)	(0.080)	(0.041)	(0.028)	(0.043)	(0.067)
Ln(Buyer income)		0.014		0.133***		-0.118**
x Year 2004		(0.071)		(0.035)		(0.053)
Ln(Buyer income)		0.031		0.242***		-0.208***
x Year 2005		(0.069)		(0.042)		(0.054)
Ln(Buyer income)		0.085		0.300***		-0.228***
x Year 2006		(0.069)		(0.037)		(0.061)
Ln(Zip code income)	-0.216*	1.214***	0.257***	0.306***	-0.475***	0.913***
	(0.127)	(0.127)	(0.027)	(0.039)	(0.113)	(0.101)
Ln(Zip code income)		-0.498***		-0.086***		-0.416***
x Year 2004		(0.062)		(0.031)		(0.046)
Ln(Zip code income)		-0.612***		-0.140***		-0.476***
x Year 2005		(0.066)		(0.038)		(0.049)
Ln(Zip code income)		-0.772***		-0.202***		-0.566***
x Year 2006		(0.063)		(0.032)		(0.052)
Zip code and year FE	Y	Y	Y	Y	Y	Y
Number of observation	36,265	36,265	36,265	36,265	36,265	36,265
R2	0.96	0.97	0.97	0.97	0.96	0.97

Table 12. Purchase and Cash-out refinancing (LPS data): Mortgage origination and income at the transaction level

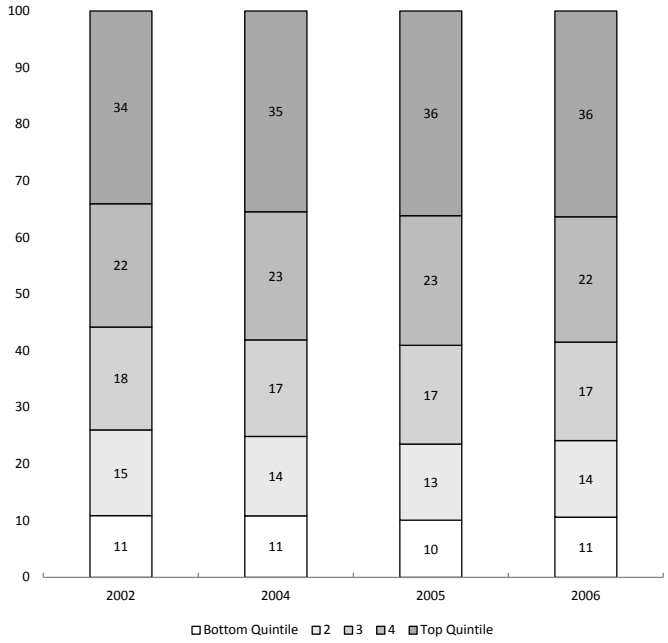
The Table shows OLS regressions of the logarithm of mortgage size at the individual level on the logarithm of average household income in the zip code (inferred using zip code household income from the IRS) and the logarithm of average borrower income (obtained from HMDA). The unit of observation is an individual loan in the LPS dataset. The first four columns include purchase mortgages, and the last four include cash-out refinancing mortgages. Standard errors are clustered by county. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Purchase mortgages				Cash-out refinancing mortgages			
Ln(Borrower income)	0.371***	0.054**	0.305***	-0.023	0.348***	-0.017	0.358***	-0.146***
	(0.026)	(0.027)	(0.027)	(0.016)	(0.032)	(0.047)	(0.030)	(0.040)
Ln(Borrower income)		0.091***		0.096***		0.087***		0.099***
x Linear trend		(0.005)		(0.005)		(0.006)		(0.007)
Ln(Zip code income)	0.405***	0.654***	0.352***	0.527***	0.356***	0.609***	0.172***	0.260***
	(0.023)	(0.028)	(0.034)	(0.037)	(0.028)	(0.043)	(0.052)	(0.054)
Ln(Zip code income)		-0.070***		-0.072***		-0.060***		-0.057***
x Linear trend		(0.005)		(0.005)		(0.007)		(0.006)
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	436,059	436,059	436,059	436,059	155,466	155,466	155,466	155,466
R2	0.59	0.59	0.64	0.64	0.54	0.54	0.61	0.61

Figure 1: Mortgage Origination by Income Level, HMDA Data

This figure shows the fraction of total dollar volume of purchase mortgages in the HMDA dataset originated by income quintile (in Panels A and B we form quintiles based on buyer income, in Panel C we use per capita income from the IRS), as well as the average debt-to-income for individuals in each quintile. Sample includes zip codes with non-missing house price data from Zillow.

Panel A: Mortgage origination (quintiles formed based on buyer income)



Panel B: Mortgage origination (quintiles formed based on IRS per capita income)

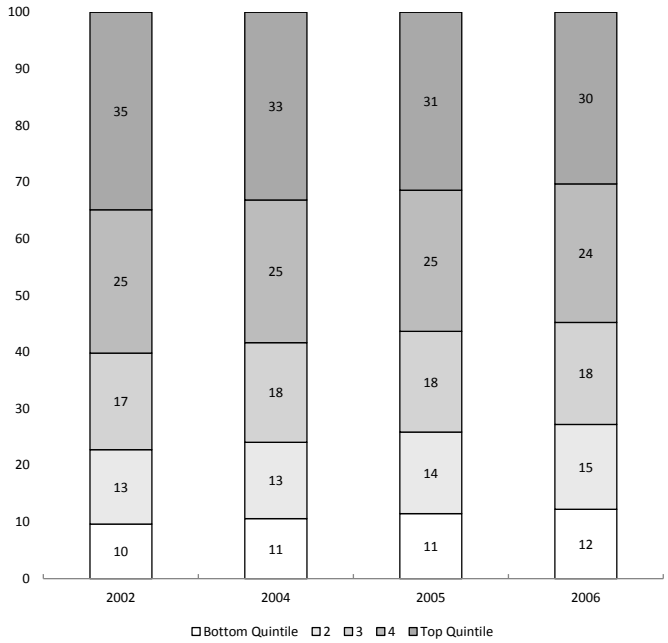
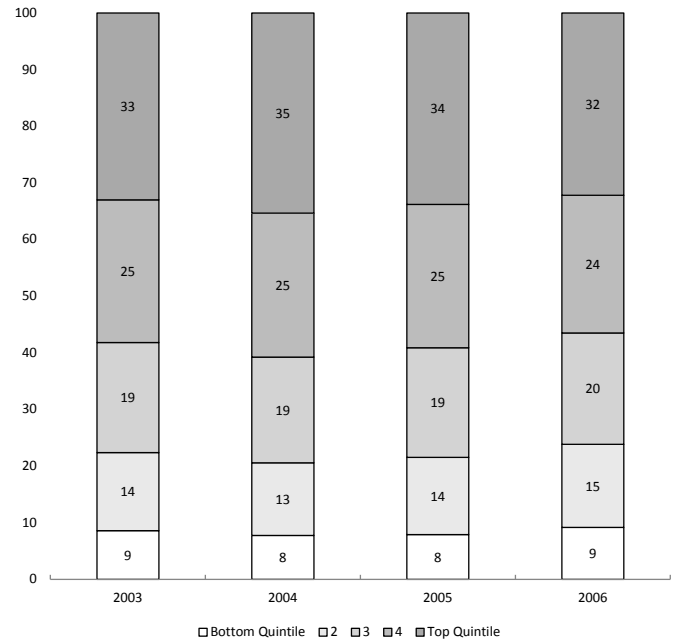


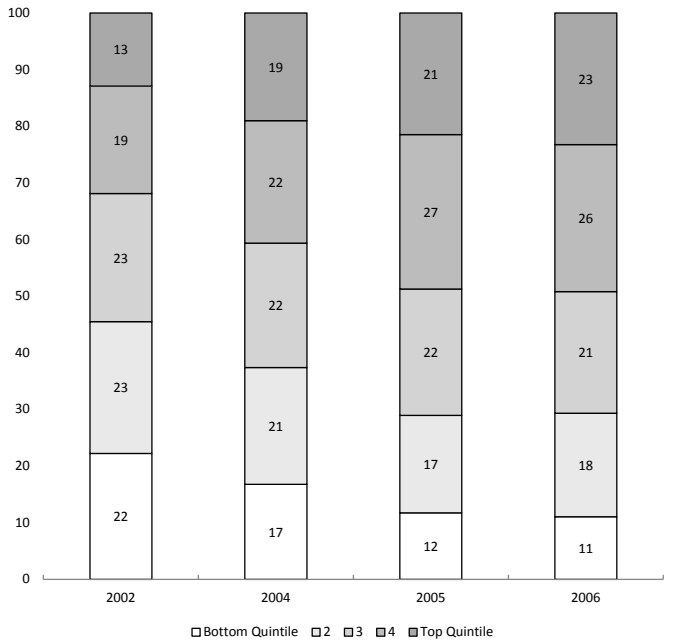
Figure 2: Mortgage Origination and Delinquency by Income Level, LPS Data

This figure shows the fraction of total dollar volume of purchase mortgages originated, as well as the total dollar volume of mortgages more than 90 days delinquent at any point during the subsequent 3 years, split by income quintile. Data is from the 5% sample of the LPS dataset. Sample includes zip codes with nonmissing Zillow house price data. In Panels A and B we form quintiles based on average buyer income, and in Panels C and D we form quintiles based on per capita income from the IRS.

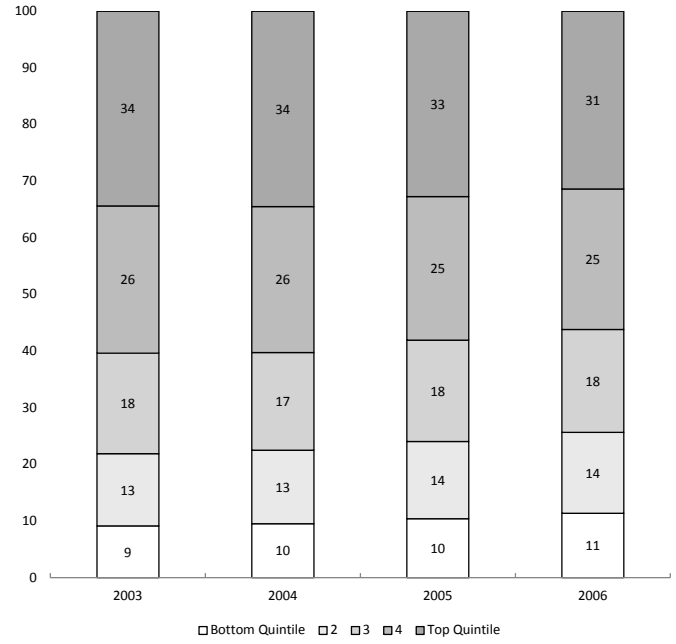
Panel A: Mortgage origination (buyer income quintiles)



Panel B: Mortgage delinquency (buyer income quintiles)



Panel C: Mortgage origination (IRS per capita income quintiles)



Panel D: Mortgage delinquency (IRS per capita income quintiles)

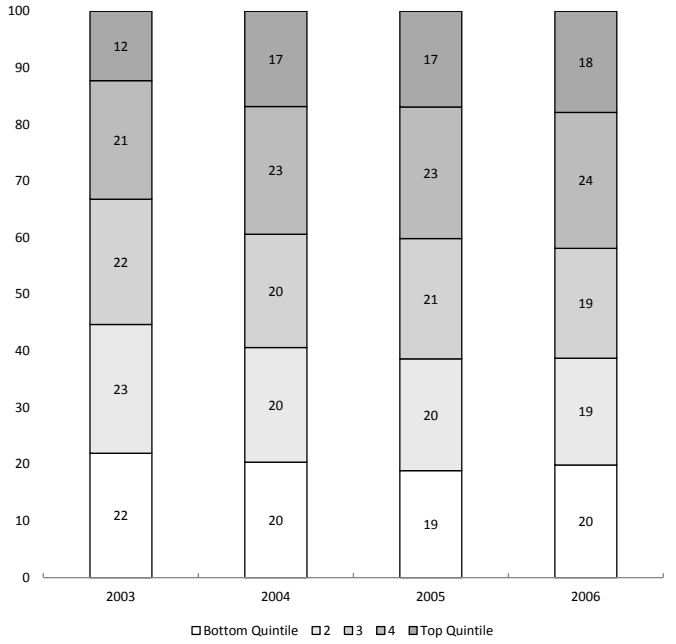
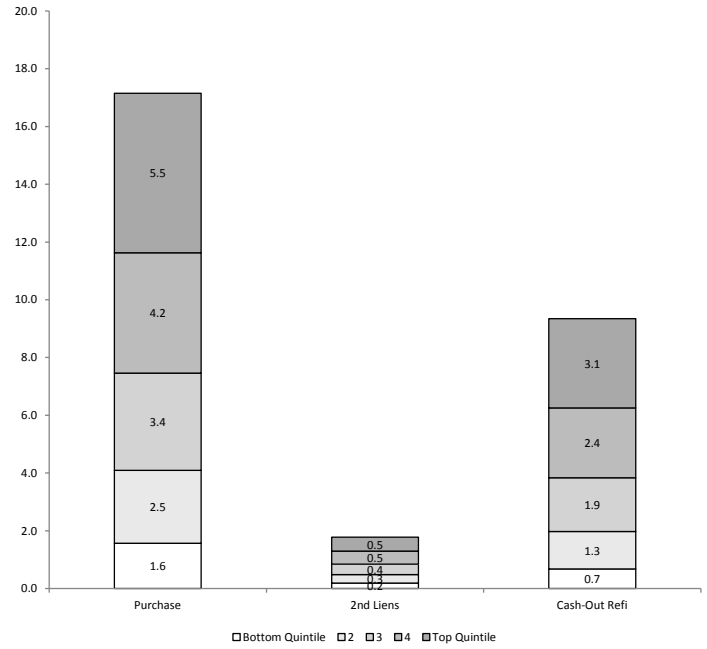


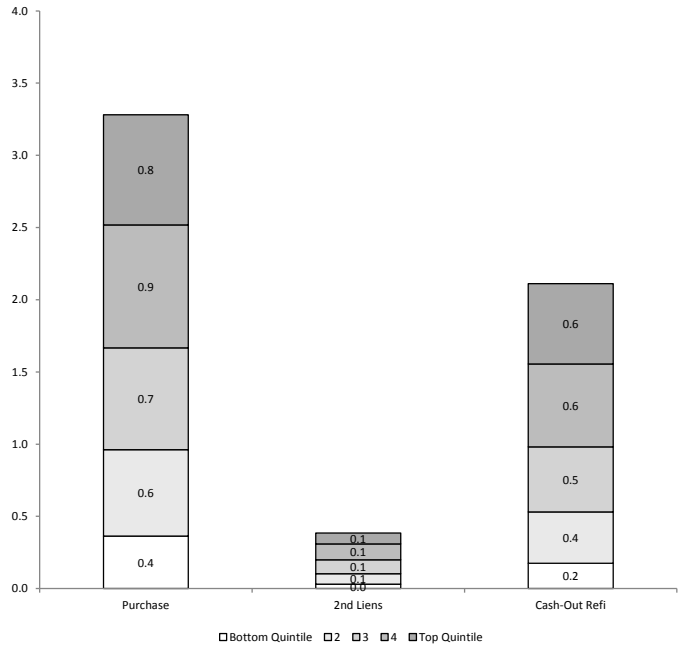
Figure 3: Mortgage origination and delinquency by product type and by income (2006)

This figure shows the total dollar volume of purchase mortgages, 2nd lien mortgages and cash-out refinance mortgages 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. Sample includes zip codes with nonmissing Zillow house price data. In Panels A and B we form quintiles based on average buyer income, and in Panels C and D we form quintiles based on per capita income from the IRS.

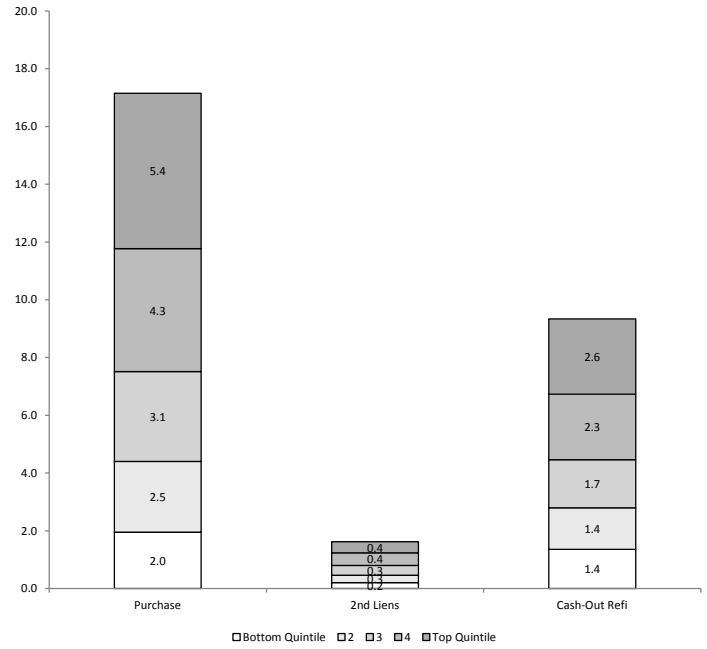
Panel A: Mortgage origination (buyer income quintiles)



Panel B: Mortgage delinquency (buyer income quintiles)



Panel C: Mortgage origination (IRS per capita income quintiles)



Panel D: Mortgage delinquency (IRS per capita income quintiles)

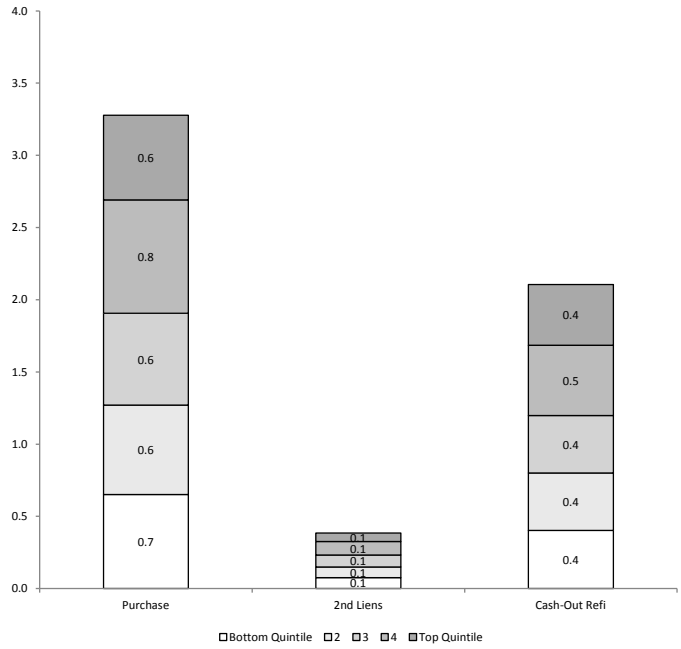
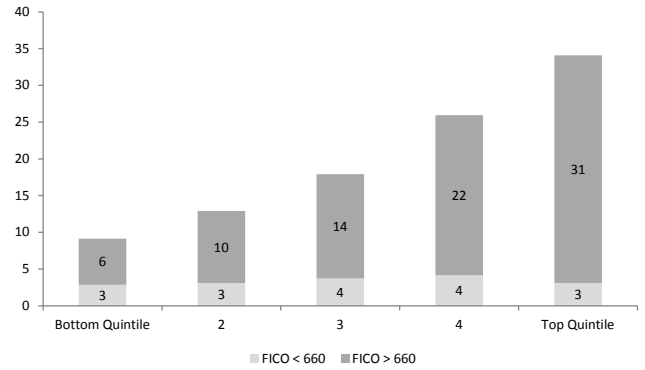


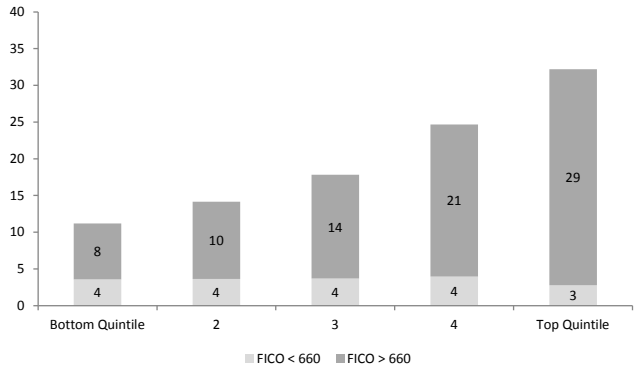
Figure 4: Mortgage origination and delinquency by credit score

This figure shows the fraction of total dollar volume of purchase mortgages, the fraction that became 90 days delinquent or more over the subsequent 3 years, and the fraction that went into foreclosure in the subsequent 3 years. Dark grey indicates the fraction of the total made up by borrowers with a credit score above (or equal to) 660, light grey indicates the fraction made up by borrowers with a credit score below 660. The fractions in each panel sum to 100%. Sample includes zip codes with nonmissing Zillow house price data. We form quintiles based on per capita income from the IRS.

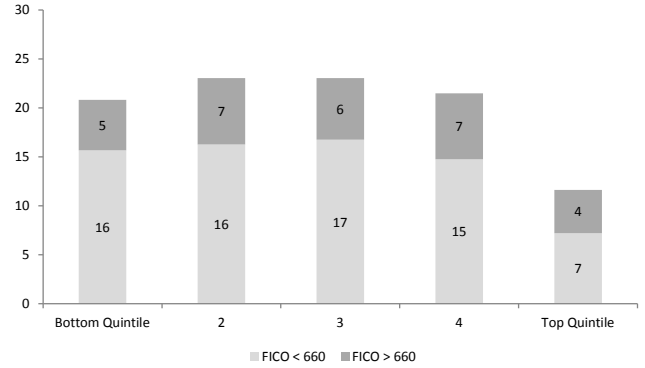
Panel A: Mortgage origination, 2003, above and below FICO=660



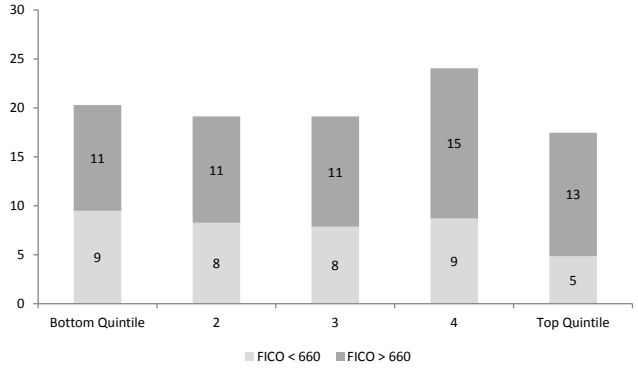
Panel B: Mortgage origination, 2006, above and below FICO=660



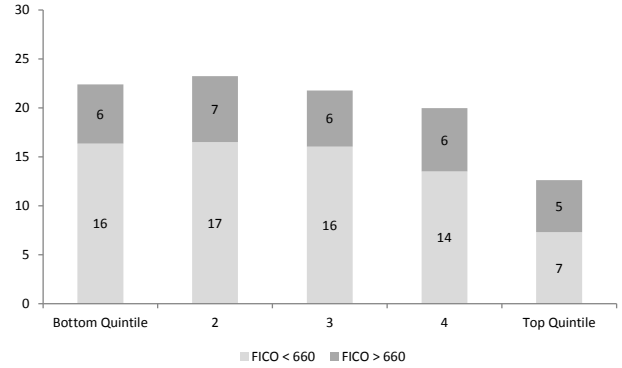
Panel C: Mortgage delinquency, 2003 loans, above and below FICO=660



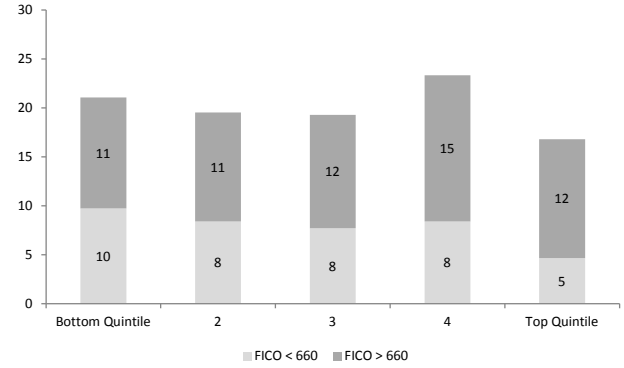
Panel D: Mortgage delinquency, 2006 loans, above and below FICO=660



Panel E: Foreclosures, 2003 loans, above and below FICO=660



Panel F: Foreclosures, 2006 loans, above and below FICO=660



APPENDIX

Appendix Table 1: Heterogeneity of the Effect. Horse Race between Income, Elasticity, GSE fraction and House Prices Growth.

	Growth in total mortgage origination		Growth in average mortgage size		Growth in number of mortgages originated	
Zip code income growth	-0.250**	0.470**	0.208***	0.370***	-0.440***	0.070
	(0.098)	(0.208)	(0.027)	(0.065)	(0.080)	(0.188)
Zip code income growth		-0.730***		-0.043		-0.638***
x High Income		(0.213)		(0.057)		(0.202)
Zip code income growth		0.303		-0.143*		0.387*
x High Elasticity		(0.228)		(0.074)		(0.202)
Zip code income growth		0.071		0.000		0.075
x High Low Documentation Loans		(0.138)		(0.034)		(0.129)
Zip code income growth		-0.132		-0.140***		0.020
x High HP Growth		(0.165)		(0.050)		(0.140)
Zip code income growth		0.156		-0.046		0.189
x High GSE Fraction		(0.160)		(0.045)		(0.141)
Buyer income growth	0.390***	0.492***	0.243***	0.204***	0.172***	0.275***
(annualized), HMDA	(0.056)	(0.120)	(0.017)	(0.030)	(0.047)	(0.099)
Buyer income growth		-0.406***		0.004		-0.353***
x High Income		(0.096)		(0.026)		(0.079)
Buyer income growth		0.055		-0.021		0.070
x High Elasticity		(0.117)		(0.035)		(0.106)
Buyer income growth		-0.043		-0.013		-0.015
x High Low Documentation Loans		(0.068)		(0.018)		(0.060)
Buyer income growth		0.206		0.030		0.157
x High HP Growth		(0.137)		(0.032)		(0.114)
Buyer income growth		-0.022		0.078***		-0.066
x High GSE Fraction		(0.091)		(0.025)		(0.078)
County FE	Y	Y	Y	Y	Y	Y
Number of observations	6,210	6,210	6,210	6,210	6,210	6,210
R2	0.34	0.36	0.74	0.75	0.30	0.32

Standard errors are in parenthesis and are clustered by county. *, **, *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Appendix Table 3. Summary Statistics using full HMDA sample

This table reports summary statistics for all zip codes covered in HMDA. Column 1 shows the summary statistics for the entire sample. Columns 2 to 4 show the summary statistics for the zip codes in the highest quartile of average income per capita in 2002 (High), second and third quartiles of average income per capita (Intermediate), and lowest quartile of average income per capita (Low). 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). *Zip code income* is the average adjusted gross income per capita by zip code from the IRS. *Buyer Income* is the average applicant income by zip code from HMDA. *Average mortgage size* is the total mortgage amount originated in a zip code used for home purchase in 2002. *Number of mortgages originated per 100 residents* is the average number of mortgages originated per 100 residents by zip code. Debt-to-income is the zip code average ratio of the mortgage balance at the time of origination over applicant income from HMDA. *Loan to value* is the average loan to value (LTV) in a zip code calculated from LPS. *Fraction of low documentation loans* is the fraction of low documentation loans originated in a zip code from LPS. *Population* is the average zip code-level population estimated from IRS returns. *Elasticity of housing supply* is the Saiz (2010) measure of housing supply elasticity at the MSA level.

Zip code income growth refers to the percentage change between 2002 and 2006 in household adjusted gross income by zip code from the IRS. *Buyer income growth* is the percentage change in average applicant income by zip code from HMDA. *Growth in total mortgage origination* refers to the percentage change in total mortgage credit originated in a zip code used for home purchase between 2002 and 2006 calculated using HMDA. *Growth in average mortgage size* refers to the percentage change in the average balance of individual mortgages in a zip code between 2002 and 2006 (also calculated using HMDA). *Growth in number of mortgages originated* is the percentage change in the number mortgages originated between 2002 and 2006. *Change in debt-to-income* and *Change in loan to value* is the change in Debt to Income and LTV by zip code between 2002 and 2006.

	Whole sample	Zip code household income, 2002			Zip code house price growth 2002-2006		
		High	Middle	Low	High	Middle	Low
	N =27385	N =6936	N =14126	N =6323	N =2020	N =4407	N =2192
Zip code household income (USD thousands), IRS, 2002	39.41	63.57	34.16	24.63	47.40	54.44	47.13
	(24.68)	(39.10)	(3.65)	(3.41)	(25.45)	(30.41)	(25.08)
Buyer income (USD thousands), HMDA, 2002	74.21	111.06	64.11	56.33	99.83	95.11	79.24
	(56.34)	(84.89)	(34.21)	(34.54)	(70.94)	(70.58)	(53.87)
Average mortgage size (USD thousands), 2002, purchase mortgages only	110.47	182.26	93.10	70.53	160.97	166.79	125.50
	(70.44)	(93.37)	(36.33)	(32.51)	(76.74)	(95.63)	(67.57)
Number of mortgages originated per 100 residents, 2002, purchase mortgages only	1.91	3.06	1.69	1.14	3.38	2.36	2.37
	(6.08)	(8.49)	(5.73)	(2.39)	(3.42)	(1.53)	(1.47)
Debt to income, 2002	1.83	2.16	1.79	1.55	2.18	2.17	2.03
	(0.48)	(0.38)	(0.42)	(0.48)	(0.36)	(0.39)	(0.36)
Loan to value, LPS, 2003 (N=13,555)	0.82	0.78	0.85	0.87	0.80	0.79	0.83
	(0.11)	(0.10)	(0.10)	(0.10)	(0.10)	(0.11)	(0.10)
Fraction of low documentation loans, LPS, 2003 (N=13,555)	0.16	0.18	0.16	0.13	0.20	0.18	0.17
	(0.25)	(0.23)	(0.27)	(0.26)	(0.22)	(0.24)	(0.23)
Population (000s), IRS, 2002	72.71	127.48	62.92	34.50	166.16	132.44	139.45
	(98.49)	(114.94)	(88.18)	(72.04)	(128.36)	(109.75)	(103.37)
Elasticity of housing supply (N=11887), Saiz (2010)	2.04	1.76	2.27	2.25	1.24	1.65	2.22
	(1.18)	(1.00)	(1.25)	(1.32)	(0.51)	(0.84)	(0.91)
Zip code income growth (annualized), IRS, 2002-2006	0.043	0.051	0.039	0.045	0.053	0.047	0.036
	(0.031)	(0.035)	(0.025)	(0.036)	(0.029)	(0.027)	(0.025)
Buyer income growth (annualized), HMDA, 2002-2006	0.058	0.062	0.055	0.059	0.108	0.062	0.032
	(0.072)	(0.064)	(0.065)	(0.090)	(0.066)	(0.050)	(0.052)
Growth in total mortgage origination (annualized), 2002-2006, purchase only	0.164	0.110	0.172	0.204	0.170	0.123	0.074
	(0.210)	(0.166)	(0.197)	(0.261)	(0.165)	(0.138)	(0.136)
Growth in average mortgage size (annualized), 2002-2006, purchase only	0.066	0.065	0.063	0.076	0.124	0.063	0.021
	(0.063)	(0.054)	(0.058)	(0.081)	(0.042)	(0.040)	(0.038)
Growth in number of mortgages originated (annualized), 2002-2006, purchase only	0.092	0.046	0.104	0.117	0.046	0.059	0.054
	(0.180)	(0.150)	(0.169)	(0.220)	(0.144)	(0.126)	(0.119)
Change in debt to income, 2002-2006	0.074	0.004	0.086	0.122	0.063	0.016	-0.107
	(0.341)	(0.289)	(0.318)	(0.422)	(0.293)	(0.274)	(0.276)
Change in loan to value, 2003-2006	-0.017	-0.007	-0.023	-0.038	-0.036	-0.010	-0.013
	(0.108)	(0.102)	(0.113)	(0.111)	(0.095)	(0.104)	(0.093)

Appendix Table 4. Mortgage origination and growth. Alternative measures

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 income per capita (from the IRS). Sample includes zip codes with house price data from Zillow. Standard errors are clustered by county. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

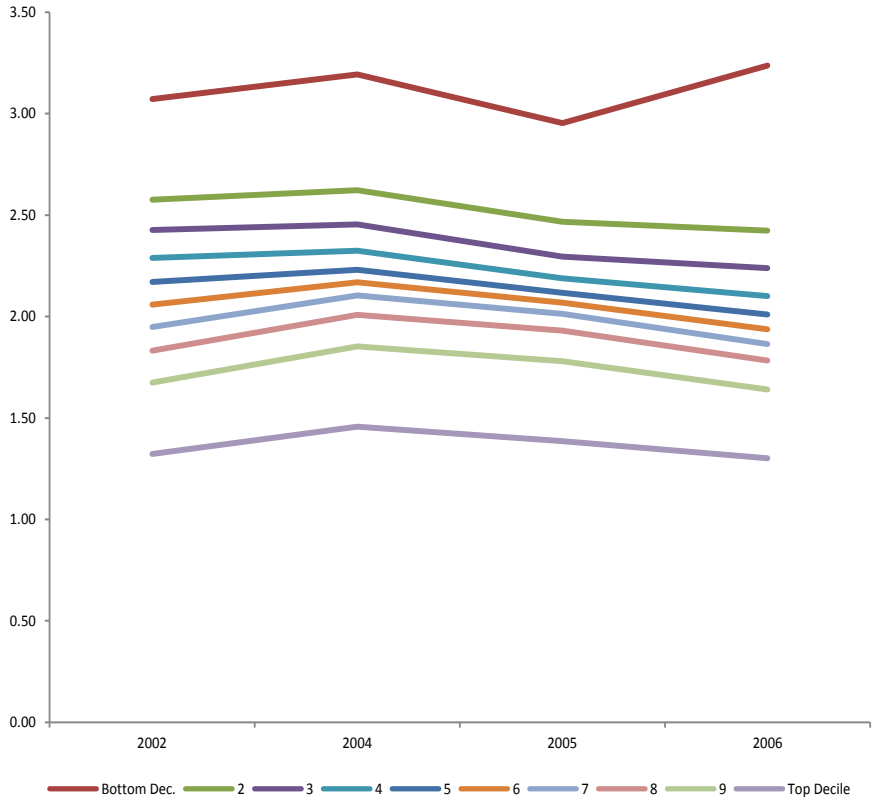
Panel A. Total income, per capita income, and population

	Mortgage growth measure		
	Total	Average	Number
Total income growth	0.415***		
	(0.041)		
Per capita income growth		0.239***	
(IRS)		(0.026)	
Population growth			1.276***
			(0.062)
County FE	Y	Y	Y
Number of observations	8,619	8,619	8,619
R2	0.34	0.53	0.34

Panel B. Population as dependent variable by IRS income per capita in 2002

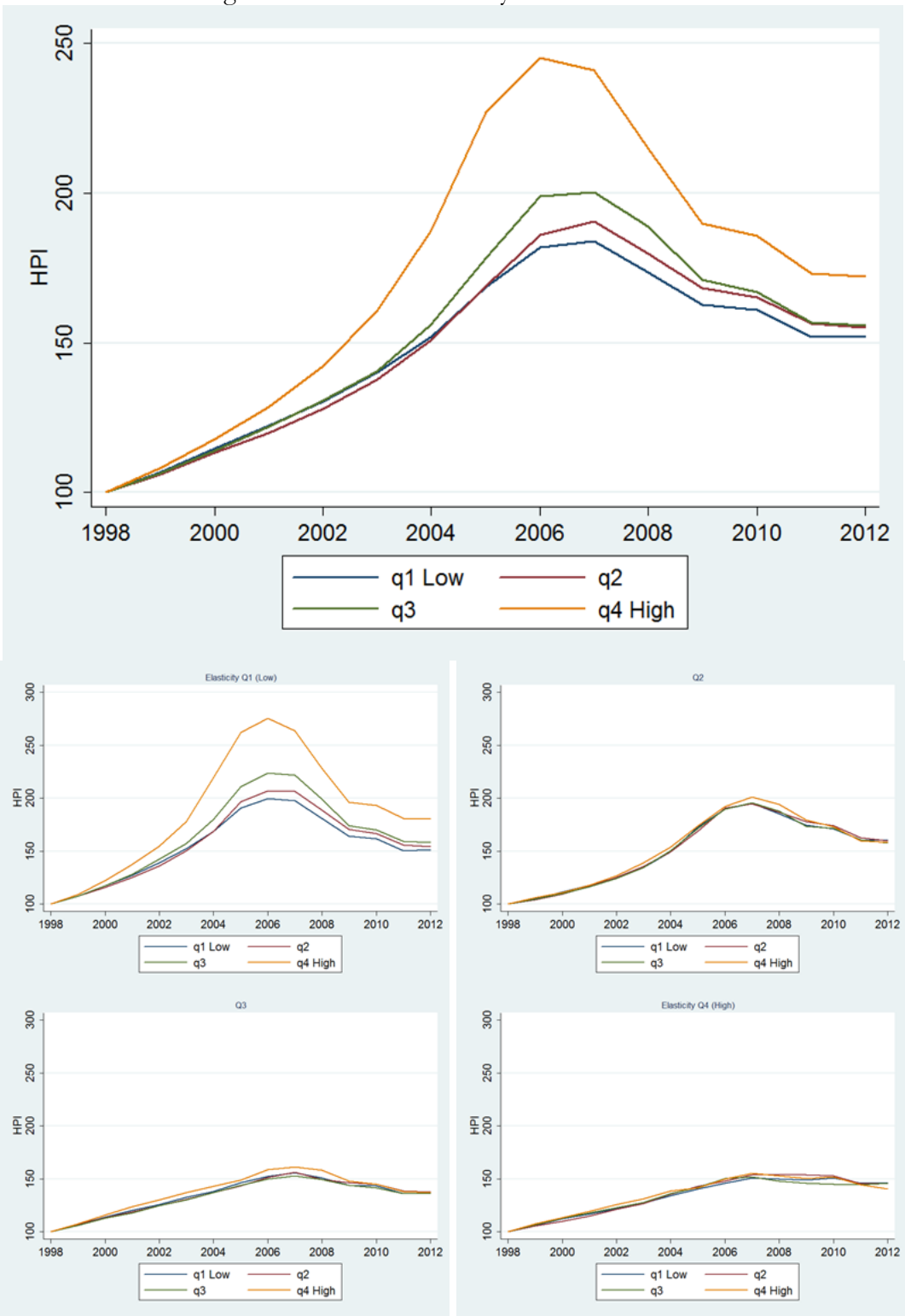
	Population Growth		
	Zip code household income (IRS), 2002		
	High	Middle	Low
Per capita income growth	-0.086***	0.034	0.014
(IRS)	(0.022)	(0.025)	(0.031)
County FE	Y	Y	Y
Number of observations	2,088	4,346	2,185
R2	0.01	0.00	0.00

Appendix Figure 1. Debt-to-Income Ratio (deciles formed based on buyer income)
Figure shows average debt-to-income for individuals in each buyer income quintile. Sample includes zip codes with non-missing house price data from Zillow.



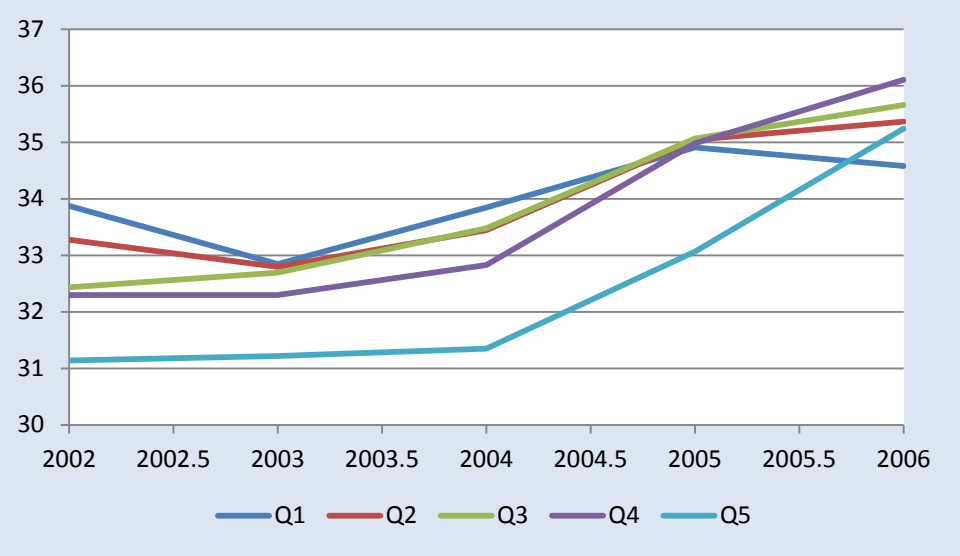
Appendix Figure 2. House Prices Time Series based on Percentage Difference between Buyers Income and IRS Income in 1998

Quartiles based on Percentage Difference between Buyers Income and IRS Income in 1998



Appendix Figure 3. DTI measured in LPS (recurring mortgage debt payments divided by monthly income)
Quintiles based on buyer income and IRS Income

Panel A: Zip codes split into quintiles based on borrower income (from HMDA)



Panel B: Zip codes split into quintiles based on IRS income (from IRS)