Fraudulent Income Overstatement on Mortgage Applications During the Credit Expansion of 2002 to 2005

Atif Mian
Princeton University and NBER

Amir Sufi
University of Chicago Booth School of Business and NBER

Abstract

Treating fraudulently overstated income on mortgage applications as true income can lead to incorrect conclusions on the nature of the mortgage credit supply expansion toward marginal borrowers from 2002 to 2005. A positive gap between zip-code-level income growth calculated from mortgage applications and income growth from the IRS likely reflects mortgage fraud, not an improvement in home-buyer income. In support of the credit supply view, mortgage credit for home purchase expanded significantly more in low-credit-score neighborhoods on both the extensive and intensive margins from 2002 to 2005, even though these neighborhoods deteriorated on many measures of income prospects. (*JEL* G21)

We thank numerous colleagues from many institutions for valuable advice and comments. We are grateful to Andrew Karolyi and two anonymous referees for detailed comments. We gratefully acknowledge financial support from the Fama-Miller Center and the Initiative on Global Markets at Chicago Booth. Seongjin Park provided excellent research assistance. Send correspondence to Amir Sufi, The University of Chicago Booth School of Business, 5807 South Woodlawn Avenue, Chicago, IL 60637; telephone: (773) 702-6148. E-mail: amir.sufi@chicagobooth.edu.

What is the source of variation over time in household debt, house prices, and macroeconomic fluctuations? In the aftermath of the Great Recession, researchers have focused on the importance of credit supply shocks, or the idea that fluctuations may be driven by changes in the willingness of lenders to supply credit.[1] One of the empirical findings used to justify such a modeling assumption is from Mian and Sufi (2009) (henceforth MS09), who show an expansion in mortgage credit for home purchase in low-credit-score zip codes from 2002 to 2005 that coincides with a decline in measures of income in these neighborhoods.[2] This is indicative of a credit supply shift: lenders became more willing to lend to borrowers that were previously routinely denied credit, even though borrower income prospects in these zip codes actually declined. The simultaneous expansion in mortgage credit for home purchase and decline in income in low-credit-score zip codes generates a negative correlation between credit growth and income growth at the zip-code level.

This finding has been called into question by Adelino, Schoar, and Severino (2016) (henceforth A2S16). Using income data from mortgage applications, they argue that the share of mortgage originations for home purchase of low-income home buyers fell from 2002 to 2005. Further, they show that the correlation between zip-code-level mortgage growth and income growth is positive when one uses the income reported on mortgage applications instead of income growth of the zip code measured by the Internal Revenue Service. They conclude that a credit supply shift toward marginal borrowers was not an important driver of household debt and house price patterns from 2002 to 2005.[3]

In this study, we argue that the A2S16 conclusion is incorrect because their analysis relies

[1]Credit supply shocks show up in many recent models. In the model of Favilukis, Ludvigson, and Van Nieuwerburgh (forthcoming), "a relaxation of financing constraints leads to a large boom in house prices." In the model of Justiniano, Primiceri, and Tambalotti (2015), "an increase in credit supply driven by looser lending constraints in the mortgage market can explain ... the unprecedented rise in home prices." In the model of Schmitt-Grohé and Uribe (2016), there is an exogenous decline in the interest rate facing borrowers in a country that leads to a boom in household borrowing.

[2]Low-credit-score zip codes in MS09 are defined as zip codes with a high fraction of individuals with a credit score below 660 as of 1996.

[3]See in particular Figure 1, Panel A, and Table 5 of A2S16. They conclude that "these results provide a new picture of the mortgage expansion before 2007 and suggest that cross-sectional distortions in the allocation of credit were not a key driver of the run-up in mortgage markets and the subsequent default crisis."

on fraudulently overstated income on mortgage applications. In particular, the dramatic rise in reported mortgage application income in low-income and low-credit-score neighborhoods is more likely to reflect fraud than an improvement in the income position of actual home buyers. The use of this fraudulently overstated income makes it appear as if mortgage credit was not disproportionately expanding toward low-income households.[4]

Before delving into the microeconomic data, this fact can be seen in the aggregate. In the left panel of Figure 1, we use data from the American Community Survey on the average income of two groups: all homeowners and homeowners with a mortgage who have moved within the past year. We refer to the latter group as recent home buyers with a mortgage. From 1980 to 2000, the real income of both groups increased. However, from 2000 to 2005, real income for recent home buyers with a mortgage actually fell. In the right panel, we show that real income fell by more than 0.5% on an annualized basis for recent home buyers with a mortgage during the mortgage credit boom, which supports the argument in MS09 that credit expanded to marginal home buyers.

The right panel also includes the annualized income growth of recent home buyers from 2002 to 2005 using mortgage application income from the Home Mortgage Disclosure Act data. This is the data relied on by A2S16. Despite the decline in income of recent home buyers with a mortgage in the ACS data, income from mortgage applications shows a large 3.7% annualized increase in real income. It is crucial to understand why the ACS and HMDA application data show such dramatic differences in income growth of recent home buyers. The ACS data support the credit supply view that credit expanded during the boom to home buyers with lower income, whereas the HMDA application data suggest home buyers during the mortgage credit boom had higher incomes.

Using zip-code-level data, we show that mortgage application income data from HMDA became fraudulently overstated during the mortgage credit boom, especially for marginal borrowers that traditionally were denied credit. As a result, mortgage application income

[4]In the online appendix, we illustrate this point using as an example Chicago neighborhoods with a large number of subprime borrowers.

data should not be used to assess the income prospects of marginal home buyers during the boom. We focus on the difference between the growth in income reported on mortgage applications between 2002 and 2005 and the growth in IRS reported income between 2002 and 2005. We refer to this difference as "buyer income overstatement," and we construct this variable at the zip-code level.

Under the analysis of A2S16, one would interpret high buyer income overstatement as evidence of high-income-growth households buying homes in otherwise low-income-growth neighborhoods. Instead, we demonstrate that high buyer income overstatement is more likely to reflect fraudulent misreporting of buyers' true income. We do so in four ways.

First, even before considering zip-code-level data, aggregate data imply that fraudulent income overstatement on mortgage applications was an important aspect of the mortgage credit boom. As shown in Figure 1, the annualized real income growth of recent home buyers implied by income reported on mortgage applications was much higher than reported by the ACS. Data from the American Housing Survey show a similar gap between income reported on mortgage applications of home buyers versus income reported to the AHS (Blackburn and Vermilyea 2012). Changes in buyer composition cannot explain this gap—it is the difference between income reported on mortgage applications and income from other data sets for the same population of recent home buyers.

Second, we show that other measures of mortgage fraud established in the literature were much higher in zip codes with high buyer income overstatement. In particular, zip codes with high buyer income overstatement witness a larger increase in the fraction of non-agency mortgages, and in particular mortgages with low or unknown documentation. We know from a large body of research that both non-agency securitized mortgages and low-documentation (low-doc) mortgages were associated with a high incidence of fraud (e.g., Ben-David 2011; Jiang, Nelson, and Vytlacil 2014; Griffin and Maturana 2016a, b; Piskorski, Seru, and Witkin 2015).

Moreover, using data compiled by Piskorski, Seru, and Witkin (2015), we show that mortgages made in high buyer income overstatement zip codes were significantly more likely to be fraudulently reported as being for an owner-occupied property, or had deliberately omitted information on second liens. Using a list of zip codes with the highest amount of mortgage fraud according to Interthinx, a mortgage fraud detection company, we show that this independent measure of fraud is also positively correlated with buyer income overstatement from 2002 to 2005.

Third, contrary to the hypothesis that buyer income overstatement represents "gentrification" of these zip codes, we show that buyer income overstatement forecasts negative income and financial outcomes. In every year of the mortgage credit boom, we calculate the difference between the average income reported on mortgage applications in a zip code and the IRS average income of all residents living in a zip code. We then show that zip codes with a large positive difference between buyer income from mortgage applications and IRS average income in the 2002 to 2005 period experienced subsequently lower IRS income growth in the following year. Further, according to IRS data, high borrower income overstatement zip codes saw a relative decline in the number of high-income individuals living in the zip code. We also use individual-level data on credit scores to show that people moving into high borrower income overstatement zip codes do not have better credit scores than residents already living there.

Looking past 2005, we find that zip codes with high buyer overstatement perform terribly. Default rates in these zip codes skyrocketed from 2005 to 2007. Using a longer horizon, the zip codes with high buyer overstatement from 2002 to 2005 experienced lower IRS income and wage growth from 2005 to 2012. The magnitude of the worse performance is very large: moving from the 10th percentile to the 90th percentile of the buyer income overstatement distribution leads to 3% lower annualized wage growth from 2005 to 2012, which is twice a standard deviation. They also saw lower median household income growth from 2000 to 2010 according to the Census. Finally, there was a jump in both poverty and unemployment rates from 2000 to 2010. These patterns are inconsistent with gentrification, but consistent with fraudulent income overstatement on mortgage applications.

Fourth, time-series evidence on buyer income overstatement over a longer horizon shows that income reported on mortgage applications was particularly "distorted" during the 2002 to 2005 period. The correlation between buyer income growth and IRS-reported income growth across zip codes is weakest during the 2002 to 2005 period relative to earlier and subsequent periods. Moreover, the weak correlation during the 2002 to 2005 period is driven entirely by zip codes with a high share of non–Government Sponsored Entity (GSE) mortgage originations. There was a decoupling of buyer income growth and IRS income growth concentrated exactly when we believe fraud was most prevalent: among mortgages originated from 2002 to 2005 sold for non-GSE securitization.

One of the key results in A2S16 that does not rely on mortgage application income focuses on the size of mortgages conditional on origination. In particular, A2S16 show that the growth in the average mortgage size conditional on origination in a zip code is positively correlated with IRS income growth in the zip code from 2002 to 2005, which they claim contradicts the credit supply view of the mortgage expansion.[5] We believe this result is incorrect because A2S16 treat first and second liens as two independent mortgages instead of combining them to obtain the total amount borrowed in a home purchase.[6] This is problematic because low-income-growth zip codes saw a large relative rise in the use of second liens to purchase homes during the credit boom. Second liens are significantly smaller than first liens, thereby generating a positive correlation between growth in the average mortgage size conditional on origination and IRS income growth when treating second liens as independent mortgages.

We account for both the first- and second-lien mortgages used in a home purchase, and

[5] As we discuss below, this finding, even if true, would not contradict the expansion of mortgage credit on the extensive margin if marginal buyers obtain smaller mortgages than the average.

[6] For example, if all households in a zip code previously obtained only a $100,000 first-lien mortgage to buy a home in 2002, but then subsequently obtain both a first-lien mortgage of $100,000 and a second-lien mortgage of $20,000 to buy a home in 2005, the analysis in A2S16 implies that the average mortgage size for purchasing a home in the zip code fell from $100,000 to $60,000 from 2002 to 2005. In this example, the average total mortgage size actually increased from $100,000 to $120,000.

we show that the growth in the size of first-lien mortgages conditional on origination in a zip code is uncorrelated with IRS income growth. Further, growth in the combined size of first-lien and second-lien mortgages per purchased housing unit, which we believe is the correct measure of the total mortgage size used in the purchase, actually increased in low-income growth neighborhoods more than high-income-growth neighborhoods. Combining first and second liens when calculating the average mortgage size leads to the opposite conclusion as found in A2S16. While MS09 focused only on the extensive margin, this finding shows stronger mortgage growth on both the extensive and intensive margins in low-credit-score zip codes seeing a decline in income growth.

This study makes three contributions to the literature. First, as discussed above, understanding the source of the expansion in mortgage credit toward low-credit-score zip codes in the early 2000s is crucial to understanding the right economic model for explaining the boom and bust. To be clear, expansion of mortgage credit for home purchase toward low-credit-score zip codes cannot directly explain the rise in aggregate household debt, and such a claim was never made in MS09. So why is a focus on these low-credit-score borrowers so important?

The reason is that what happened for the marginal borrowers during the 2000s is informative about the larger fundamental shock, which likely affected even non-marginal borrowers. Discovering what happens at the margin is crucial for developing an understanding of what happened in the aggregate. If marginal home buyers during the 2000s in low-credit-score neighborhoods saw an improvement in income prospects (as suggested by the use of income from mortgage applications), then it would cast doubt on the view that a credit supply shift was the fundamental shock that led to the rise in aggregate household debt. Expansion of credit to low-credit-score borrowers with declining income is not a sufficient condition to prove a credit supply shock in the aggregate, but it is necessary. We are certainly not the first to point out widespread fraud during the mortgage boom. But we are the first to show that use of fraudulently reported income on mortgage applications leads to the incorrect conclusion that credit did not expand disproportionately to marginal home buyers who were previously routinely denied credit (see Mian and Sufi 2016 for more on this point).

Second, we are the first to show that the average mortgage size for home purchase conditional on origination increased in low-credit-score neighborhoods from 2002 to 2005. On a related point, we are also the first to show that the average mortgage size conditional on origination once first and second liens are taken into account increased in low-income-growth zip codes from 2002 to 2005. This finding further bolsters the credit supply view by showing that credit expanded for home purchase to marginal borrowers even on the intensive margin.

Third, we add to the literature on the prevalence of fraud during the mortgage credit boom during the 2000s (e.g., Avery et al. 2012; Ben-David 2011; Blackburn and Vermilyea 2012; Garmaise 2015; Griffin and Maturana 2016a, b; Jiang, Nelson, and Vytlacil 2014; Piskorski, Seru, and Witkin 2015). We believe we are the first to show that buyer income overstatement at the zip-code level—that is, the difference between income growth according to mortgage applications and income growth according to the IRS—likely reflects fraud. Our analysis of buyer income overstatement at the zip-code level is a novel approach to measuring fraud that complements the existing literature. Further, we contribute to the mortgage finance literature by showing that income reported on mortgage applications should not be used as true income for low-credit-score individuals from 2002 to 2005.

1 Data and Summary Statistics

1.1 Sample

For the sake of comparability, the analysis here uses the same sample as was used in MS09. This sample represents zip codes where Fiserv Case Shiller Weiss (FCSW) house price data were available to us at the time MS09 was written. These 3,014 zip codes represent 35% of the population of the United States in 2000, and almost 50% of the total mortgage debt outstanding as of 2000. The appendix of MS09 replicates all results that do not require the

FCSW data using the full sample of 18,407 zip codes, and it shows that all qualitative results are similar. It is imperative when moving to the full sample of zip codes that one weights by total population of the zip code, as the full sample includes some zip codes with very few individuals.

Since the publication of MS09, we have become aware of another restriction that researchers should impose on the data when utilizing HMDA: researchers should only use zip codes located within a Metropolitan Statistical Area (MSA). A document produced by the Federal Reserve in March 2005 points out that only mortgage lenders with an office located within an MSA must report under the Home Mortgage Disclosure Act (Federal Reserve 2005). They write, "as a result, reporting of home loans made in some rural areas may be relatively low." For our house price sample, this is not an issue as only 23 zip codes are not in an MSA. For ease of comparability with MS09, we keep these 23 zip codes in the house price sample, but excluding them does not affect the results.

However, for the full sample, the restriction that a zip code must be in an MSA leads to a sample of 11,803 zip codes, instead of 18,407. In the appendix, we repeat all specifications from this study in the sample of 11,803 zip codes, and we show that all results are qualitatively similar. Once again, when moving to the larger sample, it is important to weight by total population of the zip code. None of the sources of disagreement between our analysis and A2S16 are due to sample selection, with the exception of one result in the appendix of A2S16 where they examine unweighted regressions in the full sample. We discuss this issue in Section 6.2.

1.2 Core measures, fraud, and ex post outcomes

Table 1 presents the summary statistics for data used in this study. The fraction of subprime borrowers comes from Equifax and measures the fraction of individuals with a credit score below 660 living in the zip code as of 1996. The amount for home purchase growth is the annualized growth in the total amount of mortgages originated for home purchase from 2002 to 2005. This measure is from HMDA. We also include two measures of income growth in a zip code from 2002 to 2005: growth in the average income according to the IRS and growth in income according to mortgage applications as recorded in HMDA. Both are annualized.

A critical variable in our analysis is buyer income overstatement, which is defined as the difference in the growth in income according to mortgage applications and the IRS. More specifically, for zip code z, we measure buyer income overstatement as:

$$BuyerIncomeOverstatement_{z,02-05} = \Delta BuyerIncome_{z,02-05} - \Delta IRSIncome_{z,02-05} \quad (1)$$

Buyer income overstatement is on average two percentage points from 2002 to 2005, meaning that on average the income growth on mortgage applications was two percentage points higher than average IRS income growth in the zip code on an annualized basis.[7]

We interpret buyer income overstatement as fraudulent overstatement of income on mortgage applications. Table 1 presents alternative measures of fraud. The share of non-agency securitization (or private label securitization) increased dramatically from 2002 to 2005, and many of these mortgages had low or no documentation. These data are from BlackBox Logic. We also use data from the study by Piskorski, Seru, and Witkin (2015). They study two types of fraud in the non-agency market from 2005 to 2007: misreporting of the owner-occupant status of a property and misreporting of whether a second lien is present.[8]

We also use information from Interthinx, a mortgage fraud detection company.[9] Since the second quarter of 2010, they have released a list of the zip codes with the most rampant mortgage fraud. More specifically, they focus on four types of fraud: property valuation, identity, occupancy, and employment/income. The top mortgage fraud zip codes are inclusive of all of these types of fraud. They reveal a top 10 list every quarter, and then an annual list

[7]In Figure 1, income growth on mortgage applications was 3.7%, whereas it is 6.5% in Table 1. The difference comes from three sources. Figure 1 uses 2000 to 2005 instead of 2002 to 2005, and it also uses the full sample, where nominal income growth on mortgage applications was 5.3% as opposed to 6.5%. Finally, Figure 1 uses income growth in real terms as opposed to nominal terms.

[8]We are thankful to Amit Seru for providing us with these data.

[9]Interthinx has since been purchased by First American Mortgage Solutions, and the fraud reports are available at www.firstam.com

every year of 20 or 25 zip codes. An obvious drawback is that they began releasing the list in 2010, well after the mortgage credit boom. However, there is strong persistence between 2010 and 2014 of the zip codes that make the list, which suggests that mortgage fraud is a fixed characteristic of zip codes that can be used retrospectively to examine fraud during the subprime mortgage credit boom.

We also include measures of ex post outcomes in a zip code, such as income growth, the change in the poverty rate, and mortgage defaults. Mortgage default data is from Equifax. Median income growth, change in the poverty rate, and change in the unemployment rate are from the decennial census.

1.3 Accounting for second liens

In 2004, the HMDA data began separately recording first and second liens used for home purchase. This is important because a key result in A2S16 concerns the average mortgage size conditional on origination. In our view, the correct average mortgage size per purchased home with a mortgage would include the combined value of the first- and second-lien mortgages. A2S16 treat both first and second liens as independent mortgages when constructing their average mortgage size conditional on origination. For example, if a single home purchase is financed with a first lien of $100,000 and a second lien of $20,000, A2S16 treat these as two unrelated mortgages with an average size of $60,000. The total mortgage size used to purchase the home in this example is $120,000.

In Table 1, we show that growth in the average mortgage size conditional on origination when treating both the first lien and second lien as independent mortgages is 6.6% from 2004 to 2005. This is almost exactly the same annualized growth in mortgage size reported in the summary statistics of A2S16. Therefore, we are confident that A2S16 have indeed counted both first and second liens as independent mortgages when calculating average mortgage size growth.

One issue with this approach is that the fraction of all mortgages that were second liens increased from 2004 to 2005, as Table 1 shows. Further, second liens are substantially smaller than first liens. Failure to combine the first lien and second lien as one mortgage per purchased home understates the growth in the mortgage amount for home purchase conditional on origination. As Table 1 shows, the growth in the size of first-lien mortgages from 2004 to 2005 was 12%. The full intensive margin effect is best captured by the growth in the total mortgage amount per home purchased with a mortgage, which grew at 14% from 2004 to 2005. We calculate the total mortgage amount per home purchased with a mortgage in a given year by adding both the first and second liens, and dividing through by the number of first liens.

2 A Review of the Basic Facts

In this section, we review the basic facts shown in MS09 and A2S16. We first clarify the contribution of MS09. We then show that there is substantial agreement on many facts between MS09 and A2S16. Finally, we highlight the main disagreement.

2.1 Reviewing the logic of the MS09 tests

Mortgage originations for home purchase in the United States grew substantially from 2002 to 2005. What can explain this large increase? This question is difficult to answer using aggregate data alone, which is why MS09 focused on the cross-section of zip codes by credit scores. In particular, MS09 showed two crucial facts. First, the growth in mortgage credit for home purchase was much stronger in low-credit-score zip codes, defined as zip codes with a large fraction of individuals with a credit score below 660. Second, individuals in these low-credit-score zip codes saw a relative decline in average income according to the IRS. Together with a number of other results, MS09 concluded that the larger mortgage credit growth of low- relative to high-credit-score zip codes was due to a credit supply shock, or an expansion in mortgage credit by lenders that was independent of income prospects of the borrowers.

Of course, this does not conclusively prove that strong growth in aggregate mortgage originations for home purchase was caused by a shift in credit supply, and MS09 never claimed that it does. It also does not imply that high-credit-score borrowers saw only moderate credit growth; they also experienced strong mortgage credit growth, but not as strong as low-credit-score borrowers. However, it could very well be that the rise in mortgage credit growth among even high-credit-score borrowers was caused by a shift in credit supply. But the MS09 methodology was not designed to test this hypothesis.

In this study, we only respond to results in A2S16 that appear to contradict the two facts mentioned above: (i) that mortgage credit growth was larger for low-credit-score zip codes relative to high-credit-score zip codes, and (ii) that the marginal borrowers receiving credit in low-credit-score zip codes experienced a relative decline in income growth compared with those receiving credit in high-credit-score zip codes.

As a final note, it is important to note that MS09 focused on the extensive margin of mortgage lending, and the key argument in the study was that credit expanded on the extensive margin toward low-credit-score individuals, defined as those with a credit score below 660. The article does not claim that expansion in mortgage credit for home purchase to marginal borrowers explains the rise in aggregate household debt. The aggregate household debt level is not mentioned in MS09. This point is emphasized in Mian and Sufi (2014a): recall that households in the United States doubled their debt burden to $14 trillion from 2000 to 2007. As massive as it was, the extension of credit to marginal borrowers alone could not have increased aggregate household debt by such a stunning amount. In 1997, 65% of U.S. households already owned their homes. Many of these homeowners were not marginal borrowers—most of them already had received a mortgage at some point in the past." Mian and Sufi (2011, 2014b) focus on the rise in aggregate household debt, and these studies show that home equity extraction by all but the top 20% of the distribution was the most important factor in explaining the rise in aggregate household debt.

2.2 Results agreed upon by MS09 and A2S16

The left panel of Figure 2 replicates the main result from MS09. We split zip codes by the share of subprime borrowers in the zip code as of 1996, and we show the growth in the amount of mortgages originated for home purchase from 1990 to 2008 for the top and bottom quartiles. Both groups see similar mortgage growth from 1990 to 2000, and then there is a dramatic expansion in mortgage amounts originated for home purchase in the most subprime zip codes from 2000 to 2005. Column 1 of Table 2 shows this result in a regression setting with county fixed effects.[10] Zip codes with a high fraction of subprime borrowers witnessed a dramatic relative expansion in mortgage credit for home purchase from 2002 to 2005.

A2S16 confirm this finding. In their Table 1, they sort zip codes into three groups based on average IRS household income in 2002. They show that the growth in total purchase mortgage origination amounts is stronger from 2002 to 2006 in low-income zip codes (16.8% versus 7.8%). MS09 show that the fraction of subprime borrowers in a zip code in 1996 is strongly negatively correlated with average income levels in a zip code in 2000. As a result, the A2S16 finding is similar to showing that mortgage credit growth was stronger from 2002 to 2006 in low-credit-score zip codes.

Another key finding in MS09 is that these same low-credit-score zip codes saw a relative decline in income growth during the mortgage credit boom. This result is repeated in column 2 of Table 2. A2S16 also confirm this result. In their Table 1, they show that zip codes with low average IRS household income in 2002 saw lower income growth from 2002 to 2006 relative to high-income zip codes (3.5% versus 6.4%). So both MS09 and A2S16 show that the growth in total mortgage amount originated during the credit boom was stronger in low-credit-score, low-income zip codes that saw a relative decline in income growth according to the IRS.[11]

[10] In Section 6 we address issues related to the use of county fixed effects. For now, we follow MS09 and include them in all specifications.

[11] We use the 2002 to 2005 period instead of 2002 to 2006 because both home purchase mortgages and the homeownership rate peaked in 2005. In other words, the expansion of credit availability on the extensive margin peaked in 2005. Home equity–based borrowing by existing homeowners peaked in 2006.

Given that the analysis in MS09 focused on the extensive margin of mortgage credit expansion, it does not present any results on aggregate shares of total mortgage amounts originated. We do so in the right panel of Figure 2. The share of the total amount of mortgages originated for home purchase in the lowest credit score quartile went from 20% to 27% from 2002 to 2005. The share of the bottom half of the credit score distribution went from 45% to 54%.

A2S16 also confirm this finding. In Panel B of their Figure 1, they show that the lowest 20% of zip codes sorted by 2002 IRS income went from 10% of originated amounts to 12% from 2002 to 2006. The bottom 40% went from 23% to 27%. The magnitude of the increase is smaller in A2S16, but the direction is qualitatively similar.

2.3 Point of disagreement

The main point of disagreement between MS09 and A2S16 is on the income growth of marginal buyers receiving credit during the mortgage credit boom. A2S16 critique the use of IRS income growth in a zip code, and instead use the growth in income reported on mortgage applications as a more accurate measure of true home-buyer income growth during the mortgage credit boom. In the absence of fraudulent overreporting on mortgage applications, this would be a reasonable approach. The evolution of IRS income in a zip code is not a perfect proxy for the income of the marginal buyers of homes.

Panel A of Figure 1 in A2S16 sorts individuals based on income on mortgage applications, and it shows a decline in the mortgages originated to low-income individuals. In their Table 5, Panel A, they show a positive correlation between income growth on mortgage applications and mortgage credit growth in a zip code despite the negative correlation between IRS income growth and mortgage credit growth.

In columns 3 through 6 of Table 2, we explore these issues in the MS09 data. Column 3 shows a negative correlation between mortgage credit growth and IRS income growth from 2002 to 2005 across zip codes. This is an ancillary finding to the main result in MS09:

low-credit-score zip codes saw an expansion in mortgage credit despite lower income growth, and this fact is strong enough to generate a negative correlation between credit growth and income growth across zip codes.

Column 4 replicates the A2S16 finding that there is the opposite relation when using income growth in a zip code of home buyers as measured with mortgage application income. Column 5 includes both measures together. The coefficients on each measure of income growth are similar when included together, which hints that income growth according to the IRS and mortgage applications is not strongly correlated, something we confirm in Table 6 below. Column 5 implies that mortgage credit growth was strongest from 2002 to 2005 in zip codes seeing a decline in IRS income, but an increase in income according to mortgage applications.

In column 6, we regress mortgage credit growth on buyer income overstatement, which is the difference between the mortgage application income growth from 2002 to 2005 and IRS income growth over the same time period.[12] As it shows, mortgage credit growth was particularly strong in zip codes in which mortgage applications recorded substantially higher income growth than the IRS.

We already know that mortgage credit expanded more in low-credit-score zip codes, and so a logical conclusion is that buyer income overstatement is highest in low-credit-score zip codes. Column 7 confirms this fact. This correlation is what we would expect if the gap is due to mortgage fraud. Suppose a mortgage originator and potential home buyer z with true income I_z want to close a mortgage for home purchase. The originator and applicant may work together to falsify the applicant's income $\hat{I}_z \neq I_z$ depending on the size of the mortgage relative to his income potential, and the likelihood that they can get away with misreporting income. If the potential buyer has a high credit score and can therefore easily obtain a GSE-

[12] Given that buyer income overstatement is buyer income growth minus IRS income growth, column 6 is a constrained version of column 5, where we impose that the coefficients are the same on the two income growth measures but with opposite sign. We could also define borrower income overstatement as 0.43*buyer income growth minus 0.71*IRS income growth to be consistent with column 5. For the ease of interpretation, we use the raw difference as buyer income overstatement, but all results are similar if we use this alternative definition.

qualifying mortgage, then he is not credit constrained at the margin. In such a situation, there is no incentive to misreport. However, if the potential buyer has a low-credit-score, then the originator and buyer may have an incentive to overreport income $\hat{I}_z > I_z$. Such overstatement is likely to take place in zip codes where many individuals cannot obtain a GSE-qualifying mortgage.

When examining the results in Table 2, the key question is: Why is buyer income overstatement positively related to both the fraction of subprime borrowers in 1996 and mortgage credit growth from 2002 to 2005? One explanation is that higher-income individuals were buying homes in otherwise low-credit-score, low-income-growth zip codes, and this was especially true in zip codes with high mortgage credit growth. The results in the rest of this study suggest that this view is incorrect. Instead, we show below that buyer income overstatement most likely reflects fraudulent income reporting on mortgage applications, and such fraudulent income reporting was more prominent in the same subprime zip codes that MS09 show were experiencing high mortgage credit growth from 2002 to 2005.

2.4 Core issue, visually

Figure 3 reveals visually the unusual pattern of buyer income overstatement during the mortgage credit boom. We first calculate the ratio of income of home buyers reported on mortgage applications to average IRS income in a zip code. We then plot this ratio across the zip-code-level distribution of the fraction of subprime borrowers, and we separately plot this for 1998, 2001, 2005, and 2011.

One can see how unusual the mortgage credit boom was. In 2005, the ratio of buyer income from mortgage applications to IRS income was higher than it was during previous years across almost the entire distribution. So in the aggregate, income reported on mortgage applications relative to IRS income was exceptionally high in 2005, which would imply in the aggregate that high-income individuals were marginal buyers of homes during the mortgage credit boom. We have already shown in Figure 1 that this implication is contradicted in the

American Community Survey, which shows that marginal home buyers had lower income in 2005 relative to 2000.

Figure 3 also shows that mortgage application income relative to IRS income in 2005 was highest in zip codes with many subprime borrowers. In 2005, buyer income was 2.5 times higher than the average IRS income of residents in the most subprime zip codes. The ratio was between 1.7 and 2 in prior years. Equally striking is the fact that while the ratio of mortgage application to IRS-reported income jumps for low-credit-score zip codes in 2005 relative to 2001, the jump is not sustained in subsequent years. By 2011, the pattern between the income multiple and zip-code-level credit scores reverts back to its historical trend. The mortgage credit boom was anomalous.

The large gap between the 2005 line and the lines of other years in zip codes with many subprime borrowers isolates the core issue at hand. One explanation for the tremendous jump in the ratio of mortgage application income to average IRS income in low-credit-score zip codes is that high-income individuals were buying homes in low-credit-score neighborhoods. An alternative interpretation is that the jump is driven by fraudulent reporting of income on mortgage applications that especially plagued low-credit-score neighborhoods. As we show below, the evidence supports the latter view.

3 Mortgage Fraud: Existing Research and State-Level Evidence

3.1 Existing research

The overstatement of income on mortgage applications has been well documented in various commission reports and the existing literature. For example, as the Financial Crisis Inquiry Commission reports, “Ann Fulmer, vice president of business relations at Interthinx, a fraud detection service, told the FCIC ... that about $1 trillion of the loans made during the [2005 to 2007] period were fraudulent" (FCIC, 2011, 160p). Zingales (2015) reports that \$113B of fines have been levied against lenders based on mortgage fraud during the housing boom, and he further emphasizes that this number "severely underestimates the magnitude of the problem." On the specific issue of mortgage application income, several articles and lawsuits find pervasive evidence of fraudulent income overstatement.[13]

Academic research supports the argument that fraud was endemic to mortgage markets during this period, especially among subprime mortgages (see, e.g., Ben-David 2011; Griffin and Maturana 2016a, b; Piskorski, Seru, and Witkin 2015). The three most relevant studies for buyer income overstatement are Avery et al. (2012), Blackburn and Vermilyea (2012), and Jiang, Nelson, and Vytlacil (2014).[14] The first two studies show that income reported on applications in the HMDA data set is systematically overstated. Avery et al. (2012) compare income reported on HMDA home purchase applications with income reported by recent home buyers with a mortgage in the 2000 Census and the 2005 American Community Survey. They show a large gap opens up from 2000 to 2005, with HMDA showing much higher income growth of recent home buyers than the Census/ACS.

They conclude that "users of the HMDA data should be aware that borrower income was likely significantly overstated during the peak of the housing boom, particularly in some areas of the country. One potential implication of this finding is that lending to lower-income borrowers, as measured in the HMDA data, may be attenuated around the peak

[13] An article by Matt Taibbi cites an employee at JPMorgan Chase who discovered that "around 40 percent of [mortgages] were based on overstated incomes" ("The \$9 Billion Witness: Meet JPMorgan Chase's Worst Nightmare," *Rolling Stone*, November 6, 2014). A complaint against CountryWide filed in Illinois reports that "the Mortgage Asset Research Institute reviewed 100 stated income loans, comparing the income on the loan documents with the borrowers' tax documents. The review found that almost 60% of the income amounts were inflated by more than 50%". The Federal Reserve Board assessed an \$85 million civil money penalty against Wells Fargo because employees "separately falsified income information in mortgage applications." According to an article in the *Los Angeles Times* by Mike Hudson and Scott Reckard, employees at mortgage lender Ameriquest testified that they had witnessed behavior including "deceiving borrowers about the terms of their loans, forging documents, falsifying appraisals and fabricating borrowers' income to qualify them for loans they couldn't afford" (Mike Hudson and E. Scott Reckard "Workers Say Lender Ran 'Boiler Rooms,'" *Los Angeles Times*, February 4, 2005). These practices are discussed extensively in Hudson (2011). We are grateful to Binyamin Appelbaum of the *New York Times* for providing us with many of these cites.

[14] A fourth study by Garmaise (2015) focuses on misrepresentation of assets rather than income.

of the housing market." It is worth noting that the analysis in Panel A of Figure 1 of A2S16 does what Avery et al. (2012) warn not to do: they use HMDA measures of income from mortgage applications to assert that lending did not increase disproportionately to low-income borrowers.

Blackburn and Vermilyea (2012) conduct a similar exercise but use the American Housing Survey instead of the Census/ACS. Using the AHS, they find recent home buyers who use a mortgage, and they do their best to match these home buyers with their HMDA mortgage application. They then compare AHS reported income and mortgage application reported income. They conclude: "Our results suggest a substantial degree of income overstatement in 2005 and 2006, one consistent with the average mortgage application overstating income 15 to 20%."

Jiang, Nelson, and Vytlacil (2014) use a research design focused on mortgages originated by a single bank from 2004 to 2008, and they conclude that low-documentation mortgage applications inflated incomes by an average of 28.7% relative to high-documentation mortgages. Jiang, Nelson, and Vytlacil (2014) make clear that their results should be viewed as a lower bound given the strategy. It assumes honest reporting on full-documentation loans in the non-agency market, and it is limited by the inability to see true income of the low-documentation buyers.

3.2 State-level evidence on income overstatement

Figure 1 shows that mortgage application income shows positive growth from 2000 to 2005 even though measures of home-buyer income growth from the ACS are negative. Unfortunately, the ACS does not contain enough households to examine this gap at a micro-level, such as a zip code. As a result, we use the gap between mortgage application income growth and IRS income growth to construct a zip-code-level measure of buyer income overstatement. The concern with the latter measure is that the gap between mortgage application income growth and IRS income growth reflects high-income-growth households buying homes in otherwise low-income-growth neighborhoods.

One way to alleviate this concern is to examine state-level evidence, where we have enough observations in the ACS to compare the two separate measures of buyer income overstatement. We do so in Figure 4. On the horizontal axis is HMDA income growth minus ACS home-buyer income growth from 2000 to 2005. On the vertical axis is HMDA income growth minus IRS income growth, where we use 2000 to 2005 to make the two measures comparable. We weight the states based on the number of observations in the 2005 ACS given small sample sizes.

As the figure shows, there is a strong positive correlation between the two measures. States with high buyer income overstatement using average IRS income as the benchmark are the same states with high buyer income overstatement using ACS data on recent home buyers. This supports the argument that buyer income overstatement using IRS income growth is a measure of fraud. Recall that the measure on the horizontal axis is a clear measure of potential fraud that is independent of a change in buyer composition—the HMDA data and the ACS data are both for recent home buyers.

4 Mortgage Fraud: Zip-Code-Level Evidence

4.1 Direct measures of fraud

Table 3 shows evidence that mortgage fraud was more prevalent in zip codes with high borrower income overstatement. In columns 1 and 2, we show that zip codes with high overstatement saw a larger increase in the share of mortgages originated for non-agency securitization. This is almost completely driven by an increase in the share of low-documentation mortgages originated for non-agency securitization. We know from research cited earlier that low-documentation non-agency securitization experienced elevated mortgage fraud from 2002 to 2005, and in particular fraudulent overstatement of income.

In columns 3 and 4 of Table 3, we use fraud information from Interthinx. Column 3 shows that zip codes with high buyer income overstatement from 2002 to 2005 are more likely to make the Interthinx top fraud list in 2010, and column 4 shows that they are more likely to make the Interthinx top fraud list at some point between 2010 and 2014. The magnitude of the income overstatement in top fraud zip codes is large: In the 18 zip codes that made the top fraud list at some point in 2010, buyer income overstatement from 2002 to 2005 was more than 10 percentage points, compared with an average of about two percentage points for the rest of the sample.

In columns 5 through 7 of Table 3, we use data from the study by Piskorski, Seru, and Witkin (2015) on mortgage fraud in the misreporting of owner-occupied status and the presence of second liens. As columns 5 through 7 of Table 3 show, there is a strong correlation between buyer income overstatement in a zip code and these alternative measures of fraud. In summary, columns 3 through 7 show that mortgage fraud was more likely in the zip codes where mortgage applications reported higher buyer income than local resident income.

4.2 Do high borrower income overstatement zip codes improve?

As mentioned before, fraud, by its nature, is difficult to detect. We present evidence in the previous section that fraud was prevalent in the zip codes where mortgage applications significantly overstated income growth. In this subsection, we take a different approach. If high-income-growth individuals were buying homes in low-income-growth zip codes, we should eventually see some evidence of improvement.

The results in Table 4 are from the following cross-sectional regression we estimate for each period $t-1$ to t:

$$ln(IRSInc_{czt}) - ln(IRSInc_{cz,t-1}) = \alpha_c + [ln(Buyer_{cz,t-1}) - ln(IRSInc_{cz,t-1})] * \gamma_t + \varepsilon_{zt} \quad (2)$$

In words, this specification tests whether IRS income grows faster between $t-1$ and t in zip codes where at $t-1$ buyer income is higher than IRS income. This is a simple test for gentrification: does the presence of home buyers with higher than average income reported on mortgage applications in the zip code predict future IRS income growth in the zip code? Zip-code-level IRS income data are available for 1991, 1998, 2002, 2004, 2005, and 2006, which means we run five cross-sectional regressions using the periods 1991 to 1998, 1998 to 2002, 2002 to 2004, 2004 to 2005, and 2005 to 2006.

Prior to the mortgage credit boom, we find statistically significantly positive estimates for γ: a positive difference between income on mortgage applications and average IRS income predicts higher IRS growth. But from 2002 to 2005, the relationship reverses. During the boom, a positive difference between buyer and IRS income at time t predicts a decline in IRS income growth from t to $t+1$. Zip codes with high buyer income growth relative to IRS income growth experienced subsequently worse economic performance during the mortgage credit boom, not better. The estimates are statistically significantly negative, indicating a clear structural break from the early period. This is consistent with mortgage fraud during the boom, and inconsistent with gentrification.

In column 6, we regress the change in the fraction of IRS returns in the zip code that have greater than $50,000 in adjusted gross income. High buyer income overstatement zip codes experience a relative decline in IRS returns with high incomes during the boom. Column 7 uses credit scores of individuals in the data used in Mian and Sufi (2011). We measure the credit scores of people moving into a zip code from 2002 to 2005 minus the credit scores of people living in the zip code as of 2002 according to the credit bureau. The point estimate suggests that high buyer income overstatement zip codes saw a relative decline in the credit scores of individuals moving in versus those living in the zip code, contradicting the argument that these zip codes were gentrifying. Recall that high buyer income overstatement zip codes have lower credit scores in 2002, and those moving in during the 2002 to 2005 boom did not have higher scores.

In Table 5, we take a longer view by regressing measures of future performance on buyer income overstatement from 2002 to 2005. Zip codes with high buyer income overstatement saw lower IRS income and wage growth from 2005 to 2012, a decline in Census income growth from 2000 to 2010, and an increase in both poverty rates and unemployment rates from 2000 to 2010. The poverty result is especially revealing. Poverty rates in 2000 were already significantly higher in zip codes with high buyer income overstatement from 2002 to 2005. And yet they jumped even higher from 2000 to 2010. As columns 6 and 7 show, the default rate also jumped substantially higher in zip codes with high buyer income overstatement. This latter result is consistent with Jiang, Nelson, and Vytlacil (2014), who find that fraudulent income overstatement on mortgage applications predicts default.

Figure 5 shows real income growth from 2005 to 2012 by buyer income overstatement during the boom. We plot income growth according to three measures: IRS income growth, income reported on mortgage applications for home purchase, and income reported on mortgage applications for refinancing. According to all three measures, there was an absolute decline in real income in zip codes with high borrower income overstatement during the boom. The decline is 40% for home purchase applications in high buyer income overstatement zip codes. This reflects how unusually high the income reported on mortgage applications was in 2005.

The results in Tables 4 and 5 and Figure 5 are inconsistent with the argument that high-income individuals were moving into zip codes with high buyer income overstatement. Instead, the results are more consistent with fraudulent income overstatement on mortgage applications in these zip codes from 2002 to 2005.

4.3 The decoupling of mortgage application and IRS income

How unusual is it that income growth of home buyers reported on mortgage applications deviates strongly from IRS-reported income growth of a zip code? Table 6 regresses the growth in income reported on mortgage applications of home buyers in a zip code on IRS-reported income growth of the zip code over various time periods between 1991 and 2007.

We find a significant reduction in the correlation between mortgage-application-reported income growth of home buyers and IRS income growth from 2002 to 2005. In fact, we can easily reject the hypothesis that the coefficient estimate from 2002 to 2005 is the same as it was before the mortgage credit boom. The correlation between mortgage application income growth and IRS income growth breaks down significantly during the credit boom period.

Why does the correlation between buyer income growth and IRS income growth break down from 2002 to 2005? Research cited earlier suggests that fraudulent overstatement of income was common among mortgages sold into the non-GSE securitization market. Table 7 investigates this by estimating this correlation separately for the four quartiles of zip codes by GSE share during 2002 to 2005.[15] We classify zip codes into the four quartiles according to the average share of mortgages sold to GSEs for securitization from 2002 to 2005. We keep zip codes in the same category for all time periods.

The results show that the breakdown in the correlation between buyer income growth and IRS income growth during 2002 to 2005 is entirely driven by zip codes with a low share of non-GSE mortgages. There is no change in the correlation between buyer income growth and IRS income growth in the zip codes with a high share of GSE mortgages. Further, the correlation between the growth in buyer income on mortgage applications and IRS income growth is positive in high non-GSE mortgages outside the 2002 to 2005 period. The decoupling is concentrated when and where fraud was most likely: high non-GSE-share zip codes during the mortgage credit boom.

A statistical test on equality of coefficients shows a clear structural break in the 2002 to 2005 period. More specifically, for each period, we test whether the coefficient on the IRS income growth by Quartile 1 GSE share interaction is equivalent to the coefficient on the IRS income growth by Quartile 4 GSE share interaction. We can reject equivalence only for the 2002 to 2005 period, supporting the view that the breakdown in the correlation between IRS income and mortgage application income is unique to zip codes experiencing a rise in private-label securitization during the mortgage credit boom.

[15]Sorting by 1996 zip-code-level credit scores instead of GSE share leads to qualitatively similar results.

The results in Table 7 support the view that income reported on mortgage applications in the GSE market reflected fundamental income during 2002 to 2005, whereas income reported on mortgage applications in the non-GSE market was fraudulently overstated. This supports the argument in Keys et al. (2012, 2071p), who note: “Our results suggest that the policy debate regarding securitization and lenders’ underwriting standards should separately evaluate the agency and non-agency markets.”

4.4 Tests based on stability of coefficients

While A2S16 acknowledge that fraudulent overstatement of income was prominent during the mortgage credit boom, they argue that it does not explain their results. They conduct two tests to support their argument. First, they look separately at the correlation between growth in mortgage application income and mortgage credit growth within high-GSE-share zip codes from 2002 to 2005, and they find that this correlation remains as strong as in the subsample of low-GSE-share zip codes. The authors argue that the stability of the positive correlation in the two subsamples suggests that fraud is not driving their main result. Second, they show the stability of the same coefficient between mortgage credit growth and mortgage application income growth in periods outside of the 2002 to 2005 mortgage credit boom.

The results shown by A2S16 are consistent with the view that fraud is responsible for the full-sample correlation between mortgage application income and credit growth from 2002 to 2005. In assessing their tests, it is crucial to recognize two points. First, as shown in Table 7, mortgage application income growth reflects true income growth in all other periods and in all zip codes except for within low-GSE-share zip codes from 2002 to 2005. Second, credit demand factors were more powerful in explaining mortgage credit growth outside of the 2002 to 2005 period (as shown in MS09), and even within high-GSE-share zip codes from 2002 to 2005.

As a result of these two facts, what would one expect to find in other samples? In other samples, credit demand factors such as income growth drive the variation in mortgage credit growth, and mortgage application income growth reflects true income growth. We would therefore expect to see a positive correlation during these other periods and within high-GSE-share zip codes from 2002 to 2005, which is what A2S16 show.

But can one then immediately conclude that credit demand factors are also responsible for the full-sample correlation from 2002 to 2005? The key difference from other periods is that from 2002 to 2005 within low-GSE-share zip codes, mortgage application income growth no longer reflects true income growth. We show this in Figure 3 and Tables 6 and 7. A2S16 never deny this fact. A positive coefficient outside the 2002 to 2005 period reflects normal credit demand and truthfully reported mortgage application income, but a positive coefficient from 2002 to 2005 in low-GSE-share zip codes reflects fraud.

Put differently, the stability of coefficients in regressing credit growth on mortgage application income growth is far less informative than the fact that IRS income growth and mortgage application income growth become uncorrelated from 2002 to 2005 in low-GSE-share zip codes, despite being positively related in all other time periods and within the high-GSE-share market from 2002 to 2005. This is a more direct test of fraud.

In general, it is important to note that tests conducted within the GSE market take out the most important source of variation in the data, which is the cross-comparison across GSE and non-GSE markets (or in the language of MS09, prime versus subprime zip codes). In particular, the key argument in MS09 was that subprime zip codes within a county were experiencing different trends relative to prime zip codes because of the shift in credit supply. It is not obvious that correlations within the GSE market tell us much about what happened in the non-GSE market relative to the GSE market.

5 Average Mortgage Size

MS09 show that the growth in the total amount of mortgage originations was stronger in zip codes with declining IRS income growth from 2002 to 2005. We argued that this result reflects a shift in the supply of credit that other tests in MS09 verify. A2S16 confirm the MS09 result using the growth in mortgage amounts originated as the dependent variable (see their Table 2, Panel A, column 2). However, A2S16 also use the growth in average mortgage size conditional on origination as the dependent variable and show that the growth in average mortgage size conditional on origination was positively correlated with IRS income growth from 2002 to 2005.

In Table 8, we show that the positive correlation between IRS income growth and average mortgage size growth in A2S16 is due to the manner in which they calculate mortgage sizes in their study. In particular, they treat first and second liens as independent mortgages; they do not combine them when assessing the total mortgage size used to purchase a new home.

In Panel A of Table 8, we examine the correlation between average mortgage size and the subprime share of a zip code in 1996. Information on first and second liens becomes available in the HMDA data from 2004 onward. We first confirm in column 1 that mortgage credit growth was stronger in low-credit-score zip codes from 2004 to 2005, a fact that can also be seen in the left panel of Figure 2.

In column 2 of Table 8, we construct the average mortgage size conditional on origination in the same way as A2S16. More specifically, we treat second liens as independent mortgages; for example, a $100,000 first lien and a $20,000 second lien count as two separate mortgages with an average size of $60,000. When constructed in this manner, it appears as if the average mortgage size conditional on origination is falling in zip codes with many subprime borrowers. In column 3, we show why this is problematic: the fraction of all mortgages originated that are second liens increased disproportionately in zip codes with many subprime borrowers. Second liens are smaller than first liens, which brings down the average mortgage size in low-credit-score zip codes if second liens are treated as independent mortgages.

In column 4, we show that the average mortgage size of first liens conditional on origination is uncorrelated with zip-code-level credit scores. The left-hand-side variable in column 5 is based on what we believe is the correct average mortgage size measure: the total mortgage amount per unit purchased with a mortgage. This measure combines the first and second liens, which is the total mortgage debt used in the transaction. As it shows, the growth in total mortgage size conditional on origination actually increased in low-credit-score zip codes from 2004 to 2005. Therefore, credit expanded for home purchase on both the intensive and extensive margins in low-credit-score zip codes during the mortgage credit boom.

To make the results more directly comparable to those in A2S16, Panel B uses IRS income growth from 2004 to 2005 as the right-hand-side variable. It shows similar results. Column 1 shows that total mortgage credit grew more in low-income-growth zip codes from 2004 to 2005. When following the A2S16 measure of average mortgage size, which treats first and second liens as independent mortgages, there is a positive correlation between IRS income growth and average mortgage size growth conditional on origination. This is similar to the result presented in A2S16, Table 2, column 5.

But high-income-growth zip codes saw a relative decline in second liens. As a result, when using what we believe is the correct measure that combines the first and second liens, we find that there is a negative correlation between IRS income growth and average mortgage size growth. low-income-growth zip codes experienced a relative increase in mortgage credit growth on both the intensive and extensive margins. The coefficient estimate is only marginally statistically signficantly different than zero. But there is certainly no evidence that the correlation between IRS income growth and average mortgage size is positive, as argued in A2S16.

We should note that the results in Panel A of Table 8 are even stronger than the claims made in MS09. A decline in the average mortgage size conditional on origination would not contradict the credit supply view that credit expanded on the extensive margin in low-credit-score zip codes. For example, one could easily imagine that marginal home buyers in low-credit-score zip codes that were previously denied credit obtain smaller mortgages than average, thereby pulling down the average mortgage size conditional on origination in low-credit-score zip codes.

However, the results in Table 8 show that marginal home buyers were not only more likely to obtain a mortgage, but the average mortgage size conditional on origination actually grew by more than in high-credit-score zip codes. Failure to combine first and second liens leads to what we believe is the incorrect conclusion average mortgage size growth in low-credit-score zip codes.

6 Fixed Effects, Credit Scores, Defaults, and Other Issues

6.1 Use of county fixed effects

The main test in MS09 focused on the extensive margin of mortgage credit expansion. In particular, the goal was to hold all other characteristics constant, and to see if mortgage credit disproportionately expanded toward low-credit-score individuals. The ideal test would be to have two households with different credit scores but identical in every other aspect, such as house price growth, income growth, and demographics. The test would then be to see whether credit expanded more to the low-credit-score household. Given large differences across counties in house price growth, business cycle shocks, and ex ante demographics, all of the specifications in MS09 include county fixed effects to most carefully replicate this ideal test.

While we believe that the within-county variation most closely replicates the ideal test, Table 9 shows that most of the key MS09 results are robust to using between-county variation in addition to within-county variation. In particular, as Panel A shows, mortgage credit expanded more to subprime borrowers using either within-county or between-county variation. The OLS specification, which exploits all of the variation in the data, is positive, large, and statistically significant.

Likewise, the OLS correlation between IRS income growth and the fraction of subprime borrowers is negative, indicating that subprime zip codes saw an expansion of credit despite lower income growth. The correlation is strongly negative within counties, and close to zero between counties. In the online appendix, we show that using the full sample, the correlation between income growth and the fraction of subprime borrowers is negative using between-county variation as well. Panel A of Table 9 confirms the core MS09 finding even when using the within- and between-county variation.

The difference in results when excluding county fixed effects shows up when examining the correlation between mortgage credit growth and IRS income growth. As Panel B shows, the within-county and between-county results have the opposite sign, a point made in the introduction of MS09. The MS09 interpretation of this fact is that counties seeing higher credit growth had higher income growth, but it was not the people with the high-income-growth seeing the credit growth. Recall that the goal of the analysis in MS09 was to test whether credit expanded to low-credit-score households at the margin. It is crucial to capture the income growth of the people actually obtaining the credit. This is why the county fixed effects get closer to the ideal test mentioned above.

In fact, as columns 4 and 5 of Table 9 show, one does not even need to include county fixed effects to see the reversal. State fixed effects are sufficient. In other words, the between-county positive correlation between credit growth and IRS income growth is being driven by state-level differences. State-level correlations between income growth and credit growth could be polluted by many omitted factors; we do not believe that state-level regressions isolate how income growth evolved for the actual people expanding credit.

6.2 Full sample versus house price sample

A2S16 show in their Table IA.2 in the appendix that the negative correlation between IRS income growth and mortgage credit growth becomes positive if one uses the full sample of zip codes as opposed to the sample of zip codes with house price data available. We discuss this issue at length in the appendix. The basic disagreement is due to the fact that A2S16 do not weigh zip codes by total population when they move to the full sample. At the 10th percentile of the distribution, the total number of households living in a zip code as of 2000 in the full sample is 748. At the extreme, zip codes in the full sample are as small as three or four households. Such very small zip codes lead to problems with extreme outliers. Zip codes in the full sample but not in the house price sample used in MS09 are twice as likely to have outliers in the top or bottom 1% of the credit growth distribution. In the appendix, we show that the negative correlation between IRS income growth and mortgage credit growth is robust in the full sample once zip codes are weighted by total population, or if we remove very small zip codes from the sample.

7 Conclusion

As shown in MS09, the growth in mortgage credit for home purchase was strongest in subprime zip codes from 2002 to 2005. What explains this pattern? As MS09 show, IRS income growth in these zip codes was negative in relative terms, and was even negative in real terms in many of these zip codes. However, these same subprime zip codes saw strong growth in the income being reported on mortgage applications of home buyers. One conclusion would be that subprime zip codes were seeing high-income individuals buying homes in these areas despite deteriorating conditions.

In this study, we show that such a conclusion is incorrect. Instead, we show that income reported on mortgage applications was most likely fraudulently overstated in exactly the subprime zip codes experiencing the strongest mortgage credit growth. Once we acknowledge that income on mortgage applications was likely fraudulently reported, the core result of MS09 remains clear: the expansion of mortgage credit for home purchase to subprime zip codes was unrelated to fundamental improvements in economic circumstances.

A shift in the supply of credit drove the growth in mortgage credit for the purpose of home purchase in these zip codes. This finding is important because it supports recent models arguing that shifts in credit supply are a fundamental shock driving house prices, household debt, and macroeconomic fluctuations. A number of other studies also conclude that housing dynamics from 2000 to 2007 are best described by a shift in credit supply; we summarize these other studies in a companion piece (Mian and Sufi 2016).[16]

A new result in this study is that the average mortgage size conditional on origination also grew by more in low-credit-score zip codes experiencing slower income growth. In other words, there was an expansion in mortgage credit to low-credit-score individuals from 2002 to 2005 even on the intensive margin.

Our findings point to two questions for future research. First, who committed the fraud? The borrower or the mortgage originator? Normally, it would be the duty of the mortgage originator to help stem misreporting. However, during the mortgage credit expansion from 2002 to 2005, we know originators failed to monitor and screen potential borrowers (e.g., Keys et al. 2010). In fact, there are numerous examples where mortgage brokers or originators may have falsified income information by borrowers without the borrowers' knowledge.[17]

Second, why did mortgage fraud explode from 2002 to 2005? One potential answer is that the outward shift in mortgage credit supply itself was responsible for higher fraud. For example, press reports show that fraudulent income overstatement was perpetrated by brokers originating mortgages designed to be sold into the non-agency securitization market. We look forward to more research addressing these questions.

[16]The companion piece also addresses the claim in A2S16 that the share of defaults on mortgages made to low-credit-score individuals at the time of issuance declined during the mortgage default crisis relative to the 2003 to 2005 period. As we show there, defaults among low-credit-score individuals were the main driver of the mortgage default crisis from 2007 to 2009.

[17]See, for example, the series of award-winning articles by Binyamin Appelbaum on mortgage fraud in Charlotte in the *Charlotte Observer*, available at: http://www.charlotteobserver.com/2009/06/04/762690/sold-a-nightmare-part-1-of-4.html#.VMvJXmjF9g0.

References

Adelino, M., A. Schoar, and F. Severino. 2016. Loan originations and defaults in the mortgage crisis: The role of the middle class. *Review of Financial Studies* 29:1635–70.

Avery, R., N. Bhutta, K. Brevoort, and G. Canner. 2012. The mortgage market in 2011: Highlights from data reported under the Home Mortgage Disclosure Act. *Federal Reserve Bulletin* 98:1–46.

Ben-David, I. 2011. Financial constraints and inflated home prices during the real-estate boom. *American Economic Journal: Applied Economics* 3:55–87.

Blackburn, M., and T. Vermilyea. 2012. The prevalence and impact of misstated incomes on mortgage loan applications. *Journal of Housing Economics* 21:151–68.

Favilukis, J., S. Ludvigson, and S. Van Nieuwerburgh. forthcoming. Macroeconomic implications of housing wealth, housing finance, and limited risk-sharing in general equilibrium. *Journal of Political Economy*.

Federal Reserve. 2005. Frequently asked questions about the new HMDA data. Tech. rep., https://www.federalreserve.gov/boarddocs/press/bcreg/2005/20050331/attachment.pdf.

Garmaise, M. 2015. Borrower misreporting and loan performance. *Journal of Finance* 70:449–84.

Griffin, J., and G. Maturana. 2016a. Did dubious mortgage origination practices distort house prices? *Review of Financial Studies* 29:1671–708.

———. 2016b. Who facilitated misreporting in securitized loans? *Review of Financial Studies* 29:384–419.

Hudson, M. 2011. *The monster: How a gang of predatory lenders and wall street bankers fleeced America–and spawned a global crisis.* New York: St. Martin's Griffin.

Jiang, W., A. A. Nelson, and E. Vytlacil. 2014. Liars loan? Effects of origination channel and information falsification on mortgage delinquency. *Review of Economics and Statistics* 96:1–18.

Justiniano, A., G. Primiceri, and A. Tambalotti. 2015. Credit supply and the housing boom. National Bureau of Economic Research Working Paper, #20874.

Keys, B., T. Mukherjee, A. Seru, and V. Vig. 2010. Did securitization lead to lax screening? Evidence from subprime loans. *Quarterly Journal of Economics* 125:307–62.

Keys, B., A. Seru, and V. Vig. 2012. Lender screening and the role of securitization: Evidence from prime and subprime mortgage markets. *Review of Financial Studies* 25:2071–108.

Mian, A., and A. Sufi. 2009. The consequences of mortgage credit expansion: Evidence from the U.S. mortgage default crisis. *Quarterly Journal of Economics* 124:1449–96.

———. 2011. House prices, home equity-based borrowing, and the U.S. household leverage crisis. *American Economic Review* 101:2132–56.

———. 2014a. *House of debt: How they (and you) caused the Great Recession, and how we can prevent it from happening again.* Chicago: University of Chicago Press.

———. 2014b. House price gains and U.S. household spending from 2002 to 2006. Tech. rep., National Bureau of Economic Research.

———. 2016. Household debt and defaults from 2000 to 2010: The credit supply view. Working Paper, University of Chicago Booth School of Business.

Piskorski, T., A. Seru, and J. Witkin. 2015. Asset quality misrepresentation by financial intermediaries: Evidence from the RMBS market. *Journal of Finance* 70:2635–78.

Schmitt-Grohé, S., and M. Uribe. 2016. Downward nominal wage rigidity, currency pegs, and involuntary unemployment. *Journal of Political Economy* 124:1466–514.

Zingales, L. 2015. Does finance benefit society? *Journal of Finance* 70:1327–63.

Figure 1
Income growth of marginal home buyers during credit expansion: ACS versus mortgage applications

The left panel of this figure uses data from the American Community Survey and shows income in 2005 dollars of individuals with a mortgage that bought a home within the past year and the income of all homeowners. The right panel compares income growth of individuals with a mortgage that bought a home within the past year from 2000 to 2005 according to the American Community Survey versus the growth in income reported on home-purchase mortgage applications from 2000 to 2005.

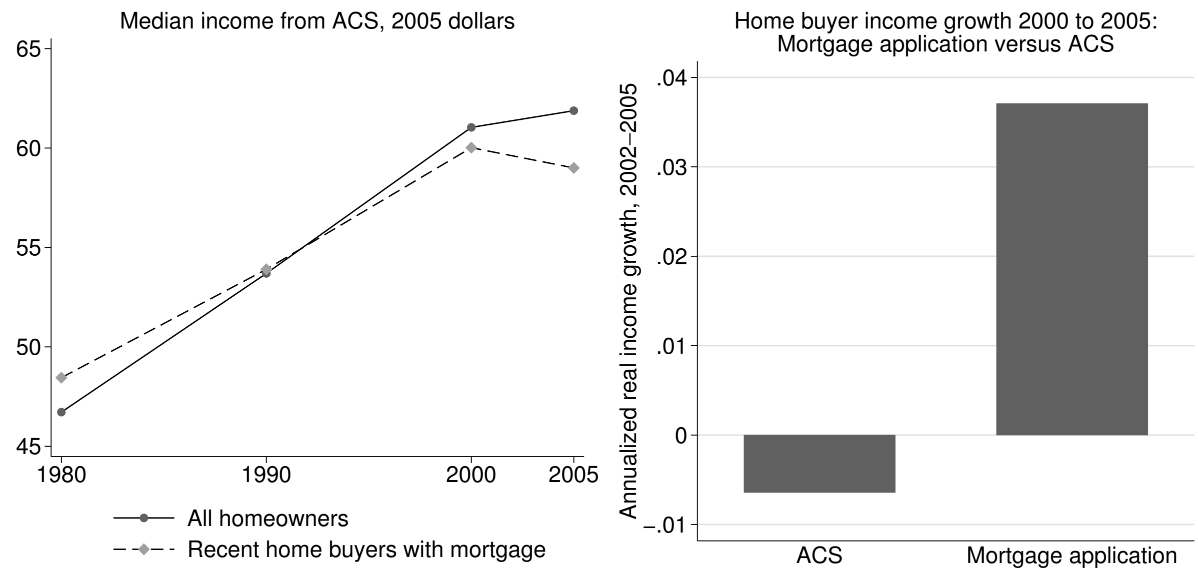

Figure 2

Home-purchase mortgage amount growth in zip codes, by 1996 credit score

The left panel of this figure shows the growth in home-purchase mortgage origination amounts in the most subprime and prime zip codes. Zip codes are sorted as of 1996 based on the share of individuals in the zip code with a credit score below 660, and the most subprime and prime zip codes are the top and bottom quartiles of the distribution, respectively. Zip codes are sorted in 1996, and remain in the same group throughout the sample. The right panel shows the share of home-purchase mortgage originations amounts in 2002 and 2005 for all four quartiles of the zip codes sorted again on the 1996 credit score.

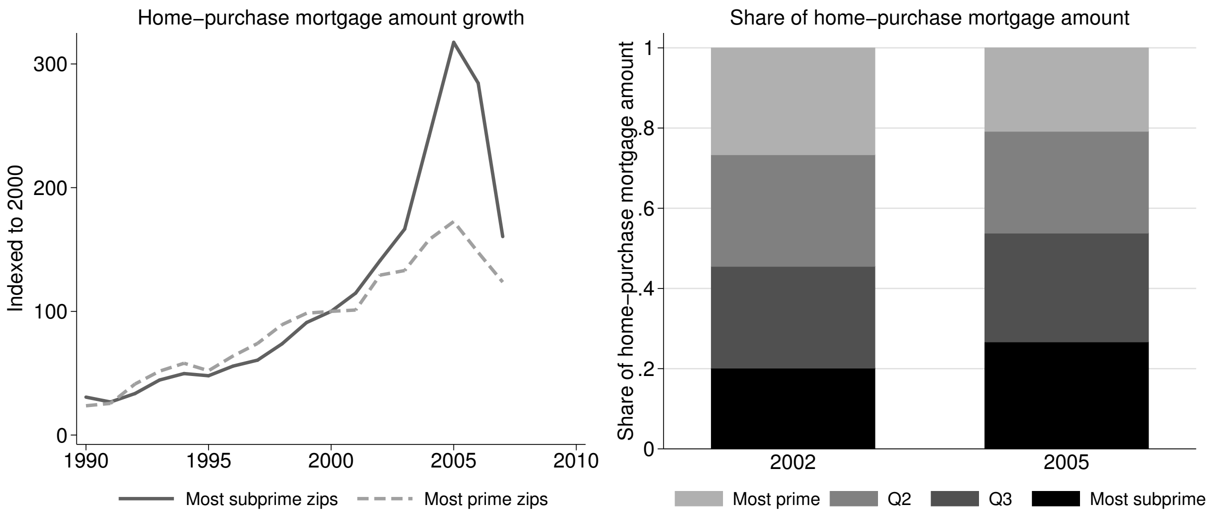

Figure 3
Ratio of income reported on mortgage applications to average IRS income of a zip code

This figure plots the ratio of income reported on mortgage applications of home buyers to average IRS income in a zip code across the distribution of the share of subprime borrowers in the zip code as of 1996. We plot this ratio for 1998, 2001, 2005, and 2011. The point is to show that 2005, the peak of the subprime mortgage credit boom, was unusual in that buyer income reported on mortgage applications was much higher than average IRS income, especially in zip codes with a large number of subprime borrowers. We believe that the gap between 2005 and the other years at the upper end of the subprime distribution reflects fraudulent reporting of income on mortgage applications.

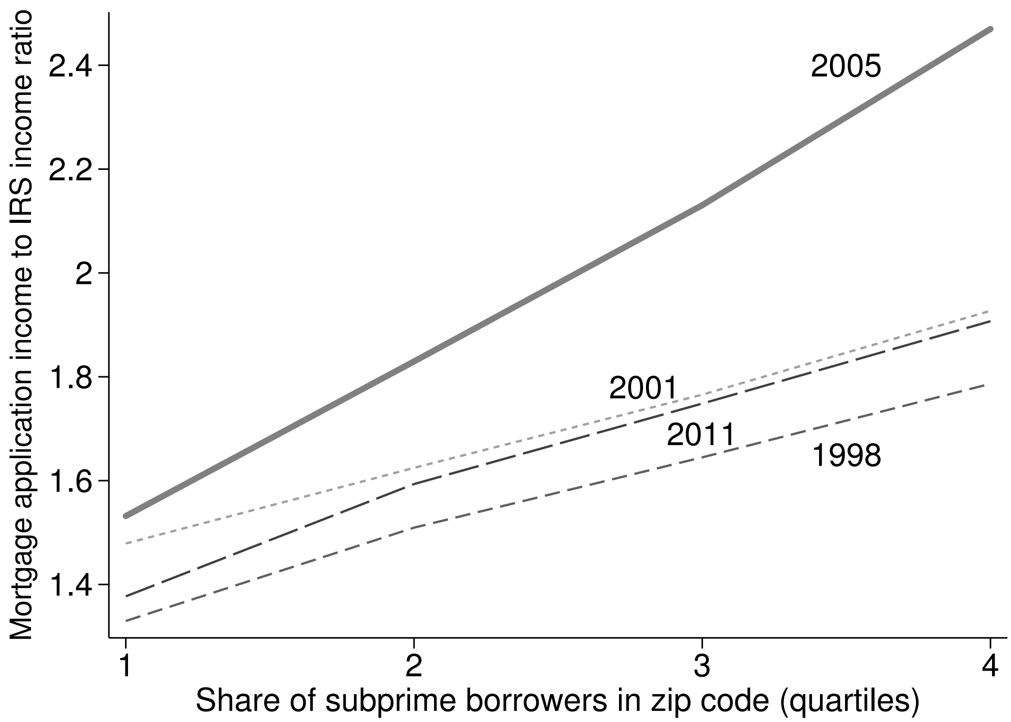

Figure 4
State-level income overstatement on mortgage applications
This figure plots buyer income overstatement at the state level from 2000 to 2005 against income overstatement on mortgage applications relative to home-buyer income from the ACS/Census. The latter variable is defined to be the annualized income growth from HMDA mortgage applications from 2000 to 2005 minus the annualized income growth of home buyers with a mortgage reporting a move within the past year in the ACS/Census. The central point is that buyer income overstatement comparing HMDA with IRS data with highly correlated with buyer income overstatement comparing HMDA with data focused only on recent home buyers with a mortgage. Each state is weighted by the number of households in the ACS in 2005, given that some states have very few households surveyed.

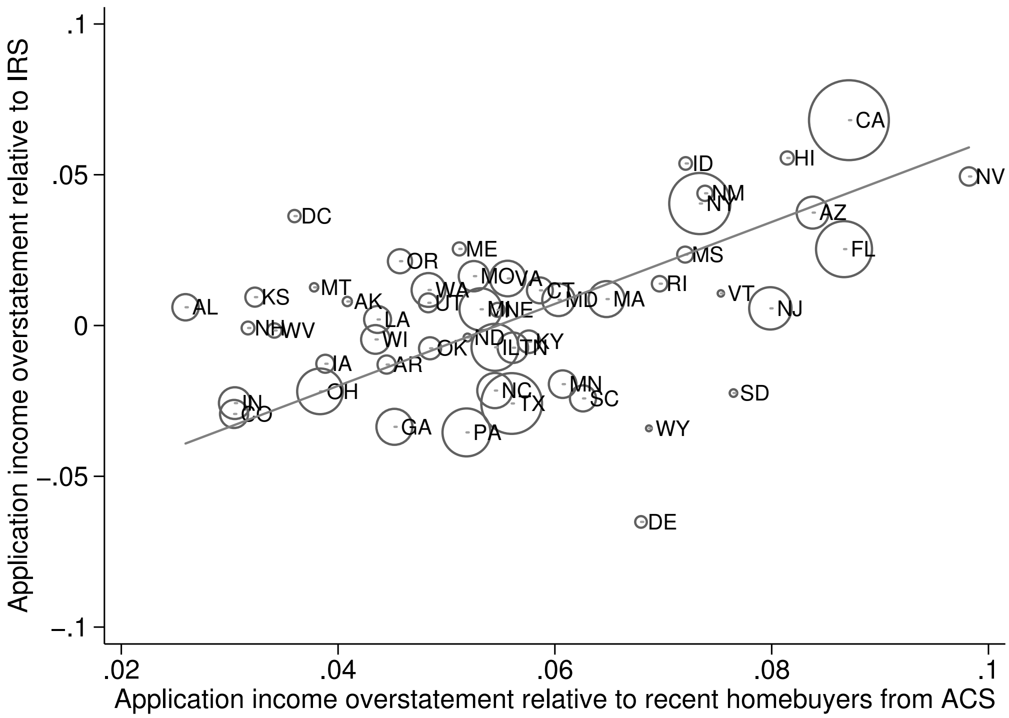

Figure 5
Real income growth from 2005 to 2012, by buyer income overstatement from 2002 to 2005

This figure shows income growth from 2005 to 2012 by the quartiles of buyer income overstatement from 2002 to 2005. More specifically, buyer income overstatement is defined to be the difference between the annualized growth in income reported on mortgage applications of home buyers from 2002 to 2005 and the annualized IRS income growth of households living in a zip code from 2002 to 2005. We show real income growth according to the IRS (left), mortgage applications for home purchase (middle), and mortgage applications for refinancing (right). The central point is that the zip codes that saw the highest growth in income reported on home purchase mortgage applications from 2002 to 2005 had the lowest real income growth from 2005 to 2012 according to many measures.

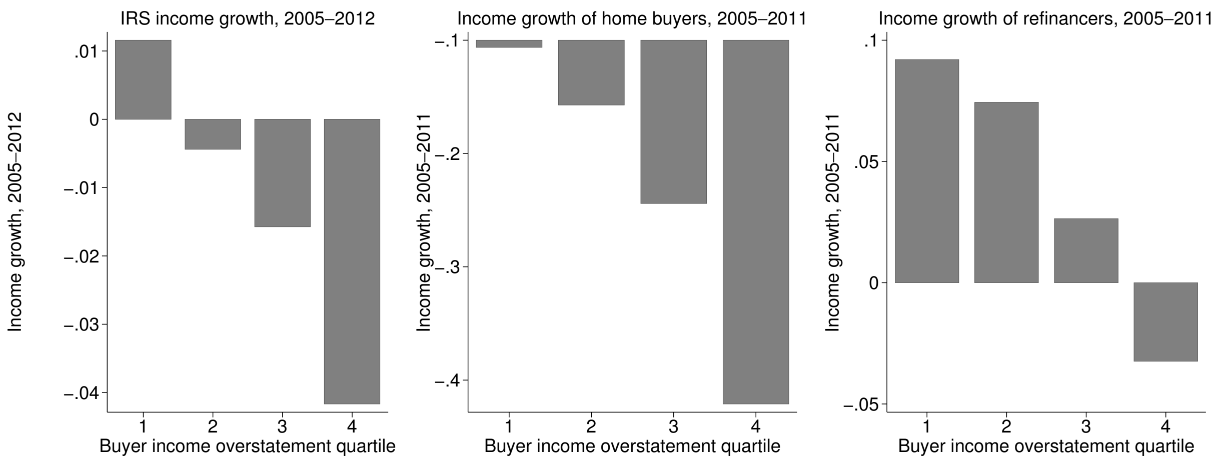

Table 1
Summary statistics
This table presents summary statistics for zip codes in our sample. The fraction of subprime borrowers is the fraction of individuals with a credit score below 660 living in the zip code. All growth rates are annualized.

	Obs	Mean	SD	P10	P50	P90
Main measures						
Fraction subprime borrowers, 1996	3,104	0.287	0.113	0.154	0.269	0.446
Amount for home purchase growth, 2002–2005	3,104	0.194	0.189	0.014	0.170	0.405
IRS income growth, 2002–2005	3,104	0.048	0.035	0.015	0.040	0.091
Buyer income growth from HMDA, 2002–2005	3,104	0.065	0.090	-0.006	0.071	0.156
Buyer income overstatement, 2002–2005	3,104	0.018	0.092	-0.067	0.024	0.112
Measure of fraud						
Change in non-agency share, 2002–2005	2,981	0.295	0.115	0.157	0.286	0.441
Change in low-doc share, 2002–2005	2,981	0.236	0.104	0.112	0.222	0.375
Misreported non-owner-occupant	2,969	0.058	0.042	0.013	0.051	0.108
Misreported second lien	2,969	0.061	0.044	0.000	0.056	0.111
Top mortgage fraud list, 2010	3,104	0.006	0.077	0.000	0.000	0.000
Top mortgage fraud list, 2010–2014	3,104	0.038	0.192	0.000	0.000	0.000
Measure of income growth after boom						
IRS income growth, 2005–2012	3,011	0.017	0.015	-0.000	0.017	0.033
IRS wage growth, 2005–2012	3,011	0.014	0.014	-0.001	0.014	0.030
Census median income growth, 2000–2010	3,011	0.022	0.012	0.006	0.023	0.036
Change in poverty rate, 2000–2010	3,011	0.026	0.036	-0.010	0.020	0.071
Change in unemployment rate, 2000–2010	3,011	0.044	0.029	0.015	0.043	0.078
Change in mortgage default rate, 2005–2007	3,104	0.034	0.036	-0.002	0.026	0.084
Change in mortgage default rate, 2005–2010	3,104	0.091	0.070	0.020	0.077	0.184
Measures of average mortgage size						
Growth in mortgage size, second liens treated as independent, 2004–2005	3,104	0.066	0.116	-0.070	0.068	0.207

Change in second lien fraction, 2004–2005	3,104	0.062	0.040	0.015	0.060	0.109
Growth in average first lien mortgage size, 2004–2005	3,104	0.119	0.111	-0.006	0.120	0.250
Growth in total mortgage size per housing unit purchased, 2004–2005	3,104	0.138	0.109	0.013	0.136	0.267

Table 2

Mortgage credit growth and income growth from 2002 to 2005

The first three columns replicate results from Mian and Sufi (2009). The amount of mortgages originated for home purchase grew more in low-credit-score zip codes from 2002 to 2005, and these same zip codes saw a decline in IRS income growth. This generates a negative correlation between income growth and mortgage amount growth. In columns 4 and 5, we follow A2S16 by using the growth in income reported on mortgage applications of home buyers in the zip code. The central point is that qualitatively we find the same result in the sample of zip codes used in MS09: IRS income growth in a zip code is negatively correlated with mortgage origination growth, whereas growth in buyer income reported on mortgage applications is positively correlated with mortgage origination growth. The right-hand-side variable in column 6 is the difference between buyer income growth and IRS income growth, which we call buyer income overstatement. All specifications include county fixed effects. **, * indicate coefficient estimates statistically distinct from 0 at the 1% and 5% levels, respectively.

	Amount for home purchase growth 2002–2005	IRS income growth 2002–2005	Amount for home purchase growth 2002–2005				Buyer income overstatement 2002–2005
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Fraction subprime borrowers, 1996	0.465**	-0.141**					0.159**
	(0.029)	(0.006)					(0.015)
IRS income growth, 2002–2005			-0.662**		-0.705**		
			(0.089)		(0.087)		
Buyer income growth from HMDA, 2002–2005				0.420**	0.433**		
				(0.038)	(0.038)		
Buyer income overstatement, 2002–2005						0.473**	
						(0.035)	
R^2	0.422	0.356	0.380	0.394	0.407	0.406	0.313
Observations	3,014	3,014	3,014	3,014	3,014	3,014	3,014

Table 3
Buyers overstating income and measures of fraud

This table shows elevated fraud in zip codes where mortgage applications overstate buyer income. More specifically, buyer income overstatement is defined to be the difference between the annualized growth in income reported on mortgage applications of home buyers from 2002 to 2005 and the annualized IRS income growth of households living in a zip code from 2002 to 2005. Columns 1 and 2 show that the zip codes with overstated income were the same zip codes seeing a large increase in the fraction of low-documentation mortgages being sold to non-GSE securitizers of mortgage pools. Columns 3 and 4 show that zip codes with overstated income are much more likely to show up on the list of top mortgage fraud zip codes put together by the mortgage fraud detection company Interthinx. Columns 5 through 7 present the correlation across zip codes between measures of fraud from Piskorski, Seru, and Witkin (2015) and income overstatement. The Piskorski et al. (2015) variables measure the fraction of mortgages for which the securitizers of non-agency mortgages misreported whether the loan was a non-owner-occupant loan or whether a second lien was present. The central point is that buyer income overstatement is highest in the same zip codes where (i) an expansion of private-label securitization of low-documentation mortgages occurred that we know was associated with fraudulent practices, and (ii) independent measures of fraud were higher. All specifications include county fixed effects. **, * indicate coefficient estimates statistically distinct from 0 at the 1% and 5% levels, respectively.

			Interthinx measures of fraud		Piskorski, Seru, and Witkins (2015) measures of fraud		
	(1) Change in non-agency share of mortgages 2002–2005	(2) Change in low-doc share of mortgages 2002–2005	(3) Zip code makes top mortgage fraud list 2010	(4) Zip code makes top mortgage fraud list 2010–2014	(5) Misreported non-owner-occupant	(6) Misreported second lien	(7) Either misreported
Buyer income overstatement, 2002–2005	0.121**	0.100**	0.051**	0.123**	0.030**	0.034**	0.051**
	(0.020)	(0.016)	(0.018)	(0.045)	(0.009)	(0.009)	(0.009)
R^2	0.483	0.598	0.071	0.067	0.271	0.321	0.245
Observations	2,981	2,981	3,014	3,014	2,969	2,969	2,969

Table 4
High buyer income overstatement zip codes become worse during the mortgage credit boom

This table presents evidence that zip codes with high buyer income overstatement did not improve during the mortgage credit boom. Columns 1 through 5 present the correlation between the growth in future IRS resident income growth of a zip code, and the log difference between home-buyer income reported on mortgage applications and IRS resident income in the lagged period. The central point is that during the mortgage credit boom, a gap between home-buyer income and resident income predicts negative relative IRS income growth going forward in the neighborhood. The left-hand-side variable in column 6 is the change in the fraction of IRS returns with greater than $50,000 in income from 2002 to 2005. The left-hand-side variable in column 7 is the credit score of people moving into a zip code from 2002 to 2005 minus the average credit score of the residents living in a zip code in 2002. All specifications include county fixed effects. **, * indicate coefficient estimates statistically distinct from 0 at the 1% and 5% levels, respectively.

	Growth in IRS income from time x to time y						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	$x = 1991$ $y = 1998$	$x = 1998$ $y = 2002$	$x = 2002$ $y = 2004$	$x = 2004$ $y = 2005$	$x = 2005$ $y = 2006$	Change in fraction of IRS returns >$50K 2002–2005	Credit score difference of residents moving in during boom
ln(Buyer income)–ln(IRS income), at time x	0.004* (0.002)	0.043** (0.002)	-0.013** (0.002)	-0.040** (0.005)	-0.023** (0.003)		
Buyer income overstatement, 2002–2005						-0.022** (0.004)	-6.961 (5.756)
R^2	0.226	0.405	0.165	0.213	0.185	0.280	0.096
N	2,590	3,013	3,014	3,014	3,014	3,014	3,013

Table 5

Zip codes where buyers overstate income become worse after the mortgage credit boom

This table presents correlations between buyer income overstatement and future measures of economic performance in a zip code. More specifically, buyer income overstatement is defined to be the difference between the annualized growth in income reported on mortgage applications of home buyers from 2002 to 2005 and the annualized IRS income growth of households living in a zip code from 2002 to 2005. All growth rates for outcome variables are annualized. The central point is that zip codes seeing the largest gap between buyer income and average income during the mortgage credit boom perform much worse going forward, which is consistent with mortgage applications overstating income from 2002 to 2005. All specifications include county fixed effects. **, * indicate coefficient estimates statistically distinct from 0 at the 1% and 5% levels, respectively.

	(1) IRS income growth 2005–2012	(2) IRS wage growth 2005–2012	(3) Census income growth 2000–2010	(4) Change in poverty rate 2000–2010	(5) Change in unemployment rate 2000–2010	(6) Change in mortgage default rate 2005–2007	(7) Change in mortgage default rate 2005–2010
Buyer income overstatement, 2002–2005	-0.094**	-0.149**	-0.121**	0.041**	0.024**	0.059**	0.126**
	(0.020)	(0.018)	(0.022)	(0.007)	(0.006)	(0.007)	(0.012)
R^2	0.325	0.372	0.412	0.333	0.252	0.331	0.529
Observations	3,011	3,011	3,011	3,011	3,011	3,014	3,014

Table 6
Correlation between mortgage application income growth and IRS income growth during subprime mortgage boom

This table presents the correlation between income growth from mortgage applications and average IRS income growth of residents living in a zip code. We present the correlation for four different time periods: 1991 to 1998, 1998 to 2002, 2002 to 2005, and 2005 to 2007. The central point is that buyer income growth and IRS income growth become much less correlated during the subprime mortgage boom of 2002 to 2005 relative to other periods. All specifications include county fixed effects. **, * indicate coefficient estimates statistically distinct from 0 at the 1% and 5% levels, respectively.

	Buyer income growth from mortgage applications, annualized			
	(1) 1991–1998	(2) 1998–2002	(3) 2002–2005	(4) 2005–2007
IRS income growth, annualized	0.383**	0.490**	0.100*	0.228**
	(0.028)	(0.084)	(0.043)	(0.038)
R^2	0.531	0.207	0.362	0.173
Observations	2,590	3,013	3,014	3,014

Table 7

Correlation between buyer income growth and IRS income growth, by GSE share

This table presents the correlation between income growth from mortgage applications and average IRS income growth of residents living in a zip code. We split the sample into four groups based on the average share of mortgages sold to GSEs for securitization from 2002 to 2005. Column 3 is the main specification that shows that the correlation between buyer income and IRS income growth from 2002 to 2005 was close to zero for low-GSE-share zip codes, but positive and significant for higher GSE share zip codes. Columns 1, 2, and 4 examine the same correlation for other time periods, where the GSE share category of a zip code is still based on 2002 to 2005. The central point is that buyer income growth and IRS income growth track each other quite well in the GSE market from 2002 to 2005, but are uncorrelated in the non-GSE market from 2002 to 2005. Further, the two income measures track each other well across the full distribution of zip codes in periods other than 2002 to 2005. This is consistent with the claim that buyer income reported on mortgage applications was fraudulent during the subprime mortgage boom from 2002 to 2005 in the non-GSE market. All specifications include county fixed effects, and the four GSE share quartile dummies. **, * indicate coefficient estimates statistically distinct from 0 at the 1% and 5% levels, respectively.

	Buyer income growth from mortgage applications, annualized			
	(1) 1991–1998	(2) 1998–2002	(3) 2002–2005	(4) 2005–2007
IRS income growth *Quartile 1 GSE share	0.481** (0.039)	0.550** (0.136)	-0.075 (0.064)	0.269** (0.052)
IRS income growth *Quartile 2 GSE share	0.298** (0.056)	0.422* (0.174)	0.134 (0.094)	0.172* (0.075)
IRS income growth *Quartile 3 GSE share	0.297** (0.060)	0.555** (0.165)	0.272** (0.087)	0.279** (0.083)
IRS income growth *Quartile 4 GSE share	0.238** (0.081)	0.330+ (0.181)	0.314** (0.107)	0.307** (0.085)
R^2	0.535	0.210	0.367	0.188
Observations	2,590	3,013	3,014	3,014

Table 8
Average mortgage size conditional on origination

This table presents results relating the growth in the average mortage size conditional on origination to measures of marginal home buyers. Separation of first liens and second liens in the HMDA data is available only from 2004 onward. Panel A uses the fraction of subprime borrowers in 1996 as the right-hand-side variable, and Panel B uses IRS income growth from 2004 to 2005. Column 2 replicates the A2S16 finding that average mortgage size increased (decreased) in low-income-growth (high-subprime-share) zip codes when treating second liens as independent mortgages. The left-hand-side variable in column 3 is the change in the fraction of mortgages for home purchase that are second liens from 2004 to 2005; in column 4 it is the growth in the average first-lien mortgage size, and in column 5 it is the growth in the total mortgage size conditional on origination, which is measured as the total amount of first- and second-lien mortgages originated divided by the number of first-lien mortgages. All specifications include county fixed effects. **, * indicate coefficient estimates statistically distinct from 0 at the 1% and 5% levels, respectively.

Panel A: Subprime share

	Amount for home purchase growth 2004–2005	Growth in mortgage size treating second liens as independent 2004–2005	Change in second lien fraction 2004–2005	Growth in average first lien mortgage size 2004–2005	Growth in total mortgage size per housing unit 2004–2005
	(1)	(2)	(3)	(4)	(5)
Fraction of subprime borrowers, 1996	0.435**	-0.108**	0.101**	-0.016	0.038*
	(0.043)	(0.017)	(0.006)	(0.017)	(0.017)
R^2	0.404	0.464	0.449	0.414	0.423
Observations	3,014	3,014	3,014	3,014	3,014

Panel B: IRS income growth

	Amount for home purchase growth 2004–2005	Growth in mortgage size treating second liens as independent 2004–2005	Change in second lien fraction 2004–2005	Growth in average first lien mortgage size 2004–2005	Growth in total mortgage size per housing unit 2004–2005
	(1)	(2)	(3)	(4)	(5)
IRS income growth,	-0.383**	0.098**	-0.126**	0.002	-0.056+
2004–2005	(0.077)	(0.031)	(0.011)	(0.030)	(0.030)
R^2	0.387	0.459	0.419	0.414	0.422
Observations	3,014	3,014	3,014	3,014	3,014

Table 9
Results with and without county fixed effects

This table presents results from MS09 with and without county fixed effects. Panel A uses the fraction of subprime borrowers in 1996 as the right-hand-side variable, and Panel B uses IRS income growth from 2002 to 2005. **, * indicate coefficient estimates statistically distinct from 0 at the 1% and 5% levels.

Panel A: Subprime share

	Amount for home purchase growth 2002–2005			IRS income growth 2002–2005		
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	Within a county	Between counties	OLS	Within a county	Between counties
Fraction of subprime borrowers, 1996	0.647**	0.465**	1.101**	-0.096**	-0.141**	-0.018
	(0.028)	(0.029)	(0.136)	(0.005)	(0.006)	(0.027)
R^2	0.151	0.085	0.286	0.095	0.181	0.003
Observations	3,014	3,014	3,014	3,014	3,014	3,014

Panel B: IRS income growth

	Amount for home purchase growth 2002–2005				
	(1)	(2)	(3)	(4)	(5)
	OLS	Within a county	Between counties	Within a state	Between states
IRS income growth, 2002–2005	0.089	-0.662**	1.823**	-0.768**	6.809**
	(0.098)	(0.089)	(0.450)	(0.090)	(1.131)
R^2	0.000	0.019	0.091	0.024	0.644
Observations	3,014	3,014	3,014	3,014	3,014