LOAN ORIGINATIONS AND DEFAULTS IN THE MORTGAGE CRISIS:
FURTHER EVIDENCE

Manuel Adelino
Antoinette Schoar
Felipe Severino

Loan Originations and Defaults in the Mortgage Crisis: Further Evidence
Manuel Adelino, Antoinette Schoar, and Felipe Severino
NBER Working Paper No. 21320
July 2015
JEL No. D30,G21,R30

ABSTRACT

This paper addresses two critiques by Mian and Sufi (2015a, 2015b) that were released in response to the results documented in Adelino, Schoar and Severino (2015). We confirm that none of the results in our previous paper are affected by the issues put forward in these critiques; in particular income overstatement does not drive any of our results. Our analysis shows that the origination of purchase mortgages increased across the whole income distribution during the 2002-2006 housing boom, and did not flow disproportionately to low-income borrowers. In addition, middle- and high-income, as well as middle- and high-credit-score borrowers (not the poor), represent a larger fraction of delinquencies in the crisis relative to earlier periods. The results are inconsistent with the idea that distortions in the origination of credit caused the housing boom and the crisis and are more consistent with an expectations-based view where both home buyers and lenders were buying into increasing housing values and defaulted once prices dropped.

Manuel Adelino
Duke University
Fuqua School of Business
100 Fuqua Drive
Durham, NC 27708-0120
manuel.adelino@gmail.com

Antoinette Schoar
MIT Sloan School of Management
100 Main Street, E62-638
Cambridge, MA 02142
and NBER
aschoar@mit.edu

Felipe Severino
Tuck School of Business
Dartmouth College
100 Tuck Hall
Hanover, NH 03755
felipe.severino@tuck.dartmouth.edu

The results presented in Adelino, Schoar and Severino (2015) [A2S2015 henceforth] provide new evidence on the increase in mortgage credit prior to 2007 and the subsequent default crisis. In contrast to some of the prior research and much popular held belief, we show that mortgage originations went up across the income distribution, not disproportionately for the poor. In fact, the majority of purchase mortgages by value were originated for middle class and richer segments of the population even at the peak of the housing boom. In addition, there was no decoupling of credit growth (for purchase mortgages at origination) from income growth as had been previously argued in Mian and Sufi (2009) [MS2009 henceforth]. In particular, debt to income levels (DTI) of purchase mortgages did not disproportionately increase for the poorer part of the income distribution relative to other borrowers. Getting this fact right is very important, since the findings in MS2009 have been used to argue that distortions in the supply of credit *caused* the run up in house prices and the subsequent financial crisis. Shedding further doubt on this interpretation, we show in A2S2015 that the share of defaults stemming from middle- and high-income borrowers and borrowers with credit scores above 660 (a common subprime threshold) *increased* significantly during the crisis. In contrast, low income borrowers and those with a credit score below 660 represent a smaller fraction of all delinquencies and foreclosures than in previous years.[1] Defaults by middle- and high-credit-quality borrowers increased in particular in zip codes that saw large house price increases in the pre-crisis period and subsequent drops in prices. These results are most consistent with an expectation-based view of the housing crisis where both home buyers and lenders were buying into the increase in house prices and defaulted when prices dropped, rather than the crisis being the result of distortions in the allocation of credit towards low income and marginal borrowers.

As a reaction to A2S2015, the authors released two documents to challenge our findings. The first, entitled "Fraudulent income overstatement on mortgage applications during the credit expansion of 2002 to 2005" (Mian and Sufi (2015a), MS2015a henceforth), challenges the findings in A2S2015 by arguing that the income provided on HMDA mortgage applications suffers from misreporting and fraud, and that the positive relation between credit and buyer income is a result of this problem. The second paper, "Household Debt and Defaults from 2000 to 2010: Facts from Credit Bureau Data" (Mian and Sufi (2015b) looks at the distribution of debt and defaults by credit score bins following the analysis in A2S2015. This paper, despite the authors' description of the results, confirms the findings in A2S2015 using credit bureau data instead of mortgage origination and delinquency data that we rely on in our analysis.

---

[1] These results refer to where leverage and mortgage defaults occurred along the income and credit score distributions, they are not meant as welfare statements. While lower income borrowers did not contribute to delinquency at a higher volume than other income groups, it is likely that lower income households (and zip codes) suffered more from defaults than higher income ones. This might be driven by more limited non-housing wealth, worse income shocks, or lack of alternative funding sources.

In what follows we will discuss the arguments presented in these critiques in turn. First, using buyer income instead of IRS household income is not our central critique of the findings in MS2009. Instead, the analysis in A2S2015 highlights the limitations of using *zip codes* instead of individuals as the unit of observation: in particular, in MS2009 the authors regress the growth in total mortgage debt for home purchase at the zip code level on the growth in per capita income from the IRS. However, we show in A2S2015 that it is important to decompose the total mortgage origination at the zip code level into the growth in the average *size* of mortgages (intensive margin) and the growth in the *number* of mortgages originated in a neighborhood (extensive margin). Given that it is households that take out mortgages, not zip codes, only the former is informative about whether there were distortions in how credit was allocated. The second dimension, the growth in the number of mortgages, captures the increase in the relative pace (or velocity) of mortgage origination but does not speak to whether the debt burden of each individual origination decision changes differentially across zip codes.

In A2S2015 we show that mortgage debt and income growth at origination are strongly positively correlated during the pre-crisis period once we use the appropriate measure of credit growth at the household level. The relation between individual mortgage size and income growth is always positive, independent of the income measure used (IRS income or buyer income from HMDA). It is the relation between the *number* of mortgages originated within a zip code (the extensive margin) and the growth in IRS income of the zip code that is negative. This means that originations increased relatively more in neighborhoods with lower income growth, likely driven by an increase in the velocity of buying and selling. This effect, once aggregated up, produces very small changes in the overall distribution of credit across zip codes, because the growth in the number of mortgages in low income (and low income growth) zip codes starts from a very small base. Our results call into question one central piece of evidence that the literature has relied on to argue that there was a shift in the allocation of credit to low income borrowers, and we expand on these points below.

While the above result holds no matter which income measure we use, there is a conceptual reason to verify that this analysis also goes through using the growth in the income of buyers rather than the average household income from the IRS: buyers have historically had much higher income levels than the average resident in a zip code, and IRS income combines the stock of existing residents with the flow of new buyers. Our previous paper shows that the results on the growth in average mortgage size hold just as strongly when we use buyer income from HMDA, but the negative correlation between income and total mortgage credit disappears, pointing to the fragility of the result in MS2009.

In A2S2015 we also conduct a host of robustness tests to show that income misreporting is not an important driver of this result. This is not to say that income misreporting did not happen. It has been shown in a

number of papers that some lenders and borrowers engaged in fraudulent practices (Ben-David, 2011, Jiang, Nelson, and Vytlacil, 2014, Griffin and Maturana, 2014, Piskorski, Seru, and Witkin, 2015, Ambrose, Conklin, and Yoshida, 2015). The central point for the results in A2S2015 is that misreporting cannot explain the estimates we obtain with respect to buyer income growth and mortgage growth, nor does it affect the results on the distribution of originations and delinquencies. We show in particular that the relation between individual leverage and income growth is unchanged when we divide areas by the propensity for income misreporting – those with more subprime lenders, and those with lower levels of originations by the government sponsored enterprises Fannie Mae and Freddie Mac (the GSEs). The correlation between buyer income growth and growth in mortgage origination is always positive and virtually unchanged if we focus on areas where income misreporting was not prevalent. We also show below that even if we drop zip codes based on measures of the discrepancy between buyer income and per capita income from the IRS (the measures suggested in MS2015a), our results still hold. Finally, the zip codes with the highest difference between HMDA and IRS income represent only a small fraction of mortgages originated in any given year. By focusing on growth rates rather than the underlying level effects, the discussion in MS2015a misses the central insight in A2S2015 that the bulk of mortgage origination was coming from middle- and high-income households even at the peak of the boom.

One consequence of the claim that credit and income decoupled during the boom has been a focus on the role that low income households played in the crisis. In fact, MS2009 argue that the increase in defaults was "significantly amplified in subprime ZIP codes," defined as those with a high fraction of borrowers with a FICO score below 660. Similarly, in MS2015a the authors present a graph (Figure VII) that shows higher shares of all delinquent mortgages coming from zip codes with more subprime borrowers. However, this analysis again suffers from the problem that aggregation at the zip code level misses important dynamics *within* zip codes. As our paper makes clear, when defaults ballooned, borrowers with low credit scores represented a *smaller* share of the total volume of delinquencies relative to earlier periods, as did those in low income zip codes. In fact, even in areas where subprime loans were most prevalent (defined as those with a high fraction of loans made by originators on the subprime lender list published by the Department of Housing and Urban Development, or HUD), it was borrowers with credit scores above the subprime cutoff (a FICO score of 660) who proportionally increased their defaults the most. As an aside, it is important to understand that zip codes with higher subprime concentration are different on many dimensions, including the fact that house prices dropped more in these areas after the crisis. These results do not square well with an interpretation that emphasizes the role of poor and low credit score borrowers in the crisis, and rather point to the role that house price expectations played in driving credit on the upside and in generating defaults after prices dropped.

These patterns are central for our understanding of the boom and bust. In A2S2015 we argue that the contribution to the overall increase in debt was shared across the distribution of borrowers and that the focus should be, instead, on understanding the drivers of defaults for middle- and high-income borrowers, as well as those with middle- and high-credit-scores, once the crisis hit. We show that dollar weighted defaults after 2007 increased most for borrowers with above median income and those with credit scores above 660 (a common subprime cutoff, used among others in MS2009). The reason for this pattern is intuitive – while default rates are higher for low income and low credit score borrowers, the proportional increase in default rates is more pronounced for borrowers who typically did not default. In their second response to our paper, MS2015b repeat this analysis using credit bureau data and show very similar patterns - first, the paper shows that the contribution to the increase in leverage (including purchase mortgages and all other mortgage-related debt) is shared broadly by borrowers between the first and fourth quintiles of the credit score distribution, fully in line with the findings in Figure 1 and 2 of A2S2015. Then, the authors show that borrowers with the lowest credit scores represent a *smaller* fraction of overall delinquencies in the crisis than they did before, and the biggest increase in the share of defaults stems from borrowers with credit scores above the subprime cutoff. We discuss below some of the apparent differences between the analysis in A2S2015 and MS2015b, and that the two papers show consistent results despite using very different datasets.

Our results are in line with a number of papers that have argued that buyers as well as investors in the mortgage market had overoptimistic beliefs about house price growth, for example, Foote, Gerardi and Willen (2012), Cheng, Raina and Xiong (2014), Shiller (2014), among many others. Similarly, Chinco and Mayer (2014) document an increased inflow of out-of-state buyers who seem to have been buying for speculative purposes. Coleman, LaCour-Little, and Vandell (2008) argue that subprime lending may have been a joint product, rather than the cause of the increase in house prices. Also Glaeser, Gottlieb and Gyourko (2010) argue that easier access to credit cannot explain the increase in house prices during the boom. Our results are also related and consistent with Haughwout, Peach and Tracy (2008), Foote, Gerardi and Willen (2008), Mayer, Pence and Sherlund (2009), Palmer (2014), and Ferreira and Gyourko (2015) who argue that the role of declining house prices and negative equity were key for explaining increased defaults.

The paper proceeds by revisiting the different arguments put forward by MS2015a and MS2015b and our responses to them. In this paper we do not repeat all the analysis put forward in A2S2015. For a description of our original results, we point the reader to our earlier paper.

2. Data description

The analysis in this paper, as in A2S2015, uses data from three primary sources: the Home Mortgage Disclosure Act (HMDA) mortgage dataset, income data from the Internal Revenue Service (IRS) at the zip code level, and a 5% random sample of all loans in the Lender Processing Services data (LPS, formerly known as McDash). The HMDA dataset contains the universe of US mortgage applications in each year. The variables of interest for our purposes are the loan amount, the applicant's income, the purpose of the loan (purchase, refinance, or remodel), the action type (granted or denied), the lender identifier, the location of the borrower (state, county, and census tract), and the year of origination. We match census tracts from HMDA to zip codes using the Missouri Census Data Center bridge. This is a many-to-many match, and we rely on population weights to assign tracts to zip codes.[2] We drop zip codes for which census tracts in HMDA cover less than 80% of a zip code's total population.[3] With this restriction, we arrive at 23,385 individual zip codes in the data.

IRS income data is obtained directly from the IRS, and we use the adjusted gross income of households that filed their taxes in a particular year in that zip code. Besides total income and per capita income, we use the number of tax filings in a zip code to construct an estimate of the population in a zip code in each year.[4]

We obtain house price indices from Zillow.[5] The zip-code-level house prices are estimated using the median house price for all homes in a zip code as of June of each year. Zillow house prices are available for only 8,619 zip codes in the HMDA sample for this period, representing approximately 70% of the total mortgage origination volume.

To identify subprime loans, we rely on the subprime and manufactured home lender list constructed by the Department of Housing and Urban Development (HUD) for the years between 1993 and 2005. This list includes lenders that specialize in such loans, and they are identified by a combination of features that include the average origination rate, the proportion of loans for refinancing, and the share of loans sold to Fannie

---

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

[3] This restriction drops 180 zip codes out of 23,565.

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

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

Mae or Freddie Mac, among others.[6] The data contain lender names, agency codes, and lender identification numbers, and we use these identifiers to match this list to HMDA.

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

For descriptive statistics of the datasets, in particular HMDA and LPS, please refer to A2S2015.

3. Income misreporting

The key point that is disputed by MS2015a stems from the fact that during the period of the housing boom there was widespread misreporting of income, which is purported to contaminate any inference made using the HMDA data. First, and as we mention above, the central points in A2S2015 hold if we use only IRS income. In particular, the conclusions from the decomposition of total mortgage origination into the average mortgage size and the number of mortgages are independent of the measure of income used. These results are presented in Tables 2 and 4 of A2S2015, and we replicate these tables in this paper (Tables 1 and 2 below). We show that the relation between growth in average mortgage size is strongly positively and significantly related to income growth no matter whether we use IRS income as in Table 1, HMDA income as in Table 2 or both (Table 3). Additionally, none of the results on the aggregate distribution of origination and delinquencies is impacted by the issue of income misreporting.

However, it is important to understand that even when we use buyer income from HMDA our results cannot be explained by widespread misreporting in this variable. This is not to say that income misreporting was not a concern during the run up to the housing crisis (as we point out above, this has been shown by a number of authors), but simply that it does not impact the estimates we obtain in A2S2015. We conduct a number of tests to show that if we focus only on zip codes where income misreporting was *not* a significant problem, our results are virtually unchanged. Panel B of Table 2 (Table 4 of A2S2015) breaks out our sample into different

---

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

quartiles based on the fraction of mortgages in the zip code that are underwritten by Fannie Mae or Freddie Mac (the government-sponsored enterprises, or GSEs), as well as the fraction of loans that were originated by subprime lenders based on the list of lenders that the US Department of Housing and Urban Development classifies as subprime (HUD, see Section 2 for details). Loans that were sold to (and then guaranteed by) the GSEs had to conform to higher origination standards than those sold to other entities, and thus were less likely to have unverified applicant income.[7] If misreporting of income were a significant driver of our results on the relation between buyer income and credit growth, then zip codes with a lower fraction of loans sold to the GSEs should exhibit a stronger relation between mortgage growth and buyer income. Similarly, loans originated by subprime lenders were much more likely to have low or no documentation. Again, if the results on mortgage growth and buyer income were driven by misreporting, we would expect zip codes with a higher fraction of subprime lenders to have a stronger correlation of mortgage growth and income.

For both of these measures of income verification we do not find that the coefficients on buyer income vary significantly. In fact, the coefficient on buyer income growth is very similar in magnitude and significance levels across all quartiles of both the GSE origination fraction and the fraction originated by subprime lenders (if anything, the coefficient is smaller for the zip codes with a relatively low fraction of GSE loans).

In Table 4 we conduct additional tests based directly on the measures of income misreporting suggested by MS2015a. In particular, we start with our whole sample (8,619 zip codes in total with nonmissing house price data from Zillow), but then drop 10% of the sample at a time, starting with the highest deciles, based on: (a) the ratio of buyer income to IRS household income as of 2006 and (b) the difference in the growth rate of buyer income and the growth rate of IRS household income between 2002 and 2006. The idea in these tests is to see to what extent the estimates are affected by dropping the more "problematic" zip codes based on the measures in MS2015a. We show the results using as the dependent variable both the growth in total mortgage origination, as well as the growth in average mortgage size (which, as we discuss above and in A2S2015, is the right measure for assessing whether there were distortions in the allocation of credit during this time period).

Panel A of Table 4 shows that removing zip codes based on the ratio between buyer income and IRS income leaves all point estimates on the buyer income unchanged for both the growth in total mortgage origination and the growth in average mortgage size. Importantly, the estimates are basically unchanged as we reduce the sample (by up to 40%). If misreporting was indeed the reason behind the positive relation between buyer income and mortgage growth, we would expect the point estimates to go to zero when we take out regions that are deemed more problematic, but this is not the case.

---

[7] Previous work, including Pinto (2010) have noted that origination standards for the GSEs dropped between 2002 and 2006, but we find similar results when we split the sample directly by the fraction of loans originated by subprime lenders.

Panel B of Table 4 shows that the same is not true about the negative relation between growth in IRS household income and total mortgage growth at the zip code level. In fact, the point estimate becomes statistically insignificant when we drop just 10% of the sample, and even smaller as we drop more zip codes. However, the results on average mortgage growth are very stable and unaffected. This supports our interpretation in A2S2015 that the result on the negative correlation between mortgage origination and income growth is very sensitive to the selection of the sample, and to the specification chosen (in particular, as we see in Table 1, without fixed effects this relation becomes positive). Once we use the same unit on both sides of the regression (that is, when we regress average mortgage size on average household income growth), the results are very stable and show no visible distortions in DTI at origination.

Panel C of Table 4 removes zip codes based on the difference in annualized growth rates of buyer income and IRS income during this time period. The effect on total mortgage origination is reduced in magnitude, but is still statistically significant as we reduce the sample by up to 40%. The relation between buyer income growth and the growth in average mortgage size is, again, positive and very stable throughout. Panel D confirms the findings in Panel B, namely that as soon as we remove a small portion of the sample based on the measures suggested in MS2015a, the relation between IRS income growth and total mortgage growth becomes insignificant, and in fact turns positive as we drop additional zip codes. The relation between average mortgage size and IRS income is strong and positive throughout.

Finally, when we look at the potential aggregate impact of income misreporting for the overall distribution of credit along the income distribution, we see that the fraction of mortgage dollars that is most affected is small, even if we assume the full difference between buyer income and IRS household income is driven by income overstatement and not merely a reflection of the fact that home buyers have always had higher income than the average household in a zip code. Figure 1 below shows the ratio of buyer income to IRS income that is also shown in Figure 1 of MS2015a. We then overlay the dollar value of loans that are originated in these areas (the bars in Figure 1). The message from this figure is that the fraction of mortgage dollars originated in the areas where overstatement was potentially largest account for about 7% of originations. By only focusing on the relative changes in growth rates rather than the underlying level effects, the discussion in MS15 (as that in MS09) again misses the important share of originations coming from middle- and high-income households.

3.1.   Low correlation between IRS income and HMDA income

MS2015a also appeal to the small correlation between IRS income growth and buyer income growth as evidence that buyer income growth is likely to reflect fraudulent reporting. The results discussed above, however, suggest that this correlation is more likely to simply reflect the fact that home buyers and residents are very different, and that they have very different income dynamics.[8] We show that this has been historically the case and is not a feature solely of the 2002-2006 period. The income of home buyers and that of the average household in a zip code have been vastly different throughout all years for which we have available data (since the 1990's) and there were no allegations that fraud was an important factor in the 1990s. In Table 1 of A2S2015, we showed that the average income of home buyers is almost twice the IRS income of residents in the same area. MS2015a confirm that this gap has existed in the data historically. This suggests that home buyers in a given year are a very different group compared to residents. A report by the Census bureau shows that only 7.4 percent of homeowners move in a given year, indicating that they are a selected sample of the population.[9] Of these 7.4 percent, about 40 percent move across counties, so growth in their IRS income cannot be accounted for in either the old or the new county. The fact that zip codes are a much smaller geographic unit than counties implies that an even larger proportion of movers move across zip codes. These facts mean that the income of residents in a zip code are likely to be a very noisy measure of home buyers in that zip code.

In Table 5 we regress the growth in the income of home buyers on the growth in IRS zip code income for the period from 1998 to 2002, and then for the boom period from 2002 to 2006. In Panel A of Table 5 we use the sample of zip codes with available house prices (Zillow sample) and Panel B of Table 5 uses the full HMDA sample. We see that the correlation between HMDA income and IRS income is similar and statistically significant, although smaller in magnitude (0.1 versus 0.2) in the pre-crisis period relative to the earlier period when we use county fixed effects. As we have argued in A2S2015, there is no convincing economic reason to include county fixed effects in this type of regressions, as we are concerned about the total variation in income, not just variation relative to the county mean. When we exclude county fixed effects, the relation between buyer income and IRS income growth is strong and again very stable in 2002 to 2006 relative to the 1998-2002 period. The regressions using the full HMDA sample confirm these results.

---

[8] Again, as we have asserted both in A2S2015 and in this paper, we do not claim that income misreporting was not a significant issue in this time period, as a large literature has shown. The concern for the purposes of our earlier work is that income misreporting cannot account for the results we obtain across the income distribution.

[9] Schachter, Jason P., 2004, "Geographical Mobility: 2002 to 2003," Census Bureau, Current Population Reports, issued March 2004.

3.2.   Englewood an Garfield Park

MS2015a cite Englewood and Garfield Park (two neighborhoods in Chicago) throughout the paper as central examples of the origins of the housing crisis and to support the income misreporting hypothesis. In fact, these neighborhoods are very good examples of the differences in the view of the housing boom and the crisis between our work and MS2009. As MS2015a point out, the zip codes that these neighborhoods are part of have very low income per capita – as of 2002, they were all in the bottom 30% of zip codes in the country, and two of them are in the bottom 10%, all significantly below the average in the Chicago metropolitan area.[10] Although intuitively appealing as examples of the effects of the boom and bust, originations to the bottom decile of zip codes by income *never* made up a large fraction of originations in the US, even at the peak of the boom, and could not produce the aggregate consequences we observed during 2008-2009. Low income zip codes had very high delinquency rates, and very poor economic outcomes in terms of employment and income, but what ultimately impacted the overall portfolio of delinquent mortgages most was the fact that middle- and high-income zip codes (and middle- and high-credit-score borrowers) experienced much higher default rates relative to earlier years.

There is another reason why these zip codes illustrate the differences in interpretation of the data between A2S2015 and the papers cited above by Mian and Sufi. Even though Englewood and Garfield Park have very low average household income, the average applicant income for purchase mortgages as of 2002 (i.e., before the boom and before income misreporting was considered a problem) was more than three times the income of average residents. By way of illustration, in zip codes 60621 and 60624, IRS income per capita was 20 thousand dollars in 2002 (both in the bottom 10% in the country), while buyers in that year reported 63 and 74 thousand dollars of income, respectively. These examples serve as an illustration of why the average income of residents in a zip code is often a poor proxy for buyer income, and, by extension, of the dynamics of buyer income over time.

3.3.   Ex post outcomes for zip codes with large differences in buyer versus resident income

MS2015a argue that the discrepancy between the buyer income and the resident (IRS) income must be an indicator of mortgage fraud and income overstatement, since the neighborhoods with a higher discrepancy are the ones that have worse economic performance ex post. They show that these are the neighborhoods

---

[10] Englewood corresponds closely to zip code 60621. West Garfield Park is part of zip code 60624, and East Garfield Park is split (almost) evenly between 60624 and 60612. Together, they make up about two-thirds of the area of 60624 and less than one third of the area of 60612.

with more subprime loans and those who ultimately had higher unemployment and more defaults in the crisis. There is, however, an important omitted variable in this analysis, since neighborhoods that had a larger difference between the income of buyers and that of residents are also those that had stronger house price growth and subsequently had a steeper house price drop. Of course, neighborhoods with large house price drops subsequently suffered large economic losses for a number of reasons, of which income misreporting is one.

To make this point more clearly, we sort zip codes into quartiles based on two characteristics, namely the fraction of loans in a zip code that are originated by subprime lenders (based on the HUD subprime lender list), as well as the house price growth in the zip code between 2002 and 2006. If reported income growth is indeed primarily driven by fraud, the fraction of loans originated by subprime lenders in an area should be an important determinant of observed buyer income growth.

Table 6 reports the summary statistics when we cross these two dimensions. The main message from the table is in Panel D, and it shows that the growth in buyer income shows relatively small variation with the fraction of subprime loans in an area (about 1 to 2 percentage point difference between the lowest and highest quartiles in the fraction of subprime loan origination). However, the variation in house price growth is associated very strongly and positively with buyer income growth, even in neighborhoods with low subprime lending. The difference in annualized growth rate in income is between 6 and 9 percentage points. Neighborhoods with strong increases in house prices subsequently also saw strong declines. As a result, it seems (at a minimum) questionable that mortgage fraud was a primary driver of poor economic performance in these areas. It is more plausible, instead, that the underlying rise and fall in house prices explains first the change in the buyer income and subsequently the poor economic outcomes of the area.

4. Distribution of delinquency

In the response to our paper, MS2015a produce a graph that shows the share of defaults by zip-code-level FICO score, and it shows that this fraction is highest in neighborhoods with a high fraction of borrowers with low credit scores (or high fraction of subprime borrowers) (see MS2015a, Figure VII). The authors argue that these results contradict the findings in A2S2015, since we show that middle- and high-income and middle- and high-FICO score individuals accounted for a larger fraction of defaults. As is clear from the figure itself in MS2015a, and the notes to the figure, the notable characteristic of the crisis period is that zip codes with a high fraction of subprime borrowers reduced, rather than increased, their share of overall delinquent mortgages. An additional issue with the figure in MS2015a is that the authors again do not look at *individual transactions* but aggregate results up to the zip code level, as was the case in MS2009. We show below that even in neighborhoods with high subprime concentration (measured, in our case, by the share of loans

originated by subprime lenders), it was middle- and high-income borrowers (and those with credit scores above the subprime cutoff) who dramatically increased their share of overall delinquencies.[11]

In Figure 2 below (and Figure 6 of A2S2015), we separate zip codes into the ones with the 50% highest and lowest concentration of subprime loans (i.e., loans originated by lenders on the HUD subprime lender list). In the median zip code in the US about 10% of loans are originated by subprime lenders, and even the top quartile of zip codes have an average of about 26% (Table 6, Panel C). This means that even for zip codes with a high concentration of subprime lending, the majority of loans are still prime loans, which again highlights that it is misleading to analyze mortgage transactions at the zip code level. We then decompose delinquencies within zip codes by the credit score of the mortgage holders who actually defaulted once the 2008 crisis sets in. In Panel A of Figure 2 we compare the cohorts of mortgage borrowers in 2003 and 2006. The first bar is the fraction of loans originated in subprime neighborhoods, broken out into those with a FICO score above and below 660 at origination. We focus on a cutoff of 660 because this is a common cutoff used for subprime borrowers by, among others, the Federal Reserve Board (FRB), the Office of the Comptroller of the Currency (OCC), the Federal Deposit Insurance Corporation (FDIC), and the Office of the Thrift Supervision (OTS), as well as several researchers, including MS2009 and Demyanyk and Van Hemert (2011). The second bar repeats the same exercise for borrowers in zip codes with low subprime concentration (that we call prime zip codes). These results are weighted by the dollar value of the mortgages.

The figure shows that about 54% of total loan value in 2003 is originated in subprime neighborhoods, of which 43 percentage points is to borrowers with a FICO score above (or equal to) 660 and only 11% to borrowers with a FICO score below 660. The rest (46%) of mortgages by value are originated in prime zip codes, 6% to borrowers with FICO scores below 660 and 40% to those with FICO scores above 660. When we repeat this analysis for 2006, we see a picture that is largely unchanged.

In Panel B of Figure 2, we then track the cohorts of borrowers three years out and calculate the dollar value of loans that defaulted within the first three years after origination. We do the same split by prime versus subprime zip codes as above, and again break out the high versus low credit score borrowers. In the 2003 cohort, only 30% of dollar values in default came from borrowers with FICO scores above 660. 20% of these had a FICO score above 660 in subprime zip codes, and 10% had a FICO score above 660 in prime zip codes. This should not be surprising, since we would expect the biggest amount of defaults to stem from low credit score (in particular, those in the subprime range). However, when we repeat the same analysis for the

---

[11] We classify individual zip codes by the concentration of loans from subprime lenders because this can be done using HMDA data, so there is enough data to reliably classify individual zip codes. Given that the only source of data we have available that has FICO score information is a 5% sample of LPS we do not classify zip codes using borrower FICO scores. We can then split *borrowers* by credit scores reliably, but not zip codes.

2006 cohort of borrowers, we see that the picture is completely reversed. Now, 61% of defaults come from borrowers with a credit score above 660. Out of these 61%, 49% come from borrowers who have a score above 660 in subprime zip codes and 12% are borrowers above 660 in prime zip codes.

These results show that even in areas where subprime loans were most prevalent, it was higher FICO borrowers who contributed most to the increase in defaults. By just comparing the proportion of delinquencies in prime versus subprime *zip codes*, the results in MS2015a miss where the increase in defaults is coming from.

4.1. Credit bureau data analysis (MS2015b)

In the second response to our paper, Mian and Sufi (2015b), written after our paper was made public and widely discussed, the authors repeat much of the analysis that we document about the contribution to the rise in debt and defaults during the house price boom and bust. In that paper, the authors find that it was not the lowest credit score individuals that contributed most to the rise in total debt (Figures 4 and 5 in MS2015b), but rather that the increase was shared broadly by all borrowers up to the 80th percentile of credit scores. Similarly, the increase in debt to income ratio is shared by all borrowers. This analysis confirms our central point that the boom in mortgage origination was widely shared, and not concentrated in the poorest or lowest credit score individuals. Importantly, these results are in stark contrast to the message in MS2009 that focuses on the growth in purchase mortgage credit to low credit score (and low income) individuals, and the implications this may have had for house price growth (or, as the authors put it in Mian and Sufi (2014), "… from 2002-2005, credit flooded into low credit score zip codes. This was the only time period in which mortgage credit expanded so much more dramatically in low credit score zip codes."

The second point made in MS2015b refers to the distribution of delinquencies by borrower credit score. Contrary to the authors' discussion of those results, they are entirely consistent with the previous findings in A2S2015, once the comparison is performed adequately. In fact, the authors show a significant decrease in the share of defaults coming from borrowers in the bottom quintile of credit scores, and an increase in the share coming from borrowers in the top three quintiles. This confirms the patterns shown in A2S2015, Figures 2-7 that delinquencies become less concentrated in the low income and low credit score individuals, not more. This is a mechanical effect that comes from the relative growth in delinquency rates. As MS2015b show in their paper, and we show for our data in Table 7 below, the default rates of low credit score individuals is high, and remains higher than those of higher credit score individuals (going from about 9% to 26% between 2005 and 2009 in the MS2015b data). But the proportional increase is much larger at the top

(going from about 2% to 8% for quintile 3, and from what appears less than 1% to 4% for quintile 4 in MS2015b). We find similar patterns for purchase mortgage default rates by cohort in our data (Panels A and B of Table 7).

The text in Mian and Sufi (2015b) appears to contradict our results, and Section 5.2 in the paper argues that there are important discrepancies. This is simply not the case. First, our paper focuses on delinquencies by cohort of purchase mortgages (which we have always been clear about), and the Mian and Sufi (2015b) results show the stock of delinquencies for all mortgage-related debt. Additionally, we show results by income bins and by FICO scores, whereas MS2015b use vantage scores (a much less frequent measure in the literature). Despite these differences, the message from both exercises is the same.

The first critique in MS2015b is that LPS under-represents low credit score individuals, which might bias our results relative to the true share of delinquencies by low income or low credit score individuals in the population. There are at least two reasons for this not to be the case. First, a simple under-representation of low credit score borrowers would not affect the changes we observe over time. It might affect the levels (i.e., the share of delinquencies from low credit score borrowers might show up as smaller than it should), but the changes by cohorts would be unaffected, unless the underrepresentation became worse for later cohorts (which there is no evidence of). Second, and more importantly, we show below that the patterns look the same in Figure 3, Panels A and B, using data from Blackbox Logic and Figure 3, Panels C and D, for a sample of borrowers from the public Freddie Mac data (Figure A4 of A2S2015). Blackbox logic is focused specifically on loans backing subprime mortgage backed securities, so it heavily *over*-represents subprime loans. We obtain the same the results even with these very different sets of borrowers. Finally, Ferreira and Gyourko (2015) also find that the crisis is, in fact, primarily a crisis of prime borrowers using panel data from Dataquick. Therefore, issues with the representativeness of LPS cannot account for the patterns we show in this paper.

MS2015b also argue that using credit scores at the time of origination introduces biases because credit scores might increase endogenously in high house price appreciation areas (leading us to classify low credit quality borrowers as high credit score ones). This may happen because individuals are able to either sell their homes or to obtain lines of credit instead of defaulting. The authors argue that this was particularly severe for low credit score individuals in high house price appreciation areas. This argument is not relevant for our results, and this is clear from the data shown in Table 8 of MS2015b. That table shows that, between 2000 and 2006, there is an increase in the vantage scores for the bottom quintile of individuals by credit score of 30 points in the lowest house price appreciation areas, whereas this increase is 43 points in the highest house price appreciation areas. The difference between the two is *13 points*. The "endogenous" increase in credit scores

that can be attributed to higher house prices causing high credit scores (which might, then, lead to misclassification of good borrowers as bad borrowers) can account for, at most, 2 to 18 points in the credit score according to this table. This a small change economically that has no material impact on any measurement we perform.

In fact, the increase in vantage scores shown by MS2015b is a result simply of the passage of time, as they use credit scores for borrowers as of 1997. All borrowers in the MS2015b sample below the 80th percentile experience a 20-40 point increase in credit score, irrespective of the house price increase in their area.

The argument that misclassification of borrowers drives our results is also entirely inconsistent with the data. If it were, in fact, the case that our results reflected some upward drift in quality of all borrowers, we should see an "emptying out" of the bin of borrowers below a FICO score of 660 in terms of mortgage originations. That is not what we see. In fact, both by income and by credit score, the share of originations to the low end of either distribution was stable, or experiences small increases, it never drops. The surprising fact we have pointed out in A2S2015 is that the share of *delinquencies* that these borrowers make up becomes smaller over time, not larger.

Finally, it is worth comparing the fractions of delinquencies shown in A2S2015 and MS2015b directly, since the latter only focuses on level results rather than looking at the changes in shares of total delinquency over the crisis period.[12] In Figure 4 below, we construct quintiles by FICO score, to match the data sorting in MS2015b.[13] This is the same picture shown in Figure 3 of A2S2015, but constructed with quintiles of FICO scores rather than a cutoff of 660 and 720. The patterns are exactly the same: the bottom quintile makes up two-thirds of delinquencies in the 2003 mortgage cohort, but drops to only 36% for the 2006 cohort. The second quintile grows from a share of 18% to 29%, the third increases from 9% to 20%, and the top two increase from a combined fraction of 7% to about 15%. Panel B of Figure 4 shows the results for all mortgage types, instead of just purchase mortgages. The quantities are essentially unchanged.

Even when focusing on the cross section of credit scores, the two analysis are very similar: MS2015b show in Figure 9 that the bottom quintile by credit score represented about 31% of all delinquent mortgage debt as of 2009, the second quintile was at about 33%, the third represented about 20%, the fourth about 9% and the top quintile about 5%. Given the differences between the datasets, these numbers are surprisingly consistent

---

[12] MS2015b measure the stock of delinquencies for all mortgage-related debt, whereas A2S2015 show results primarily by purchase mortgage cohort. As Figure 7 of A2S2015 shows, however, the differences in the shares of delinquencies by income when we include all mortgage types are very small.
[13] The cutoffs are as follows: the bottom quintile includes borrowers with a FICO score up to 653, the second quintile those with a score up to 698, the third goes up to 736, the fourth goes to 770, and the fifth includes all borrowers with a credit score above 770.

with those we present with LPS data, in particular for the later cohorts in Figure 4. This corroborates the insight in A2S2015 that middle-income and credit score borrowers dramatically increased their share of delinquent debt as the crisis progressed relative to earlier cohorts.

5. Conclusion

In sum, the results presented in this paper confirm that the results in A2S2015 are not driven by "fraudulent income overstatement". We show that there was no decoupling of mortgage growth from income growth at origination over the 2002 to 2006 period once we correctly decompose zip code leverage into the extensive margin (number of loans transacted in a zip code) versus the intensive margin (average loan size). Instead, our results document that mortgage debt at origination grew proportionally across the income distribution, and especially middle- and high-income and borrowers with a FICO score above 660 represented a larger share of defaults once the crisis hit. We also show that the second response to our paper by Mian and Sufi, MS2015b, while written as if the results were contradictory to ours, is entirely consistent with our findings. In addition, these results are hard to reconcile with the earlier view of the crisis that unprecedented levels of lending to low income and low credit score neighborhoods set off the crisis.

REFERENCES

Adelino, Manuel, Antoinette Schoar, and Felipe Severino. 2015. Loan Originations and Defaults in the Mortgage Crisis: The Role of the Middle Class. National Bureau of Economic Research working paper No. 20848.

Ambrose, Brent W., James Conklin, and Jiro Yoshida. 2015. Reputation and exaggeration: Adverse selection and private information in the mortgage market. Working paper.

Ben-David, Itzhak. 2011 Financial constraints and inflated home prices during the real-estate boom. American Economic Journal: Applied Economics 3(3), 55–78.

Cheng, I., Shail Raina, and Wei Xiong. 2014. Wall Street and the housing bubble. American Economic Review 104(9).

Chinco, A., and C. Mayer. 2014. Misinformed speculators and mispricing in the housing market. Working Paper.

Coleman IV, Major, Michael LaCour-Little, and Kerry D. Vandell. 2008. Subprime lending and the housing bubble: Tail wags dog? Journal of Housing Economics 17(4), 272–290.

Demyanyk, Y., and O. Van Hemert. 2011. Understanding the subprime mortgage crisis. Review of Financial Studies 24(6), 1848–1880.

Ferreira, Fernando, and Joseph Gyourko. 2015. A New Look at the U.S. Foreclosure Crisis: Panel Data Evidence of Prime and Subprime Borrowers from 1997 to 2012. National Bureau of Economic Research working paper No. 21261.

Foote, Christopher L., Kristopher S. Gerardi, and Paul S. Willen. 2008. Negative equity and foreclosure: Theory and evidence. Journal of Urban Economics 64(2), 234–245.

Foote, Christopher L., Kristopher S. Gerardi, and Paul S. Willen. 2012. Why did so many people make so many ex post bad decisions: The causes of the foreclosure crisis. FRB Boston Public Policy Discussion Paper Series, paper no. 12-2.

Glaeser, Edward L., Joshua D. Gottlieb, and Joseph Gyourko. 2013. Can cheap credit explain the housing boom? Housing and the Financial Crisis (2013), 301.

Griffin, John, and Gonzalo Maturana. 2014. Who facilitated misreporting in securitized loans? Journal of Finance, forthcoming.

Haughwout, Andrew, Richard Peach, and Joseph Tracy. 2008. Juvenile delinquent mortgages: Bad credit or bad economy? Journal of Urban Economics 64.2: 246-257.

Jiang, W., A. A. Nelson, and E. Vytlacil. 2014. Liar's loan? Effects of origination channel and information falsification on mortgage delinquency. Review of Economics and Statistics, forthcoming.

Keys, B. J., T. Mukherjee, A. Seru, and V. Vig. 2010. Did securitization lead to lax screening? Evidence from subprime loans. Quarterly Journal of Economics, 125(1), 307–362.

Mayer, Chris, and Karen Pence. 2009. Subprime mortgages: What, where, and to whom? In Edward L. Glaeser and John M. Quigley, eds., Housing Markets and the Economy: Risk, Regulation, and Policy. Cambridge, MA: Lincoln Institute of Land Policy.

Mayer, Christopher, Karen Pence, and Shane M. Sherlund. 2009. The rise in mortgage defaults. Journal of Economic Perspectives 23(1), 27–50.

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

Mian, Atif, and Amir Sufi. 2014. House of Debt. Chicago: University of Chicago Press.

Mian, Atif, and Amir Sufi. 2015a. Fraudulent Income Overstatement on Mortgage Applications during the Credit Expansion of 2002 to 2005. National Bureau of Economic Research working paper No. 20947.

Mian, Atif, and Amir Sufi. 2015b. Household debt and defaults from 2000 to 2010: Facts from Credit Bureau Data. National Bureau of Economic Research working paper No. 21203.

Palmer, C. 2014. Why did so many subprime borrowers default during the crisis: Loose credit or plummeting prices?

Piskorski, Tomasz, Amit Seru, and James Witkin. 2015. Asset quality misrepresentation by financial intermediaries: Evidence from RMBS market. Journal of Finance, forthcoming.

Shiller, Robert J. 2014. Speculative asset prices. American Economic Review 104(6), 1486–1517.

Table 1 (Table 2 of Adelino, Schoar and Severino, 2015) Purchase mortgage origination and income

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

Panel A. Mortgage growth measures between 2002 and 2006

	Total Purchase Mortgage Origination		Average Mortgage Size		Number of Mortgages	
Growth of IRS Household Income	-0.182**	0.368***	0.239***	0.587***	-0.402***	-0.218**
	(0.090)	(0.109)	(0.026)	(0.038)	(0.075)	(0.091)
County FE	Y	N	Y	N	Y	N
Number of observations	8,619	8,619	8,619	8,619	8,619	8,619
R2	0.33	0.00	0.68	0.09	0.31	0.00

Panel B. Panel specification

	Total Purchase Mortgage Origination		Average Mortgage Size		Number of Mortgages	
Ln(IRS Household Income)	0.378***	1.104***	0.442***	0.451***	-0.068	0.654***
	(0.080)	(0.095)	(0.033)	(0.048)	(0.073)	(0.075)
Ln(IRS Household Income) x Year 2004		-0.137***		0.010		-0.152***
		(0.021)		(0.010)		(0.018)
Ln(IRS Household Income) x Year 2005		-0.246***		0.020		-0.270***
		(0.026)		(0.014)		(0.021)
Ln(IRS Household Income) x Year 2006		-0.382***		-0.015		-0.369***
		(0.027)		(0.015)		(0.023)
Zip Code FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Number of observations	36,299	36,299	36,299	36,299	36,299	36,299
R2	0.97	0.97	0.97	0.97	0.97	0.97

Table 2 (Table 4 of Adelino, Schoar and Severino, 2015). Mortgage origination and growth in buyer income

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

Panel A. Mortgage growth measures between 2002 and 2006

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

Panel B. Heterogeneity by propensity for income misreporting

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

Panel C. Alternative time periods

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

Table 3 (Table A2 of Adelino, Schoar and Severino, 2015). Mortgage origination and income, IRS and buyer income

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

Panel A. With county fixed effects

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

Panel B. Without county fixed effects

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

Table 4. Mortgage origination and income: dropping zip codes based on MS2015a measures

The table shows OLS regressions of annualized growth in total mortgage credit at the zip code level (purchase mortgages only) on the annualized growth rate of average household income (from the IRS) and the annualized growth rate of average buyer income in the zip code (obtained from HMDA). The table shows the regressions for the whole sample and then dropping the top 10%, 20%, 30% and 40% of zip codes based on the ratio of buyer income to IRS income as of 2006 (Panels A and B), as well as the difference in the annualized growth rate of buyer income and IRS household income between 2002 and 2006. All specifications include county fixed effects. Standard errors are clustered by county. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Sample reduced based on ratio of buyer income to IRS household income

	Growth in Total Purchase Mortgage Origination					Growth in Average Mortgage Size				
	All	< 90th	< 80th	< 70th	< 60th	All	< 90th	< 80th	< 70th	< 60th
Growth of Buyer Income (HMDA)	0.368***	0.342***	0.320***	0.305***	0.308***	0.282***	0.269***	0.269***	0.272***	0.267***
	(0.047)	(0.048)	(0.050)	(0.054)	(0.060)	(0.015)	(0.015)	(0.016)	(0.017)	(0.018)
Number of observations	8,620	7,755	6,893	6,032	5,170	8,620	7,755	6,893	6,032	5,170
R2	0.02	0.02	0.01	0.01	0.01	0.17	0.16	0.16	0.16	0.16

Panel B: Sample reduced based on ratio of buyer income to IRS household income

	Growth in Total Purchase Mortgage Origination					Growth in Average Mortgage Size				
	All	< 90th	< 80th	< 70th	< 60th	All	< 90th	< 80th	< 70th	< 60th
Growth of IRS Household Income	-0.182**	-0.102	-0.059	-0.061	-0.086	0.239***	0.257***	0.264***	0.264***	0.257***
	(0.090)	(0.084)	(0.086)	(0.086)	(0.098)	(0.026)	(0.027)	(0.029)	(0.030)	(0.033)
Number of observations	8,619	7,755	6,893	6,032	5,170	8,619	7,755	6,893	6,032	5,170
R2	0.00	0.00	0.00	0.00	0.00	0.03	0.04	0.04	0.04	0.04

Panel C: Sample reduced based on growth in buyer income minus the growth in IRS household income

	Growth in Total Purchase Mortgage Origination					Growth in Average Mortgage Size				
	All	< 90th	< 80th	< 70th	< 60th	All	< 90th	< 80th	< 70th	< 60th
Growth of Buyer Income (HMDA)	0.368***	0.255***	0.207***	0.175***	0.132**	0.282***	0.258***	0.250***	0.237***	0.224***
	(0.047)	(0.049)	(0.052)	(0.053)	(0.055)	(0.015)	(0.017)	(0.018)	(0.018)	(0.019)
Number of observations	8,620	7,758	6,896	6,034	5,172	8,620	7,758	6,896	6,034	5,172
R2	0.02	0.01	0.01	0.00	0.00	0.17	0.14	0.13	0.12	0.11

Panel D: Sample reduced based on growth in buyer income minus the growth in IRS household income

	Growth in Total Purchase Mortgage Origination					Growth in Average Mortgage Size				
	All	< 90th	< 80th	< 70th	< 60th	All	< 90th	< 80th	< 70th	< 60th
Growth of IRS Household Income	-0.182**	-0.076	-0.003	0.040	0.069	0.239***	0.261***	0.272***	0.270***	0.270***
	(0.090)	(0.080)	(0.083)	(0.084)	(0.085)	(0.025)	(0.025)	(0.027)	(0.029)	(0.030)
Number of observations	8,619	7,758	6,896	6,034	5,172	8,619	7,758	6,896	6,034	5,172
R2	0.00	0.00	0.00	0.00	0.00	0.03	0.04	0.05	0.05	0.05

Table 5. Buyer income (HMDA) and household income from the IRS

The table shows the historical correlation between buyer income from HMDA and per capita income from IRS. Columns 1 and 2 show the correlation including county fixed effect and columns 3 and 4 show the estimates without county fixed effect. Panel A shows the results for a sample of zip codes with available house prices from Zillow, and Panel B show the results for all zip codes in the HMDA dataset. Standard errors are clustered by county. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Zillow sample

	Growth of Buyer Income (HMDA)			
	1998-2002	2002-2006	1998-2002	2002-2006
Growth of IRS Household Income	0.220***	0.113***	0.321***	0.426***
	(0.039)	(0.037)	(0.041)	(0.038)
County FE	Y	Y	N	N
Number of observations	8,605	8,619	8,605	8,619
R2	0.26	0.43	0.02	0.04

Panel B. HMDA sample

	Growth of Buyer Income (HMDA)			
	1998-2002	2002-2006	1998-2002	2002-2006
Growth of IRS Household Income	0.036*	0.072***	0.112***	0.272***
	(0.021)	(0.020)	(0.024)	(0.023)
County FE	Y	Y	N	N
Number of observations	27,308	27,385	27,308	27,385
R2	0.36	0.41	0.00	0.01

Table 6. Summary statistics by house price growth and subprime origination

This table shows descriptive statistics by zip code split by quartiles of the proportion of purchase mortgages originated by subprime lenders (subprime lenders are defined by the HUD subprime lender list), as well as quartiles of house price growth between 2002 and 2006. Data is from HMDA and the sample includes zip codes with non-missing house price data from Zillow.

Panel A. Distribution of zip codes

	Low HP growth	2	3	High HP growth
Low subprime	535	651	646	338
2	639	610	522	398
3	583	604	484	483
High subprime	435	351	539	801

Panel B. Annualized growth in total purchase mortgage origination

	Low HP growth	2	3	High HP growth
Low subprime	7.1%	11.3%	10.2%	14.8%
2	7.5%	11.0%	12.1%	15.3%
3	7.2%	12.1%	13.2%	17.9%
High subprime	8.1%	13.6%	15.9%	18.2%

Panel C. Fraction of total purchase mortgages originated by subprime lenders

	Low HP growth	2	3	High HP growth
Low subprime	3.8%	3.8%	3.8%	4.3%
2	8.0%	8.0%	8.1%	8.0%
3	12.6%	12.4%	12.5%	12.7%
High subprime	22.8%	22.1%	23.3%	26.1%

Panel D. Annualized growth in buyer income (HMDA)

	Low HP growth	2	3	High HP growth
Low subprime	3.8%	5.0%	7.2%	9.9%
2	3.2%	5.1%	7.4%	9.7%
3	3.0%	4.8%	6.9%	10.4%
High subprime	2.8%	5.1%	8.2%	11.9%

Table 7. Delinquency rates by income and credit score bin

The table shows delinquency rates by IRS household income quintiles, as well as credit score at origination, for the purchase mortgage cohorts between 2003 and 2006. Delinquency rates are defined as the fraction of mortgages that become 90 days delinquent or more at any point during the three years after origination. Data are from a 5% random sample of the LPS dataset and include zip codes with non-missing house price data from Zillow.

Panel A. Quintiles formed based on IRS household income as of 2002

	Bottom Quintile	2	3	4	Top Quintile
2003	7.5%	5.4%	3.8%	2.5%	1.4%
2004	8.8%	6.6%	4.7%	3.6%	2.1%
2005	15.6%	11.5%	9.6%	7.5%	4.7%
2006	28.3%	21.5%	17.3%	15.4%	10.4%

Panel B. Credit score (FICO) cutoffs

	FICO < 660	660 ≤ FICO < 720	FICO ≥ 720
2003	12.6%	2.4%	0.6%
2004	15.0%	3.5%	0.8%
2005	24.7%	9.2%	2.8%
2006	40.2%	21.7%	7.3%

Figure 1. Mortgage origination and the ratio of buyer income to IRS income

The table shows the distribution of the ratio of buyer income from HMDA to the IRS household income in the zip code on the left axis (red line), as well as the proportion of total dollar amount of origination for purchase mortgages on the right axis (bars). Panel A includes zip codes with non-missing house prices from Zillow, and Panel B includes all zip codes in the HMDA dataset.

Panel A. Zillow sample
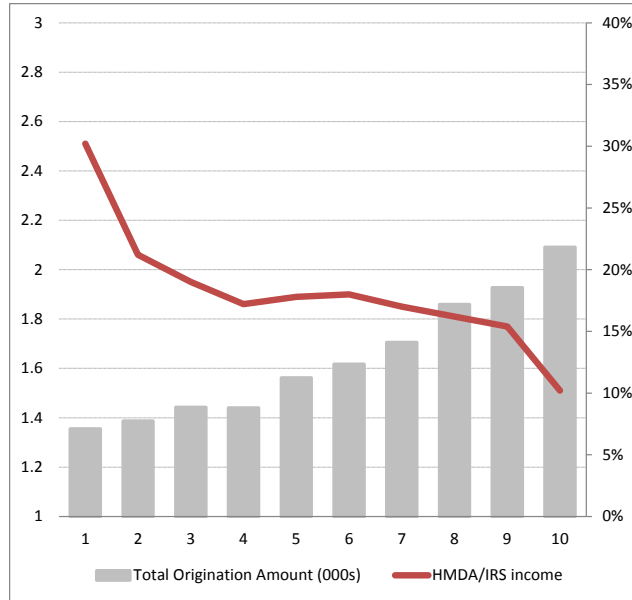
Panel B. HMDA sample
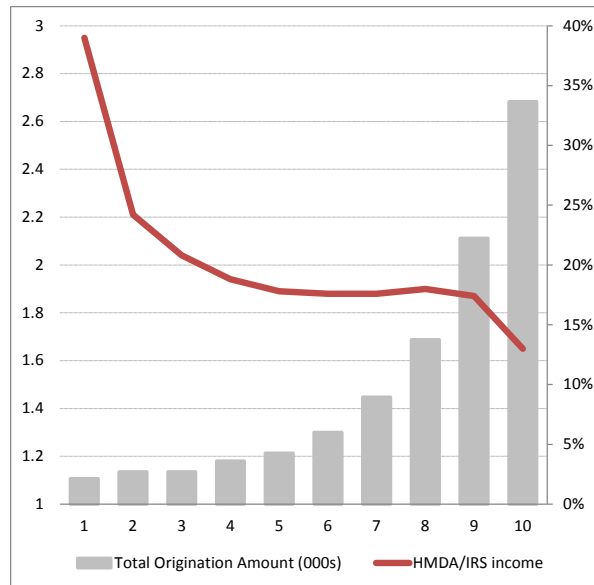
Figure 2. Mortgage origination and delinquency by credit score in prime and subprime neighborhoods

This figure shows the fraction of the dollar volume of purchase mortgages more than 90 days delinquent at any point during the 3 years after origination for the 2003 and 2006 origination cohorts. Panels show splits by above and below the median of subprime origination (defined as the percentage of loans made by lenders on the HUD subprime lender list), as well as by whether the borrower is above or below a credit score of 660 (a common FICO cutoff for subprime borrowers). In each year fractions sum to 100 (the total amount of delinquent mortgages for each cohort), up to rounding error. Sample includes zip codes with non-missing Zillow house price data. Data is from the 5% sample of the LPS dataset and the sample includes zip codes with non-missing Zillow house price data.
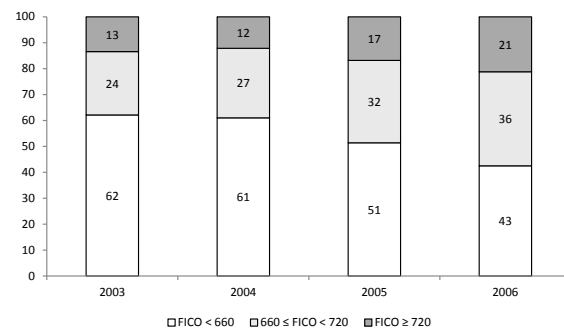
Panel A. Origination
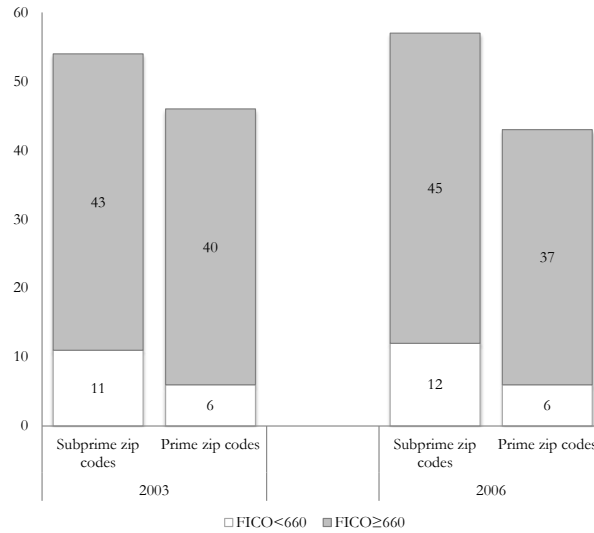
Panel B. Delinquency
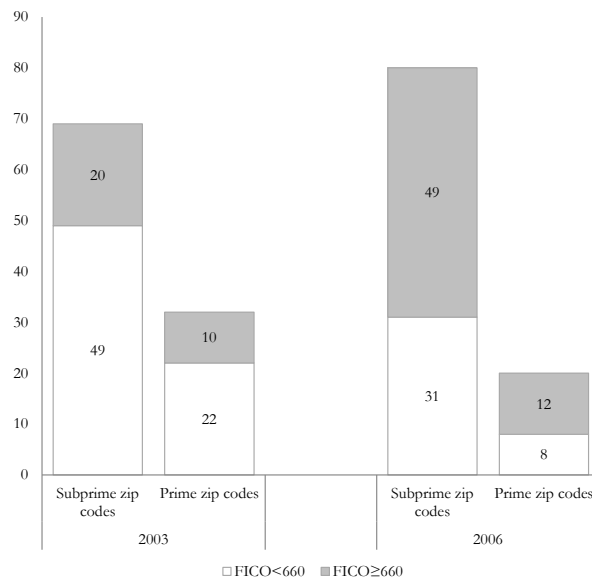
Figure 3 (Figure A4 of Adelino, Schoar and Severino 2015) Mortgage Origination and Delinquency by Credit Score, Alternative Datasets

This figure shows the fraction of total dollar volume of purchase mortgages originated, as well as the total dollar volume of delinquent mortgages by cohort, split by income quintile. Data in Panel A is from Blackbox Logic, a dataset of private-label securitized mortgages, and data in Panel B comes from the public Freddie Mac single-family home dataset. The sample includes zip codes with non-missing Zillow house price data. A mortgage is defined as being delinquent if payments become more than 90 days past due (i.e., 90 days, 120 days or more, in foreclosure or REO) at any point during the 3 years after origination.
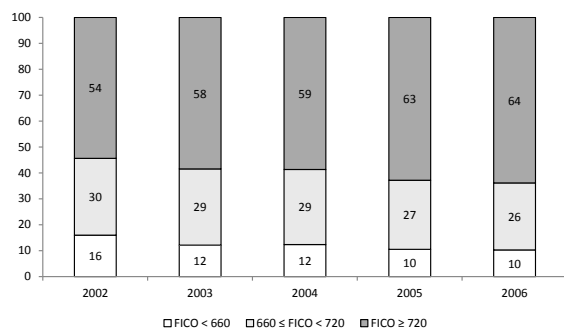
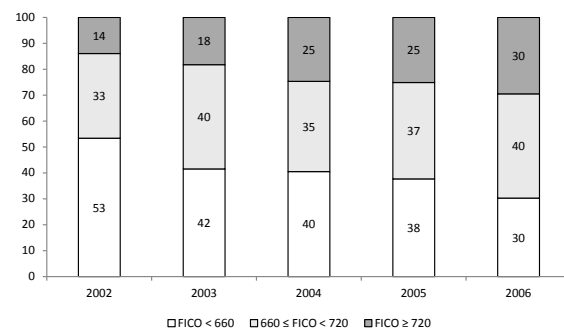**Panel A. Origination (Blackbox Logic data)**
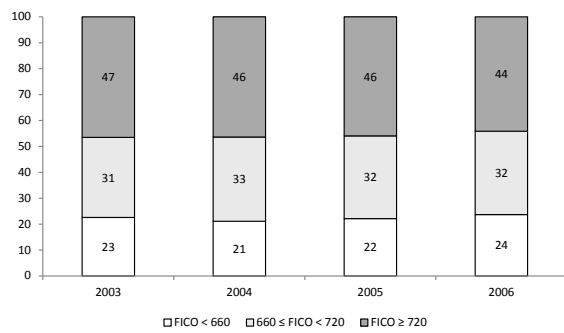
Panel B. Delinquency (Blackbox Logic data)

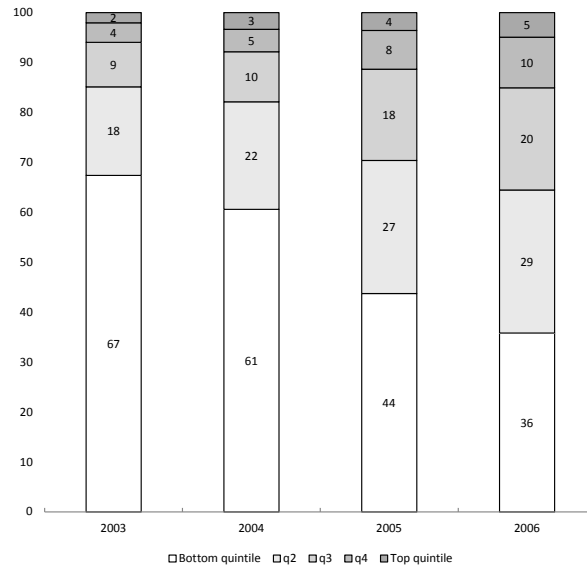**Panel C. Origination (Freddie Mac data)**

Panel D. Delinquency (Freddie Mac data)

Figure 4: Delinquency by credit score quintiles

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

Panel A. Purchase mortgages

Panel B. All mortgage types