Credit Growth and the Financial Crisis: A New Narrative*

Stefania Albanesi, University of Pittsburgh, NBER and CEPR
Giacomo DeGiorgi, GSEM-University of Geneva, BREAD, CEPR
Jaromir Nosal, Boston College

December 27, 2021

Abstract

This paper studies mortgage debt and defaults between 2001 and 2012 using a large, nationally representative panel of credit reports. We show that credit growth between 2001 and 2006 was concentrated in the prime segment, challenging the broadly accepted view that the financial crisis was caused by an expansion of credit to subprime borrowers. We also find that the 2007-2009 spike in mortgage defaults was concentrated among prime borrowers. Among those borrowers, we find that real estate investors played a critical role in the rise of delinquencies and foreclosures. We also examine the geographical variation in mortgage debt and defaults. We find that everywhere prime borrowers accounted for most of the rise in mortgage debt during the 2001-2006 boom and the subsequent rise in defaults during the crisis. Demographic factors such as age distribution, racial composition and educational attainment may account for the greater exposure to the 2007-2009 recession of areas with a large concentration of subprime borrowers. Our findings suggest that, borrower characteristics, such as credit scores, may be less important than behavior, such as investment activity, in driving mortgage defaults, thus providing an alternative narrative that challenges the large role of subprime credit for the crisis.

*We are grateful to Christopher Carroll, Ambrogio Cesa-Bianchi, Gauti Eggertsson, Joel Elvery, Nicola Gennaioli, Richard Harrison, Douglas McManus, Virgiliu Midrigan, Ned Prescott, Giorgio Primiceri, Joe Tracy, Eric Swanson, Paul Willen and many seminar and conference participants for useful comments and suggestions. We also thank Matthew Ploenzke, Jakob Fabina, Richard Svoboda and Harry Wheeler for excellent research assistance. De Giorgi acknowledges financial support by the SNF no. 100018-182243, and the EU FP7-631510. Correspondence to: stefania.albanesi@gmail.com.

1 Introduction

Understanding the causes of the 2007-2009 financial crisis is one of the key questions in modern economics and it is critical to designing policies and regulation to prevent similar episodes in the future. The rise in mortgage defaults in the United States starting in late 2006 is viewed by most as the precipitating factor, with the expansion of subprime mortgages in 2001-2006 as its leading cause. The broadly accepted narrative about the financial crisis is based on the findings in Mian and Sufi (2009) and Mian and Sufi (2015) suggesting that most of the growth in credit during the 2001-2006 boom was concentrated in the subprime segment, despite the fact that income did not rise over the same period for this group of borrowers. The expansion of subprime credit is then seen as the leading cause of the rise in mortgage delinquencies and foreclosures, which caused the housing crisis and subsequently the 2007-2009 recession.[1]

This paper studies the evolution of household borrowing and default between 2001 and 2012 using a large administrative panel of anonymous credit files from the Equifax credit reporting bureau. The data contains information on individual debt holdings, defaults and credit scores. We examine the evolution of mortgage debt and defaults during the housing boom and throughout the financial crisis and its aftermath. Our findings suggest an alternative narrative that challenges the view that an expansion of the supply of mortgage credit to subprime borrowers in 2001-2006 played a large role in the housing and financial crisis. Specifically, we show that mortgage balance growth between 2001 and 2006 is concentrated in the middle and at the top of the credit score distribution. We also find that the same borrowers are responsible for the rise in defaults during the 2007-2009 crisis. Borrowers with subprime credit scores typically have higher default rates than those with higher credit scores, however, during the financial crisis, the share of mortgage delinquencies attributable to borrowers with prime credit scores increased from 50% to 60% and their share of foreclosures from 20% to 50%.

Mian and Sufi (2009) and Mian and Sufi (2015) identify subprime borrowers based on their credit score in the late 1990s. We show that low initial credit score individuals exhibit future credit score growth over time, while that is not true for higher credit score borrowers. These dynamics are mostly driven by the fact that low credit score individuals are disproportionally young. This approach therefore is at risk of confounding an expansion of the supply of mortgages to high risk borrowers with the life cycle demand for credit of borrowers who were young at the start of the boom. To avoid this pitfall, our approach estimates future growth in mortgage balances and mortgage delinquencies based on borrowers' recent lagged credit score. This is closer to industry practices and prevents joint endogeneity of credit scores with borrowing and delinquency behavior, while ensuring that the ranking best reflects the borrower's likely ability to repay debt at the time

[1] See Mian and Sufi (2009), Mian and Sufi (2010), Mian and Sufi (2011), Mian, Rao, and Sufi (2013) and Mian, Sufi, and Trebbi (2015). More recently, Adelino, Schoar, and Severino (2016) and Foote, Loewenstein, and Willen (2016) have made the case that the growth in mortgage balances was concentrated in the middle of the income distribution.

of borrowing.

We use a variety of approaches to establish that growth in mortgage balances was concentrated among borrowers with a relatively high credit score. First, we calculate the one-year-ahead growth in mortgage balances by four-quarter-lagged credit score quartile. To account for the impact of differences in the age distribution across credit score quartiles, we estimate a time effect regression where age and time effects are interacted with the lagged credit score quartiles. Finally, we consider the distribution of new mortgage originations and of the fraction with a mortgage by lagged credit score quartile. All these approaches confirm that mortgage borrowing grew more for borrowers in the middle and at the top of the credit score distribution between 2001 and 2006, both on the intensive and the extensive margin. Further, using payroll data for 2009, merged with the credit bureau files for a subset of individuals, we show that there is a positive correlation between income levels in 2009 and growth in mortgage balances between 2001 and 2006. Using PSID data, we also show that growth in mortgage balances in 2001-2006 is positively associated with growth in income over the same period.

To directly compare our results to Mian and Sufi (2009), we also conduct an analysis at the zip code level. Specifically, we sort zip codes into four quartiles based on the share of subprime borrowers in 1999.[2] We find that zip codes with the largest share of subprime borrowers exhibit stronger growth in per capita mortgage balances over the 2001-2006 boom, confirming previous findings. However, in all zip codes, *prime* borrowers are responsible for most of the credit growth, with the prime share of mortgage balances increasing substantially in all zip codes. Additionally, we find that the excess growth in mortgage balances for prime borrowers compared to subprime borrowers is the highest in zip codes with the highest share of subprime borrowers. This finding calls into question the practice of proxying the behavior of individual subprime borrowers with that of locations with a high concentration of such borrowers for the purpose of measuring mortgage activity. Additionally, we show that in all zip codes it was prime borrowers who drove the rise in defaults during the crisis. The prime share of foreclosures goes up in all zip codes by 10 to 25 percentage points, and it grows more in zip codes with a high share of subprime borrowers, reflecting the behavior of mortgage balances.

The sharp rise in defaults for mid-to-high credit score borrowers is puzzling, as these borrowers historically exhibit very low default rates on any type of debt and even lower mortgage default rates. It is therefore critical to examine why borrowers with good credit histories disproportionately contributed to the foreclosure crisis. We show that mortgage defaults by mid-to-high credit score borrowers were predominantly driven by real estate investors, which we identify as borrowers with two or more first mortgages, following Haughwout et al. (2011). We show that real estate investors played a critical role in the rise in mortgage balances for borrowers in the middle and at the top of the credit score distribution. Investment activity surged dramatically for mid-to-high credit

[2]Subprime borrowers are those with Equifax Risk Score below 660. The share of subprime borrowers at the zip code level is quite stable, so the results are not sensitive to the sorting year.

score borrowers starting in 2004, resulting in the investor share of mortgage balances for prime borrowers rising from around 20% in 2001 to above 30% in 2007. Most importantly, we show that the rise in mortgage defaults is also driven to a large extent by real estate investors. The fraction of investors with new mortgage delinquencies grew by a factor of 5 or more for the top two credit score quartiles between 2005 and 2009, while it grew by less than 50% for non-investors. We find a similar but magnified pattern for foreclosures. This striking result shifts the focus from borrowers' *characteristics*, such as credit scores, to borrowers' *behavior*, and provides guidance to policy makers interested in understanding the cause of the housing crisis and designing interventions to mitigate and prevent future such episodes.

We also explore the broader aggregate implications of our findings. In response to the 2007-2009 housing crisis, a large empirical literature has exploited geographical variation to establish a positive correlation between mortgage debt growth and the severity of the recession,[3] to provide evidence on the role of the collateral channel in the transmission of financial shocks to real economic activity.[4] Our empirical findings on the zip code level call into question the mechanism linking the size of mortgage debt growth during the credit boom to the depth of the recession due to the tightening of collateral constraints on marginal borrowers with high marginal propensity to consume. To provide some insight into other potential mechanisms, we explore additional factors that may explain the zip code level patterns. We show that zip codes with higher fraction of subprime borrowers are younger, have lower levels of educational attainment and have a disproportionately large minority share of the population. It is well known that younger, less skilled and minority workers suffer larger and more persistent employment loss during recessions (see Mincer (1991) and Shimer (1998)). Taken together, our findings suggest that using geographically aggregated data does not provide a good approximation of the distribution of debt across different types of borrowers. Moreover, the positive correlation between credit growth during the boom and the depth of the ensuing recession may be due to other factors, such as the prevalence of young, unskilled and minority workers who are particularly sensitive to the business cycle.

1.1 Relation to the Literature

Our primary contribution is to show that both the growth in mortgage balances during the 2001-2006 boom and the rise in mortgage defaults during the 2007-2009 crisis was driven by borrowers in the middle and at the top of the credit score distribution. Our analysis clarifies a number of conflicting results in the literature. Early work on the housing crisis, such as Mian and Sufi (2009),

[3]Some examples include Mian and Sufi (2009), Mian and Sufi (2011), Mian, Sufi, and Trebbi (2015), Mian, Rao, and Sufi (2013), Mian and Sufi (2010), Midrigan and Philippon (2011), Kehoe, Pastorino, and Midrigan (2016), Keys et al. (2014).

[4]There is a large theoretical literature on the role of collateral constraints in causing or amplifying swings in economic performance, following the pioneering work of Kiyotaki and Moore (1997). Some recent contributions include Iacoviello (2004), Berger et al. (2015), Corbae and Quintin (2015), Mitman (2016), Justiniano, Primiceri, and Tambalotti (2016), Kaplan, Mitman, and Violante (2017).

had identified subprime borrowers as primarily responsible for both the rise in mortgage debt in 2001-2006 and the rise in defaults in the subsequent crisis, using zip code level data and credit score distributions from before the housing boom. Such borrowers are interpreted as being low income and at high risk of default. Our findings on mortgage balance growth differ from those in Adelino, Schoar, and Severino (2016) and Adelino, Schoar, and Severino (2017), who show that the rise in mortgages is equal across borrowers in different credit score bins. By contrast, we show that it is borrowers in the middle and at the top of the credit score distribution that exhibited the highest growth. The difference between our conclusions comes from the fact that those papers base their analysis on the distribution of new mortgages by credit score and income at origination, which is a measure of gross changes in mortgages, whereas with our data we are able to isolate net changes in mortgage balances. Our results are consistent with Foote, Loewenstein, and Willen (2016), who find that the geographical relation in mortgage debt growth and income does not change relative to previous periods during the 2001-2006 credit boom, and there is no relative growth in defaults for areas with higher concentration of marginal borrowers. Our analysis combines individual level data and geographically aggregated data and shows that in areas with a high concentration of subprime borrowers both the increase in mortgage debt and the increase in defaults is driven by prime borrowers.[5]

We show that the discrepancy in the findings on the distribution of mortgage debt growth is driven by the fact that Mian and Sufi (2009) and Mian and Sufi (2015) classify borrowers by their credit scores in the late 1990s, and thus identify low credit score borrowers with those who were young at the start of the boom. We show that low initial credit score borrowers exhibit the fastest subsequent growth in credit score, and that there is a strong positive relation between the life cycle evolution of income, mortgage balances and credit score. To avoid this pitfall, our approach estimates future growth in mortgage balances based on borrowers' recent lagged credit scores. This is closer to industry practices and ensures that the ranking best reflects the borrower's likely ability to repay debt at the time of borrowing. We show that the cross sectional variation of credit scores is mostly explained by variation in labor income, conditional on age. We also clarify that in individual level data growth in mortgage balances is positively linked to growth in labor income, though this relation does not hold in geographically aggregated data. Additionally, based on estimates with the individual data, we show that there is a sizable life cycle pattern in mortgage balance growth, with a hump in middle age, but only for mid-to-high credit score borrowers, consistent with the interpretation that the life cycle 'need' to borrow must be accompanied by a sufficiently high credit score to qualify for mortgage debt.

We contribute to the literature on real estate investors by attributing the disproportionate rise in mortgage defaults during the crisis for borrowers with relatively high credit scores to real

[5]Ferreira and Gyourko (2015) also find that default activity by prime borrowers intensifies during the crisis, however, their definition of prime/subprime borrowers is based on lender characteristics, not on the individual characteristics of the borrower.

estate investors, and document the importance of investor activity at the individual and zip code level. Bhutta (2015) examines the contribution of real estate investors to the growth of mortgage balances during the boom but does not document their default behavior and does not differentiate by credit score. While Mian and Sufi (2018) emphasizes the link between investor activity and private label securitization, we emphasize the critical role of investors in driving the foreclosure crisis, as these borrowers were primarily responsible for the rise in defaults. Haughwout et al. (2011) also show that investors account for large increases in the share of home purchases and delinquencies, but they focus on selected geographical areas with large real estate booms. Albanesi (2018) provides a comprehensive empirical analysis of the behavior of real estate investors between 2004 and 2012 and explores the factors behind their higher default rates, such as high leverage and the prevalence of strategic default. The rise in investor activity can reconcile the fact that though the number of mortgage originations rose during the boom, homeownership and the fraction of the U.S. population with a mortgage did not vary substantially over the same period. Bhutta and Keys (2016) and Chen, Michaux, and Roussanov (2013) show that in 2001-2003 the surge in originations was mainly driven by refinancing activity, whereas based on our analysis, investment activity mostly contributed to the ongoing but slower pace of originations growth starting in 2004. Our findings shift the emphasis from borrower characteristics, such as the credit score, to borrower behavior, such as investment activity, as the main contributor to the rise in defaults during the 2007-2009 mortgage crisis.

The rest of the paper is organized as follows. Section 2 describes the data, existing evidence, and our new findings on the behavior of debt and defaults by recent credit score. Section 2.4 explains the role of life cycle in the dynamics of credit scores, income and mortgage debt. Section 2.5 presents results at the zip code level, showing that prime borrowers drive the growth in mortgage balances in all zip codes. Section 3 discusses the role of real estate investors in defaults by high credit score borrowers and examines factors affecting the geographical variation in mortgage debt and defaults. Section 4 concludes.

2 Distribution of Debt and Defaults

Our analysis is based on the Federal Reserve Bank of New York's Consumer Credit Panel/Equifax Data (CCP), an anonymous longitudinal panel comprising a 5% random sample of all individuals who have a credit report with Equifax. The data allows us to track all aspects of borrowers' financial liabilities, including bankruptcy and foreclosure, mortgage status, detailed delinquencies, various types of debt, with number of accounts and balances. Apart from the financial information, the data contains individual descriptors such as age, ZIP code and credit score. We provide more detail on the data in Appendix A. For 2009, we also have access to payroll data for a subset of approximately 11,000 borrowers merged with their credit files from Equifax Workforce Solutions, referred to as Worknumber data. These data are described in detail in Appendix A.1. We use the

merged data to explore the relationship between income, credit scores and debt in Section 2.4.

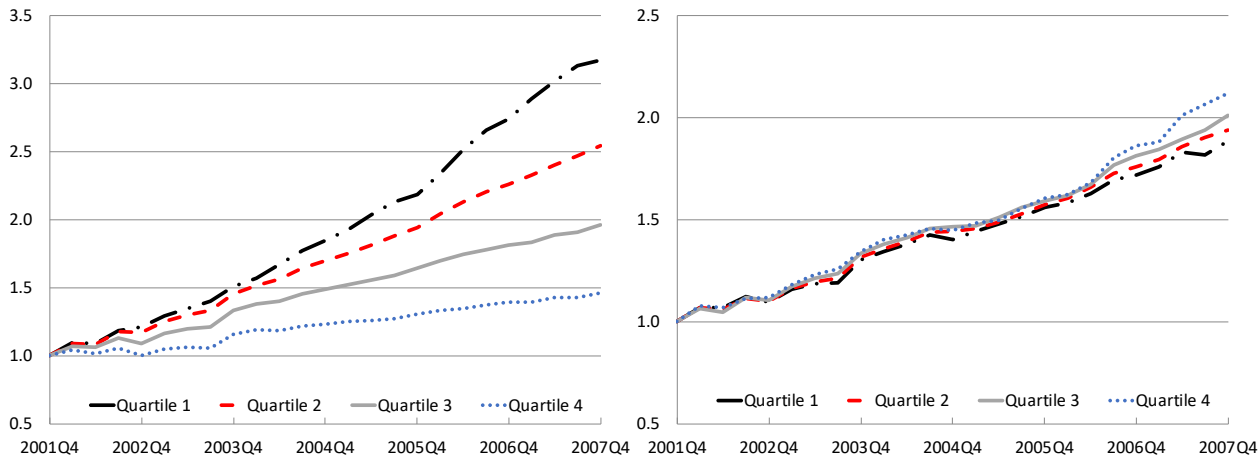
2.1 Existing Evidence

Most of the existing literature has focussed on the role of subprime borrowers both in fueling the growth in mortgage balances during the 2001-2006 boom and in the subsequent foreclosure crisis. However, a new consensus has recently been emerging on the fact that subprime and low income borrowers were not the primary drivers of the expansion of mortgage balances in 2001-2006 (see Adelino, Schoar, and Severino (2016), Adelino, Schoar, and Severino (2017) and Foote, Loewenstein, and Willen (2016)). Our findings are consistent with this new consensus and we also explain why our results differ from prior analyses based on credit scores. The original narrative on the role of subprime borrowers is exemplified by the seminal work of Mian and Sufi (2009). They rank zip codes in selected Metropolitan Statistical Areas (MSAs) by the fraction of residents with Equifax Risk Score below 660 in 1996 and show that zip codes with a larger fraction of these borrowers, referred to as *subprime*, exhibit stronger credit growth during the credit boom. Mian and Sufi (2015) refine this analysis with individual level data. They rank individuals by their 1997 Vantage Score and show that borrowers in the first quartile of the 1997 credit score distribution show a larger increase in mortgage balances during the boom. We show that these findings are a consequence of using *initial* credit scores to rank borrowers. Low credit score individuals are disproportionately young and experience subsequent increase in their credit scores compared to higher credit score borrowers, who are typically older. That growth in credit scores is associated with growth in income and followed by growth in mortgage balances. Hence, the rise in mortgage balances for borrowers who have low credit score in the late 1990s does not necessarily reflect an expansion in the supply of credit to risky borrowers during the 2001-2006 boom, but simply the typical life cycle demand for borrowers who were young just before the start of the boom.

We replicate the results in Mian and Sufi (2015) and Mian and Sufi (2009) by ranking individuals and zip codes by their credit score at the earliest available date in our sample. For individuals, we consider quartiles of the Equifax Risk Score distribution in 1999, while for the zip code level analysis, we rank zip codes by the fraction of individuals with Equifax Risk Score lower than 660 in 1999. Figure 1 displays the growth of per capita mortgage debt balances relative to 2001Q4, which is the last quarter of the 2001 recession, according to the NBER business cycle dates. The left panel displays the individual data, where borrowers are ranked based on their average credit score in 1999. The first quartile contains the individuals with the lowest credit score.[6] The right panel presents zip code level evidence. Here, quartile 1 corresponds to the zip codes with the *lowest* fraction of subprime borrowers in 1999, where subprime borrowers are identified as having an Equifax Risk Score lower than 660. The median fraction of subprime borrowers in 1999 is 8.8%

[6]The Equifax Risk Score cut-offs for the individual ranking are 615 for quartile 1, 710 for quartile 2, 778 for quartile 3, and 836 for quartile 4. The typical cut-off used to identify subprime borrowers is 660, therefore, quartile 1 comprises only subprime borrowers and quartile 2 contains mainly prime individuals.

for zip codes in quartile 1, 29.5% for those in quartile 2, 44% in quartile 3 and 71.8% in quartile 4 (see Table 7).



(a) Individuals: Ranked by 1999 Equifax Risk Score (b) Zip Codes: Ranked by Fraction of Subprime in 1999

Figure 1: Per capita real mortgage balances, ratio to 2001Q3. Source: Authors' calculation based on Federal Reserve Bank of New York's Consumer Credit Panel/Equifax Data.

For the individual data, the net growth in per capita mortgage balances by initial credit score between 2001Q4 and 2007Q4 is 217% for quartile 1, 154% for quartile 2, 96% for quartile 3, and 46% for quartile 4. At the zip code level, the growth of per capita mortgage balances by share of subprime borrowers during the expansion is 88% for quartile 1 (lowest share), 94% for quartile 2, 101% for quartile 3, and 112% for quartile 4 (highest share). While at the individual level there is much more variation by credit score, both the individual and the zip code level data suggest a stronger growth in mortgage balances for borrowers with low initial credit score and zip codes with a large initial fraction of subprime borrowers.

In what follows, we provide evidence that the growth in mortgage balances and the subsequent defaults during the crisis are driven by borrowers with credit scores in the middle and at the top of the distribution at the time of borrowing. Then, we relate our findings to the ones in figure 1 by showing that conditioning on initial credit scores selects relatively young borrowers who experience subsequent credit score growth and mortgage debt growth due to life cycle factors. Finally, we explore the potential drivers behind the rise in mortgage defaults among high credit score borrowers and show that real estate investors play a disproportionate role.

2.2 Mortgage Growth by Recent Credit Score

We now present our approach to characterizing the distribution of debt and defaults during the boom and defaults during the crisis based on default risk at the time of borrowing, as measured by the individual credit score.

The credit score is the most common proxy for individual credit risk, giving an ordinal ranking of borrowers by their predicted default risk, and is widely used by the financial industry. The credit score can be accessed by lenders together with the borrower's credit report and is often a key determinant of loan terms, such as the interest rate, the downpayment and size of the loan. We have access to the Equifax Risk Score, which is a proprietary measure designed to capture the likelihood of a consumer becoming 90+ days delinquent within the subsequent 24 months. The measure has a numerical range of 280 to 850, where higher scores indicate lower default risk. Since credit scores are proprietary models, it is useful to relate credit scores to other observables that are critical for default risk. Using the Equifax credit files with payroll income information available in 2009, we find a strong positive relation of credit score and income, particularly for younger borrowers. Detailed information on credit scores and their relation to age and income is provided in Appendix D.

Our goal is to document the relation between credit score at the time of borrowing and subsequent mortgage debt growth. Most mortgage applications in the United States are approved within a few weeks, and mortgage lenders use the credit score at the time of application and sometimes additional information in the credit report to determine the borrower's risk profile. This suggests that a recent credit score should be used to rank borrowers based on their perceived default risk at the time of contracting a new loan. The credit score may itself influenced by a borrower's behavior, such as inquiries, originations and balance changes. To avoid any influence of a borrower's current behavior on the credit score used for ranking, we use a 4 quarter lagged credit score. Specifically, at quarter t, we rank borrowers by their credit score at $t-4$ and calculate the net percentage change in mortgage balances per capita between quarter $t+4$ and t. Panel (a) in Figure 2 shows that borrowers in quartile 2 of this recent credit score ranking, exhibit the strongest annual percentage growth in mortgage balances, followed by quartiles 3 and 1, and then by quartile 4. The lowest credit score quartile exhibits the fastest growth only in 2006Q2 and Q3. To compare to the prior literature, panel (b) of Figure 2 presents the same growth rates for quartiles of the 1999 credit score ranking. Most notably, sorting by initial credit scores substantially increases the growth rate in the lowest credit score quartile after 2002. This comparison illustrates the importance of the timing of credit score rankings.

Another important consideration pertains to the use of growth rates. Following the prior literature, Figure 2 presents the net percentage rates of growth across credit score quartiles with the base given by the average per capita balances in each quartile in each quarter. As we show in Figure 14 in Appendix B, the level of mortgage balances is increasing in the credit score. At the start of the sample, average mortgage balances for quartile 1 are approximately \$10,000, whereas they are close to \$25,000 for quartiles 2-4. Even with higher percentage growth rates for lower credit scores, the resulting distribution of mortgage balances becomes increasingly skewed towards higher credit scores. This is because, even though the growth rate of balances in quartiles 3 and 4 is the smallest, it applies to the highest level of mortgage balances. By 2007, average mortgage balances for

(a) Ranking based on 4-quarter-lagged credit score

(b) Ranking based on 1999 credit score

Figure 2: 4-quarter ahead growth in credit score quartile-level average mortgage balances. 3QMA. Source: Authors' calculations based on FRBNY CCP/Equifax Data.

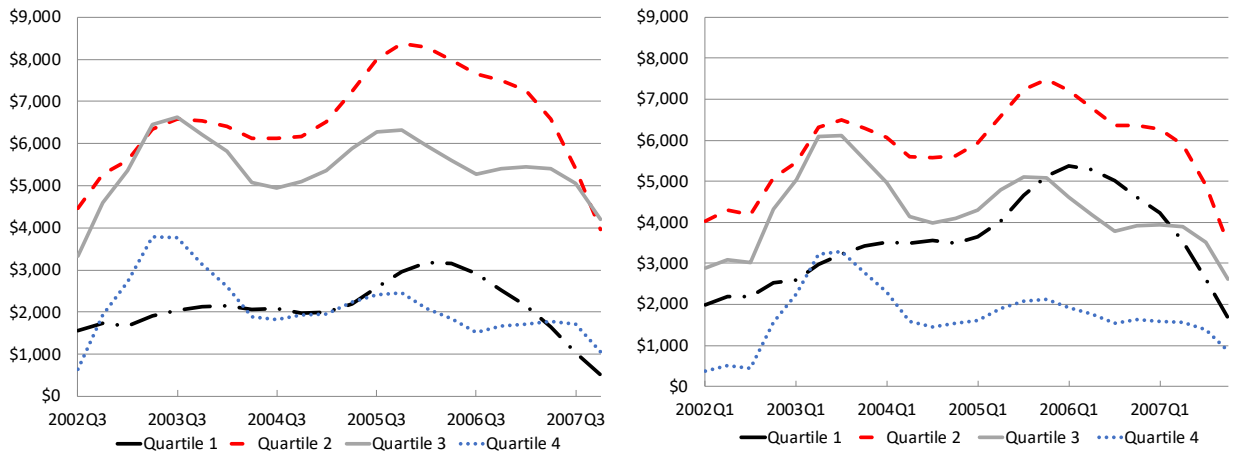
quartile 1 have risen only to $15,000 while they reach $45,000-50,000 for the higher quartiles. For this reason, Figure 3 reports the 4-quarter ahead change in the average balances by quartile. Panel (a) reports this measure by 4-quarter lagged credit score ranking and shows that the actual change in average balances in each quartile is highest in quartiles 2 and 3, which average approximately $7,000 and $5,000 yearly growth in mortgage balances, respectively, followed by quartiles 4 and 1, which average approximately $1,500 of yearly growth in mortgage balances. Panel (b) plots the same measures by 1999 credit score ranking. Even with this ranking, quartile 2 and 3 exhibit the largest yearly growth in mortgage balances, followed by quartile 1 and quartile 4. Figure 3 more accurately illustrates the shift in the distribution of new debt in the population towards high credit score individuals.

2.2.1 Regression Analysis

Debt growth can also be influenced by other factors, most notably the borrowers' life cycle, as we will show in Section 2.4. To isolate the effect of credit scores on mortgage balance growth, we use a regression approach.

We relate yearly growth in mortgage balances to the credit score lagged by one quarter. The one quarter lag in the credit score matches information available to lenders when reviewing loan applications and prevents joint endogeneity between the credit score and future borrowing behavior.[7]

[7]For mortgages, lenders typically also verify a borrower's recent income history. We do not have access to income, though, as we show in Appendix D.2, income and recent credit score are positively related, conditional on age.



(a) Ranking based on 4-quarter lagged credit score

(b) Ranking based on 1999 credit score

Figure 3: 4-quarter ahead change in credit score quartile-level average mortgage balances. 3QMA. Source: Authors' calculations based on FRBNY CCP/Equifax Data.

Our baseline specification is:

$$\Delta b_t^i = \sum_{j=1,2,3,4} \alpha(j_{-1}) + \text{ quarter fe} + \text{ age fe} + \text{ interactions} + \varepsilon_t^i, \tag{1}$$

where Δb_t^i denotes the change in mortgage balances between quarters t and $t+4$ for individual i. The main explanatory variable is $\alpha(j_{-1})$ which is a fixed effect for the 1-quarter lagged quartile of the credit score distribution.

The baseline specification includes quarter effects, age effects and their interaction with the 1-quarter-lagged credit score quartile. The interaction between quarter effects and the 1-quarter-lagged credit score captures the heterogeneity in borrowing behavior that we seek to identify. Additionally, the interaction between age and the 1-quarter-lagged score are meant to introduce flexibility to the profiles of mortgage balance growth with respect to the starting score for borrowers of different age.

To report the findings from estimating equation 1, we plot the interaction between time effects and the quartile effects, adjusted by adding the estimated quartile effect and the average quartile-specific age effect, where such average is calculated using the quartile-specific age distribution, reported in Table 1, and the estimated quartile specific age effects presented in Figure 5.[8] This adjustment affects levels only and not the time series variation of our results. Figure 4 presents these estimates for the baseline specification. Table 9 in Appendix C reports additional coefficient estimates.

Quartile 1 borrowers experience a steady $1,000-2,000 yearly increase in balances between 2001

[8]The regressions are estimated for the population of 20-85 year old borrowers.

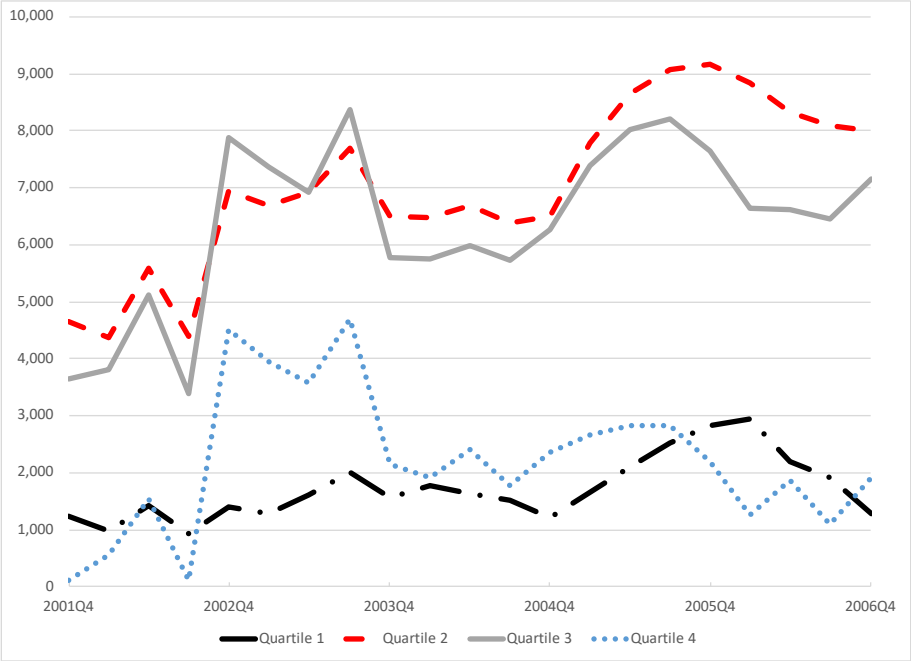


Figure 4: Estimated age-adjusted time effects by 1Q lagged Equifax Risk Score quartile from balance change regressions obtained from the time×credit score quartile interactions. Dependent variable is the 4Q ahead change in per capita mortgage balances in USD. Sample period 2001Q1-2011Q4. Source: Authors' calculations based on FRBNY CCP/Equifax Data.

and 2005, rising to approximately $3,000 by the end of 2006, while the growth for quartiles 2 and 3 is three to five times larger over the same period. Quartile 4 borrowers experience similar mortgage balance growth as quartile 1 for most of the sample period, except in 2002-2003, corresponding to the refinancing boom (Chen, Michaux, and Roussanov (2013) and Bhutta and Keys (2016)). In this two year period, mortgage balance growth also rose for quartile 2 and 3 borrowers, but not for quartile 1 borrowers. In late 2004 and in 2005, all borrowers experience a sizable increase in mortgage balance growth, with the peak in the 4-quarter-ahead change in mortgage balances reaching $9,000 for quartile 2 borrowers, $8,000 for quartile 3 borrowers and approximately $3,000 for quartile 1 and quartile 4 borrowers. The level of the estimated changes as well as their dynamics are consistent and confirm the findings for simple averages in Figure 3. They can be put in perspective by the credit quartile-specific mean debt holdings reported in Figure 14 in Appendix B. Since quartile 1 balances are significantly lower than other quartiles, the findings of low dollar growth in balances reinforces further the conclusion that mortgage debt growth was driven by borrowers in the middle of the credit score distribution.

Figure 5 presents the estimated age effects for each credit score quartile. The estimates suggest a strong life cycle component in mortgage balance growth, that varies substantially by credit score. For quartile 1 borrowers, yearly mortgage balance growth is highest at age 20-23, at around $3,000, and drops to approximately 0 by age 48, hovering at -$1,000 at age 60 and higher. For borrowers

in quartiles 2-4, mortgage balance growth peaks at age 27-35, and monotonically declines with age after that. Quartile 3 borrowers experience the highest peak mortgage balance growth at approximately $10,000, followed by quartile 2 borrowers at $8,000 and quartile 4 borrowers at $6,000. As above, given that the lowest credit score quartile borrowers also have by far the lowest balances, this implies very little life cycle growth in the lowest quartile. Intuitively, in order to see a pronounced life cycle pattern in mortgage balances, the need to borrow due to life cycle reasons has to be matched with the ability to borrow determined by a sufficiently high credit score.

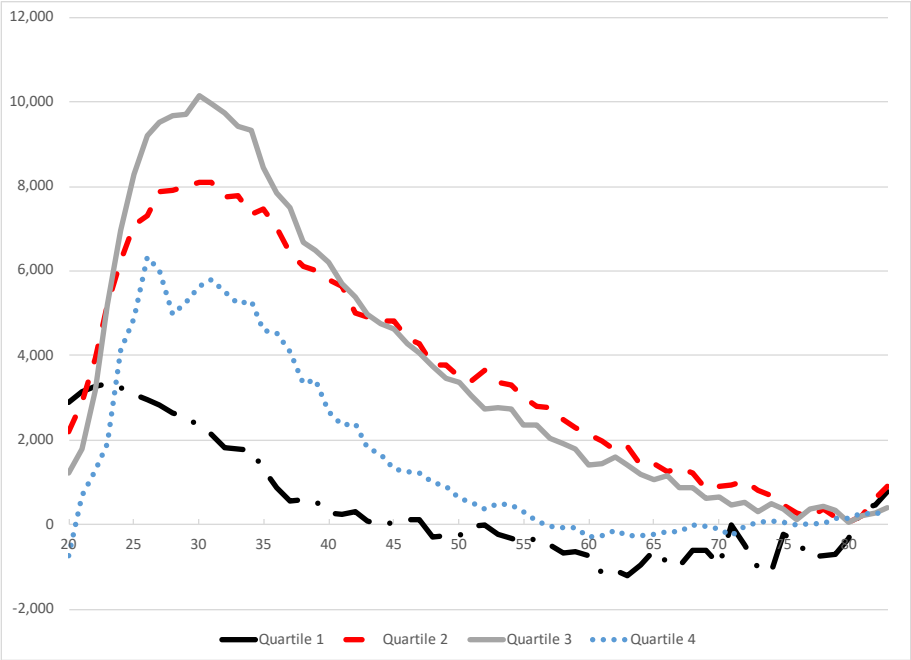


Figure 5: Estimated age effects from balance change regressions obtained from age×credit score quartile interactions. Dependent variable is the 4Q ahead change in per capita mortgage balances in USD, omitted category is age 20. Sample period 2001Q3-2011Q4. Source: Authors' calculations based on FRBNY CCP/Equifax Data.

2.2.2 Mortgage Originations

To corroborate our analysis of mortgage balances, we also examine borrowing behavior on the extensive margin, ranking borrowers by their 4-quarter-lagged credit score.

Figure 6 presents the fraction of individuals with a new mortgage origination, by 4-quarter-lagged credit score quartiles. The fraction of new mortgage originations for borrowers in quartile 1 of the credit score distribution remains stable throughout the sample period, exhibiting a slight downward trend, from just above 10% in 2001Q4 to 6.7% in 2007Q4. Quartiles 2 through 4 peak in 2003Q4 (at 19.5%, 24% and 17.8%, respectively) and then drop significantly through the rest of the sample period. These findings show that new mortgage originations were concentrated in the

middle and at the top of the credit score distribution, which is consistent with the pattern we found for mortgage balances. As previously noted, the sizable rise in mortgage originations for quartiles 3 and 4 in 2003 and 2004 reflects the rise in refinancing activity over that period.

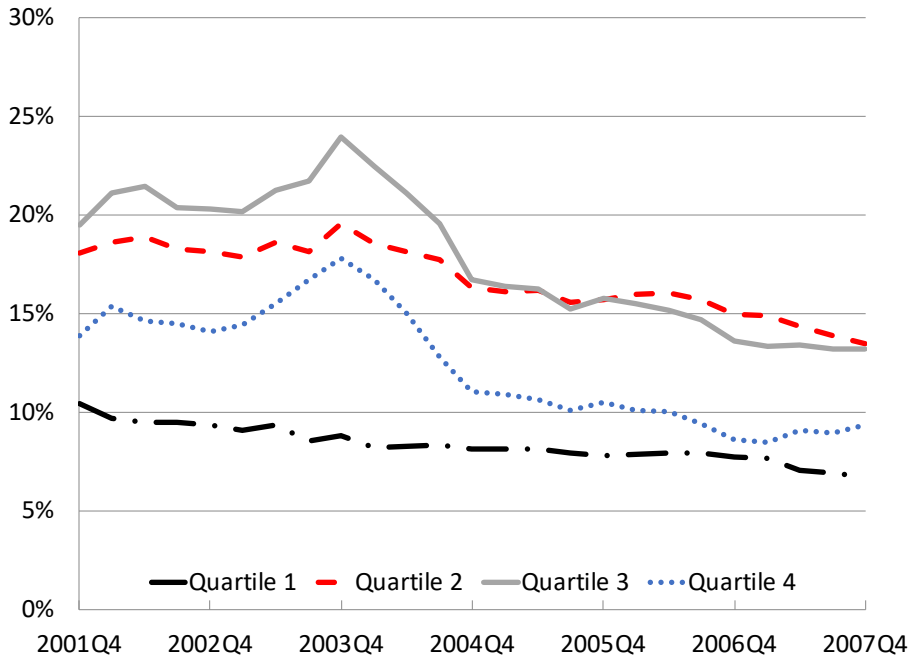


Figure 6: Borrowers with a new mortgage origination. Fraction in each quartile of the 4Q lagged Equifax Risk Score distribution. Source: Authors' calculations based on FRBNY CCP/Equifax Data.

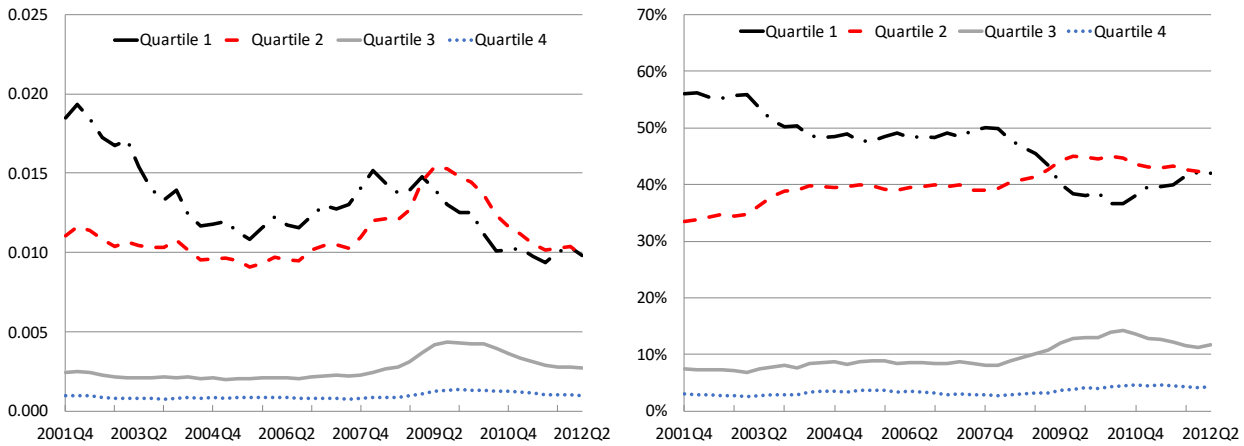
To summarize, we find that from 2001 to 2006, mortgage debt growth was highest for borrowers in the middle and at the top of the credit score distribution. Next, we examine the consequences of this finding for mortgage delinquencies and foreclosures.

2.3 Defaults

We now examine default activity by recent credit score. Specifically, we concentrate on mortgage delinquencies and foreclosures by recent credit score, with the baseline results based on 4-quarter-lagged credit score quartiles.

Figure 7 presents the distribution of new mortgage delinquencies. The fraction of borrowers with a new 90+ days mortgage delinquency in the last 4 quarters, displayed in panel (a), is highest for borrowers in quartile 1 in 2001-2004. During this period, it drops from 1.9% to 1.1%, and by 2004Q1, the fraction with a new mortgage delinquency in quartile 1 is very similar to the fraction for quartile 2. The delinquency rate starts rising for both quartile 1 and 2 in 2005Q2, though the rise for quartile 2 is much bigger than for quartile 1, so that the fraction with new delinquencies peaks at 1.5% in 2008Q1 for quartile 1 and at 1.5% in 2009Q2 for quartile 2. The fraction with

new delinquencies hovers around 0.2% for quartile 3 and 0.08% for quartile 4 during the boom. During the crisis, it rises to a peak of 0.43% in 2009Q3 for quartile 3, with a very modest rise for quartile 4 over the same period. As a result of the large rise in the fraction of new delinquencies for borrowers in quartile 2 and 3, the quartile 1 share of new delinquencies, displayed in panel (b), falls from 50% to 38% during the crisis. The share of new delinquencies rises from 40% to 45% for quartile 2 borrowers, and from 8% to 14% for quartile 3 borrowers during the crisis. In 2009-2010, quartiles 2 to 4 accounted for over 65% of all mortgage delinquencies.



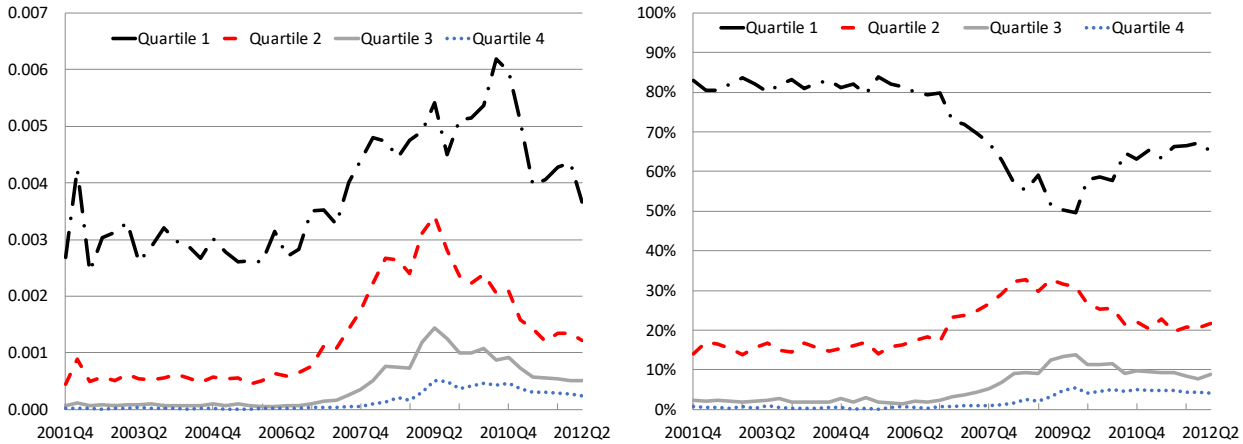
(a) Fraction with new delinquencies

(b) Share of new delinquencies

Figure 7: New 90 days+ delinquencies by credit score quartile, 4Q lagged Equifax Risk Score. Source: Authors' calculations based on FRBNY CCP/Equifax Data.

Figure 8 presents the same statistics for new foreclosures. The fraction of borrowers with new foreclosures in the last 4 quarters among those in quartiles 1 and 2 of the 4-quarter-lagged credit score distribution, displayed in panel (a), average to 0.26% and 0.1%, respectively, for the period ending in 2005Q2. For quartiles 3 and 4, this fraction is very close to zero until 2006Q3. In mid 2006, new foreclosures start rising for all quartiles, and the rise is particularly pronounced for borrowers in quartile 2 and 3 of the 4-quarter lagged credit score distribution. As a result, the share of new foreclosures, displayed in panel (b), for quartile 1 borrowers drops from around 80% during the boom to a low of 49% in 2009Q3. By contrast, the share of new foreclosures for quartile 2 borrowers rises from around 15% during the boom to a peak of 32% in 2009Q1. The share of foreclosures for quartile 3 also rises noticeably from around 2% during the boom, to a peak of 14% in 2009Q32, and the share for quartile 4 also experiences a 5 percentage point rise over the same period.

In summary, these results suggest that the rise in defaults during the 2007-2009 mortgage crisis is concentrated in the middle of the credit score distribution, with a sizable growth in mortgage defaults even at the top of the credit score distribution. The share of new mortgage delinquencies and foreclosures to low credit score borrowers drops considerably during the crisis, challenging the



(a) Fraction with new foreclosures

(b) Share of new foreclosures

Figure 8: New foreclosures by credit score quartile, 4Q lagged Equifax Risk Score. Source: Authors' calculations based on FRBNY CCP/Equifax Data.

notion the increase debt and defaults by low credit score borrowers was the main determinant of the housing crisis.

2.4 Explaining the Discrepancy: The Role of Life Cycle

We now explain the discrepancy between the results on the distribution of debt based on initial and recent credit score rankings, focussing on the link between age, debt and credit scores. This analysis illustrates the problems associated to using initial credit scores and rationalizes our approach based on recent credit scores.

Credit Scores by Age We begin by noting that low credit score individuals are disproportionately young. Table 1 reports the frequency distribution of age by credit score quartile for 20-85 year old borrowers. The median age for borrowers in quartile 1 is 39 years, rising to 58 years for borrowers in quartile 4. While 44% of borrowers in quartile 1 are younger than 35, only 4% of borrowers in quartile 4 are in that age group.

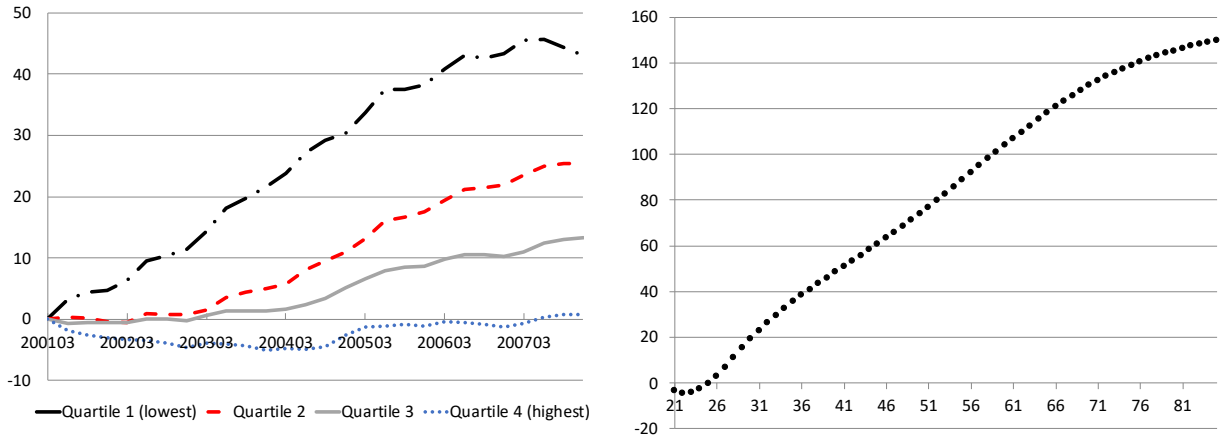
Given their relatively young age, low credit score individuals exhibit credit score growth over time. This is illustrated in panel (a) of Figure 9, which plots the difference between the credit score in each quarter of our sample relative to its value in 2001 for borrowers in different quartiles of the 1999 credit score distribution. For individuals in quartile 1, the credit score grows by 45 points between 2001 and 2007. The credit score grows by approximately 20 points for individuals in quartile 2, by 10 points for individuals in quartile 3 and is essentially constant for those in quartile 4. To more precisely assess the relation between age and the credit score, panel (b) plots the estimated age effects for the Equifax Risk Score, from a regression specification that includes time effects and

Table 1: Median Age by Credit Score Quartile

Age	Quartile 1	Quartile 2	Quartile 3	Quartile 4
Median	39	44	48	58
20-24	0.13	0.11	0.03	0
25-34	0.31	0.22	0.16	0.04
35-44	0.27	0.24	0.23	0.2
45-54	0.17	0.2	0.22	0.19
55-64	0.07	0.11	0.15	0.19
65-85	0.06	0.1	0.2	0.41

Age distribution by credit score quartile, 2004-2013. Fraction in each age bin and median age by credit score quartile. Source: Authors' calculation based on FRBNY CCP/Equifax Data.

state fixed effects. The credit score rises by approximately 20 points from age 25 to age 30 and by an additional 20 points from age 30 to age 35.[9] These results illustrate that individual credit scores vary systematically with age, and are not a constant individual characteristic. In particular, young borrowers have the lowest credit scores and experience the strongest future growth in credit scores.



(a) Evolution of credit scores over time

(b) Age effects in credit scores

Figure 9: Panel (a) Difference between current and 2001 credit score by Equifax Risk Score quartile in 1999, 3Q moving average. Panel (b) Estimated age effects for Equifax Risk Score 1999-2013. Source: Authors' calculation based on Federal Reserve Bank of New York's Consumer Credit Panel/Equifax Data.

[9]While US law prevents age from being used directly in credit scoring models, length of the credit history is one of the most important factors in credit score variation, and the estimated age effects capture this property.

Mortgage Balance Growth, Credit Scores and Income by Age We also exploit the panel dimension of our data to explore the relation between age, credit scores, mortgage balances and income. Specifically, we relate the evolution of credit scores and mortgage balances from 1999 to 2009 to total labor income in 2009 by age in 1999. Using labor income data linked to Equifax, described in Appendix D, we can examine credit score growth between 1999 and 2009 in relation to income levels in 2009 for individual borrowers who are young in 1999, and we also consider the evolution of their mortgage balances. We find that young borrowers in 1999 with high income in 2009 exhibit the largest growth in credit scores between 1999 and 2009.

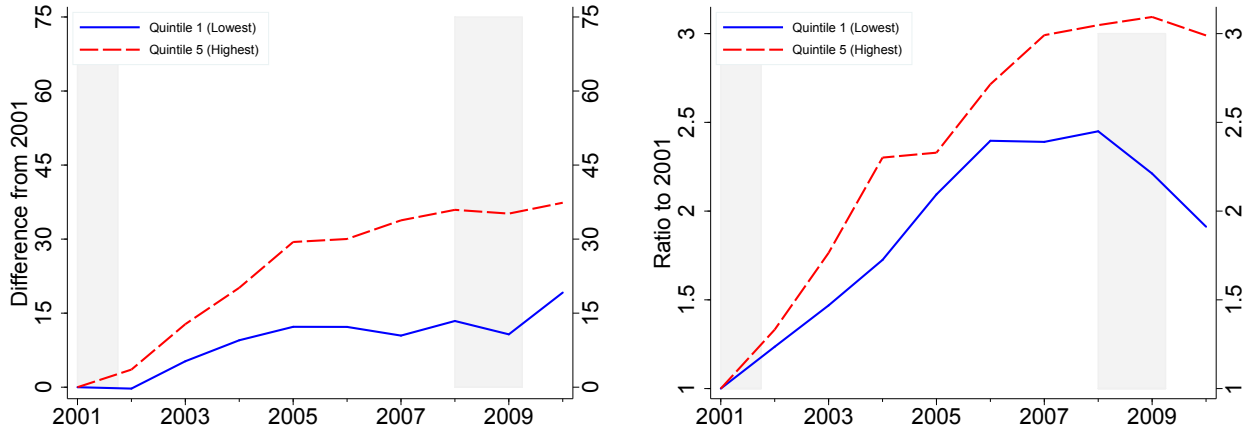
The results are displayed in Figure 10. The charts clearly show that 25-34 year olds in 1999 who are in the top quintile of the 2009 labor income distribution exhibit a much stronger growth in credit scores. For those in the bottom quintile, the credit score rises by only 10 points between 2001 and 2009, while it grows by 40 points for those in the top quintile. Similarly, mortgage balances grow by a factor of 3.3 between 2001 and and 2007 for the top quintile, and by a factor of 2.4 for the bottom quintile.[10] While we do not have panel data on income, it is well documented that lifecycle income growth is steepest for individuals in the 25-34 age group, and dispersion of income increases with age.[11] Given the positive relation between income and credit score documented in Appendix D and the positive relation between credit score growth and income in 2009, these results suggests that the growth in mortgage balances between 2001 and 2006 for 25-34 year olds in 1999 is driven by lifecycle growth in their income over the same period. The same qualitative patterns hold for 35-44 year olds in 1999 and 45-54 year olds in 1999, as reported in figures 16 and 17 in Appendix D.3. However, the magnitude of the increase in both credit scores and mortgage balances is much smaller, as these groups start with higher initial credit scores and, as shown in Figure 5, growth in mortgage balances slows substantially at this point in the life cycle. Moreover, individuals at this age also experience smaller subsequent growth in income.

Appendix D.4 also reports estimates of the relation between the growth in total debt balances and total income using the Panel Study of Income Dynamics (PSID) over the 1999-2007 period. The PSID analysis confirms the positive relation between income growth and growth in debt balances in 2001-2006 at the individual level, and also shows that mortgage debt accumulation slows with age and is largest for borrowers who are young in 1999.

To summarize, the individual level data suggests a strong positive relation between growth in mortgage balances, credit score and income during the 2001-2006 boom. The relation between these variables is a function of life cycle factors that affect not just levels, but future dynamic paths that vary across age groups, and thus cannot be captured by cohort effects, as suggested in Mian and Sufi (2015). Ranking borrowers by initial credit score confounds these life cycle factors for young borrowers at the start of the boom with variation in behavior low credit score borrowers

[10] The growth in both credit scores and mortgage debt balances is monotonically increasing in 2009 income quintile. We report only quintile 1 and 5 for brevity.

[11] See, for example Gourinchas and Parker (2002).



(a) Credit score

(b) Mortgage balances

Figure 10: Equifax Risk Score, difference with 2001 in credit score points, and mortgage balances, ratio to 2001, for 25-34 year olds in 1999 in relation to their 2009 Worknumber total annual labor income quantile. Source: Authors' calculations based on FRBNY CCP/Equifax Data.

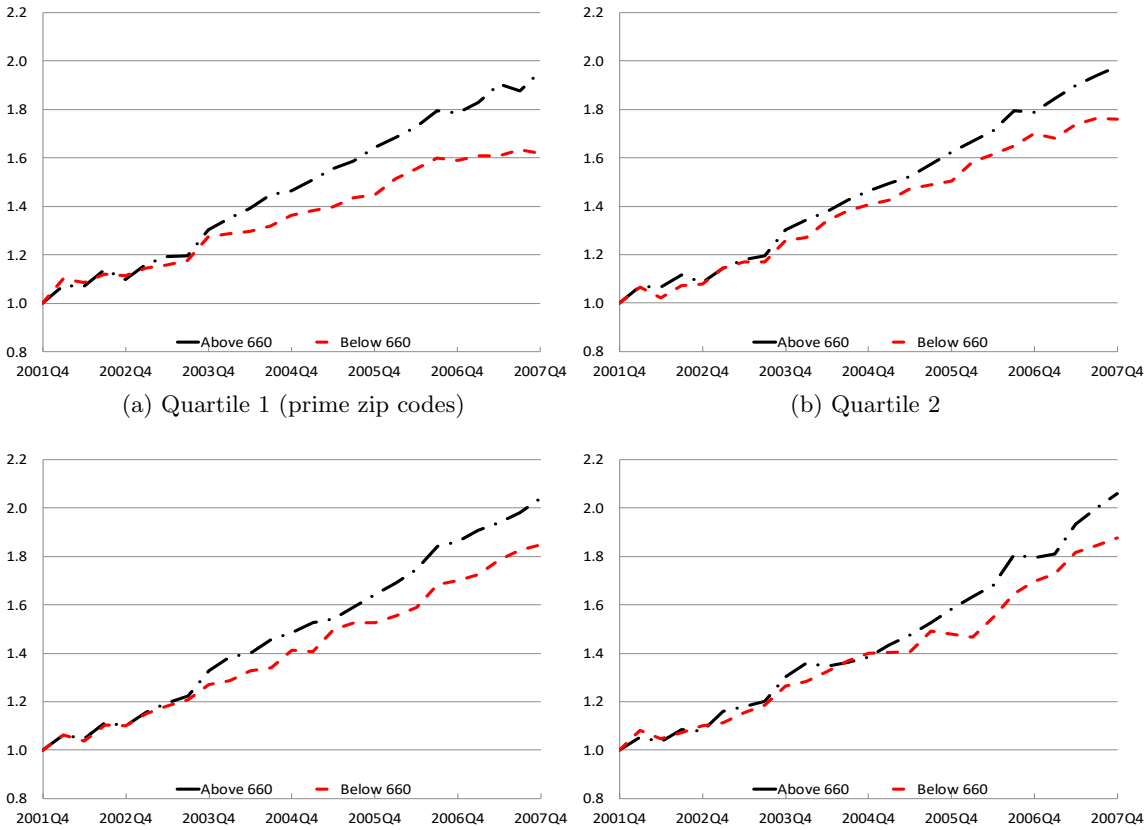
over the boom. Our results on the positive relation between mortgage balance growth in income in individual data also counter the negative relation between income growth and debt growth found in zip code level data by Mian and Sufi (2009). In the next section, we examine zip code level data to explore this discrepancy.

2.5 Geographical Variation: Prime versus Subprime

Starting with the seminal work of Mian and Sufi (2009), the macroeconomic literature has used geographical variation to link mortgage debt growth to the severity of the housing crisis and of the ensuing 2007-2009 recession. As shown in Figure 1, ranking zip codes by the fraction of subprime borrowers in 1999, suggests that mortgage debt growth in 2001-2006 was stronger in zip codes with high initial fraction of subprime borrowers. In this section, we study disaggregated patterns of borrowing across these zip code quartiles, distinguishing between the behavior of prime borrowers, those with credit scores above 660, following the standard definition, and subprime borrowers with credit scores below 660.

Figure 11 presents zip code level mortgage balance growth starting in 2001Q4 for prime and subprime borrowers by quartile of the fraction of subprime borrowers in the zip code. Two observations stand out. First, prime borrowers experience much higher growth in mortgage balances during the boom relative to subprime borrowers *in all zip code quartiles*. Second, in zip codes with the highest fraction of subprime borrowers, mortgage balances grow more than in other zip codes for *both* prime and subprime borrowers. Between 2001Q4 and 2007Q4, prime individuals balances grew by 96% in zip codes in quartile 1 of the share of subprime borrowers, and by 106% in quartile

4. For sub-prime borrowers, the corresponding numbers are 61% and 87%. The second observation is consistent with the prevailing view that areas with a large fraction of subprime borrowers experienced a greater growth in mortgage balances, however, contradicts the notion that it is primarily subprime borrowers driving this behavior.



(a) Quartile 1 (prime zip codes)

(b) Quartile 2

(c) Quartile 3

(d) Quartile 4 (sub-prime zip codes)

Figure 11: Zip code level per capita mortgage debt growth for prime (Equifax Risk Score above 660) and subprime (Equifax Risk Score below 660) borrowers by quartile of share of subprime in 1999. Based on 8Q lagged individual credit scores. Source: Authors' calculations based on FRBNY CCP/Equifax Data.

Table 2 presents the share of mortgage balances held by prime borrowers for zip codes ranked by the fraction of subprime borrowers in 1999. The share of mortgage balances held by prime borrowers is larger than their share in the population, particularly so in zip codes with a high share of subprime borrowers, ranging from 84% in zip codes in quartile 1 by share of subprime borrowers, to 55% in zip codes in quartile 4 of the share of subprime borrowers in 2001Q4. The share of mortgage balances held by prime borrowers increased in zip codes with a relatively high share of subprime borrowers, growing most, from 55% in 2001Q4 to 62% in 2007Q4, in zip codes with the largest share of subprime borrowers. This confirms that, especially in zip codes with a

large fraction of subprime borrowers, it was prime borrowers who experienced a larger increase in mortgage balances.

Table 2: Prime Share of Mortgage Balances

Quartile of subprime in 1999	Quartile 1 (low)	Quartile 2	Quartile 3	Quartile 4 (high)
share of prime borrowers in 1999	80.4%	68.6%	56.7%	38%
prime share of mortgage balances 2001Q4	83.8%	77.5%	67.6%	54.7%
prime share of mortgage balances 2007Q4	82.7%	78.2%	71.1%	61.9%

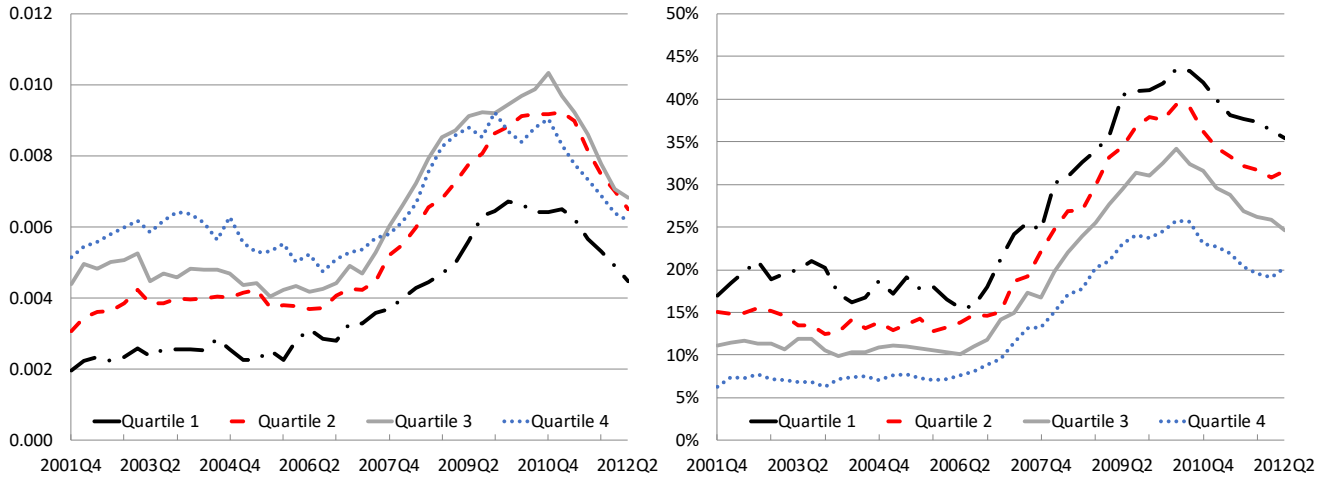
Share of borrowers with Equifax Risk Score above 660 (prime borrowers) and share of mortgage balances help by these borrowers by quartile of the 4Q lagged Equifax Risk Score. Source: Authors' calculations based on FRBNY CCP/Equifax Data.

Turning to defaults, Figure 12 presents the fraction with new foreclosures in the last 4 quarters and the prime share of new foreclosures by zip code ranked by the share of subprime borrowers in 1999. Not surprisingly, zip codes with higher fraction of subprime borrowers exhibit higher foreclosure rate throughout the sample period, though the variation is modest and the biggest difference is between quartile 1, comprising zip codes with the lowest fraction of subprime borrowers, and the higher quartiles. The increase in the foreclosure rate during the crisis is sizable for all quartiles. Foreclosure rates for zip codes in quartiles 2-4 converge during the crisis, whereas the rate for quartile 1 remains lower, despite its increase. Turning to the share of new foreclosures for *prime* borrowers, it is clear that prime borrowers contribute more to the growth in foreclosures during the crisis in all zip codes. The share of prime borrowers' foreclosures rises approximately by 10-25 percentage points between 2006Q2 and 2009Q4,[12] reaching a peak share of almost 25% in the subprime quartile 4 up to 45% in quartile 1. These findings confirm the pattern uncovered with individual data, that foreclosures grow comparatively more for higher credit score borrowers during the crisis.

3 Interpreting the Evidence

The findings presented in the previous section are puzzling given the typically very low default rates for mid-to-high credit score borrowers. It is then natural to ask why individuals with good credit histories experienced such abnormally high default rate during the crisis. In this section, we document the rise in real estate investors and we show the increase in defaults for prime borrowers is primarily driven by these borrowers.

[12] Mortgage delinquencies follow a similar pattern and are discussed in detail in Albanesi, De Giorgi, and Nosal (2017).



(a) Fraction with new foreclosures

(b) Prime share of foreclosures

Figure 12: Panel (a): fraction with new foreclosures. Panel (b): share of new foreclosures for prime borrowers. Zip code level averages by quartile of the share of subprime borrowers in 1999. Source: Authors' calculations based on FRBNY CCP/Equifax Data.

3.1 The Role of Real Estate Investors

There are a number of factors that may render mortgages for investment properties more prone to default than those for owner occupied housing. First, the financial and psychological costs of default for resident owners are typically quite substantial, including moving and storage costs, longer commute times and so on. Real estate investors are not subject to these costs. Second, only the primary residence is protected in personal bankruptcy. While a financially distressed borrower could potentially file for bankruptcy to stay foreclosure procedures and possibly restructure their mortgage on their primary residence, this option is not typically available for investment properties. Finally, mortgages for investment properties must meet stricter credit standards and are usually charged an additional premium to qualify for GSE insurance.[13] This makes it more likely that investors will contract non-standard mortgages with shorter maturity or variable rates, which are intrinsically more risky.[14]

We follow Haughwout et al. (2011) and define *investors* as borrowers who hold 2 or more first mortgages, *non-investors* as borrowers with only 1 first mortgage. Table 3 presents statistics measuring investor activity by individual borrowers ranked by credit score quartile. The share

[13]See `http://www.freddiemac.com/singlefamily/pdf/ex19.pdf` for a classification and quantification of the additional fees charged by the GSEs for mortgages with special attributes, such as investment properties.

[14]Keys et al. (2012) document the sizable increase of Alt-A mortgages, that have low standard for income documentation and would be particularly appropriate for real estate investors who may have variable and hard to document income. Further, Foote and Willen (2016) also discuss the role of alternative mortgage products and the fact that their structure may increase the risk of default. However, Elul and Tilson (2015) present evidence of substantial misrepresentation of home purchases as primary residences, for the purpose of qualifying for GSE sponsored mortgages.

of investors among mortgage borrowers is increasing in credit score, ranging between 6.6% for quartile 1 to 10.9% for quartile 4 in 2001Q4. The investor share of mortgage balances follows the same pattern but is larger, ranging from 12.1% for quartile 1 to 22.7% for quartile 4. Both statistics increased substantially during the boom, with larger increases for higher credit score borrowers. In 2007Q4, the share of investors among mortgage borrowers reached 8.1% for quartile 1 credit scores and 14.2% for quartile 4 credit scores, while the investor share of mortgage balances ranged from 18.3% for quartile 1 to 31.7% for quartile 4.[15]

Table 3: Investor Activity

	Quartile 1	Quartile 2	Quartile 3	Quartile 4
		2001Q4		
Share of investors	6.6%	10.8%	11.5%	10.9%
Investor share of mortgage balances	12.1%	20%	21.5%	22.7%
		2007Q4		
Share of investors	8.1%	15.6%	16.2%	14.2%
Investor share of mortgage balances	18.3%	33.3%	35%	31.7%

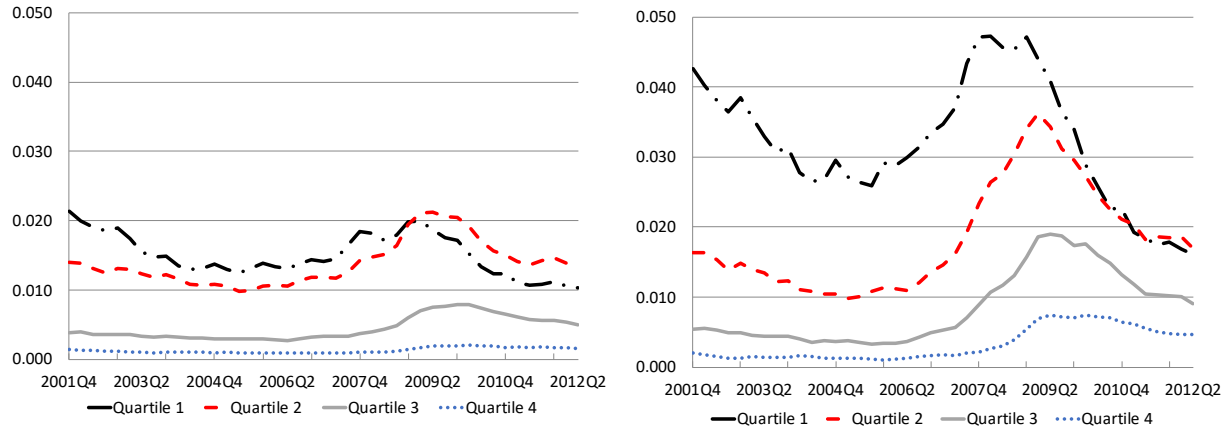
Fraction of borrowers with 2 or more first mortgages and share of first mortgage balances of these borrowers among all borrowers with at least 1 first mortgage, by quartile of the 8Q lagged Equifax Risk Score. Source: Authors' calculations based on FRBNY CCP/Equifax Data.

Figure 13 reports the fraction of borrowers with a 90+ day mortgage delinquency by investor status. Between 2002 and 2006, delinquency rates are similar, and very low, for investors and non investors for borrowers in quartiles 2-4, but more than twice as high for investors relative to non-investors for borrowers in quartile 1. For non-investors, the fraction of borrowers with mortgage delinquencies grows by 50% in quartile 1 between 2005 and 2009, doubles in quartiles 2 and 3 and increases by 40% in quartile 4. Strikingly, the fraction with new delinquencies rises much more for investors than for non-investors over the same period. It roughly doubles for quartile 1, grows 3-fold for quartile 2, and exhibits an almost 5-fold increase for quartiles 3 and 4.[16]

As a consequence of the greater rise of defaults for investors relative to non-investors, the share of defaults accounted for by investors rises during the crisis. Table 4 presents the statistics. The investor share of delinquencies rises from 10%-15% in 2001 to more than 40% in quartile 4, 35% in quartile 3, 23% in quartile 2 and 16% in quartile 1 at the end of 2008. The increase is similarly dramatic for foreclosures, and also concentrated among high credit score individuals, with investors

[15] Ferreira and Gyourko (2015) find that the fraction of investors is very similar for prime and subprime borrowers. However, their definition of investors includes only businesses and borrowers with a tax address different from their mortgage address. Chinco and Mayer (2014) also identify real estate investors using the difference between property address and tax address.

[16] Analogous patterns hold for foreclosure rates, presented in Figure 18 in Appendix E.



(a) Non-investors

(b) Investors

Figure 13: Fraction with new 90+ days mortgage delinquency in the last 4 quarters for borrowers only 1 (Non-investors, panel (a)) and with 2 or more (Investors, panel (b)) first mortgages by quartile of the 8Q lagged Equifax Risk Score. 3QMA. Source: Authors' calculations based on FRBNY CCP/Equifax Data.

accounting fore more than 50% of new foreclosures in quartiles 3 and 4. While real estate investors typically have higher default rates than non-investors, during the 2007-2009 crisis this gap grew substantially, so at high credit scores, investors who accounted at most for 15% of all mortgage borrowers were responsible for over 50% of all foreclosures.[17]

The disproportionate contribution of high credit score investors to defaults during the 2007-2009 mortgage crisis suggests that *borrower behavior*, such as investment activity, may have been more critical than *borrower characteristics*, such as credit scores, for the mortgage crisis.

Investors at the Zip Code Level Table 5 presents the fraction of investors at the zip code level for prime and subprime borrowers. Investor activity is mostly concentrated among prime borrowers, though more prevalent in zip codes with lower share of subprime borrowers. More importantly, the rate of growth in investor activity between 2001Q4 and 2007Q4 is quite similar across zip codes. Among prime borrowers, the share of investors grows by 60% in quartile 1, 50% in quartile 2, then 63% in quartile 3 and 70% in quartile 4. Among subprime borrowers, the corresponding growth rates are 13%, 10%, 19%, 19%. This evidence confirms that the increase in investor activity is concentrated among prime borrowers.

Given the large rise in the share of defaults attributed to prime borrowers and the link between investor activity and defaults seen in the individual data, we provide more detail on investor activity for prime borrowers at the zip code level in Table 6. Mortgage balances grow more for investors

[17] Albanesi (2018) shows that investors have much higher leverage compared to non-investors and the payment to income ratio on real estate loans is twice as high as for non-investors, which may account for their higher default rates.

Table 4: Investor Share of Defaults

	Quartile 1	Quartile 2	Quartile 3	Quartile 4
Investor share		2001Q4		
of 90+ days delinquencies	12.1%	12.2%	15.2%	15.4%
of foreclosures	9%	17%	25.6%	25.4%
Investor share		2008Q4		
of 90+ days delinquencies	16.3%	22.6%	33.1%	40.1%
of foreclosures	14.5%	37%	52%	55%

Investor share of 90+ days delinquencies and foreclosures by quartile of the 8Q lagged Equifax Risk Score. Source: Authors' calculations based on FRBNY CCP/Equifax Data.

Table 5: Investor Activity by Zip Code

Share of subprime borrowers	Quartile 1 (low)	Quartile 2	Quartile 3	Quartile 4 (high)
		2001Q4		
Share of investors, prime borrowers	8.8%	10.2%	8.1 %	5.6%
Share of investors, subprime borrowers	3.8%	3.6%	2.8%	1.8%
		2007Q4		
Share of investors, prime borrowers	14.2%	15.4%	13.3%	9.6%
Share of investors, subprime borrowers	4.2%	4%	3.3%	2.2%

Share of borrowers with 2 or more first mortgages among prime and subprime borrowers at the zip code level, with zip codes ranked by the share of subprime borrowers in 1999. Source: Authors' calculations based on FRBNY CCP/Equifax Data.

compared to non-investors in all zip codes, however, the additional growth in mortgage balances for investors relative to non-investors during the boom is higher in zip codes with larger share of subprime borrowers. This matches the pattern for defaults. The rise in the foreclosure rate during the crisis for zip codes in quartiles 3-4 is approximately double the rise in quartiles 1-2 for both investors and non-investors, though the difference in investor and non-investor foreclosure rates is largest in zip codes with a high share of subprime borrowers.

Table 6: Investor Activity for Prime Borrowers by Zip Code

Share of subprime borrowers	Quartile 1 (low)	Quartile 2	Quartile 3	Quartile 4 (high)
2001Q3-2007Q4 mortgage balance growth				
1 first mortgage	59%	62%	66%	69%
2+ first mortgages	88%	90%	103%	109%
crisis rise in foreclosure rate				
1 first mortgage	0.008	0.012	0.016	0.017
2+ first mortgages	0.029	0.037	0.057	0.070

Selected zip code level indicators of investor activity by quartile of the share of subprime borrowers in 1999. All indicators for prime borrowers, defined as those with Equifax Risk Score above 660. The crisis rise in the foreclosure rate is the difference in the 2002Q1-2005Q4 average foreclosure rate and the crisis peak. The crisis peak varies by quartile, with 2007Q4 the most common date. Source: Authors' calculations based on FRBNY CCP/Equifax Data.

Summarizing, though the fraction of investors with prime credit score is similar across zip codes, prime investors exhibit larger increases in mortgage balances during the boom and a more severe increase in foreclosures during the crisis in zip codes with high share of subprime borrowers.

3.2 Geographical Variation

Our analysis in Section 2.5 suggests that using geographically aggregated data fails to provide accurate evidence on the pattern of borrowing at the individual level. Zip codes with large fraction of subprime borrowers experience larger growth in mortgage balances, however, in all zip codes, it was prime borrowers who mostly contributed to this growth. Marginal borrowers experienced a smaller rise in mortgage balances everywhere, even as areas with a large concentration of marginal borrowers in the aggregate showed higher mortgage growth. We now argue that in addition to providing misleading evidence on the distribution on mortgage borrowing, geographically aggregated data can also impair our understanding of the resulting macroeconomic implications.

Several studies find a positive relation between the size of the increase in mortgage debt growth during the 2001-2006 credit boom– often instrumented with Saiz (2010) house price elasticities–

and the severity of the 2007-2009 recession.[18] These studies attribute this correlation to a fall in consumption by *marginal* borrowers driven by tightening of collateral constraints, resulting from the decline in housing values. Since this causal mechanism is not consistent with our findings, we explore additional economic indicators at the zip code level to shed light on this correlation.

Table 7 reports several economic indicators by quartile of the share of subprime borrowers in 2001. Many population characteristics critical to business cycle sensitivity are related to the share of subprime borrowers. Zip codes with higher fraction of subprime borrowers are younger, have lower levels of educational attainment and have a disproportionately large minority share in the population. It is well known that younger, less educated, minority workers suffer larger employment losses during recessions (see Mincer (1991) and Shimer (1998)). Indeed, zip codes with large subprime population have higher unemployment rates both during the boom and during the crisis. The average unemployment rate for 2001-2007 was 4.94% in quartile 1 and 5.72% in quartile 4. In 2007-2010, the average unemployment rate rose to 6.93% in quartile 1 and 7.81% in quartile 4.

Zip codes with a large fraction of subprime borrowers also exhibit lower per capita income levels in both the boom and the recession. In 2001-2007, the average real per capita income was $41,045 in quartile 1 and only $21,019 in quartile 4, whereas in 2007-2010 it was $46,341 for quartile 1 and $21,898 for quartile 4. Consistent with Mian and Sufi (2009), income growth during the boom was lower in zip codes with higher fraction of subprime. Average per capita income grew by 35% between 2001 and 2007 for quartile 1 and only 4% for quartile 4. However, as we show in Section 2.4, *individual* debt and credit score growth is positively related to *individual* income. This discrepancy in the relation between income and debt growth at the individual level and at the zip code level may be driven by the fact that zip codes with a large subprime population also exhibit higher income inequality. We measure this with the ratio of average income for individuals with incomes above $200,000 over average income for the entire population, based on IRS data. Since top incomes were growing at a much faster rate than bottom incomes during the sample period (see for example Saez (2018)), the positive relation between income growth and debt growth in individual data need not be matched in aggregate data.

One motivation for considering zip code level evidence is the scarcity of information on individual borrowers in credit report data. Geographical aggregation provides access to a number of additional indicators, such as income, housing values and so on. Very often, geographical patterns are interpreted as reflecting individual behavior. Our findings caution against this interpretation

[18] For example, Mian, Sufi, and Trebbi (2015) find that states with higher foreclosure rates experienced a larger decline in consumption, while Mian and Sufi (2014) use county level data and show that a larger decline in household net worth during the crisis experience a more pronounced decline in non-tradable employment. Mian, Rao, and Sufi (2013) exploit geographic variation in house price declines over the period 2006-2009 and household balance sheets in 2006, to estimate the elasticity of consumption expenditures to changes in the housing share of household net worth, and find a positive and sizable elasticity. Kaplan, Mitman, and Violante (2016) refine this analysis and find that, once the direct effect of the fall in local house prices has been controlled for, household balance sheets do not have an effect on durable consumption.

Table 7: Zip Code Level Indicators

	Quartile 1	Quartile 2	Quartile 3	Quartile 4
Share of subprime in 1999	8.8%	29.5%	44%	71.8%
		Demographics		
Median age	49	46	45	42
Associate+ degree (2012)	45%	31%	23%	17%
Percent white	93%	90%	83%	63%
Percent black	1.7%	3.6%	7.6%	24.6%
		Economy		
Average UR 2001-2007	4.94%	5.19%	5.38%	5.72%
Average UR 2007-2010	6.93%	7.30%	7.51%	7.81%
Per-capita Income 2001-2007	\$41,045	\$30,442	\$25,692	\$21,019
Per-capita Income 2007-2010	\$46,341	\$33,224	\$27,491	\$21,898
Per-capita Income Growth 2001-2007	25%	16%	10%	4%
Per-capita Income Growth 2007-2010	10%	10%	11%	10%
$\frac{\text{Mean Income} \geq \$200K}{\text{Mean Income}}$ (2006-11)	6.4	7.9	9.4	11.8
		House Prices		
HP Growth 2001-2007	29%	37%	42%	47%
HP Growth 2007-2010	-21%	-30%	-27%	-36%

Selected zip code level indicators by quartile of the share of subprime borrowers in 1999. Per-capita income expressed in 2012 USD, adjusted by CPI-U. UR (unemployment rate) is the U3 official rate. Source: Authors' calculations based on FRBNY CCP/Equifax Data, IPUMS and IRS data.

and suggest that the positive correlation between credit growth during the boom and the depth of the recession may be due to other characteristics that vary with geography, such as demographics.

4 Conclusion

Our analysis suggests a new narrative on the 2001-2006 housing boom and the 2007-2009 mortgage crisis. We find that most of the increase mortgage defaults during the crisis is driven by mid-to-high credit score borrowers, who also contributed most to the growth in mortgage balances growth during the boom. Additionally, we show that the growth in defaults is mostly accounted for by real estate investors, whose numbers surged starting in early 2004, especially among high credit score borrowers. We also show that there is a positive relation between income growth and mortgage balances at the individual level during the 2001-2006 boom. Our findings shift the focus from borrower characteristics, such as credit scores, to borrower behavior, particularly investment

activity. This can provide important insights into regulatory reforms and policy actions to mitigate the risks of similar episodes in the future.

Our analysis also suggests that using geographically aggregated data does not provide a good approximation of the patterns of borrowing at the individual level. As shown in the individual level analysis, it was prime borrowers who experienced the largest growth in mortgage balances, but aggregation masks this pattern at the zip code level.

References

Adelino, Manuel, Antoinette Schoar, and Felipe Severino. 2016. "Loan originations and defaults in the mortgage crisis: The role of the middle class." *The Review of Financial Studies* 29 (7): 1635–1670.

———. 2017. "Dynamics of Housing Debt in the Recent Boom and Great Recession." Working paper 23502, National Bureau of Economic Research.

Albanesi, Stefania. 2018, September. "Investors and the 2007-2009 Housing Crisis: An Anatomy." Technical Report, University of Pittsburgh.

Albanesi, Stefania, Giacomo De Giorgi, and Jaromir Nosal. 2017, August. "Credit Growth and the Financial Crisis: A New Narrative." Working paper 23740, National Bureau of Economic Research.

Berger, David, Veronica Guerrieri, Guido Lorenzoni, and Joseph Vavra. 2015, October. "House Prices and Consumer Spending." Working paper 21667, National Bureau of Economic Research.

Bhutta, Neil. 2015. "The ins and outs of mortgage debt during the housing boom and bust." *Journal of Monetary Economics* 76:284 – 298.

Bhutta, Neil, and Benjamin J. Keys. 2016. "Interest Rates and Equity Extraction during the Housing Boom." *American Economic Review* 106 (7): 1742–74 (July).

Chen, Hui, Michael Michaux, and Nikolai Roussanov. 2013, September. "Houses as ATMs? Mortgage Refinancing and Macroeconomic Uncertainty." Working paper 19421, National Bureau of Economic Research.

Chinco, Alex, and Christopher Mayer. 2014, January. "Misinformed Speculators and Mispricing in the Housing Market." Working paper 19817, National Bureau of Economic Research.

Corbae, Dean, and Erwan Quintin. 2015. "Leverage and the Foreclosure Crisis." *Journal of Political Economy* 123 (1): 1–65.

Elul, Ronel, and Sebastian G Tilson. 2015. "Owner occupancy fraud and mortgage performance."

Ferreira, Fernando, and Joseph Gyourko. 2015, June. "A New Look at the U.S. Foreclosure Crisis: Panel Data Evidence of Prime and Subprime Borrowers from 1997 to 2012." Working paper 21261, National Bureau of Economic Research.

Foote, Christopher L., Lara Loewenstein, and Paul S. Willen. 2016. "Cross-sectional patterns of mortgage debt during the housing boom: evidence and implications." Technical Report 22985, National Bureau of Economic Research.

Foote, Christopher L., and Paul S. Willen. 2016. "The Subprime Mortgage Crisis." In *Banking Crises*, 324–336. Springer.

Gourinchas, Pierre-Olivier, and Jonathan A Parker. 2002. "Consumption over the life cycle." *Econometrica* 70 (1): 47–89.

Haughwout, Andrew, Donghoon Lee, Joseph S Tracy, and Wilbert Van der Klaauw. 2011. "Real estate investors, the leverage cycle, and the housing market crisis." *Federal Reserve Bank of New York Staff Reports*, no. 514.

Iacoviello, Matteo. 2004. "Consumption, house prices, and collateral constraints: a structural econometric analysis." *Journal of Housing Economics* 13 (4): 304–320.

Justiniano, Alejandro, Giorgio E Primiceri, and Andrea Tambalotti. 2016. "Quantitative Modeling of the Financial Crisis A Simple Model of Subprime Borrowers and Credit Growth." *The American Economic Review* 106 (5): 543–547.

Kaplan, Greg, Kurt Mitman, and Giovanni L. Violante. 2016, May. "Non-durable Consumption and Housing Net Worth in the Great Recession: Evidence from Easily Accessible Data." Working paper 22232, National Bureau of Economic Research.

———. 2017, August. "The Housing Boom and Bust: Model Meets Evidence." Working paper 23694, National Bureau of Economic Research.

Kehoe, Patrick, Elena Pastorino, and Virgiliu Midrigan. 2016. "Debt constraints and employment." Technical Report, National Bureau of Economic Research.

Keys, Benjamin J, Tomasz Piskorski, Amit Seru, and Vikrant Vig. 2012. "Mortgage financing in the housing boom and bust." In *Housing and the Financial Crisis*, 143–204. University of Chicago Press.

Keys, Benjamin J, Tomasz Piskorski, Amit Seru, and Vincent Yao. 2014. "Mortgage rates, household balance sheets, and the real economy." Technical Report, National Bureau of Economic Research.

Kiyotaki, Nobuhiro, and John Moore. 1997. "Credit Cycles." *Journal of Political Economy*, no. 2:211–248.

Lee, Donghoon, and Wilbert van der Klaauw. 2010. "An Introduction to the FRBNY Consumer Credit Panel." FRBNY Staff Report 479.

Mian, Atif, Kamalesh Rao, and Amir Sufi. 2013. "Household Balance Sheets, Consumption, and the Economic Slump." *The Quarterly Journal of Economics* 128 (4): 1687–1726.

Mian, Atif, and Amir Sufi. 2009. "The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis." *The Quarterly Journal of Economics* 124 (4): 1449–1496.

———. 2010. "The Great Recession: Lessons from Microeconomic Data." *The American Economic Review* 100 (2): 51–56.

———. 2011. "House Prices, Home Equity–Based Borrowing, and the US Household Leverage Crisis." *The American Economic Review* 101 (5): 2132–2156.

———. 2014. "What explains the 2007–2009 drop in employment?" *Econometrica* 82 (6): 2197–2223.

———. 2015, May. "Household Debt and Defaults from 2000 to 2010: Facts from Credit Bureau Data." Working paper 21203, National Bureau of Economic Research.

———. 2018, July. "Credit Supply and Housing Speculation." Working paper 24823, National Bureau of Economic Research.

Mian, Atif, Amir Sufi, and Francesco Trebbi. 2015. "Foreclosures, house prices, and the real economy." *The Journal of Finance* 70 (6): 2587–2634.

Midrigan, Virgiliu, and Thomas Philippon. 2011. "Household Leverage and the Recession." *NBER Working Paper 16965*.

Mincer, Jacob. 1991. "Education and unemployment." Technical Report, National Bureau of Economic Research.

Mitman, Kurt. 2016. "Macroeconomic effects of bankruptcy and foreclosure policies." *The American Economic Review* 106 (8): 2219–2255.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek. 2019. "IPUMS USA: Version 9.0 [dataset]." Technical Report, Minneapolis, MN: IPUMS.

Saez, Emmanuel. 2018. "Striking it richer: The evolution of top incomes in the United States." In *Inequality in the 21st Century*, 39–42. Routledge.

Saiz, Albert. 2010. "The geographic determinants of housing supply." *The Quarterly Journal of Economics* 125 (3): 1253–1296.

Shimer, Robert. 1998. "Why is the US unemployment rate so much lower?" *NBER Macroeconomics Annual* 13:11–61.

A Data

The Federal Reserve Bank of New York's Consumer Credit Panel/Equifax Data (CCP), an anonymous longitudinal panel comprising a 5% random sample of all individuals who have a credit report with Equifax. The data is described in detail by the Center for Microeconomic Data at the Federal Reserve Bank of New York. The main reference is Lee and van der Klaauw (2010). A technical note with a description of the dataset is available here: `https://www.newyorkfed.org/medialibrary/interactives/householdcredit/data/pdf/Technical_Notes_HHDC.pdf`. The data dictionary is available at `https://www.newyorkfed.org/medialibrary/interactives/householdcredit/data/pdf/data_dictionary_HHDC.pdf`. We use a 1% sample for the individual analysis, which includes information for approximately 2.5 million individuals in each quarter. We use the full 5% sample for the zip code level analysis. Our quarterly sample starts in 1999Q1 and ends in 2013Q3.

A.1 Individual Income Data for 2009

Our analysis also relies on supplementary payroll data, which, merged with our credit panel data, allows us to map individuals' incomes for 2009 to their credit files. The Equifax Workforce Solutions data, also known as Worknumber data, provided by Equifax is a nationally-representative random sample of individuals containing employment and payroll verification information provided directly from the employers. The information provided for each employee includes the last three years of total income, the date of first hire, tenure, and for the current year status (part time/full time), weekly hours, pay rate and pay frequency.

Income Measure Description There are various income measures provided in the Worknumber dataset. For each year of data available variables are given for the total 12-month base, bonus, overtime, and commission compensation in year t, $t-1$, and $t-2$. This information however is only available for a little over $\frac{1}{3}$ of the sample. The other measure of income, which is widely available across the sample, is rate of pay and pay frequency. We therefore impute total income using a simple *rate* $\times$ *frequency* approach to account for the lack of representation found in the sample regarding the total 12-month income variables. This yields about 11,000 observations for 2009. The sample of records is nationally representative, both in terms of geographical and age distribution.

Comparison with the CPS To gauge the accuracy of the imputed income measure in our data, we performed a simple comparison with the income levels reported in the Consumer Population Survey. We present results based on income quintiles below.

We conduct a similar analysis, comparing the distribution of income and age by state in the Worknumber sample and compare it to the American Community Survey. We also find that the sample is consistent with this survey. These results are available upon request.

A.2 Zip Code Level Data

Demographic and Economic Indicators We obtain zip code level demographic and economic indicators for the 2000 and 2000 Census from Ruggles et al. (2019). Per-capita income corresponds to Total Personal Income (INCTOT). Educational attainment measured with HIGHGRADE.

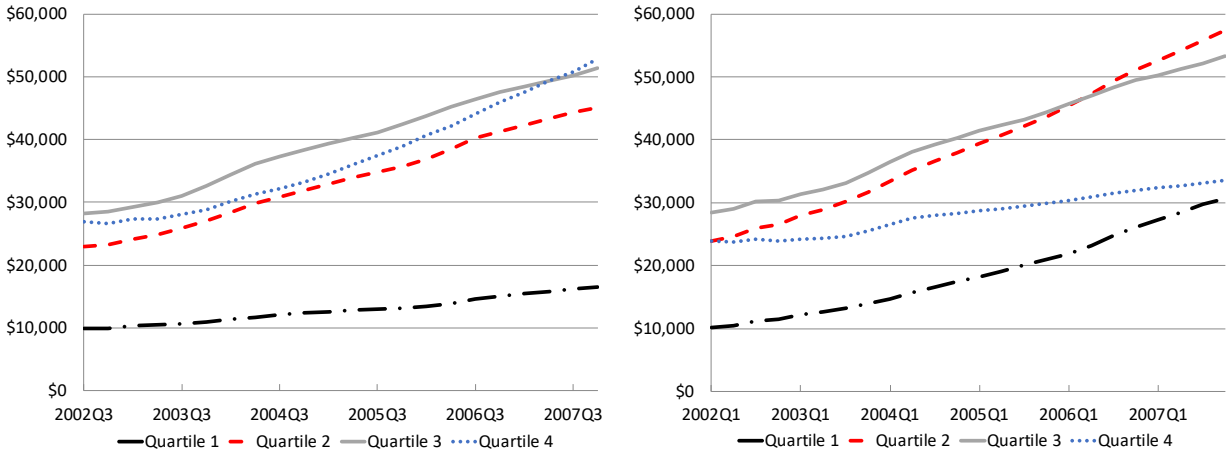
Table 8: Income Distribution Comparison by Quintile

Calculation	Dataset	1	2	3	4	5
Mean	CPS	11058.67	24791.32	36584.61	51872.45	110192.2
	Worknumber	17078.07	26565.46	39589.76	58510.22	117260.1
Median	CPS	12000	25000	36000	50000	85000
	Worknumber	16640	27040	39520	57512	99990

Source: IPUMS, Equifax Worknumber. Worknumber income calculations made using proxied income from pay periods and pay rate. CPS income calculations made using total wage and salary income.

IRS Income Data We use zip-code level data on Adjusted Gross Income published by the IRS to construct our measure of inequality. Starting in 2006, the IRS splits the income returns into 7 brackets: (i) under $10,000, (ii) $10,000 under $25,000, (iii) $25,000 under $50,000, (iv) $50,000 under $75,000, (v) $75,000 under $100,000, (vi) $100,000 under $200,000 and (vii) $200,000 or more. For each of the income brackets, the IRS provides the number of returns, the number of joint returns, and total AGI. To compute our measure of inequality, we first calculate the average per-return AGI for the over $200,000 bracket and the per-return AGI for the pooled brackets (i)-(vi). We make an adjustment for joint returns by counting joint returns as two returns to get the per-filer AGI. Our inequality measure is the ratio of the over-$200,000 per-filer income to the under-$200,000 per-filer income for each year. The data is available for download at the following link: `https://www.irs.gov/statistics/soi-tax-stats-individual-income-tax-statistics-zip-code-data-soi`.

B Mortgage Balance Levels by Credit Score



(a) Quartiles based on 4 quarter lagged credit score

(b) Quartiles based on 1999 credit score

Figure 14: Average mortgage balances by credit score quartile. 3QMA. Source: Authors' calculations based on FRBNY CCP/Equifax Data.

C Mortgage Balance Growth by Recent Credit Score

We report additional results for the estimates for delinquent balances described in Section 2.2.1. Table 9 reports coefficient estimates for the baseline specification.

Table 9: Mortgage Balance Growth

Dependent Variable: 4-Quarter Ahead Mortgage Balance Change				
	1-Quarter Lagged Credit Score Quartile Effects			
	Quartile 1	Quartile 2	Quartile 3	Quartile 4
raw	2,907	2,207	1,213	-745
age adjusted	1,155	4,703	4,243	1,006

Estimated 1-quarter lagged Equifax Risk Score quartile effects for balance change regressions, in USD. Age adjusted quartile effects obtained from estimated age effects displayed in Figure 5 and quartile specific age distribution. All estimates significant at 1% level. Sample period 2001Q1-2011Q4. Number of obs. 64,588,488. Source: Authors' calculations based on FRBNY CCP/Equifax Data.

D Age, Credit Scores, Income and Debt

D.1 Credit Scores

The most widely known credit score is the FICO score, a measure generated by the Fair Isaac Corporation, which has been in existence in its current form since 1989. Each of the three major credit reporting bureaus– Equifax, Experian and TransUnion– also have their own proprietary credit scores. Credit scoring models are not public, though they are restricted by the law, mainly the Fair Credit Reporting Act of 1970 and the Consumer Credit Reporting Reform Act of 1996. The legislation mandates that consumers be made aware of the 4 main factors that may affect their credit score. Based on available descriptive materials from FICO and the credit bureaus, these are payment history and outstanding debt, which account for more than 60% of the variation in credit scores, followed by length of the credit history, which explains 15-20% of the variation, followed by new accounts and types of credit used (10-5%) and new "hard" inquiries, that is those resulting from a borrower initiated credit application. U.S. law prohibits credit scoring models from considering a borrower's race, color, religion, national origin, sex and marital status, age, address, as well as any receipt of public assistance, or the exercise of any consumer right under the Consumer Credit Reporting Reform Act. The credit score cannot be based on information not found in a borrower's credit report, such as salary, occupation, title, employer, date employed or employment history, or interest rates being charged on particular accounts. Finally, any items in the credit report reported as child/family support obligations are not permitted, as well as "soft" inquiries and any information that is not proven to be predictive of future credit performance. Soft inquiries include consumer-initiated inquiries, such as requests to view one's own credit report, promotional inquiries made by lenders in order to make pre-approved credit offers, or administrative inquiries

made by lenders to review open accounts. Requests that are marked as coming from employers are also not counted.

D.2 Cross-Sectional Evidence

To quantify the impact of income and age on credit scores, we regress the credit score on income, income square, age, age square, and interactions between age, income, including state fixed effects. Since the credit score is bounded above, we use a truncated regression approach. Specifically, we estimate the following specification:

$$CS^i_{2009} = \alpha + \beta_1 y^i_{2009} + \beta_2 \left(y^i_{2009}\right)^2 + \gamma_1 \text{age}^i_{2009} + \gamma_2 \left(\text{age}^i_{2009}\right)^2 + \text{interactions} + \varepsilon^i_{2009} \tag{2}$$

where i denoted individual borrowers, CS^i_{2009-h} is a borrower's credit score in quarter $2009-h$, and h denotes the leads/lags in the credit score relative to income, with $h \in \{-8Q, -4Q, 0, 4Q, 8Q\}$. The coefficient α corresponds to the constant and y^i_{2009} is a borrower's total labor income in 2009. Figure 15 displays the in-sample predicted relation between the credit score and income for different age levels. The range of income levels varies by age as they do in our sample. The intercept of the relation at the lowest income level identifies age effects, while the slope of the relation measures the age specific relation between income and credit scores. The fact that this slope declines with age suggests that for older borrowers variation in credit scores are less dependent on income.

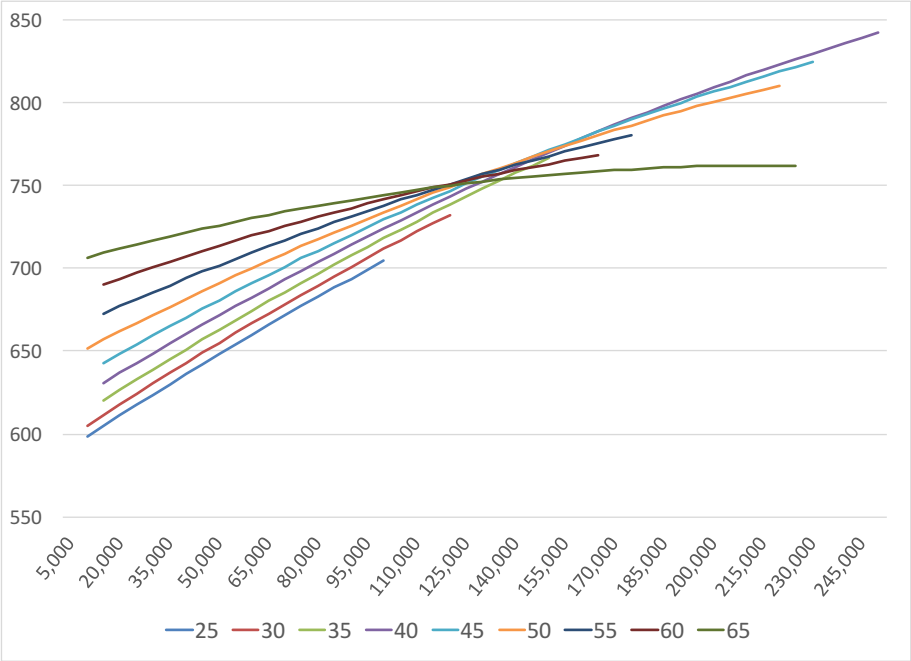
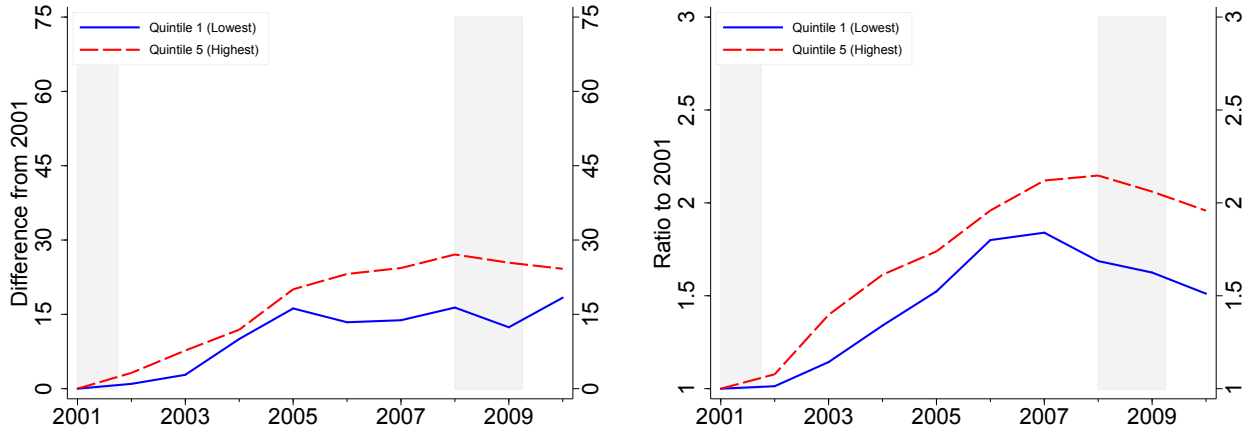


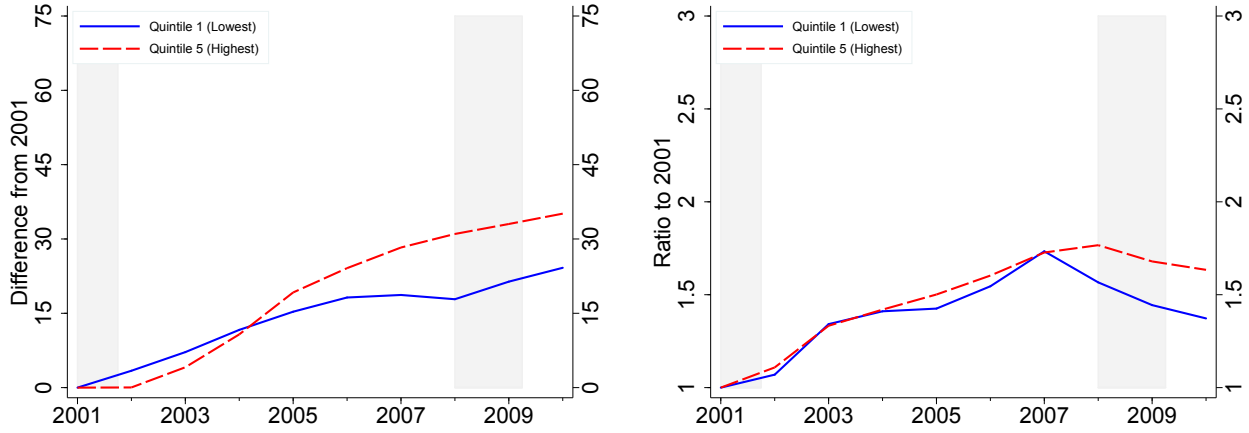
Figure 15: In sample estimated relation between 8Q lagged Equifax Risk Score by age and 2009 total annual labor income, for age specific 1st-99th percentiles of income range. Source: Authors' calculations based on FRBNY CCP/Equifax Data.

D.3 Credit Scores and Mortgage Balances by Age in 1999



(a) Credit score (b) Mortgage balances

Figure 16: Equifax Risk Score, difference with 2001 in credit score points, and mortgage balances, ratio to 2001, for 35-44 year olds in 1999 in relation to their 2009 Worknumber total annual labor income quantile. Source: Authors' calculations based on FRBNY CCP/Equifax Data.



(a) Credit score (b) Mortgage balances

Figure 17: Equifax Risk Score, difference with 2001 in credit score points, and mortgage balances, ratio to 2001, for 45-54 year olds in 1999 in relation to their 2009 Worknumber total annual labor income quantile. Source: Authors' calculations based on FRBNY CCP/Equifax Data.

D.4 PSID Evidence on Income and Debt

To assess the generality of the relation between income, age and debt described in Section 2.4, we use the PSID to estimate the relation between debt growth and income during the boom period.

Using the panel structure of the PSID, we can directly assess the relation between income and debt growth at the individual data. While debt is poorly measured in the PSID relative to the Consumer Credit Panel that we use for our main analysis, we have income at a yearly or bi-yearly frequency.

Table 10: Relation Between Debt Growth and Income Growth

Dependent Variable: 2007-1999 change in log total debt (real USD)				
Δlog(income)	0.066**	0.068**	0.21	0.081
1999 age		-0.064***	-0.01***	-0.070**
1999 age sq		0.001***		
1999 age $\times \Delta log$(income)			-0.003	-0.001
$log(\text{income}_{1999})$			0.001	-0.270
1999 age $\times$ $log(\text{income}_{1999})$				0.006*

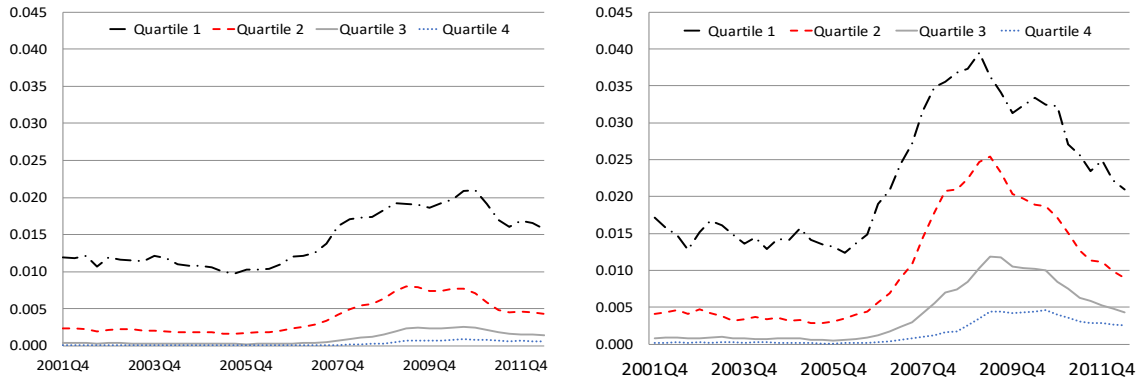
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ No. obs. 1,395. Source: Authors' calculations based on PSID Data.

The estimates for various specifications are displayed in Table 10. The dependent variable is the change in real log total debt between 2007 and 1999, and the baseline specification includes the change in log income over the same period as a dependent variable. The coefficient is positive and highly significant, with a 1 log point change in income corresponding to a 0.066 log point increase in the change in debt over the period. This coefficient implies that a 10,000$ increase in income from a value of 50,000$ in 1999 is associated with a 1$ increase in debt. The second column includes 1999 age and 1999 age squared. The coefficient on the change in income changes little, and the coefficient on age is negative and significant, consistent with our previous finding on the fact that debt accumulation slows with age, and debt accumulation is strongest for borrowers who are young in 1999. The third column includes a an interaction between 1999 age and the change in income, log income in 1999 and no squared age term. In this case the coefficient on the change in log income is positive but much smaller and not significant, while the coefficient on age is still negative and significant, but smaller in magnitude. The coefficient on log income in 1999 is positive but not significant. The last column also adds an interaction between log income in 1999 and age in 1999. In this case the coefficient on the change in income is positive and larger in magnitude relative to previous specifications, but not significant. The other coefficients are similar, with a larger magnitude of the negative coefficient on age. The interaction between age and log income in 1999 is positive and significant, suggesting that higher initial income is associated with larger growth in debt conditional on age. These results confirm our findings based on the Equifax data, suggesting that income growth and debt growth are positively related over the 2001-2006 period.

E Investors: Additional Figures

Figure 18 presents the fraction of borrowers with new foreclosures in the last 4 quarters. Similar to delinquencies, during the 2002-2006 housing boom the foreclosure rate is very similar for investors and non-investors for all quartiles. However, during the crisis, the foreclosure rate diverges, with investors experiencing much higher foreclosure rates than non-investors, especially for higher credit score quartiles. For investors, foreclosure increases by a factor of 4 for the lowest quartile, and by

more than a factor of 10 for quartiles 2-4. For non-investors, the foreclosure rate roughly doubles in quartile 1-2, and rises very modestly for quartiles 3-4.



(a) Non-investors (b) Investors

Figure 18: Foreclosure rates for borrowers with only 1 (Non-investors, panel (a)) and 2 or more (Investors, panel (b)) first mortgages by quartile of the 8Q lagged Equifax Risk Score. Source: Authors' calculations based on FRBNY CCP/Equifax Data.

F Stability and Consistency of Zip Code Rankings

Mian and Sufi (2009) ranks zip codes by the fraction of subprime in 1996. Mian and Sufi (2011) ranks zip codes by initial personal disposable income or initial leverage, which they define as total debt balances per capita over average personal disposable income. Mian and Sufi (2014) rank counties by the decline in household net worth during the crisis, which is instrumented by the Saiz (2010) house prime elasticities to capture the rise in house prices during the boom and the associated rise in leverage. Here, we examine the relation between these measures at the zip code level.

We first consider the stability of each ranking. Table 11 reports the fraction of zip codes that remain in the same quartile of each ranking in the subsequent year. We consider three indicators: the fraction of subprime borrowers, average personal disposable income (PDI) and average leverage, defined as total balances per capita over average personal disposable income. All rankings are very stable, with approximately 70% of all zip codes remaining in the same quartile of the fraction of subprime borrower distribution year to year, over 90% for personal disposable income and 59-75% for leverage. We also examine the correlation between various rankings. The Spearman correlation between fraction of subprime and PDI ranges from -0.46 and -0.58, and decreases over the sample period. The Spearman correlation between fraction of subprime and leverage is negative, ranging between -0.03 at the end of the sample and -0.15 at the height of the credit boom. This is consistent with a greater growth in leverage for zip codes with low fraction of subprime during the boom.

Table 11: Stability and Correlation of Zip Code Rankings

	Fraction in same quartile			Correlation with % subprime	
	% sub-prime	PDI	Leverage	PDI	Leverage
2001	0.68	0.88	0.59	-0.46 ***	-0.04 ***
2002	0.71	0.91	0.62	-0.50 ***	-0.05 ***
2003	0.73	0.92	0.66	-0.51 ***	-0.06 ***
2004	0.70	0.90	0.63	-0.53 ***	-0.10 ***
2005	0.71	0.90	0.67	-0.53 ***	-0.15 ***
2006	0.72	0.89	0.67	-0.55 ***	-0.15 ***
2007	0.72	0.87	0.69	-0.58 ***	-0.09 ***
2008	0.72	0.92	0.73	-0.58 ***	-0.11 ***
2009	0.72	0.95	0.74	-0.58 ***	-0.04 ***
2010	0.73	0.95	0.75	-0.58 ***	-0.03 ***
2011	0.72			-0.57 ***	-0.03 ***

Fraction of zip codes in same quartile in subsequent year, by fraction of subprime borrowers, PDI and leverage. Correlation (Spearman ρ) of fraction of subprime borrowers in 2001 and PDI or leverage in each sample year. Leverage is the ratio of total debt balances to PDI. *** denotes significance at the 1% level. Source: Authors' calculations based on FRBNY CCP/Equifax Data, IPUMS, IRS, BLS, ACS data.