# Cross-Sectional Patterns of Mortgage Debt During the Housing Boom: Stocks and Flows*

Christopher L. Foote[†] Lara Loewenstein[‡] Paul S. Willen[§]

April 1, 2016

**Preliminary: Please do not quote or cite without permission.**

**Abstract**

High rates of mortgage-debt growth among low-income households play a central role in many explanations of the early 2000s housing boom. We show that growth rates of debt for higher-income households were equally large. The similarity of growth rates meant that the distribution of debt with respect to income changed little during the boom. Moreover, because high-income borrowers always account for a disproportionately large share of outstanding mortgage debt, uniform rates of debt growth imply that high-income borrowers took out far more debt in dollar terms: the richest quintile of U.S. ZIP codes accounts for about $1.5 trillion of new mortgage debt from 2001 to 2006, as compared to about $320 billion for the lowest quintile. The equality of debt growth rates across income groups is consistent with subsequent foreclosures, as defaults across income categories rose in rough proportion as well. Previous research purporting to show that the distribution of debt shifted toward low-income borrowers was based on the *flow* of new mortgage originations alone, so this research could not detect offsetting shifts in mortgage terminations that left the distribution of debt constant over time.

*The views in this paper are not necessarily those of the Federal Reserve Bank of Boston or the Federal Reserve System. We have received helpful comments from Manuel Adelino, Neil Bhutta, Kris Gerardi, Kristoph Kleiner, Antoinette Schoar, Rosen Valchev, and seminar participants at the Atlanta Fed, Brandeis, and the Homer Hoyt Institute. We also thank Brigitte Madrian and Stephen Zeldes, who invited one of us to discuss Adelino, Schoar, and Severino (2015b) at the NBER's Summer Institute. Work on that discussion encouraged us to write this paper.

[†]Federal Reserve Bank of Boston. Email: `Chris.Foote@bos.frb.org`.
[‡]Federal Reserve Bank of Boston. Email: `Lara.Loewenstein@bos.frb.org`.
[§]Federal Reserve Bank of Boston. Email: `Paul.Willen@bos.frb.org`.

# 1 Introduction

The early 2000s saw a large expansion of mortgage debt in the United States. As measured by the Federal Reserve System's Flow of Funds accounts, the aggregate stock of home mortgages on the liability side of household balance sheets doubled from $5.3 trillion in 2001 to $10.6 trillion in 2007. As Figure 1 shows, this growth in debt was much greater than the growth of income over the same period, as the aggregate debt-to-income ratio rose to unprecedented levels. This paper analyzes the cross-sectional allocation of debt during the boom, with particular attention to how the new debt was allocated across the income distribution. As the title of the paper suggests, we analyze both the *stocks* of outstanding mortgage debt on household balance sheets and *gross flows* of debt, which are mortgage originations and terminations. We find that outstanding debt stocks rose at equal rates across the income distribution, a finding that contradicts common explanations of the boom that rely on disproportionate borrowing by low-income individuals or communities.[1] To be sure, low-income borrowing expanded during the boom, with much of this debt packaged into the subprime mortgage-backed securities that caused so many problems during the financial crisis. Yet borrowing by high-income individuals rose at similar rates. Moreover, because mortgage debt rises with income in the cross-section, high-income borrowers were responsible for a large majority of the additional mortgage debt in dollar terms.

The widespread nature of the mortgage boom has yet to be appreciated for at least two reasons. First, most empirical work on the boom has searched for interesting borrowing patterns at the low end of the income distribution, ignoring the massive amount of borrowing at the top. Additionally, most previous research has focused not on the stock of debt but on one gross flow, new originations. The main reason for this limited focus is probably data availability. The datasets traditionally used for housing analysis cover originations alone, with the best example being the data collected under the Home Mortgage Disclosure Act (HMDA). For many questions, such as the possibility of racial discrimination in lending, a sole focus on originations is appropriate.[2] Yet when the analysis concerns household balance sheets, studying originations alone is problematic. HMDA data provide no information on

[1] A summary of existing academic work on the mortgage boom is found in the introductory paragraph of Amromin and McGranahan (2015): "A voluminous literature that [has] analyzed [pre-Great-Recession] developments noted that the pre-recessionary period was characterized by the liberalization of credit access to households that had previous found it difficult to qualify because of poor credit records, insufficient income, or both. The liberalization of credit access has largely been ascribed to financial innovation through securitization markets that allowed loan originators to offload credit risk to a broad set of private investors" (p. 147, [insertions added]).

[2] This is not to argue that the borrower-level information available in public-use HMDA datasets is comprehensive. For the Boston Fed's study of racial discrimination (Munnell et al. 1996), the authors supplemented HMDA data with additional information about mortgage applicants that lenders used to evaluate potential borrowers.

mortgage terminations, so inferring changes in the stock of debt from them is analogous to inferring changes in employment from data on new hires alone, with no adjustments for layoffs and quits. In fact, using HMDA data to study balance sheets is even more dangerous than using only hiring data in an employment study, because an offsetting relationship between mortgage originations and terminations is often hard-coded into housing transactions. Every time a mortgage is refinanced, one mortgage is originated at the same time another is terminated. In purchase transactions, simultaneous origination and termination occurs whenever the buyer borrows to buy the home and the seller pays off an existing mortgage (or mortgages) with the proceeds of the sale. In both cases, the ultimate change in the stock of debt is not the value of the newly originated mortgage but the difference between the value of the new mortgage(s) and the value of the terminated one(s), which can be positive or negative.

The main dataset used below comes from individual-level outstanding mortgage balances collected by the Equifax credit bureau and assembled into the Federal Reserve Bank of New York's Consumer Credit Panel, a database developed some years after the housing boom. Because these data carry no information on income, we aggregate the Equifax records to the ZIP-code level, and then combine them with ZIP-level income data from Internal Revenue Service (IRS). An additional household-level analysis uses data on both mortgage debt and income from the Federal Reserve's Survey of Consumer Finances (SCF). While the SCF provides comprehensive information on individual-level income and balance sheets, it is much smaller than the Equifax/IRS dataset and is available only at three-year intervals. The higher frequency and the geographic detail of the ZIP-level Equifax/IRS data allow for an integrated analysis of stocks and flows both within and across housing markets.

There are four major results. The first is that, as noted earlier, the aggregate increase in the stock of debt was broad-based, with borrowers across the income distribution raising their debt levels by similar percentage amounts. We find that a ZIP code at the mean of the 2001 income distribution would be expected to experience an increase in mortgage debt of about 53 log points from 2001 to 2006.[3] For a ZIP code at the 90th percentile of 2001 income, the expected change in debt is only three log points higher, while debt growth for ZIP code at the 10th percentile is only three log points lower. Similar stability in growth rates with respect to income is also found in household-level data from the SCF. The near-equality of debt-growth rates across the distribution resulted in no reallocation of mortgage debt toward low-income ZIP-codes of households during the boom, as seen in the top two panels

[3]This calculation is based on long-difference regressions reported in Table 3. Both debt and income in these regressions are normalized by the number of tax returns in the ZIP code. Due to a peculiarity in IRS income-data collection in 2007, discussed further below, these regressions are based on debt changes from 2001 to 2006. Because Figure 1 suggests 2007 as the end of the boom, and because SCF data are available in 2007, in the Internet Appendix we verify that all of the ZIP-level results remain robust to using 2007 as the last year of the mortgage boom instead.

of Figure 2. Another implication of equal debt-growth rates is that because rich people tend to account for more debt, in dollar terms the growth of debt among the rich dwarfed debt growth for lower-income borrowers. The bottom left panel of Figure 2 shows that from 2001 to 2006, ZIP codes in the highest income quintile accounted for about \$1.5 trillion of the total \$4.0 trillion increase in total mortgage debt, while debt for the lowest income quintile rose by only \$320 billion. The lower right panel shows similar dollar-value patterns in the household-level data of the SCF.

The second major finding of the paper concerns the two gross flows of debt, originations and terminations. Previous research has used HMDA data to document a shift in within-county patterns of mortgage originations that appear to send more credit to ZIP codes with relatively low-incomes (Mian and Sufi 2009). We show that these changes in origination flows are fully consistent with equal growth in debt stocks due to offsetting shifts in mortgage terminations. This finding highlights the importance of using data on debt stocks rather than gross flows when analyzing household balance sheets. It also sheds light on the debate between Mian and Sufi (2009) and Adelino, Schoar, and Severino (2015b) on whether the HMDA origination data really do show increases in relative debt burdens during the housing boom.[4]

The claim that debt grew equally across the income distribution may seem odd, given the salience of subprime lending in common narratives about housing boom. We use a separate source of comprehensive data on securitized subprime loans to confirm that subprime loans were in fact more common among low-income borrowers. But the third finding of the paper is that in relation to the stock of all outstanding mortgage debt, alternative mortgage products like subprime or Alt-A loans were dwarfed by prime loans, which were favored by richer borrowers. Even among the alternative products, Alt-A loans—which were generally low-documentation mortgages made to borrowers with high credit scores—experienced higher growth than the subprime loans favored by lower-income borrowers. In light of the first two findings of the paper, the quantitative importance of prime and Alt-A lending suggests that subprime loans did not *cause* a reallocation of debt toward low–income borrowers. Rather, subprime loans *prevented* a reallocation of mortgage debt toward the wealthy.

The final finding concerns the consequences of the housing bust for borrowers in different income classes. Using the comprehensive Equifax data, which also has information on de-

[4]As explained below, Adelino, Schoar, and Severino (2015b) split the total dollar value of purchase-mortgage originations in each ZIP code into the average value of each mortgage and the number of new purchase mortgages originated. Their finding that the origination patterns highlighted in Mian and Sufi (2009) are generated by relative changes in numbers of mortgages—not by changes in average amounts—leads them to argue that the data do not support the significant change in household-level balance sheets that is claimed in Mian and Sufi (2009). However, Adelino et al. are unable to determine how many new mortgages reflect new ownership experiences, so they cannot evaluate the possibility that credit expanded along the extensive margin (Mian and Sufi 2015b).

faults, we confirm that in *absolute* terms, increases in foreclosures were larger in low-income areas. But foreclosures were not concentrated in low-income areas, because in *relative* terms increases in foreclosures in richer communities were just as high.[5] There is of course an analogy between the growth of mortgage debt during the boom and the growth of defaults during the bust. High-income communities always account for a disproportionately large share of mortgage debt, so scaling up mortgage debt by equal rates during the boom generates larger dollar-value changes in debt in high-income areas. Similarly, low-income communities always account for an outsize share of foreclosures, so scaling up defaults equally during the bust brings about large absolute increases in foreclosures in low-income communities.

Other cross-sectional facts can undoubtedly be generated from disaggregated data on mortgage debt stocks. Some of these facts may highlight unique housing-market experiences for specific groups of people. But all of these facts must be consistent with the broad patterns outlined below—growth rates of debt were similar throughout the income distribution, and absolute increases in debt were largest among high-income borrowers. It is hard to reconcile these facts with the common claim that expanded low-income borrowing set off a housing bubble, for the simple reason that the dollar amounts of low-income borrowing represented such a small fraction of overall debt accumulation. The data are more supportive of theories that allow for the broad patterns found in this paper. One such speculative story is that some fundamental determinants of housing prices caused them to move higher early in the boom, where these determinants are one or more of a long list of possibilities: accommodative monetary policy used to fight the 2001 recession; higher savings rates among developing countries, that pushed U.S. interest rates lower; a perception that housing was a safer investment vehicle after the 2001 stock market crash; etc.

What is unknown is how any modest increases in house prices brought about by developments like these morphed into a full-blown housing bubble, in which prices continued to rise under their own momentum to levels that far exceeded their fundamental values. Perhaps people simply noticed the original price increases and, swept along by behavioral biases such as group-think and optimism, expected them to continue indefinitely. These exaggerated price expectations encouraged buyers to offer high prices for houses, rendering those expectations self-fulfilling: the hallmark of an asset bubble. Unfortunately, the study of bubbles is too young to provide much guidance on this point. For now, we have no choice but to plead ignorance, and we believe that all honest economists should do the same. But acknowledging what we do not know should not blind us to what we do know: the bursting of a massive and unsustainable price bubble in the U.S. housing market triggered the ensuing financial crisis, which led to the most severe recession the U.S. has seen since the Great Depression.

[5]If anything, credit bureau data indicate larger percentage increases in foreclosures in high-income communities, relative to corresponding increases in low-income communities.

# 2 Data

## 2.1 Debt and Income Data from Equifax and the IRS

The main measure of mortgage debt used below comes from the Federal Reserve Bank of New York's Consumer Credit Panel, a quarterly, longitudinal 5-percent sample of individual credit histories supplied by the Equifax credit bureau. The dataset begins in 1999, and because individual-level credit histories are included in the sample based on the last two digits of the individual's social security number, the dataset can be updated to incorporate new entrants over time.[6] Among other debt variables, the Equifax data contain detailed information on mortgage debt, including the amounts and dates associated with the origination of new loans, and outstanding balances for first mortgages, subordinate mortgages, and home equity lines of credit (HELOCs). We can also measure mortgage terminations. We specify that a termination has occurred in the first quarter that a mortgage balance goes to zero, and the value of that termination is defined as the balance of the mortgage in the previous quarter.

The ability to paint a comprehensive picture of both stocks and flows of mortgage debt is a unique characteristic of credit-bureau data. The HMDA data used in previous research follow a law passed in 1975 that requires certain financial institutions to report individual-level data relating to mortgage applications and originations, including the dollar amount of each new mortgage and the census tract of the house backing the loan. As far as originations go, HMDA is an appropriate and near-comprehensive data source, but as noted earlier HMDA data cannot be used to measure mortgage terminations or debt stocks.[7] Data from public deeds registries suffer from a similar limitation, in that they provide good coverage of originations but problematic coverage of terminations.[8]

Loan-level datasets generated by mortgage securitizers or mortgage servicers provide information on both originations and terminations, yet neither type of dataset is comprehensive. The CoreLogic Private Label Securities ABS Database provides loan-level data only for mortgages that have been packaged into non-agency securities (that is, securities not

[6] As discussed below, we will aggregate the Equifax records by ZIP code in order to match them with available income data from the IRS. When we do so, we multiply the aggrgated debt data by 20, because the data come from a 5-percent sample of individuals.

[7] HMDA's coverage of originations is very good but still incomplete. Only mortgage companies and depository institutions with offices in metropolitan areas are required to report, and the reporting of home equity lines of credit (HELOCs) is optional. There is also limited information about the individuals applying for mortgages (only race, income, and gender), and some researchers have questioned the accuracy of the borrower-level income data reported on HMDA forms (Mian and Sufi 2015b).

[8] The dataset in Ferreira and Gyourko (2015) is based on public-records data supplied by the DataQuick company. The lack of precise information on mortgage terminations in that dataset makes it hard for the authors to know whether a new, non-purchase mortgage represents the refinance of an existing loan or a new mortgage that adds to the homeowner's existing stock of debt. The authors assume that a new non-purchase mortgage is a refinance if it its value is more than half of either the imputed current price of the home or of the total mortgage balance taken out when the home was purchased.

backed by the government-sponsored entities Fannie Mae, Freddie Mac, and Ginnie Mae). The CoreLogic dataset includes an expansive set of variables for each loan, but these data cannot measure the aggregate stock of debt, because even at the peak of the boom, subprime and other types of non-agency loans made up a small share of the overall market.[9] Yet CoreLogic data can be used to measure cross-sectional patterns in the use of securitized subprime and Alt-A debt, and we will do so below.[10] The loan-level dataset from McDash Analytics has broader coverage than CoreLogic, because it is based on data supplied by mortgage servicers (typically banks) and therefore includes agency and portfolio loans as well as non-agency loans. Unfortunately, the collection of servicers in McDash is generally not considered representative of the entire mortgage market until at least 2005.

A disadvantage of the Equifax dataset is that it contains no information on income. As a result, we follow previous research and construct aggregates of debt at the ZIP-code level, to be merged with ZIP-level income data published by the IRS. ZIP-level data is available on a host of income variables, including adjusted gross income (AGI) and salary and wage income, for the years 1998, 2001, 2002, and 2004–2012.[11] In addition to the income variables, we also use the numbers of exemptions and returns in the IRS dataset to measure ZIP-level population and households, respectively.

The IRS data are comprehensive, because they are based on the universe of tax returns filed in a given year. Even so, the data are not perfect. For one thing, the IRS uses suppression rules to ensure that no individual information can be backed out of the published ZIP-level data, and these suppression rules change from year to year. Additionally, measurement error in the IRS income data can arise from changes in the share of earners who file income taxes. In 2007, the number of filers rose as people were encouraged to file in order to receive a stimulus payment. Figure 3 compares the aggregate number of returns from the ZIP-code data (red dots) to the aggregate number of returns published by IRS (blue line); the latter series omits any return filed for the sole purpose of receiving a stimulus payment. In most years, the total number of returns in ZIP-level data is smaller than IRS's published total, in part because of the suppression rules. But in 2007, the ZIP-level data imply many more returns, because these data include returns filed to for the sole purpose of receiving stimulus checks. In the Internet Appendix, we show that the additional filers

[9]The CoreLogic database was originally called the LoanPerformance database after the company that developed it. The CoreLogic data include the loan-to-value ratio, the debt-to-income ratio, the credit score at origination, and the level of documentation used to originate the loan. The CoreLogic company also supplies a separate dataset of repeat-sales house-price indexes, which is explained more fully below.

[10]Alt-A loans were loans to prime borrowers that did not qualify for standard prime pools, typically because of reduced documentation. The name is derived from the fact that lenders referred to prime borrowers as "A" borrowers, as opposed to the "B" and "C" borrowers who were considered subprime.

[11]The IRS income data come from the Statistics of Income Program. See `http://www.irs.gov/uac/SOI-Tax-Stats-Individual-Income-Tax-Statistics-ZIP-Code-Data-%28SOI%29` for details.

have little effect on income aggregates, which implies that these filers reported low (or zero) income. However, by distorting our measure of the number of households in each ZIP code, the 2007 spike in returns might also distort our results if we define the boom as ending in 2007. Consequently, when using the ZIP-level data below we choose 2006 as the ending year instead. Fortunately, robustness checks in the Internet Appendix show that the distortion induced by the extra 2007 filers is not severe, as our main ZIP-level results go through even when ending the boom in 2007.

Another measurement issue related to the IRS data is what type of income to use. In the empirical work below, income is defined as salary and wage income, which is likely to be the most important type of income considered by lenders when underwriting mortgage loans. A type of income that lenders are *not* likely to consider is capital gains, which is included in AGI. Figure 4 shows that capital gains drives a non-trivial wedge between AGI and wage income in the mid-2000s, as strong growth in capital gains caused AGI to grow faster than wages and salaries in the early 2000s. The figure also shows that the ZIP-level relationship between AGI and wage income weakened during this period, as evidenced by a decline in coefficients generated by annual cross-sectional regressions of ZIP-level wages on AGI. Here again, however, the measurement issues are not too much of a concern. The Internet Appendix shows that our main results are robust to defining income as either as salary and wages or as AGI.

Table 1 presents summary statistics for the ZIP-level Equifax/IRS dataset. The values are medians within each IRS return-weighted income quintile in two main years of interest: 2001 and 2006. Because the quintiles are return weighted, the number of ZIP codes in the lowest quintile is higher than the highest quintile, because each of those ZIP codes has fewer returns. As expected, the value of mortgages (with the exception of second mortgages), household mortgage debt, and the median house value increase as income increases. It is also worth noting that house prices grew slightly more in lower-income zip codes, and that the proportion of mortgaged households grew modestly for all income groups. It is also apparent from this table that the Equifax risk score is correlated with income, as it increases monotonically across the quintiles and that this is not driven purely by age, as the median age does not vary much across the quintiles.

## 2.2 Household-Level Data from the Survey of Consumer Finances

Although the Equifax/IRS dataset allows a detailed look at the cross-sectional evolution of mortgage debt, its ZIP-level nature means that results could be influenced by the migration of households across ZIP-code boundaries. We therefore supplement the ZIP-level data with individual-level data from the SCF, a triennial survey of households conducted by the Federal Reserve. Sample sizes in the SCF range from just over 3,000 households in 1989 to more than 6,000 by 2013, so the SCF is much smaller than the ZIP-level dataset. Because we will focus on households headed by persons aged 64 or younger, our effective sample sizes are even smaller. Yet what the SCF lacks in size it makes up for in quality, because it provides a complete characterization of household-level balance sheets, including data on mortgage debt.[12] The SCF also measures income, with separate information on total income and wage and salary income. The SCF includes only the most basic geographic identifier (Census region), so detailed geographic breakdowns are not possible. But the SCF does include a host of demographic variables, including the age of the household head, so we can perform some analyses conditioning on age. We use the Combined Summary Extract data of the SCF that pulls together key variables from the 1989 through 2013, which are downloadable from the University of California at Berkeley.[13]

Summary statistics for SCF data in 2001 and 2007 appear in Table 2. The table makes clear that the mortgage debt variable in the SCF is a comprehensive measure, including debt on properties other than the primary residence as well as HELOCs. The top panel uses data from all households and defines income as AGI. As we saw in the opening barcharts of Figure 2, similar growth rates of mortgage debt across the income distribution generate much larger dollar increases in debt for quintiles with the highest incomes. The last two columns of the panel illustrate the rapid rise in housing values during the early 2000s. The lower panel of Table 2 defines income as salary and wages, excluding households with zero values of that variable. Patterns for both debt and housing values are similar to results that use AGI as the classification variable instead.[14]

## 2.3 Aggregate Debt Comparisons

Figure 5 compares estimates of the aggregate stock of mortgage debt from the Flow of Funds, the Equifax dataset, and the SCF in years when the SCF is available. The Equifax totals are close to, but somewhat smaller than, the SCF and Flow-of-Funds totals. Yet our Equifax debt totals are essentially identical to some unreported Equifax totals calculated by Brown

[12]The SCF includes separate information on debt secured by the household's primary residence as well as any other real estate debt. We always combine these two measures. Like the total debt measure in the Equifax data, the SCF debt discussed below encompasses first mortgages, subordinate mortgages, and HELOCs.

[13]Variables included in the Combined Summary Extract data are those used in the regular analyses of the SCF published in the *Federal Reserve Bulletin*. See Bricker et al. (2014) for the most recent *Bulletin* article, which discusses the 2013 wave of the SCF, and `http://sda.berkeley.edu/Abstracts/SCF.html` for details regarding the Combined Summary Extract of the SCF maintained at Berkeley.

[14]The second column of figures in the table shows the number of SCF observations for each quintile. When these observations are weighted, they generate equal numbers of households in each quintile. The number of unweighted observations is much largest for the richest quintile to allow the SCF to accurately characterize the long right tail of the wealth distribution.

et al. (2015), who compare Equifax data to the SCF along a number of dimensions.[15] Two measures of SCF aggregates are presented in the figure. The first comes from Henriques and Hsu (2014), who compare various SCF aggregates to their Flow of Funds counterparts. Even though Flow of Funds data are typically constructed from administrative records supplied by financial institutions and government agencies, rather than from surveys, Henriques and Hsu (2014) show that most balance-sheet measures in the SCF are close to corresponding Flow of Funds estimates. This comparability is particularly true for mortgage debt, a pattern that the authors attribute to the clarity of the mortgage-debt concept and the stability of mortgage-data collection procedures in both the SCF and the Flow of Funds over time. Figure 5 replicates the close correspondence between mortgage debt in the Flow of Funds and Henriques and Hsu's SCF measure. Gratifyingly, the figure also shows that our SCF aggregates, based on the public-use Combined Extract Data, are essentially identical to Henriques and Hsu's, with the small differences between them probably resulting from the fact that the Combined Extract dataset is a multiply imputed version of the SCF.[16] Note that comparability of the SCF data to the mortgage measure in the Flow of Funds requires the use of all mortgage data available, including HELOCs. This is why we include HELOCs and other types of secondary mortgages when using either the Equifax dataset or the SCF.[17]

# 3 Cross-Sectional Patterns in Stocks of Debt

In this section, we first look at unconditional distributions of debt at the ZIP-code level, asking whether the most important movements in these distributions occurred within or between housing markets. We then bring income into the analysis, by asking how movements in the debt stocks correlated with income levels and income growth. We conclude the section by confirming the central ZIP-level results on debt stocks and income using household-level data from the SCF.

[15] Specifically, in billions of 2010 dollars, Brown et al. (2015) estimate that total mortgage debt in Equifax in 2004, 2007 and 2010 to be $7,631, $10,034 and $9,282, respectively. Our Equifax totals expressed in the same units and years are $7,741, $9,728, and $9,074. In addition, unreported work shows that our totals are close to those reported Bhutta (2015), which also analyzes mortgage debt in the New York Fed Consumer Credit Panel.

[16] The SCF's multiple imputation procedure creates five copies, or "implicates," of the data, with missing data imputed differently across each implicate. Users of the SCF are instructed to perform statistical tests on each implicate separately, using sample weights, and then combine the resulting parameter estimates and variance-covariance matricies using the Repeated-Imputation Inference (RII) of Rubin (1987). For a good summary of of RII, see Montalto and Sung (1996).

[17] Analysis of the components of mortgage balances in Equifax, such as first mortgages and HELOCs, is available in the Internet Appendix. Recent work by Amromin and McGranahan (2015) and Amromin, McGranahan, and Schanzenbach (2015) also uses the Equifax dataset but splits mortgage debt into non–HELOC mortgage debt and HELOCs. Though these papers do not emphasize the point, they also find broadly similar growth rates of mortgage balances across the income distribution, even when HELOC balances are excluded.

## 3.1 Unconditional Distributions of Debt

The most basic way of using cross-sectional data to study the housing boom is to plot unconditional micro-level distributions of mortgage debt. The top panel of Figure 6 depicts the returns-weighted kernel distributions of (the log of) mortgage debt-per-return for ZIP codes in 2001 and 2006. The large increase in aggregate mortgage debt during this period is reflected by the rightward shift in this distribution over time. The ZIP-level distribution also widened, however, indicating that mortgage debt did not rise in all ZIP codes equally. The two panels in the lower row depict the variation in mortgage debt within and across housing markets. Here, housing markets as defined as cities, or more formally as Core Based Statistical Areas (CBSAs).[18] By construction, the within-CBSA distributions in the lower left panel are both centered at zero, because they are distributions of ZIP-level debt *relative* to CBSA means. The stable shape of the within-CBSA distributions indicates that much of the increased dispersion in total debt arises from an increase in the between-city dispersion depicted in the lower right panel.

The three distributions in Figure 6 argue against the common claim that the housing boom reallocated debt to areas or individuals that had low levels of debt before. That type of reallocation toward low-debt borrowers would have narrowed the debt distributions, but no such narrowing is evident across the nation as a whole (top panel) or within individual housing markets (lower left panel). A related claim is that the boom reallocated debt toward low-*income* communities. Yet if these low-income communities were also low-debt communities, then the same critique applies: there should have been a narrowing of the debt distribution that did not in fact occur. However, we must be careful about using the *unconditional* distributions in Figure 6 to make statements about the allocation of debt *conditional* on income. The unconditional distributions will obviously be affected by changes in the debt-income relationship, but these distributions are formally determined by how the debt-income relationship interacts with the distribution of income across communities.[19] The same point applies to the introductory barcharts in Figure 2. The stability of those debt distributions does not rule out a shift in the relationship between income and debt, because those distributions are also affected by shifts in the distribution of income across households

[18] The government defines CBSAs as groups of counties or county equivalents that are integrated around an urban core with at least 10,000 residents. Those based on urban cores with between 10,000 and 50,000 people are called micropolitan statistical areas, and CBSAs based on larger urban cores are called metropolitan statistic areas. The CBSA classification system replaced the government's previous urban classification system, based on metropolitan statistical areas only, in 2003.

[19] To see this, note that $f_1(d) = \int_0^\infty f(d|y)g(y)dy$, where $f_1$ is the marginal (or unconditional) distribution of debt $d$, $f(d|y)$ is the distribution of debt conditional on income $y$, and $g(y)$ is the distribution of income. This equation makes clear that changes in the distribution of income $g(y)$ also matter for the marginal distributions $f_1(d)$. The potential impact of $g(y)$ means that the effects of changes in the conditional debt-income relationship $f(d|y)$ may be not be directly evident in the unconditional distributions.

or communities. As a result, in order to learn about the debt-income relationship we have to estimate it directly. We take up that task next.

## 3.2 Debt Conditional on Income: Levels Regressions

We first specify a *conditional expectation function* for debt and income. A parametric form for this function is

$$E(d_{cit}|y_{cit}) = \alpha_t + \beta_t \cdot y_{cit}, \tag{1}$$

which assigns debt $d$ to unit $i$ in housing market $c$ in year $t$ as a function of income $y$.[20] The parameters of this function, $\alpha$ and $\beta$, have time subscripts to allow the function to change. Though simple, the conditional expectation function helps formalize a number of theories about the mortgage boom. One theory, noted above, is that credit flowed disproportionately to borrowers with low incomes. As suggested by the introductory bar charts in Figure 2, the cross-sectional relationship between debt and income is strongly positive: richer people and communities have higher debt levels. A reallocation of debt toward low-income borrowers would this reduce this positive correlation over time: $\beta_{2006} < \beta_{2001}$. Alternatively, if debt and income are specified in natural logs (as they will be below), a uniform percentage increase in debt at each level of income would be expressed as rising values of the intercept $\alpha_t$ across time periods, with no changes in the relationship between debt and income that is summarized by $\beta_t$.

To evaluate these alternatives against the data, we follow papers such as Chetty, Friedman, and Rockoff (2014) and first provide a nonparametric estimate of the conditional expectation function. The top left panel of Figure 7 depicts a binned scatter plot constructed by dividing all ZIP codes into 20 returns-weighted bins on the basis of income-per-return.[21] This division is performed separately for 2001 and 2006. We then take averages of the log of debt-per-return and the log of income-per-return within each bin. Plotting the debt and income averages against each other suggests that the debt-income relationship is close to linear in logs in both 2001 and 2006. Additionally, debt-per-return shifted up significantly and nearly uniformly among all income groups, resulting in similar slopes for the two lines of points. Because this slope reflects the importance of income in the allocation of debt, the nonparametric estimate suggests that the impact of income on debt allocation changed little during the boom.

The slopes of the scatter plots are summarized parametrically by the $\beta_t$s in equation

[20] Here, unit $i$ could refer either to a ZIP code (in Equifax/IRS data) or to a household (in the SCF). For households, the relationship between mortgage debt and income is also dependent on age, in part because older households have had time to pay off their debts. Therefore, we when we analyze debt at the household level we will always condition on age, as well as some other demographic factors discussed below.

[21] Recall that income is specified as salary and wage income, not AGI.

1, and the top right panel of Figure 7 presents estimates of these coefficients for all but one year between 2001 and 2006.[22] The estimates, which can be interpreted as elasticities, lie in a fairly tight range between about 1.45 and 1.55.[23] The income effect at the end of the sample period is about .05 higher than the income effect at the start, a difference that is significant statistically, but economically small. In any case, the point estimates in this panel provide no evidence that income became less important during the boom.[24] Because the binned scatter plot suggests that the elasticity of mortgage debt with respect to income is slightly higher for low-income households (that is, the slope of the scatter plot is steeper for at low incomes), we ran a number of unreported regressions that also include the square of income-per-return, to allow for a nonlinear relationship between debt and income. Our results are robust to this specification, as the debt-income elasticities remain stable at both high and low incomes. In other words, even though the implied relationship between debt and income may not be perfectly linear in logs, the relationship shifted up uniformly across the income distribution, as the binned scatter plot suggests.

We next bring city-level factors into the analysis, because a variant of the conventional thinking about the boom is that it saw debt flow to ZIP codes with low incomes *relative* to others in the same housing market. The intercept $\alpha_t$ in the parametric model now replaced with year-specific city fixed effects,

$$E(d_{cit}|y_{cit}) = \alpha_{ct} + \beta_t \cdot y_{cit}, \tag{2}$$

so that a finding of $\beta_{2006} < \beta_{2001}$ reflects a reallocation of debt to areas with low incomes *relative* to others in the same city. A model with city-level effects could also characterize an alternative story in which the relationship between relative income and relative debt did not change ($\beta_{2006} = \beta_{2001}$) and any changes in the distribution of debt are due to between-city movements, reflected in increased dispersion in the $\alpha_{ct}$s.

The lower two panels of Figure 7 investigate these alternatives by focusing on the relationship between debt and income relative to local-market means. The binned scatter plot in the lower left panel is constructed by first deviating both the debt and income variables from CBSA means. We then separate the ZIP codes into 20 bins based on their incomes relative to these means and construct the required averages. Because debt and income are both measured as deviations, the scatters go through the origin. It is remarkably difficult to spot any

[22]Recall that IRS data for 2003 is not available. The estimates in the top right panel are not generated from separate regressions, but rather by a pooled regression in which the constant and the income terms in equation 1 are interacted with yearly dummies. The two methods are equivalent statistically, though the pooled regression turns out to be easier to run. Like the scatter plots, the regressions are weighted by the number of returns in the ZIP code.

[23]These regressions are generated by the actual ZIP-level data, not by the averages in the binned scatter plots.

[24]The standard error on the difference is .014 and the t-statistic is 3.4.

significant shift in the slope of the relationship between relative debt and relative income. The lower right panel investigates relative variation in debt and income parametrically. The regression has the same form as the regression that generates the panel immediately above it, but also allows for CBSA $\times$ year effects.[25] The addition of CBSA fixed effects reveals a somewhat different pattern in the coefficients. There is an increase in the importance of income from 2001 to 2002, and then a decline in the coefficients thereafter. By 2006, the income coefficient has returned to essentially its 2001 value. The difference between the 2006 and 2001 coefficients is –.011, a gap that is neither economically nor statistically significant.

## 3.3 Debt Conditional on Income: Long-Difference Regressions

Both the binned scatter plots and the debt-income regressions reveal a remarkable stability in the relationship between debt and income during the housing boom, with or without including city-level effects. These results are estimated by correlating levels of income with levels of debt in different years, so it is worth asking what happens when we estimate the debt-income relationship in differences instead. It is not hard to devise a long-difference specification that allows for potential changes in the levels relationship over time. Consider any time-1 and time-2 relationship between the generic variables $y$ and $x$, summarized by $\beta_1$ and $\beta_2$:

$$y_1 = \beta_1 x_1 \quad \text{and} \quad y_2 = \beta_2 x_2.$$

The difference between $y_2$ and $y_1$ can be written as $\Delta y = \beta_2 x_2 - \beta_1 x_1$. By adding and subtracting either $\beta_2 x_1$ or $\beta_1 x_2$ from this expression, some algebra shows that $\Delta y$ can be written in two ways:

$$\Delta y = \beta_2 \Delta x + (\beta_2 - \beta_1) x_1 \tag{3}$$

$$\Delta y = \beta_1 \Delta x + (\beta_2 - \beta_1) x_2. \tag{4}$$

Both of these equations suggest a regression of $\Delta y$ on the change in $x$ and a level of $x$. If the first-period level $x_1$ is used, as in equation 3, the coefficient on $\Delta x$ is the levels relationship in the second period $\beta_2$. The opposite situation occurs in equation 4, where the use of the second-period level $x_2$ causes the coefficient on $\Delta x$ to equal $\beta_1$ instead. Regardless of whether $x_1$ or $x_2$ is included, the levels coefficient measures the change in $\beta$ over time $(\beta_2 - \beta_1)$.

Including a level term in the long-difference regressions makes intuitive sense, because this term will tell us whether (say) poorer ZIP codes experienced higher growth in mortgage debt than richer ones, after conditioning on ZIP-level income growth. If the low-income

[25] As noted in footnote 22 the regressions are run as a single pooled regression, so the introduction of CBSA $\times$ year effects merely requires interacting CBSA dummies with yearly dummies.

ZIP codes did experience relatively high debt growth, then the estimate on the levels term will be negative (low income correlates with high debt growth), a finding that implies the importance of income is declining over time ($\beta_2 < \beta_1$). On the other hand, if no difference in relative debt growth is found, then the level of $x$ drops out of the regression, $\beta$ does not change, and the traditional differenced panel specification emerges: $\Delta y = \beta \Delta x$.

Results of the long-difference regressions appear in Table 3. Due to increased concern about potential measurement error in the differenced data, we estimate the regressions on three samples with increasingly strict criteria for inclusion. Columns 1 and 2 present results using a sample with no trims of outlying right- or left-hand-side variables. Column 3 uses a sample in which the observations with the highest and lowest 1% of values for debt growth, income growth, and/or initial income levels are deleted before estimation, and Column 4 uses a sample in which observations with the outlying 5% of values for these three variables are trimmed.[26] In each regression, both the income-growth and income-level regressors are deviated from sample means. This normalization has no effect on the income coefficients, but it allows the constant term to equal the expected growth in mortgage debt for a ZIP code that has average income growth from 2001 and 2006 as well as the average income level in 2001.

Panel A presents the results using the entire sample of ZIP codes. The first column includes only income growth on the right-hand-side, and the resulting coefficient (0.930) is somewhat lower than the $\beta$'s generated from the levels regressions in Figure 7, which ranged from about 1.45 to 1.55. Some difference is to be expected, however, because the long-difference regressions are identified solely by within-ZIP-code variation. Column 2 adds the 2001 income level. The coefficient on income growth remains essentially the same (0.951) and the income-level coefficient is positive, suggesting a *rising* importance of income levels over time ($\beta_{2006} > \beta_{2001}$). The same pattern emerges from the regressions using the trimmed samples in columns 3 and 4: the income growth coefficient is close to 1, while the income-level coefficient is positive.

Positive values of the income-level coefficients in Panel A provide no support for the claim that income became a less important determinant of debt during the boom. However, these coefficients do imply that richer ZIP codes saw modestly higher debt growth rates during the boom. To see how much higher, recall that the normalization of the income-growth and income-levels terms cause the constant terms to reflect growth in debt for the average ZIP-code in terms of 2001-06 income growth and the 2001 income level. All of the constants Panel A in Table 3 are in the neighborhood of .53, indicating that the average ZIP code experiences mortgage-debt growth of about 53 log points. In the bottom three rows of Panel A, we use the income-level coefficients and the empirical distribution of 2001 income levels

[26] See Appendix Table A.8 for the distributions of income growth, initial income levels, and debt growth.

to calculate expected debt growth for ZIP codes at the 90th and 10th percentiles of the 2001 income distribution, as well as the difference between them. In all cases, the high-income ZIP code experiences growth of 56 to 57 log points, while the low-income ZIP code experiences growth rates of about 50 log points. In other words, the difference in expected growth is only about 6 to 7 log points, slightly more than 10 percent of the average growth in debt experienced by ZIP codes in this period.

We can also use this regression framework to ask about the allocation of debt within cities. Panel B estimates the same regressions as Panel A, but the sample is now limited to ZIP codes that lie within CBSA boundaries. Using the new sample has only minor effects on the estimates. Panel C uses the same CBSA sample but also includes a full slate of CBSA fixed effects. Clearly, these effects matter, as adding them to the long-difference specifications cause the R-squareds to rise from the .07-.11 range in Panel B to the .37-.53 range in Panel C. Yet the inclusion of these fixed effects causes only small changes in the the first three income-growth coefficients, although the last income-growth coefficient, corresponding to the 5-percent trimmed sample, is now somewhat higher than the corresponding estimate in the previous panel (.921 vs. .825). More closely related to the question of how credit allocation changed over the boom is the effect of the CBSA fixed effects on the income-level coefficients. These effects cause the level coefficients to decline from small but significantly positive coefficients to essentially zero.[27]

Putting the levels and long-different regressions together, a consistent story emerges. The long-difference regressions generate implied income effects that are somewhat smaller than those emerging from the levels regressions. However, both sets of regressions indicate that debt-allocation patterns changed little during the boom. To the extent that there were any differences in debt-growth rates across ZIP codes, both sets of regressions imply that debt grew a little more for those ZIP codes located in cities with higher incomes. This is seen by noting that the levels regressions without the CBSA fixed effects implied a small increase in $\beta$ from 2001 to 2006 of 0.05; in the long-difference regressions, the corresponding change in $\beta$, measured as the coefficient on the 2001 income level—ranges from near 0.06 to near 0.08. Yet when city-fixed effects are included in either type of regression, even these small differences fall to zero, indicating a remarkable stability in the way that income mattered for the allocation of debt within cities. It is also worth noting that the unconditional debt distributions also make sense in light of the regression results above. Figure 6 showed that the city-level distribution of debt levels widened during the boom, which is consistent with the larger debt growth in cities that had high debt levels to begin with (and thus presumably had higher initial incomes as well). Figure 6 also showed that there was no change in the

[27] The interpretation of the constant terms in Panel C as expected debt-growth rates for a ZIP code with both average 2001-06 income growth and average 2001 income remains valid in the presence of the CBSA fixed effects, because the fixed effects are constrained to have mean of zero.

unconditional distributions of debt within cities, consistent with the stable within-city effect of income on debt that both sets of regressions imply.[28]

### 3.4 Correlates of City-Level Growth in Mortgage Debt

Most of this paper investigates debt at a high level of disaggregation, using either ZIP codes or households. But given their quantitative importance, the city-level determinants of debt deserve some analysis in their own right. The map in Figure 8 is based on the CBSA-level effects from 1%-trimmed regression of Table 3, Panel C.[29] It shows the large growth in debt in the so-called sand states of California and Florida, and to a lesser extent Arizona and Nevada. Debt also grows a great deal throughout the northeastern United States and in some cities of the Midwest, such as Chicago and Minneapolis. On a broader level, the map also shows that the CBSA definition is quite expansive. There are close to 1,000 CBSAs available, with both large "metropolitan" and smaller "micropolitan" areas included.

Table 4 regresses the adjusted CBSA fixed effects on a number of local variables. As in the long-different regressions, both the income-level and income-growth regressors are deviated from sample averages, so that the constant terms in the table are the expected city-wide debt-growth rates for localities with both average income in 2001 and average income growth from 2001 to 2006. The first column of Table 4 includes only a constant term, so it recovers the estimated constant of .53 from the original long-difference regression. As we would expect from the earlier regression results, the initial income of the city is a positive determinant of subsequent debt growth, as column 2 generates a initial-income coefficient of 0.25. Although this coefficient is statistically significant, initial income explains less than 15 percent of the total variation in citywide debt growth. Column 3 adds average income growth in each city from 2001 to 2006, which does not enter significantly. Perhaps the lack of significance for city-level income growth should not be that surprising, because we already controlled for ZIP-level income growth in the long-difference regression that generated the left-hand-side variable.

[28] In the Internet Appendix, we show how the regression results are affected by defining the boom as ending in 2007 rather than 2001. The small positive impacts of ZIP-level income apparent in the baseline levels and long-difference regressions falls to zero, while a small negative effect of ZIP-level income emerges in the within-city analysis. The negative within-city effects in the 2001-07 regressions are a little smaller in absolute value than the positive effects in the baseline regression without city fixed effects. In other words, the effect of the 2001-07 sample period on income effects is similar in sign to their effect in the baseline results. Whereas in the baseline sample the fixed effects turn small positive effects into zeros, in the 2001-07 sample the fixed effects turn zero effects into very small negative ones.

[29] The fixed effects in that panel were constrained to have a mean of zero, and as in the other panels we defined both the initial-income and income-growth regressors as deviations from sample averages. Consequently, adding the estimated constant term from that regression to the estimated fixed effects before mapping them allows us to investigate city-level mortgage debt growth conditional on ZIP-level initial income and income growth.

Column 4 enters city-level growth in house prices from 2001 to 2006, as measured by repeat-sales indexes from CoreLogic.[30] Unlike the income variables, the price-change variable is not deviated from sample means. The estimated coefficient is positive and statistically significant. Just as important, the R-squared with price-growth as the only regressor is close to twice the size of the corresponding statistic when the two income variables are entered together (.27 vs. .15). The remaining columns add the income variables back to the specification. In these regressions, the coefficient on price growth remains close to its previous value and the initial-income variable remains positive. The income-growth variable, however, enters negatively when all three regressors are included.

Clearly, these regressions should be interpreted as conditional correlations rather than structural estimates. Determining the direction of causality among changes in mortgage debt, house prices, and income at the city-level is difficult. An exogenous increase in mortgage debt may send house prices higher if it allows borrowers to bid up the price of homes. Alternatively, an exogenous increase in house prices may induce mortgage lenders to expand their mortgage portfolios if they expect the rising values of homes, not the characteristics of borrowers, to ensure repayment of the loans. Yet the results above may be useful for future research, because they help characterize the most important cross-sectional shifts in mortgage debt during the boom. These city-level shifts were modestly correlated with initial income of cities and much more strongly with subsequent growth in prices. A theory designed to explain the housing cycle should be consistent with these facts, just as it should explain the lack of debt shifts within housing markets.

## 3.5 Debt and Income in the Survey of Consumer Finances

In this subsection, we confirm the central results of the ZIP-level analysis with household-level data from the SCF. Figure 9 is the household-level corollary to the levels analysis for ZIP codes presented in Figure 7.[31] Because of smaller sample sizes, the upward-sloping relationships in the binned scatter plot of the top panel are not as smooth as those in the ZIP-level data. Yet the same type of upward shift in mortgage debt throughout the income distribution is evident. The lower panel of Figure 9 displays the estimated income effects from a household-level poisson regression of mortgage debt on wage income and other demographic variables.[32] The estimated income effects generally decline from the 1989 through 2001 and

[30] When there are insufficient repeat-sales counts to generate a CBSA index for a particular month, CoreLogic updates the CBSA index using the relevant state index. As a result, price indexes for small CBSAs are often less reflective of truly local conditions than indexes for large CBSAs.

[31] There are no within-CBSA plots because the SCF is not available at that level of detail.

[32] A poisson regression of $y_i$ on $x_i$ is specified as $y_i = exp(\alpha + \beta x_i + \epsilon_i)$. For the SCF regressions, the left-hand-side variable is the level (not log) of household debt and the regressor of interest is the log of household wage income. The poisson specification is preferred to a log-log specification because the latter

then rise thereafter. There is no evidence of a sustained decline in the importance of income on debt during the housing boom of the early 2000s, which would be expressed by a decline in estimated income coefficients during this period. The demographic detail in the SCF allows us to estimate debt-income relationships between age groups.

Figure 10 depicts income coefficients from a regression in which the four age-group dummies are interacted with the income regressor.[33] Income coefficients for the youngest households are depicted in the top left panel. These coefficients are typically larger than those for older age groups in the other three panels (note the difference in vertical scales across panels). The large income effects estimated for young households suggests that wage income has a larger effect on the allocation of mortgage debt among those who are on the margin of homeownership. More to the point of this paper is the lack of any evidence for any age group that income effects decline during the housing boom. In fact, for young households, point estimates effects suggest a rising importance of income over time, though the yearly confidence intervals are large.

# 4 Cross-Sectional Patterns in Gross Flows of Debt

We now turn to the cross-sectional relationship between income and gross flows of mortgage debt. This analysis can generate new facts to shed light on the mortgage boom, just as the construction of gross job and worker flows over the past 25 years has aided the study of labor markets.[34] Additionally, a discussion of gross mortgage-debt flows is helpful in understanding previous research on mortgage debt, in particular the different ways that Mian and Sufi (2009) and Adelino, Schoar, and Severino (2015b) interpret changes in HMDA purchase-mortgage flows during the housing boom.

would exclude households with zero levels of debt. Households with zero levels of wage income are excluded from the regressions, as are households with heads aged 65 years or older. In addition to the log of household income, the regressions also include dummies for the age group of the household head (younger than 35, 35-44, 45-54 and 55-64), the number of children, and dummies for nonwhite and marital status. As with the ZIP-level regressions, the SCF regressions are run as a single pooled regressions, in which the right-hand-side variables are all interacted with yearly dummies.

[33]As noted in footnote 32, the yearly SCF poisson regressions are estimated as a pooled regression, so the age-income interactions are also interacted with yearly dummies.

[34]The job-finding rate plays a central role in the benchmark search and matching model of the labor market (Pissarides 2000). And job creation and destruction rates have been used to study both the cyclical behavior of the labor market as well as potential structural declines in business dynamism and labor market fluidity (Davis, Haltiwanger, and Schuh 1996; Davis and Haltiwanger 2014).

## 4.1 Measuring Gross Flows of Mortgage Debt

A full decomposition of the stock of mortgage debt is:

$$\begin{aligned}\Delta \text{ Stock} \quad = \quad & \underbrace{\text{Purchases \& Other Originations}^{35} + \text{Increases in existing balances}^{36}}_{\text{Gross Inflows}} \\ & - \underbrace{\text{Sales \& Other Terminations}^{37} - \text{Decreases in existing balances}^{38}}_{\text{Gross Outflows}}\end{aligned}$$

For each individual mortgage in the Equifax data we have the origination date and amount, and the outstanding balance over time. While Equifax does not explicitly supply a termination date, we can see when the mortgage is removed from the data, and what its balance was when it was removed. This allows us to create measures of gross inflows and outflows to the stock of mortgage debt by year. For each mortgage, we calculate the change in the outstanding balance between the fourth quarters of consecutive years. If the oustanding balance of the mortgage went up—which is perfectly feasible for home equity lines of credit, for example—it is counted as a contribution to the inflow of mortgage debt. Similarly, if the oustanding balance went down, it is classified as a contribution to the outflow of mortgage debt. If the mortgage was originated during the course of the year, its outstanding balance as of the fourth quarter of that year is counted as a contribution to the inflow. If a mortgage was terminated, its balance as of the fourth quarter of the previous year is classified as an outflow. The gross inflows and outflows are constructed as the sum of the individual contributions of each mortgage.

We do not net out gross debt flows at the individual level before aggregating them, as is done in a recent paper on gross debt flows that also uses the Equifax data. In Bhutta (2015), households with rising debt balances contribute to so-called inflows of mortgage debt, while households with stable or declining balances contribute to debt outflows. A refinance of one mortgage into another one with the same debt balance would be affect neither inflows nor outflows in Bhutta (2015), because the refinance would not affect the individual household's debt balance. Such a refinance would contribute to both originations and terminations as we measure them, however, because we do not net out originations and terminations at the individual level before aggregating.

Yearly values for total originations and total terminations as we measure them are depicted in Figure 11. Even though these flows are measured differently than in Bhutta (2015),

[35]Other originations include interest and cash-out refinances, home equity loans, and home equity lines of credit, where the latter is only included only if it is originated with a positive balance.

[36]Increases in balance mainly refers to increases in HELOC balances.

[37]Other terminations includes mortgages that have been refinanced.

[38]Decreases in balances account for standard amortization and any prepayments.

the patterns in the figure match up with some broad lessons of that paper. For example, one of Bhutta's headline findings is that recent changes in the stock of aggregate mortgage debt have resulted mostly from changes in inflows rather than outflows, a pattern suggested by Figure 11 as well.

One disadvantage of the Equifax debt data is that it is difficult to accurately identify purchase mortgages from refinances in the early years of the sample. As a result, the purchase series in Figure 11 is derivied from HMDA data. In the early 2000s, there is a large gap between total originations as measured in Equifax and purchase mortgages as measured by HMDA. This gap is consistent with a major refinancing boom during that time, a topic we return to below.

## 4.2 Income and Gross Flows of Debt at the ZIP-Code Level

How do these gross flows correlate with income in the cross-section? The same statistical methodology used to investigate debt stocks can be applied to gross flows as well. Recall from the levels-regression discussion in section 3.2 that a simple way to parametrize the conditional expectation function for the stock of debt $d$ in ZIP code $i$ is $E(d_{cit}|y_{cit}) = \alpha_t + \beta_t \cdot y_{cit}$, where $y$ denotes income and $t$ denotes the year. Earlier, we saw that that the yearly $\beta$s generated by this levels regression on debt stocks were stable over time, a finding that was later confirmed by the long-difference regressions.[39] The two panels of Figure 12 present the estimated $\beta$s from levels regressions that have either total originations (top panel) or total terminations (bottom panel) on the left-hand-side. In contrast to the regressions that use the stock of debt, there are large year-to-year changes in the $\beta$s for both flows. For example, the large estimated $\beta$s for 2002 suggest that both gross originations and terminations were appreciably larger in high-income ZIP codes. The lack of IRS income data for 2003 prevents precludes estimation for that year, but in 2004 and thereafter the estimated $\beta$s decline over time. Yet even though there are pronounced changes in income effects during the boom, the similarity of these changes across both gross flows means that the changes cancel each other out in their effect on the stock, a pattern that explains the lack of significant income-effect shifts in the previous section. Additionally, even though the gross-flow $\beta$s decline after 2002, they remain positive. In every year of the sample, high-income ZIP codes originate and terminate higher values of mortgage debt than low-income ZIP codes do. This fact mirrors the previous section's finding that high-income ZIP codes are always responsible for higher stocks of mortgage debt.

Figure 13 performs the same type of analysis on a within-county basis. We choose counties rather than CBSAs as the geographic area so that the results can be easily compared to some

[39]See the top right panel of Figure 7 for the levels results and Table 3 for the long-difference results.

previous research using HMDA data, as explained further below. Each of the three rows of Figure 13 contains a binned scatter plot and a plot of estimated income effects for either the flow of mortgage originations (top row), the flow of mortgage terminations (middle row) or the stock of debt (bottom row). The binned scatter plots in the left panels are generated by data from 2002 to 2006, so they are constructed in the same way as the CBSA-deviated scatter plots in Figure 7. The only differences between these plots and the previous ones is that these plots use flows rather than stocks and they reflect deviations from county means rather than CBSA means.

The within-county results essentially replicate those from the previous figure. For both originations and terminations, the scatter plots and the estimated $\beta$s indicate a decline in the positive relationship between income and gross flows. In the scatter plots on the left, this drop is reflected by a shallower slope of points for the 2006 data relative to that for 2002 data, while the income coefficients on the right show the decline as a reduction in the estimated $\beta$ from 2002 to 2006. As before, the co-movement between the gross flows cancel out, so the relationship between income and debt stocks, seen in the bottom row, remains stable.

## 4.3 Refinances and Purchases

The origination and termination patterns depicted in the previous two figures can be explained in part by disproportionate participation of high-income borrowers in the early-2000s refinancing boom. The reasons behind this boom are well known. As the effects of the 2001 recession became apparent, the Federal Reserve aggressively reduced the Federal Funds rate, from 6½ percent at the end of 2000 to 1¾ percent by late 2001. The so-called jobless recovery that followed encouraged further rate cuts, with the funds rate reaching a low of 1 percent in mid-2003. Mortgage rates followed short-term rates lower, with the 30-year rate falling from around 8½ percent in early 2000 to about 5½ percent in mid-2003.[40] Figures 12 and 13 suggest that high-income borrowers were more likely to participate in this boom, consistent with the empirical literature on determinants of refinancing propensities.[41] Because the mortgage rate is determined in a national market and thus is constant across local housing markets, the higher propensity of richer borrowers to refinance shows up in both the

[40] The interest rate cited is the 30-year contract rate for conventional 30-year mortgages as measured by Freddie Mac.

[41] In his Presidential Address to the American Finance Association, Campbell (2006) highlighted three major financial mistakes that are often made by U.S. households, one of which is the failure to refinance a fixed-rate mortgage when declining interest rates make it advantageous to do so. He presents empirical analysis of early-2000s data from the American Housing Survey indicating that "younger, smaller, better educated, better off, white households with more expensive houses were more likely to refinance their mortgages between 2001 and 2003. These patterns suggest that prompt refinancing requires financial sophistication" (p. 1581).

overall regressions in Figure 12 and the within-county regressions in Figure 13.

Reduced rates of refinancing in high-income communities over the course of the boom may not be the only reason that the importance of income on gross flows declined over time. A potential complementary reason for this pattern is that low-income ZIP codes were receiving higher dollar values of purchase mortgages as the boom progressed. Unfortunately, as noted earlier we cannot distinguish refinances from purchases for all years of the Equifax sample. However, we can construct a measure of "purchase-mortgage intensity" by relating the number of HMDA purchase-mortgage originations in a ZIP code to its number of first liens from Equifax. If purchase-mortgage intensity grew relatively more for low-income ZIP codes, then we could infer that the waning of the early-2000s refinancing boom was not the only reason that low-income communities began to see relatively higher originations over time. The low-income communities would be taking on more purchase mortgages at the same time that refinancing at high-income communities was falling off, with both effects contributing to the declining importance of income in total originations apparent in Figures 12 and 13.

Figure 14 shows that relative purchase-mortgage intensity did in fact rise in low-income communities from 2001 to 2006. In both panels of this figure, the blue bars measure purchase-mortgage intensity across income quintiles in 2001, with the red bars depicting this measure in 2006. Looking across either the country has a whole (top panel), or within CBSAs (bottom panel), purchase-mortgage intensity rises more in low-income areas. What does this pattern imply for the allocation of credit? It is tempting to infer from this figure that there was a relative expansion of credit along the extensive margin to borrowers in low-income communities; that is, an expansion in the number of low-income persons who were able to borrow money to buy homes, compared to the number of new homeowners in high-income communities. This type of credit expansion can be distinguished from an expansion along the intensive margin, where the number of borrowers stays constant but the borrowers tend to take out larger mortgages.

Yet Figure 14 is not compelling evidence of a significant extensive-margin credit expansion for at least two reasons. To begin with, the expansion is not very big; the increase in purchase-mortgage intensity in the lowest-income quartiles is only a few percentage points. But a much more fundamental reason that Figure 14 does not prove an extensive-margin credit expansion is that it does not tell us how many of the *new purchase mortgages* in the low-income communities resulted in *new homeowners*. An alternative story consistent with Figure 14 is that the number of homeowners remained the same in low-income communities, but that as the housing boom progressed there were more likely to be bought and sold quickly, or "flipped." If sales turnover increased relatively more in low-income areas, then we would see a relative increase in purchase-mortgage originations in these areas. But because one homeowner would exit every time a house was flipped to a new owner, the total number of homeowners would not change, and there would be no relative expansion of credit along the extensive margin.

One way to distinguish a true extensive-margin credit expansion from the sales-turnover alternative is to look back at the individual-level relationship between income and debt from the SCF. If the new homeowners posited in the extensive-margin theory had had low incomes, then income should have become a progressively less-important determinant of mortgage debt in the SCF. That was not the case. A second way to distinguish between the alternative stories is to use the Equifax data. Even though these data cannot distinguish purchase mortgages from refinances for all years in the sample, we can always distinguish households who have mortgages from those who do not. If we equate an increase in the share of mortgaged households with an extensive-margin expansion of credit, then the Equifax data can be used to study this aspect of credit allocation.

Figure 15 shows how the shares of mortgaged households changed across the income distribution from 2001 to 2006. The bar charts in the top two panels simply report these shares by income quintile for those two years, with the top right panel deviating ZIP-level incomes from CBSA means. As we might expect, there is a positive relationship between income and shares of mortgaged households in both panels. A large part of this positive correlation undoubtedly flows from higher rates of homeownership in high-income communities, but the shares of mortgaged households depicted in Figure 15 also depend on the share of homeowners who do not own their homes free and clear. More important for our purposes are the relative changes mortgaged-household shares over time. Both panels indicate that these shares rose in all quintiles, and the increases appear comparable across the income distribution.

The relative sizes of the share increases are further examined by binned scatter plots in the lower panels of Figure 15, where the vertical axes depict the logs of mortgaged-household shares. Consider first the bottom left panel, which examines ZIP codes without regard to their CBSA location. The implied relationship between mortgaged-household shares and income shifts up from 2001 to 2006, but not in the way that relative extensive-margin expansion for low-income communities would suggest, because the shift appears greater at the top end of the income distribution. For low-income ZIP codes, the shares of mortgaged-households rise during the boom, as seen in the earlier bar chart, but this rise is accomplished by the lower dots in the scatter plot moving up along the conditional expectation that is traced out by the 2001 data. The story is different for high-income ZIP codes, where the increase in mortgage shares owes more to a shift up in this conditional expectation. The lower right panel of Figure 15 repeats the analysis on a within-CBSA basis. The conditional expectations have essentially the same shape in both 2001 and 2006, suggesting that high-income shifts in the previous panel were primarily due to between-CBSA shifts in mortgaged-households shares. The bottom line is that no panel in Figure 15 points to a relative extensive-margin expansion of mortgage debt in low-income communities.

Putting all the results together suggests an explanation of gross-flow patterns that is consistent with earlier results on debt stocks. At the start of the boom, total originations and terminations were especially high in richer communities, no doubt in part because of the refinancing boom. Because refinancing involves simultaneous origination and termination of mortgages, changes in the pace of refinancing over time did not have a first-order effect on relative debt stocks in high- and low-income areas. But declining refinancing in high-income ZIP codes is only one reason that relative originations rose in low-income areas as the boom progressed. HMDA data show that purchase-mortgage originations increased more in low-income ZIP codes as the boom progressed. Even here, however, changes in one gross flow were offset by similar changes in the other. Equifax data indicate that relative shares of mortgaged households did not rise much in low-income areas, independent of income. The implication is that purchase-mortgage intensity in low-income areas was driven by increased sales turnover, another situation in which mortgages are typically originated and terminated at the same time. Because all of these shifts in gross flows tended to cancel each other out, the end result is that debt stocks rose in high- and low-income communities at similar rates, as we found in the previous section.

## 4.4 Relation to Previous Research on Mortgage Originations

A recognition that shifting patterns of gross debt flows can be consistent with stable patterns of debt stocks is critical for understanding the existing academic literature on the nature mortgage boom. A central finding in the influential paper by Mian and Sufi (2009, henceforth MS09) is that ZIP-level growth in income and growth in HMDA purchase-mortgage originations were *negatively correlated* between 2002 and 2005.[42] Looking across other sub-periods of their data, which begin in 1991, the authors find no other negative correlations between income growth and growth of purchase-mortgage credit. MS09 therefore interprets the negative correlation for 2002-05 as a fundamental change in lending patterns at the height of the boom—a change that allocated a disproportionate amount of mortgage credit to borrowers with the poorest income prospects.

A closer look at the negative correlation shows how it is statistically consistent with the findings above. The correlation is generated by regressing the flow of purchase-mortgage originations on growth in AGI and county-level fixed effects. For the years of the negative

[42]The negative-correlation finding appears in both Figure III and Table IV of MS09. As noted below, the negative correlation involves ZIP-level income and credit growth relative to county means.

correlation, 2002 and 2005, this regression is

$$\Delta Purchase\ Originations_{i,2002-05} = \delta \Delta Income_{i,2002-05} + \phi_{county} + \epsilon_i, \tag{5}$$

where $i$ indexes ZIP codes. The fixed effects $\phi_{county}$ cause the the coefficient $\delta$ to be identified on relative growth rates of income and purchase originations within counties. This regression coefficient turns out to be negative, and is reported as the negative correlation in MS09.

While the regression coefficient $\delta$ is intended to reflect lending standards between 2002 and 2005, it does not measure the true relationship between purchase originations and income in 2002, that relationship in 2005, or the change in the relationship between those two years. Recalling the earlier discussion of long-difference regressions in section 3.3, any investigation of a change in lending standards must also include the *level* of income in one of the years being tested:

$$\Delta Purchase\ Originations_{i,2002-05} = \beta \Delta Income_{i,2002-05} + \gamma Income_{i,2005} + \phi_{county} + \epsilon_i. \tag{6}$$

As implied by equation 4 in section 3.3, including the 2005 income level will cause $\beta$ to measure the debt-income relationship in 2002, while $\gamma$ will measure the change in this relationship from 2002 to 2005.

Table 5 shows what happens to the negative correlation when an income-level term is added. Each row in this table is a separate regression. The first regression, designated as Model 1, follows MS09 by using growth in purchase-mortgages originations from HMDA on the left-hand side. Model 1 replicates the negative sign of the MS09 coefficient on income growth when no income-level term is included. We do not replicate MS09's exact estimate because we do not use annualized growth rates; we also use growth rates between 2002 and 2006, rather than 2002 and 2005.[43] The next row, Model 2, adds the income level of each ZIP code in 2006. This regression shows that what MS09 characterizes as a *negative correlation* is in reality *a decline in a positive correlation*. Specifically, the regression coefficients of Model 2 imply that there is a positive and highly significant relationship of 0.88 between levels of purchase-mortgage originations and income in 2002, which declines by 0.38 to an implied 0.50 in 2006.[44] This decline is qualitatively similar to the decline in within-county income effects for *total* originations seen earlier in Figure 13. It is also consistent with the rising level of purchase-mortgage intensity seen in Figure 14.

[43] Our choice of years is motivated by the large change in the originations-income relationship between 2002 and 2006 that is evident in Figures 12 and 13. Regressions using other years generate similar results and are available from the authors on request.

[44] Recall that from equation 4 that when the latter-year income level is entered in a long-difference regression, as in Model 2, the coefficient on the income-growth term is the levels relationship in the first year, and the income-level coefficient is the change in the relationship from the first year to the last year.

The question of what we should learn from a decline in the correlation between purchase mortgages and income is the subject of debate. Adelino, Schaor, and Severino (2015, henceforth A2S15) perform some additional econometric work to argue that the declining correlation tells us little about the driving forces behind the mortgage boom.[45] Like MS09, A2S15 uses HMDA data on purchase-mortgage originations, but the paper also exploits the fact that HMDA data are available at the borrower level. This granularity allows a split of the aggregate dollar value of originations in a ZIP code into two parts: the *number* of new purchase mortgages and the *average amount* of each new mortgage. The authors then correlate the two components of total purchases independently with income. They find that the decline in the purchase-mortgage correlation is driven exclusively by a change in how income matters for the number of new purchase mortgages, not in the effect of income on the average size of mortgages. A replication of their results expressed in our long-difference framework appears as Model 3 in Table 5. Like A2S15, this model shows that the most significant change in origination patterns relates to higher numbers of new mortgages in low-income areas, not to higher average mortgage amounts. In particular, virtually all of the 0.38 decline in the estimated income effect in Model 2 occurs via a relative increase in the number of mortgages in low-income ZIP codes. A much smaller portion is due to larger mortgage sizes.

To A2S15, the importance of mortgage numbers in driving MS09's results, as opposed to mortgage amounts, is enough to refute that paper's claim that credit and income became decoupled during the boom. "Given that households, not ZIP codes, take on mortgages, only the relation between individual mortgage size and income can inform us about changes in the debt burden across households" (p. 3). Mian and Sufi (2015b) argue that this characterization of credit expansions is too narrow. As discussed earlier, credit expansions can occur through the extensive margin (more borrowers in low–income areas) as well as through the intensive margins (larger mortgages in low-income areas).[46] A much stronger critique of MS09's decoupling hypothesis is that most of the additional mortgages originated in the low-income areas did not in fact result in more homeowners. Rather, these mortgages simply reflected a higher rate of sales turnover for houses located in low-income areas, and

[45] A2S15 do not estimate a long-difference regression with a levels term included, like Model 2. However, one set of regressions in A2S15, which is estimated in levels, suggests a flattening out of the relationship between purchase mortgages and income. These regressions project ZIP-level of HMDA purchase-mortgage originations from 2002, 2004, 2005, and 2006 on AGI. Income is interacted with yearly dummies, and fixed effects for year and ZIP code are also included. As noted by the authors, the coefficients from this regression indicate that the positive relationship between originations and income becomes progressively flatter.

[46] (Mian and Sufi 2015b). note that an expansion along the extensive margin would also restrain the growth of average mortgage amounts if the mortgages awarded to new borrowers are relatively small. Additionally, Mian and Sufi (2015b) discusses other work in A2S15 that uses borrower-level income data from HMDA, contending the ZIP-level income data from the IRS is to be preferred to borrower-level income data from HMDA because of concerns about fraud.

were thus cancelled out by higher mortgage terminations.

This possibility, of course, is exactly the issue examined in the previous subsection. There, we used both HMDA and Equifax data to show that while relative purchase-mortgage intensity rose in low-income ZIP codes, the number of mortgaged households did not. This finding suggests that if we were to run the long-difference models with mortgage *stocks* rather than purchase-mortgage *flows*, we should find relative differences in neither mortgage numbers nor in mortgage amounts. If we did, then rather than debate about whether an increase in the number of purchases mortgages implies that such an extensive-margin credit expansion occurred, we could answer definitely whether an extensive margin credit expansion occurred.

The debt-stock models appear in the lower rows of Table 5. Model 4a runs the long-difference regression using the growth rate of the standard debt variable in this paper, which is the stock of mortgage debt normalized by the number of tax returns in the ZIP code. The small coefficient on the income-level term (–0.05), suggests no large changes in the stock of debt across the income distribution, consistent with the earlier results of this paper. Model 4b is in the spirit of A2S15, in that it splits the debt-growth variable into two parts: the growth rate of mortgage debt per household (upper row) and the growth rate of the total number of mortgaged households (lower row).[47] Ignoring rounding, the two income-level coefficients in Model 4b must add up to the corresponding level coefficient of the aggregated model immediately above it, Model 4a. It not therefore not surprising that both of the income-level coefficients in Model 4b are small.

The models so far in Table 5 have used AGI as the measure of ZIP-level income, so as to be consistent with both MS09 and A2S15. Models 5a and 5b follow the earlier regressions of this paper and use wage-and-salary income instead. Doing so has two effects. First, the implied relationship in 2002 between the level of mortgage debt and the level of income is now stronger. The relevant coefficient in the AGI model (Model 4a) is 0.42, while the coefficient in the wage-and-salary model (Model 4b) is 0.74. This increase is consistent with the idea that the true measure of income used in mortgage-lending decisions is wages, for which the AGI measure is a noisy proxy. A second and more important effect of using wages rather than AGI is that changes in the 2002-06 changes in the debt-income relationships, which are captured by the coefficients on the income-level terms, decline to essentially zero.

## 4.5 Corroborating Evidence

The debt-stock regressions in the lower rows of Table 5 may puzzle many readers. Wasn't the mortgage boom characterized by an increase in the homeownership rate, accomplished

[47] This decomposition interprets the number of tax returns as the number of households, as we have throughout this paper.

in large part by the extension of mortgages to previously disadvantaged borrowers? If so, then why is there no evidence of a credit expansion along the extensive margin, let alone the intensive one?

In fact, the mortgage boom of the early 2000s saw small or negative changes in homeownership rates among previously marginalized borrowers. Figure 16 uses published data from the Census Bureau to illustrate this point. The top right panel shows that recent U.S. history does indeed include a period in which the aggregate homeownership rate in the United States rose sharply, but most of this period predates the mortgage boom. The gray vertical lines in this panel and elsewhere in Figure 16 denote the years 2001 and 2007, so they demarcate the years in which mortgage debt rose rapidly relative to income, as seen in Figure 1. The homeownership rate began to rise in 1995, several years before the 2001 onset of the mortgage boom. The homeownership rate did continue to rise for a few years therefore, but this rate ended the boom about where it began. The next two panels of Figure 16 use published Census data to show that the pre-2001 period of rising homeownership featured disproportionate ownership increases among households with below-median incomes (top right panel) as well as African-American households (bottom left panel). Yet both panels also show that these trends reversed themselves in 2001, when mortgage debt began growing must more rapidly than income. Finally, the lower right panel uses published Census tabulations from yearly American Housing Surveys to plot average homeownership rates disaggregated by household income relative to the poverty line.[48] From 1995 to 2001, each of the four groups with incomes no higher than 200 percent of the poverty line saw its homeownership rate either rise or stay essentially flat. For the poorest two groups of households, who had incomes at or below the poverty line, ownership increases during this early period were especially large. Yet after the mortgage boom begins in 2001, low-income ownership rates begin to fall, with the steepest declines experienced by the households with the lowest incomes. These patterns are not consistent with a broad extensive-margin credit expansion among poor households during the mortgage boom. But they are consistent with Equifax results showing that any relative increase in purchase-mortgage originations in low-income communities over the course of the boom was offset by higher terminations.[49]

Additional corroborating evidence on the lack of a stock response to changes in debt

[48] These figures comes from Table 2-12 ("Income Characteristics: Occupied Units") of the Annual Housing Survey tabulations published by Census for 1995, 2001, 2005, and 2007. These reports are available at `http://www.census.gov/content/census/en/programs-surveys/ahs/data.html/`. See Segal and Sullivan (1998) for an early econometric analysis of homeownership increases by income and race.

[49] Gerardi and Willen (2009) link HMDA data to property-level deed records to study the effect of subprime lending on urban neighborhoods in Massachusetts. They find that during the boom, African-Americans accounted for a disproportionately large share of buyers in urban neighborhoods. But African-Americans also accounted for an equally high percentage of sellers. The implication is that subprime increased sales turnover without affecting minority homeownership rates.

flows comes from MS09 itself. Although virtually all of credit analysis in MS09 involves the flow of HMDA purchase-mortgage originations, one regression makes use of debt-stock data from the same source we use, the Equifax credit bureau.[50] Column 5 of Table V in MS09 (p. 1472) regresses annualized growth in ZIP-level mortgage debt on the share of borrowers with subprime credit scores in 1996 (also from Equifax) as well as covariates reflecting ZIP-level growth in income, establishments, employment, and crime from 2002 to 2005. The coefficient on the subprime share is positive, suggesting higher debt growth in ZIP codes with many subprime borrowers. It is shown elsewhere in the paper that these ZIP codes also have lower income levels, as we would expect.[51] But the effect of subprime shares on the implied differences in debt-growth is small. Holding other covariates constant, MS09's regression implies that the debt-stock growth rate would be about 1.4 percentage points higher each year if a ZIP code moved from the bottom to the top quartile of subprime shares.[52] This estimate, which is small relative to the large increase in mortgage debt that occurred over this period, is probably biased up due to the omission of home-equity debt from the stock measure, given the high growth rates of home-equity debt in high-income areas.[53] The coefficient is also undoubtedly affected by the inclusion of the non-income variables, in ways that are hard to determine.[54] In any case, this regression indicates a much smaller responsiveness of debt stocks to income, relative to the effect of income on gross debt flows.

# 5 Cross-Sectional Evidence on Income and Defaults

This paper has focused on the borrowing patterns of different income groups during the mortgage boom, but a complete understanding of the boom also involves what happened to different borrowers during the subsequent bust. One reason why so much attention has been paid to low-income borrowing is the large absolute number of foreclosures that would later occur among low-income borrowers. How can we square the uniform growth rates of debt during the boom with the apparent concentration of foreclosures at the bottom of the income distribution? It turns out that what appeared to some as a concentration of

[50] MS09 generally use the Equifax debt data to identify so-called subprime ZIP codes, which they define as ZIP codes with a high proportion of residents with low credit scores. In a later section, we discuss the relationship between credit scores and debt, as opposed to the relationship between income and debt that has been the main focus so far.

[51] See Table II of MS09

[52] The subprime-share coefficient estimate in MS09's debt-stock regression is 0.05. Table II of MS09 indicates that the share of subprime borrowers in first-quartile ("prime") ZIP codes is 0.159, while the share in last-quartile ("subprime") ZIP codes is 0.444. The 1.4 percentage-point figure is generated from by multiplying the estimated coefficient times the difference in subprime shares.

[53] Recall the discussion of home-equity debt at the end of section 2.

[54] MS09 does not discuss the quantitative predictions of the debt-stock model, nor do they compare its results to debt-flow regressions that appear in the same table. Instead, the stock regression serves as a comparison to a separate debt-stock regression with non-housing debt on the left-hand-side.

foreclosures among poorer borrowers is no such thing. Just as debt was scaled up equally across the income distribution during the boom, foreclosures were similarly scaled across the distribution as well. As a result, foreclosure patterns not only fail to argue against the uniform nature of the mortgage boom—they actually help confirm it.[55]

To show this, we again turn to the Equifax data, which allow us to calculate default rates by ZIP code. Individual mortgages in the Equifax data are noted as either current or delinquent, with the latter set further delineated by length of delinquency: 30, 60, 90 or 120+ days. The dataset is quarterly, so we can define the default rate in quarter $t$ as the share of all active first liens in quarter $t-1$ that transition to 90-days delinquent or quarter $t$.[56] The top two panels in Figure 17 present binned scatter plots of the log of this default rate against income per return categories in 2001:Q4 and 2009:Q4. We choose 2009 as the latter year for this comparison because the foreclosure crisis did not peak until well after house prices had begun to fall. Because we use natural logs when measuring defaults, a uniform percentage increase in defaults across the income distribution will show up as a uniform shift upwards in the implied relationship, as was the case earlier with mortgage debt.

The upper left panel of Figure 17 shows the relationship across all ZIP codes in the country, without regard to CBSA location. As we would expect, there is a strong negative relationship between default rates and income, as defaults are always higher in ZIP codes at the bottom of the income distribution. But this plot also shows that over the course of the housing bust, defaults rose in both high- and low-income communities. If anything, the default rate grew more *in percentage terms* in high-income income ZIP codes, as the slope of the conditional expectation function flattens from 2001 to 2009. The upper right panel makes the within–CBSA comparison. As was the case for within-CBSA comparisons of mortgage debt, both defaults and income are measured relative to CBSA means. A modest flattening in the default–income relation is evident in this plot as well.

The lower panel of Figure 17 presents time-series evidence on the default-income relationship over the entire course of the housing cycle. Both defaults and income are measured relative to quarterly CBSA means. The top two lines in this panel are simple plots of default rates for the lowest-income and highest-income quintiles; consistent with the scatter plots, default rates are always higher in poorer communities. The third line in the panel shows the simple difference between these two rates. Because rates in the low-income quintile are always relatively high, the near-equal percentage increases in defaults shows up as a larger

[55]Adelino, Schoar, and Severino (2015b) also examine foreclosures across the income distribution, but their emphasis is on the dollar value of defaults for a specific vintage of mortgages. Additionally, that paper uses the McDash data generated by mortgage servicers, so the results may be less representative of the entire mortgage market than results based on Equifax data.

[56]The resulting ratio is similar to a sample hazard. We define the number of active first liens in the previous quarter as all liens that are less than 90 days delinquent. These liens therefore comprise the "risk set" for loans that can become 90 days delinquent in the current quarter.

absolute increase in defaults that grows over time. However, the near-equal percentage increases seen in the earlier binned scatter plots imply that the ratio of default rates will not change much, as is confirmed by the bottom line in the panel. In fact, because percentage increases in defaults are a little bit higher for richer quintiles, the ratio of default rates declines a little over time. Figure 17 makes clear that the expansive run-up in mortgage debt during the boom phase of the housing cycle was mirrored by an expansive increase in foreclosures during the bust. This finding provides additional evidence against the claim that the housing cycle was driven by factors unique to a specific corner of the mortgage market.

Like the earlier results on mortgage debt, the default results illustrate a broader lesson about using cross-sectional data to evaluate macroeconomic phenomena. When different cross-sectional units have different underlying values of the variable being studied, *equal percentage* changes this variable—a pattern that would suggest an aggregate force acting equally on all cross-sectional units—will show up as *different absolute* changes in the variable. This is certainly the case for the increase in mortgage debt, where equal percentage changes resulted in very large dollar values of additional debt taken on by wealthy borrowers and communities. This section has shown that the pitfall is also relevant for foreclosures, because near-equal percentage increases in that variable resulted in a larger absolute increase for poorer communities.[57] Of course, absolute differences are often informative. Policymakers with limited anti-foreclosure resources would do well to target those resources to low-income communities, because of the large number of foreclosures there.[58] But when economists study the causes and consequences of the housing cycle, a simple focus on higher absolute increases in default rates in low-income communities misses the most interesting part of the story.[59]

[57] Mian and Sufi (2009) performs a diffs-in-diffs analysis on foreclosures using absolute differences, which leads them to argue that foreclosures were "concentrated" in ZIP codes with large shares of subprime borrowers (p. 1449). Figure 17 shows that foreclosures were not concentrated in low–income communities, if we define concentration in the usual economic sense. That is, there was no increase in the *share* of foreclosures in low-income ZIP codes relative to foreclosures in all ZIP codes.

[58] A separate reason to target anti-foreclosure efforts in low-income communities include the larger effects that foreclosures might have on low-income borrowers who default, relative to high-income borrowers who do so. Also, there is probably a higher likelihood that the typical foreclosure in a low-income area is a "double-trigger" foreclosures, sparked by a combination of negative equity and adverse life event like job loss, rather than a "strategic" default, in which the borrower could stay current on his loan but chooses not to.

[59] The Internet Appendix contains an analysis of default rates that employs the same regression approach used to analyze growth in mortgage debt.

# 6 What About Subprime?

## 6.1 Subprime's Role the Crisis vs. the Amount of Subprime Debt

Another way in which a uniform increase in debt appears at odds with the conventional wisdom on the boom involves subprime lending. Losses on subprime investments played a critical part in the financial crisis, which to this day is often referred to as a subprime crisis. A natural conjecture is that because subprime was so important in the housing bust, subprime lending to low-income borrowers must have played a large part in the housing boom. As we will confirm below, it is true that subprime mortgages were especially popular in low–income areas. It is of course also true that losses tied to privately securitized loans, including subprime mortgages, played a key role in the financial crisis of 2008. However, the reason that subprime mortgages were important in the financial crisis was not because the stock of subprime debt had become especially large. A much more important reason for the importance of subprime was that these loans were generally were not insured by the government, as was the case with the prime loans packaged into securities by the government-sponsored agencies, Fannie Mae, Freddie Mac, and Ginnie Mae. The lack of government insurance on subprime left private investors on the hook for subprime losses. Another reason that losses on subprime mortgages had a disproportionate impact on the financial system was that this impact was amplified by the creation of synthetic collateralized debt obligations, or synthetic CDOs. These instruments were collections of credit default swaps that referenced underlying tranches of bonds backed by subprime mortgages. By creating synthetic CDOs, Wall Street firms allowed investors to bet on the performance of the underlying subprime mortgages without having to originate additional subprime loans.[60] The famous Abacus deal, which was arranged by Goldman Sachs and on which the investor John Paulson made billions of dollars, was a synthetic CDO. Like other synthetic CDOs, this deal had both winners and losers, but the losers were investment banks and other firms inside the financial system, while the winners were mortgage-market outsiders (Foote, Gerardi, and Willen 2012). By concentrating additional losses on the financial system, the synthetic CDOs helped amplify the effect of subprime defaults in the crisis.

A further illustration of subprime's modest size is the fact that subprime was not even the largest type of privately securitized debt. Data on outstanding non-agency residential mortgage-backed securities (RMBS) is available from Securities Industry and Financial Markets Association (SIFMA), an industry trade group collects and publishes data on a variety

[60] In the movie *The Big Short*, synthetic CDOs are illustrated with a blackjack game played by the singing star Selena Gomez and the behavioral economist Richard Thaler. Observers to the game take side bets on how the wages placed by Gomez and Thaler will turn out. The actual mortgage-backed securities are analogous to the original bets played by Gomez and Thaler, while the synthetic CDOs are similar to the side bets made by the spectators.

of securities.[61] According to SIFMA, the total amount of outstanding private-label RMBS rose from $819 billion in 2001 to $2.7 trillion in 2007. The subprime component of RMBS grew from $345 billion to $772 billion over the same period, an increase of about $428 billion, or a little more than one-fifth of the total increase in private-label RMBS. A much larger share of the total increase private-label debt was due to Alt-A lending. These "alternative-A" mortgages were usually given to people with prime credit scores who did not fully document their incomes, or who wanted mortgages with other elevated risk characteristics. From 2001 to 2007, the value of Alt-A bonds grew at a much faster rate than subprime, mushrooming from about $55 billion to just over $1.0 trillion. Another important source of loans for private-label RMBS was the jumbo prime market, which produced prime mortgages too large to be packaged into agency MBS. Jumbo participation in non-agency RMBS rose by about $380 billion from 2001 to 2007, comparable to the increase in securitized subprime debt.[62]

Even when they are combined, subprime and Alt-A loans made up a small part of the overall boom in mortgage debt. The top panel of Figure 18 depicts the total stock of outstanding mortgage debt over time, as measured by the Federal Reserve's Flow of Funds accounts, along with two counterfactuals. The red line in this panel uses the SIFMA data to show how total mortgage debt would have evolved had the only net additions to aggregate debt been securitized subprime loans, while the dashed blue line includes net additions of Alt-A loans as well.[63] Neither counterfactual accounts for much of the total increase in mortgage debt.

While most debt incurred during the financial crisis consisted of prime loans, subprime debt did help prevent a reallocation of mortgage debt toward high-income borrowers. The bottom panel of Figure 18 shows the contributions to mortgage debt of subprime, Alt-A, and prime debt to mortgage-debt growth rates across the income distribution of ZIP codes from 2001 to 2006. The horizontal axis in this figure is delineated by 20 return-weighted bins of income per tax return. Because SIFMA does not publish information on debt at the ZIP-code level, we calculate the cross-sectional subprime and Alt-A contributions using loan-level data from the CoreLogic ABS Private Label Securities ABS Database. As noted earlier, the CoreLogic dataset contains information about subprime and Alt-A mortgages that were packaged into non-agency securities.[64] The estimates of the implied contributions

[61] The main SIFMA statistics page is `http://www.sifma.org/research/statistics.aspx`. The spreadsheet with data on the amount of U.S. mortgage-related issuance and outstanding debt is found at: `http://www.sifma.org/uploadedFiles/Research/Statistics/StatisticsFiles/SF-US-Mortgage-Related-SIFMA.xls`.

[62] Other types of privately securitized RMBS included HELOCs, junior liens, resecuritizations, and manufactured housing. The Internet Appendix contains a complete listing.

[63] The two counterfactuals add net contributions of subprime and Alt-A to the stock of total mortgage debt in 2001:Q1.

[64] The CoreLogic data, formerly known as the LoanPerformance dataset, also includes prime jumbo loans.

of subprime debt in this panel are overstated to some degree, because CoreLogic's coverage of the non-agency market grows from 2001 and 2006.[65] However, the cross-sectional pattern is clear. Because subprime debt was more prevalent in low-income areas, it helped these areas keep up with the prime-driven growth of mortgage debt in richer areas. In other words, subprime debt did not cause a reallocation of mortgage debt—it prevented one.

## 6.2 Creditworthiness and Mortgage Debt

Because subprime mortgages had elevated risk characteristics, any discussion of subprime leads naturally to the broader question of how cross–sectional patterns of mortgage debt were affected by creditworthiness. The focus in this paper has been on mortgage debt and income, but one can also correlate changes in mortgage debt over time to observable measures of creditworthiness, such as credit scores. Such a study would have to confront two important issues, the first being endogeneity. When a borrower purchases a home and then makes a series of on-time payments, her credit score typically rises. Reverse causation therefore influences the correlation between mortgage debt and the borrower's current credit score. A second reason that studying the correlation of credit scores and debt is tricky is that deriving precise theoretical predictions for the sign or magnitude of this correlation is difficult. Typically, people borrow to buy homes early in their adult lives, after they have incurred other debts and before they have built up substantial savings or established long histories of paying bills on time. The life-cycle borrowing pattern therefore exerts a negative influence on the cross-sectional relationship between credit scores and debt, regardless of the current state of lending standards.

A rough attempt to address both the endogeneity problem and the life-cycle issue is to compare ZIP-level correlations of mortgage debt with lagged averages of credit scores at different points in time. Using lags of credit scores reduces the endogenous feedback between debt and creditworthiness, while the time-period comparison addresses the age issue, as long as the age distributions of ZIP codes do not change much. Figure 19 presents these correlations for 2001 and 2006 using the Equifax data, on both an overall and within–CBSA basis. Distributions of debt with respect to lagged credit–score quintiles are also presented.[66] The figure shows that debt rose across the credit-score distribution equally, and that the implied correlations between debt and creditworthiness were stable over time. These results mirror the earlier findings on the relationship between debt and income.

Figure 19 should not be interpreted as the last word on the relationship between credit

[65]See Figure A.9 in the Internet Appendix.

[66]The credit score used for both the distributions and the correlations is the Equifax Risk Score, which is generated from a proprietary model by Equifax and which is similar in spirit to the FICO score used by many lenders.

scores and debt. Indeed, a forthcoming study by Albanesi et al. (2016) exploits the individual age data in the Equifax data to address both the endogeneity and the life-cycle issues more carefully. However, it is gratifying that Figure 19 is consistent with the bottom line of this paper: the mortgage boom was not concentrated at the low end of the socioeconomic distribution, but rather generated by a surge of borrowing across a wide swath of the American population.[67]

# 7 Conclusions

The main message of this paper is that mortgage debt grew rapidly across the income distribution during the early 2000s, a pattern that resulted in very large dollar amounts of debt taken on by high-income borrowers. Looking within individual housing markets, the pattern is similar, as booming cities saw debt rise at equal rates in its rich and poor neighborhoods. These patterns are hard to square with any theory of the crisis that relies on a reallocation of credit toward low-income borrowers. Rather, the findings imply that the housing cycle was just that—a housing cycle—not a "subprime" cycle that was sparked by disproportionate borrowing at the lower end of the income distribution (Ferreira and Gyourko 2015).

Had we found instead that low-income borrowers and communities experienced higher-than-average rates of debt growth, then theorists might have had an interesting puzzle to explain: how could agents at the bottom end of the income distribution have had such a big effect on the aggregate housing market? As illustrated in the seminal paper of Krusell and Smith (1998), formal models of asset markets with heterogeneous agents can allow poorer agents to behave differently than richer ones, so that a complete characterization of asset markets requires a sophisticated model that tracks wealth distributions over time. However, because poor people do not have much wealth, their behavior does not strongly influence these distributions. As a result, it is hard to see how a formal model could explain a $1.5-trillion increase in mortgage debt for the richest income quintile with a relaxation of borrowing constraints among poor borrowers. Perhaps it is fortunate that this puzzle does not need to be solved, because mortgage debt grew at similar rates across the income

[67] Mian and Sufi (2015a) uses individual-level data from Equifax to analyze the relationship between creditworthiness and debt growth. They find that "individuals in the 20th to 60th percentile of the initial credit score distribution contributed most to the total dollar rise in household debt, having both high initial debt levels and relatively strong growth in debt" (p. 2). Aside from their basis in individual-level rather than ZIP-level data, these results are not strictly comparable to ours because the authors fix individual credit scores at their 1997 levels. This particularly strong way of solving the endogeneity problem prevents entrants after 1997 from contributing to the results. It also prevents events after 1997 from influencing credit scores for individuals in the sample. Finally, this method makes it hard to determine whether aggregate increases in mortgage debt in time periods besides the early 2000s looked different. Adelino, Schoar, and Severino (2015a) includes an additional discussion of Mian and Sufi (2015a).

distribution.

In our view, the results above are more consistent with an alternative story, in which exogenous borrowing constraints play no role and the causality runs from house prices, or house-price expectations, to the widespread accumulation of mortgage debt. During the boom, optimistic views of house-price growth were widely shared by potential home buyers (in all income classes) as well as mortgage lenders.[68] If we start from the presumption that these price expectations were overly optimistic, then the decisions of the nation's borrowers and lenders make sense. Everyone would have tried to profit from the housing boom, either by buying the houses that were rising in price, or by making loans that were backed by this rapidly appreciating collateral. Of course, the key theoretical hurdle here is figuring out where the optimistic beliefs came from. In the last several years, economists have begun work on models of so-called distorted beliefs, in which rational expectations are augmented or replaced by beliefs that are influenced by expressly psychological factors.[69] A model in which beliefs about future house prices are widely distorted would appear to be consistent with the broad-based nature of the mortgage boom. A model in which distorted lending incentives or misguided government policy leads to disproportionate borrowing among low-income individuals does not.

[68]For average borrower-level expectations, see Case, Shiller, and Thompson (2012). For the price expectations of Wall Street analysts, see Gerardi et al. (2008).

[69]Papers that explore the formation of beliefs include Gennaioli and Shleifer (2010), Gennaioli, Shleifer, and Vishny (2012), Barberis (2013), Brunnermeier, Simsek, and Xiong (2014), Simsek (2013), Fuster, Laibson, and Mendel (2010), Geanakoplos (2009), and Burnside, Eichenbaum, and Rebelo (2015). In addition to Adelino, Schoar, and Severino (2015b), empirical papers supporting the price-expectations theory include Cheng, Raina, and Xiong (2014) and Bayer, Mangum, and Roberts (2016).

## References


Adelino, Manuel, Antoinette Schoar, and Felipe Severino. 2015a. "Loan Originations and Defaults in the Mortgage Crisis: Further Evidence." National Bureau of Economic Research Working Paper No. 21320.

———. 2015b. "Loan Originations and Defaults in the Mortgage Crisis: The Role of the Middle Class." National Bureau of Economic Research Working Paper No. 20848. Forthcoming in *Review of Financial Studies*.

Albanesi, Stefania, Giacomo DeGiorgi, Jaromir Nosal, and Matthew Ploenzke. 2016. "Credit Growth and the Financial Crisis: A New Narrative." *In process* .

Amromin, Gene and Leslie McGranahan. 2015. "The Great Recession and Credit Trends across Income Groups." *American Economic Review* 105 (5):147–153.

Amromin, Gene, Leslie McGranahan, and Diane Whitmore Schanzenbach. 2015. "Consumer Credit Trends by Income and Geography in 2001-12." Federal Reserve Bank of Chicago. Available at `https://www.chicagofed.org/publications/chicago-fed-letter/2015/342`.

Barberis, Nicholas. 2013. "Psychology and the Financial Crisis of 2007–2008." In *Financial Innovation: Too Much or Too Little?*, edited by Michael Haliassos. Cambridge, MA: MIT Press.

Bayer, Patrick, Kyle Mangum, and James W. Roberts. 2016. "Speculative Fever: Investor Contagion in the Housing Bubble." National Bureau of Economic Research Working Paper No. 22065.

Bhutta, Neil. 2015. "The Ins and Outs of Mortgage Debt During the Housing Boom and Bust." *Journal of Monetary Economics* 76 (1):284–298.

Bricker, Jesse, Lisa J. Dettling, Alice Henriques, Joanne W. Hsu, Kevin B. Moore, John Sabelhaus, Jeffrey Thompson, and Richard A. Windle. 2014. "Changes in U.S. Family Finances from 2010 to 2013: Evidence from the Survey of Consumer Finances." *Federal Reserve Bulletin* 100 (4). Available at `http://www.federalreserve.gov/pubs/bulletin/2014/pdf/scf14.pdf`.

Brown, Meta, Andrew Haughwout, Donghoon Lee, and Wilbert van der Klaauw. 2015. "Do We Know What We Owe? Consumer Debt as Reported by Borrowers and Lenders." *FRBNY Economic Policy Review* 21 (1):19–44. Available at `http://www.newyorkfed.org/research/epr/2015/EPR_2015_comparisons_brown.pdf`.

Brunnermeier, Markus K., Alp Simsek, and Wei Xiong. 2014. "A Welfare Criterion For Models With Distorted Beliefs." *Quarterly Journal of Economics* 129 (4):1753–1797.

Burnside, Craig, Martin Eichenbaum, and Sergio Rebelo. 2015. "Understanding Booms and Busts in Housing Markets." Northwestern University Working Paper. Available at http://www.kellogg.northwestern.edu/faculty/rebelo/htm/booms-busts.pdf.

Campbell, John Y. 2006. "Household Finance, Presidential Address to the American Finance Association." *Journal of Finance* 61 (4):1553–1604.

Case, Karl E., Robert J. Shiller, and Anne K. Thompson. 2012. "What Have They Been Thinking? Homebuyer Behavior in Hot and Cold Markets." *Brookings Papers on Economic Activity* 45 (2):265–298.

Cheng, Ing-Haw, Sahil Raina, and Wei Xiong. 2014. "Wall Street and the Housing Bubble." *American Economic Review* 104 (9):2797–2829.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review* 104 (9):2593–2632.

Davis, Steven J. and John C. Haltiwanger. 2014. "Labor Market Fluidity and Economic Performance." Paper presented at the Federal Reserve Bank of Kansas City's Jackson Hole Conference, *Re-Evaluating Labor Market Dynamics*. Available at https://www.kansascityfed.org/~/media/files/publicat/sympos/2014/2014davis_haltiwanger.pdf.

Davis, Steven J., John C. Haltiwanger, and Scott Schuh. 1996. *Job Creation and Destruction*. Cambridge, MA: MIT Press.

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

Foote, Christopher L., Kristopher S. Gerardi, and Paul S. Willen. 2012. "Why Did So Many People Make So Many *Ex Post* Bad Decisions: The Causes of the Foreclosure Crisis." In *Rethinking the Financial Crisis*, edited by Alan S. Blinder, Andrew W. Lo, and Robert M. Solow. The Russell Sage Foundation and the Century Foundation: New York, N.Y. and Washington, D.C., 136–186.

Fuster, Andreas, David Laibson, and Brock Mendel. 2010. "Natural Expectations and Macroeconomic Fluctuations." *Journal of Economic Perspectives* 24 (4):67–84.

Geanakoplos, John. 2009. "The Leverage Cycle." In *NBER Macroeconomics Annual 2009*, edited by Daron Acemoglu, Kenneth Rogoff, and Michael Woodford. Chicago, IL: University of Chicago Press, 1–65.

Gennaioli, Nicola and Andrei Shleifer. 2010. "What Comes to Mind." *Quarterly Journal of Economics* 125 (4):1399–1433.

Gennaioli, Nicola, Andrei Shleifer, and Robert Vishny. 2012. "Neglected Risks, Financial Innovation, and Financial Fragility." *Journal of Financial Economics* 104 (3):452–468.

Gerardi, Kristopher, Andreas Lehnert, Shane Sherlund, and Paul S. Willen. 2008. "Making Sense of the Subprime Crisis." *Brookings Papers on Economic Activity* 39 (2):69–159.

Gerardi, Kristopher and Paul S. Willen. 2009. "Subprime Mortgages, Foreclosures, and Urban Neighborhoods." *B.E. Journal of Economic Analysis & Policy* 9.

Henriques, Alice M. and Joanne W. Hsu. 2014. "Analysis of Wealth Using Micro- and Macrodata: A Comparison of the Survey of Consumer Finances and Flow of Funds Accounts." In *Measuring Economic Sustainability and Progress*, *National Bureau of Economic Research Studies in Income and Wealth*, vol. 72, edited by Dale W. Jorgenson, J. Steven Landefeld, and Paul Schreyer, chap. 9. Chicago: University of Chicago Press, 245–274.

Internal Revenue Service. 2007. *Statistics of Income 2007: Individual Income Tax Returns (Publication 1304, Revised 07-2009).* Washington, DC. Available at `https://www.irs.gov/pub/irs-soi/07inalcr.pdf`.

Krusell, Per and Anthony A. Smith, Jr. 1998. "Income and Wealth Heterogeneity in the Macroeconomy." *Journal of Political Economy* 106 (5):867–896.

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

———. 2015a. "Household Debt and Defaults from 2000 to 2010: Facts from Credit Bureau Data." National Bureau of Economic Research Working Paper No. 21203.

Mian, Atif R. and Amir Sufi. 2015b. "Fraudulent Income Overstatement on Mortgage Applications during the Credit Expansion of 2002 to 2005." National Bureau of Economic Research Working Paper No. 20947.

Montalto, Catherine Phillips and Jaimie Sung. 1996. "Multiple Imputation In the 1992 Survey Of Consumer Finances." *Journal of Financial Counseling and Planning* 7:133–147.

Munnell, Alicia H., Geoffrey M. B. Tootell, Lynne E. Browne, and James McEneaney. 1996. "Mortgage Lending in Boston: Interpreting HMDA Data." *American Economic Review* 86 (1):25–53.

Pissarides, Christopher A. 2000. *Equilibrium Unemployment Theory.* Cambridge, MA: The MIT Press, second ed.

Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys.* Wiley Series in Probability and Mathematical Statistics. New York: John Wiley & Sons.

Segal, Lewis M. and Daniel G. Sullivan. 1998. "Trends in Homeownership: Race, Demographics, and Income." *Economic Perspectives* 22 (1998-2):53–72. Federal Reserve Bank of Chicago. Available at `https://www.chicagofed.org/publications/economic-perspectives/1998/2qepart4`.

Simsek, Alp. 2013. "Belief Disagreements and Collateral Constraints." *Econometrica* 81 (1):1–53.

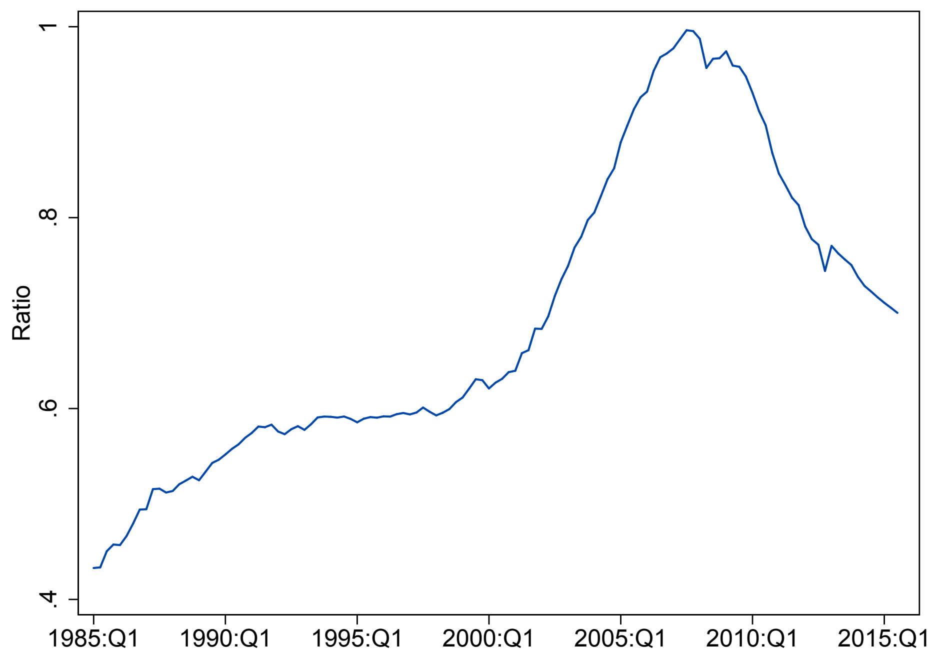



**Figure 1.** THE RATIO OF OUTSTANDING MORTGAGE DEBT IN THE UNITED STATES TO PERSONAL DISPOSABLE INCOME. *Source:* Board of Governors of the Federal Reserve System (Flow of Funds).

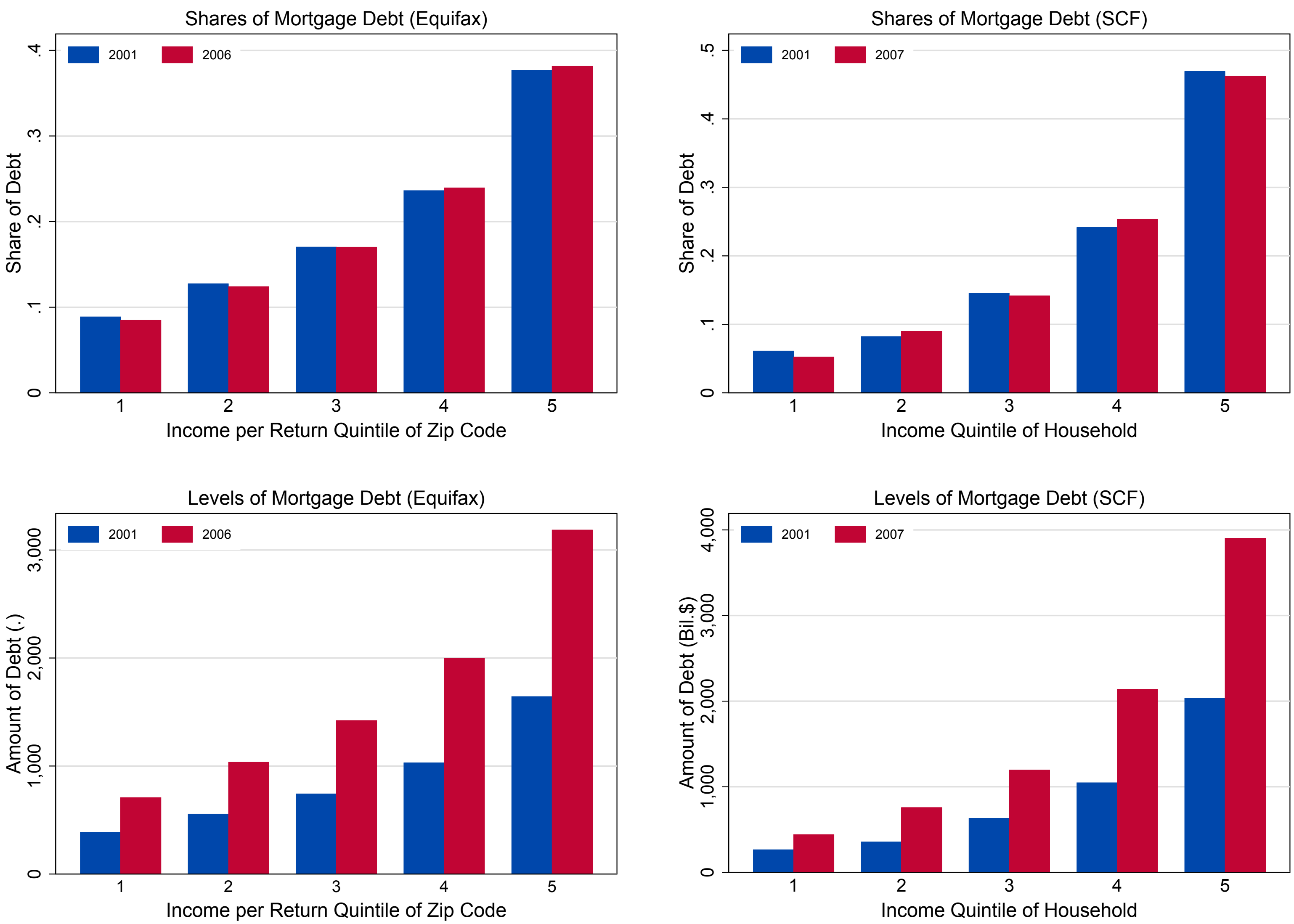



**Figure 2.** Distributions of Mortgage Debt with respect to Income among U.S. Zip Codes (left panels) and Households (right panels). *Note:* For zip codes, mortgage debt data are from the Equifax credit bureau and income is defined as aggregate IRS wage and salary income per tax return. For individual households, both mortgage debt and wage income are from the Survey of Consumer Finances. Households with no wage income in the SCF and zip codes with no reported wage and salary income from the IRS are not included in the distributions. *Source:* NY Fed Consumer Credit Panel/Equifax, IRS Statistics of Income, and Survey of Consumer Finances.

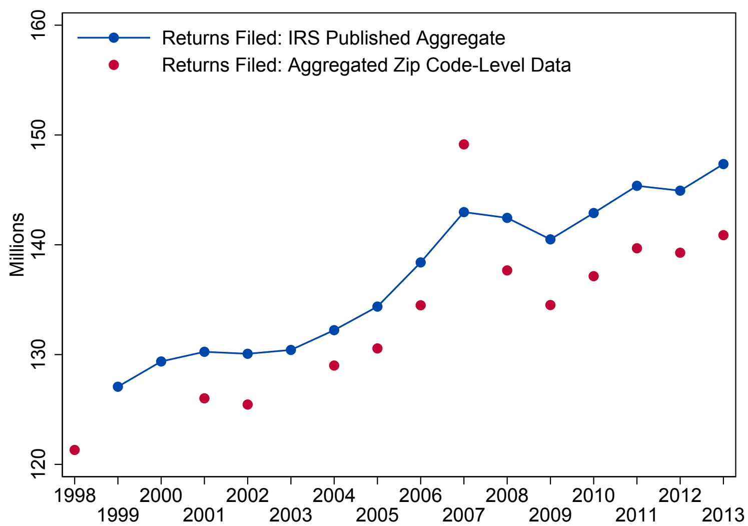



**Figure 3.** Two Measures of Aggregate Individual Income Returns Filed. *Note:* The blue line depicts the total number of individual income returns filed for the given tax year as published by the IRS. The 2007 value for this series omits returns filed by individuals for the sole purpose of receiving the 2007 economic stimulus payment. The red dots depict annual aggregates implied by the ZIP-level IRS data; the 2007 value for this series includes all filers. *Source:* Internal Revenue Service, Statistics of Income Historical Table 1 (available at `https://www.irs.gov/uac/SOI-Tax-Stats-Historical-Table-1`), Internal Revenue Service (2007).

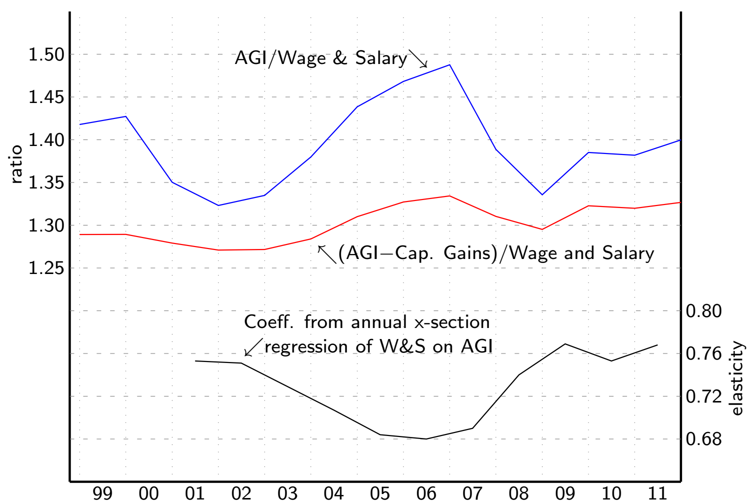



**Figure 4.** The Relationship Between Adjusted Gross Income (AGI) and Wage and Salary Income (W&S). *Note:* The blue line at top is the ratio of aggregate AGI to aggregate wage and salary income. The red line is the same ratio after capital gains are excluded from AGI. The black line at bottom depicts coefficients from yearly cross-sectional regressions of wage and salary income on AGI at the zip code-level. Aggregates are generated by summing zip code-level data. *Source:* IRS Statistics of Income.

| | Income per Return Quintile | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| **2001** | | | | | |
| Zip Codes (#,000) | 17 | 9 | 6 | 4 | 4 |
| S&W per Return ($,000) | 22 | 26 | 31 | 38 | 52 |
| AGI per Return ($,000) | 28 | 34 | 39 | 49 | 71 |
| Avg. Mortgage Debt ($,000) | 50 | 60 | 73 | 90 | 127 |
| Avg. 1st Mortgage ($,000) | 43 | 50 | 60 | 74 | 99 |
| Avg. 2nd Mortgage ($,000) | 6 | 5 | 4 | 4 | 4 |
| Avg. HELOC ($,000) | 2 | 3 | 3 | 3 | 5 |
| Mortgaged Households (%) | 28 | 35 | 40 | 46 | 52 |
| Median Age | 45 | 45 | 44 | 44 | 45 |
| Median Risk Score | 657 | 684 | 700 | 721 | 742 |
| Median House Price ($,000) | 79 | 97 | 119 | 156 | 243 |
| **2006** | | | | | |
| Zip Codes (#,000) | 16 | 8 | 6 | 4 | 4 |
| S&W per Return ($,000) | 24 | 30 | 35 | 42 | 59 |
| AGI per Return ($,000) | 32 | 39 | 46 | 57 | 87 |
| Avg. Mortgage Debt ($,000) | 73 | 87 | 111 | 145 | 212 |
| Avg. 1st Mortgage ($,000) | 58 | 66 | 82 | 101 | 138 |
| Avg. 2nd Mortgage ($,000) | 10 | 8 | 10 | 11 | 13 |
| Avg. HELOC ($,000) | 6 | 8 | 10 | 12 | 19 |
| Mortgaged Households (%) | 32 | 41 | 46 | 52 | 59 |
| Median Age | 47 | 47 | 47 | 46 | 47 |
| Median Risk Score | 656 | 689 | 707 | 729 | 754 |
| Median House Price ($,000) | 133 | 148 | 189 | 249 | 390 |
| House Price Apprec. 2001–2006 | 51 | 41 | 42 | 43 | 44 |

**Table 1.** SUMMARY STATISTICS FOR ZIP CODES IN THE NEW YORK CONSUMER CREDIT PANEL. *Note:* Values at the zip code-level are summarized by return-weighted salary and wages per return quintiles from the IRS, meaning that there are approximately the same number of returns in each quintile. The reported values are the return-weighted median within each quintile. Average mortgage debt is the total stock of mortgage debt divided by the number of people holding a mortgage, after correcting for joint mortgages. The average value of each type of mortgage is the total stock of debt of that mortgage type divided by the number of outstanding mortgages of that type in each zip code. The percentage of mortgaged households is the number of couples or individuals holding a mortgage divided by the number of returns from the IRS. The median house price is from Zillow, and house price appreciation at the zip code-level is calculated from the CoreLogic zip code-level house price index. *Source:* NY Fed Consumer Credit Panel/Equifax; IRS Statistics of Income; Zillow.

| | | | | | Debt on Primary Residence | | | | | | Real Estate Assets | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Year | Income Quintile | No. of Unweighted Obs. | Mortgaged Households (% of Hholds) | Total Mortgage Debt | Total | Non-HELOC | HELOC | Other Mortgage Debt | Income | Home Ownership Rate (%) | Value of Primary Residence | Value of All Resid. Real Estate |
| Panel A: Income Defined as AGI (Zero Incomes Included) | | | | | | | | | | | | |
| 2001 | 1 | 683.6 | 14 | 5,294 | 5,219 | 5,090 | 129 | 75 | 10,167 | 41 | 31,051 | 32,175 |
| | 2 | 659.4 | 28 | 13,044 | 12,510 | 12,166 | 344 | 535 | 24,453 | 58 | 63,403 | 69,047 |
| | 3 | 719.6 | 47 | 30,539 | 28,670 | 28,196 | 474 | 1,869 | 41,142 | 67 | 81,736 | 91,029 |
| | 4 | 705.2 | 64 | 55,464 | 52,439 | 51,451 | 988 | 3,025 | 66,705 | 82 | 136,648 | 153,193 |
| | 5 | 1,674.2 | 79 | 122,314 | 110,457 | 105,987 | 4,470 | 11,858 | 211,252 | 93 | 311,906 | 389,058 |
| 2007 | 1 | 664.2 | 15 | 10,795 | 9,661 | 9,321 | 340 | 1,134 | 12,690 | 41 | 56,960 | 64,258 |
| | 2 | 616.8 | 32 | 22,170 | 20,809 | 19,686 | 1,123 | 1,360 | 28,977 | 56 | 87,176 | 95,520 |
| | 3 | 648.8 | 52 | 56,299 | 54,035 | 52,965 | 1,070 | 2,264 | 47,872 | 70 | 137,874 | 152,275 |
| | 4 | 685.6 | 72 | 106,882 | 96,519 | 92,614 | 3,905 | 10,363 | 77,131 | 84 | 227,398 | 263,010 |
| | 5 | 1,801.6 | 81 | 219,228 | 184,652 | 174,227 | 10,425 | 34,576 | 257,914 | 94 | 537,018 | 714,545 |
| Panel B: Income Defined as Salary and Wages (Zero Incomes Excluded) | | | | | | | | | | | | |
| 2001 | 1 | 590.8 | 26 | 15,581 | 14,953 | 14,451 | 502 | 628 | 10,713 | 44 | 52,955 | 66,497 |
| | 2 | 556.2 | 34 | 22,068 | 19,777 | 19,380 | 396 | 2,291 | 27,045 | 49 | 56,305 | 63,681 |
| | 3 | 563.2 | 55 | 38,015 | 35,834 | 35,159 | 676 | 2,180 | 43,065 | 69 | 85,389 | 94,221 |
| | 4 | 581.8 | 70 | 63,617 | 59,797 | 58,578 | 1,219 | 3,820 | 67,481 | 81 | 127,975 | 142,824 |
| | 5 | 1,091.8 | 84 | 127,374 | 117,407 | 112,853 | 4,553 | 9,968 | 168,217 | 92 | 292,371 | 349,305 |
| 2007 | 1 | 550.4 | 26 | 25,039 | 23,108 | 21,831 | 1,278 | 1,930 | 11,717 | 43 | 78,536 | 86,943 |
| | 2 | 522.8 | 39 | 43,960 | 39,013 | 37,286 | 1,727 | 4,948 | 30,618 | 52 | 93,208 | 106,750 |
| | 3 | 503.6 | 62 | 70,726 | 67,731 | 64,087 | 3,645 | 2,995 | 49,518 | 72 | 153,338 | 176,989 |
| | 4 | 560.0 | 77 | 123,914 | 110,677 | 106,423 | 4,255 | 13,237 | 77,340 | 83 | 234,546 | 271,525 |
| | 5 | 1,117.2 | 87 | 231,376 | 199,942 | 191,041 | 8,901 | 31,434 | 197,649 | 94 | 501,207 | 655,585 |

**Table 2.** Summary Statistics for Households in the Survey of Consumer Finances. *Note:* All variables are calculated as simple means of weighted averages of the five multiple implicates of the public-use Combined Extract Data. Figures are nominal dollar values unless otherwise noted. *Source:* Combined Extract Dataset of the Survey of Consumer Finances.

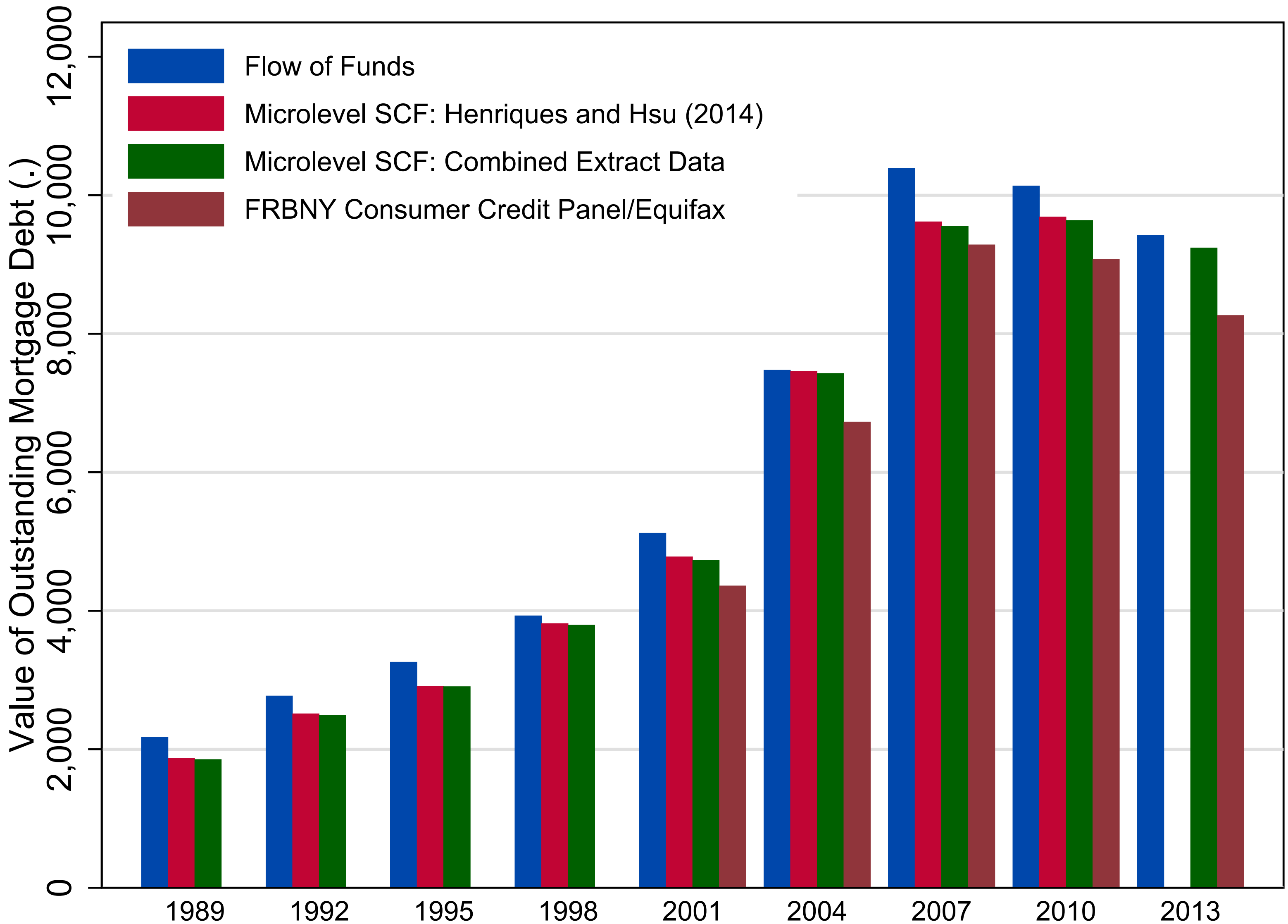



**Figure 5.** Alternative Measures of Aggregate U.S. Mortgage Debt. *Source:* Board of Governors of the Federal Reserve System (for Flow of Funds); Table 9.1 (p. 250) of Henriques and Hsu (2014); authors' calculations using the Combined Extract Data of the Survey of Consumer Finances; and authors' calculations using the NY Fed Consumer Credit Panel/Equifax.

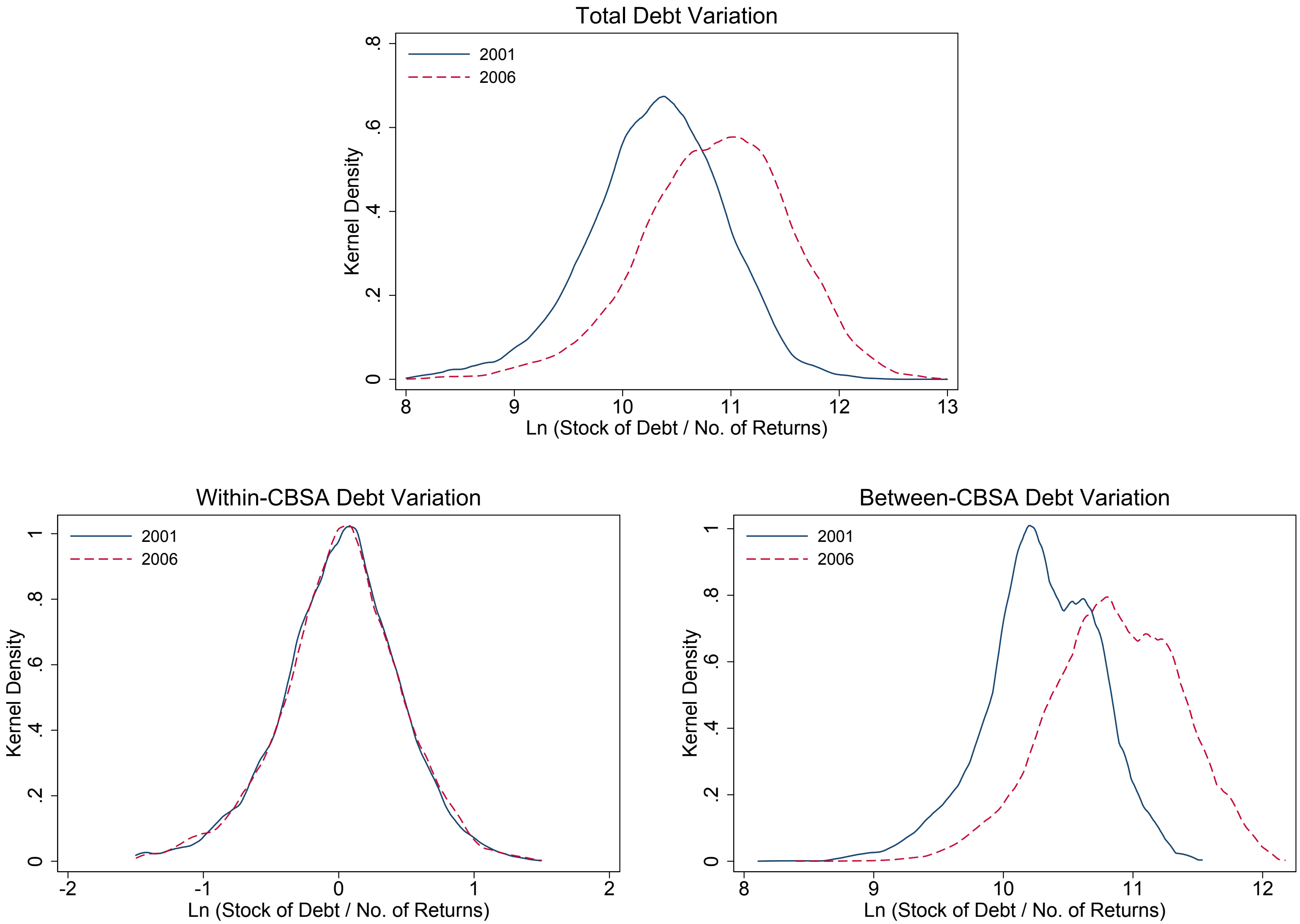



**Figure 6.** Densities of Zip Code-Level Mortgage Debt. *Note:* The top panel depicts the 2001 and 2006 returns-weighted kernel densities of (the log of) total zip code-level mortgage debt divided by the number of tax returns in the zip code. The bottom left panel shows the densities after the log of zip code-level debt per return is deviated from means corresponding to Core Based Statistical Areas (CBSAs). The bottom right panel depicts the kernel densities of CBSA averages of debt. *Source:* NY Fed Consumer Credit Panel/Equifax and IRS Statistics of Income.

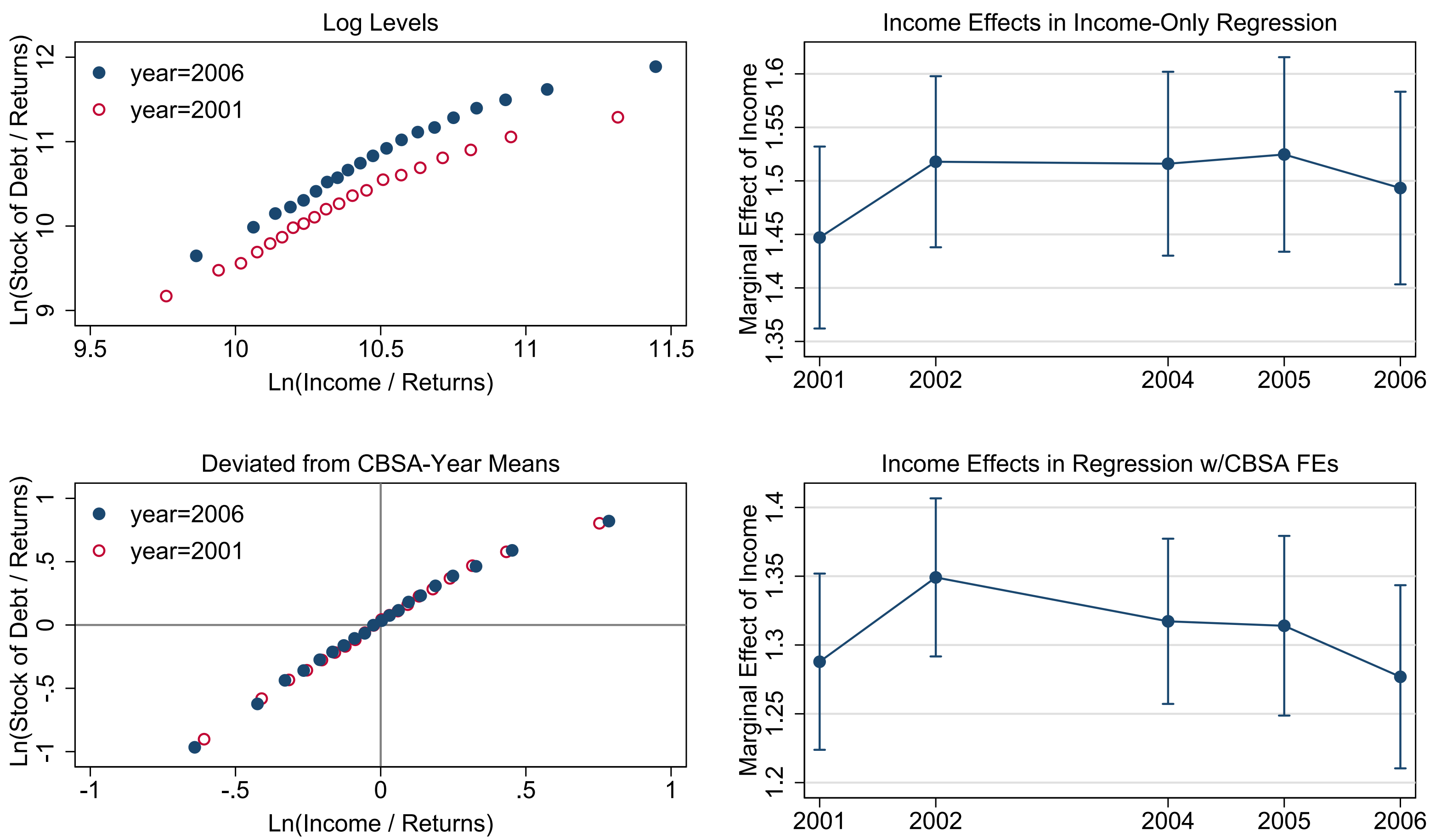



**Figure 7.** The Relationship between Levels of Mortgage Debt and Levels of Wage and Salary Income across U.S. Zip Codes. *Note:* The top left panel is a binned scatter plot of zip code-level debt and wage and salary income in 2001 and 2006. Both debt and income are expressed as natural logs of per-return values. The top right panel graphs the income coefficients (and 95-percent confidence intervals) from the implied returns-weighted regression of mortgage debt on income for all years between 2001 and 2006, save for 2003 (when IRS income data are not available). Coefficients are generated from a single pooled regression that includes interactions of the income variable with yearly dummies; and standard errors are clustered by county (not county-year). The lower left panel is a binned scatter plot of zip code-level debt and income after both variables have been deviated from returns-weighted CBSA-year means. The lower right panel depicts the income coefficients from a returns-weighted debt regression that includes CBSA × year fixed effects as well as income × year interactions. Standard errors are clustered by CBSA. *Source:* NY Fed Consumer Credit Panel/Equifax and IRS Statistics of Income.

Dependent Variable: 2001–2006 Zip Code-Level Change in Ln Mortgage Debt per Return

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Sample Restriction | None | None | 1% Trim | 5% Trim |
| | Panel A: All Zip Codes | | | |
| 2001-06 Change in Ln Income per Return | 0.930*** | 0.951*** | 1.054*** | 0.857*** |
| | (0.051) | (0.051) | (0.066) | (0.066) |
| 2001 Ln Income per Return Level | | 0.064*** | 0.071*** | 0.084*** |
| | | (0.011) | (0.011) | (0.013) |
| Constant | 0.532*** | 0.532*** | 0.532*** | 0.532*** |
| | (0.007) | (0.007) | (0.007) | (0.007) |
| R-sq. | 0.073 | 0.081 | 0.098 | 0.070 |
| Observations (No. of Zip Codes) | 35,710 | 35,710 | 27,564 | 18,251 |
| Expected Debt Growth: | | | | |
| 90th 2001 Income Pctile | | 0.563 | 0.566 | 0.572 |
| 10th 2001 Income Pctile | | 0.506 | 0.503 | 0.498 |
| Difference | | 0.057 | 0.063 | 0.074 |
| | Panel B: CBSA Zip Codes without Fixed Effects | | | |
| 2001-06 Change in Ln Income per Return | 0.945*** | 0.965*** | 1.063*** | 0.825*** |
| | (0.053) | (0.054) | (0.070) | (0.068) |
| 2001 Ln Income per Return Level | | 0.077*** | 0.075*** | 0.086*** |
| | | (0.012) | (0.012) | (0.014) |
| Constant | 0.528*** | 0.528*** | 0.530*** | 0.530*** |
| | (0.008) | (0.008) | (0.008) | (0.007) |
| R-sq. | 0.088 | 0.100 | 0.111 | 0.072 |
| Observations (No. of Zip Codes) | 27,664 | 27,664 | 21,630 | 14,993 |
| Expected Debt Growth: | | | | |
| 90th 2001 Income Pctile | | 0.564 | 0.565 | 0.570 |
| 10th 2001 Income Pctile | | 0.497 | 0.500 | 0.496 |
| Difference | | 0.067 | 0.065 | 0.075 |
| | Panel C: CBSA Zip Codes with CBSA Fixed Effects | | | |
| 2001-06 Change in Ln Income per Return | 0.906*** | 0.905*** | 1.017*** | 0.921*** |
| | (0.056) | (0.056) | (0.064) | (0.076) |
| 2001 Ln Income per Return Level | | 0.006 | −0.012 | −0.013 |
| | | (0.015) | (0.015) | (0.012) |
| Constant | 0.528*** | 0.528*** | 0.529*** | 0.529*** |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| R-sq. | 0.37 | 0.372 | 0.495 | 0.525 |
| Observations (No. of Zip Codes) | 27,664 | 27,664 | 21,630 | 14,993 |
| Expected Debt Growth: | | | | |
| 90th 2001 Income Pctile | | 0.531 | 0.524 | 0.525 |
| 10th 2001 Income Pctile | | 0.526 | 0.534 | 0.534 |
| Difference | | 0.005 | −0.010 | −0.011 |

Standard errors in parentheses

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

**Table 3.** Long-Difference Regressions. *Note:* The 2001–2006 change in log income per return (top coefficient in each panel) and the 2001 log level of income per return (second coefficient) are each deviated from sample means, so the corresponding constant terms can be interpreted as the expected value of the dependent variable when both the income-growth and the income-level covariates are equal to their sample means. All regressions are weighted by the number of zip code-level tax returns in 2001. The CBSA fixed effects in Panel C are constrained to have a weighted mean of zero. The 1 percent trimmed samples deletes outliers above the 99th percentiles of the distributions of debt growth, income growth, and 2001 income level, as well as observations below the 1st percentiles of those three distributions. The 5 percent trims are defined analogously. The percentile cutoffs are calculated from returns-weighted distributions; because returns are distributed unequally across zip codes, these trims delete higher fractions of zip codes. Standard errors are clustered by county in Panels A and B and by CBSA in Panel C. *Source:* NY Fed Consumer Credit Panel/Equifax and IRS Statistics of Income.

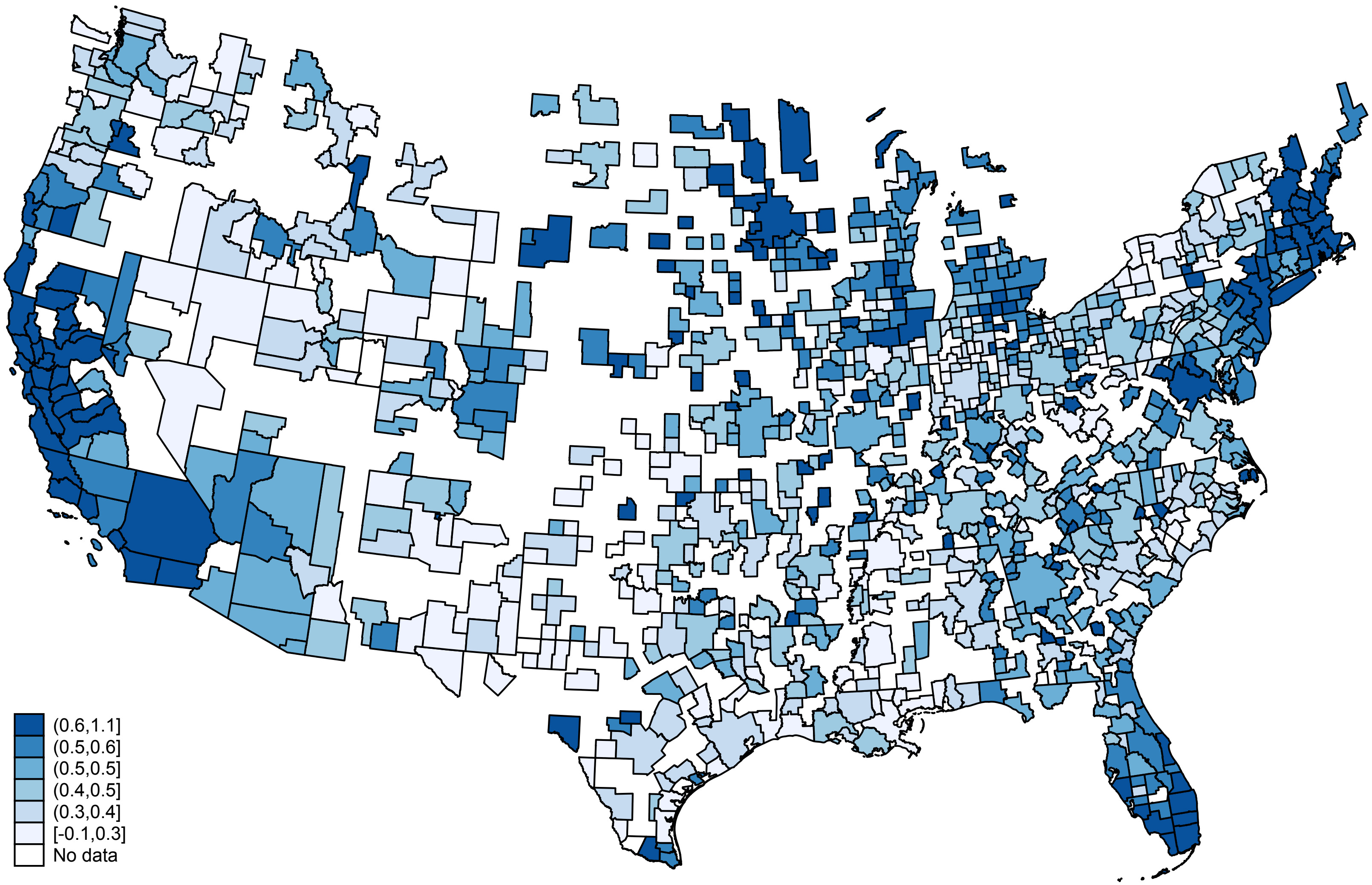



**Figure 8.** CBSA-Level Fixed Effects from Long-Difference Regression. *Note:* This map graphs the CBSA-level fixed effects estimated in the long-difference regression in the third column of Panel C in Table 3. The estimated constant has been added back to these fixed effects, because these effects are constrained in the regression to have a mean of zero. *Source:* NY Fed Consumer Credit Panel/Equifax and IRS Statistics of Income.

Dependent Variable: Estimated CBSA-Level Fixed Effects from Zip Code-Level 2001–2006 Long-Difference Debt Regressions

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| 2001 Ln CBSA Income Level | | $0.25^{***}$ | $0.28^{***}$ | | $0.23^{***}$ | $0.16^{***}$ |
| | | (0.04) | (0.04) | | (0.03) | (0.03) |
| 2001-06 Change in Ln CBSA Income | | | 0.38 | | | $-0.81^{**}$ |
| | | | (0.22) | | | (0.26) |
| 2001-06 Change in Ln CBSA House Price | | | | $0.30^{***}$ | $0.29^{***}$ | $0.36^{***}$ |
| | | | | (0.03) | (0.03) | (0.04) |
| Constant | $0.53^{***}$ | $0.53^{***}$ | $0.53^{***}$ | $0.41^{***}$ | $0.41^{***}$ | $0.38^{***}$ |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| Observations (No. of CBSAs) | 932 | 932 | 932 | 932 | 932 | 932 |
| R-sq. | 0.000 | 0.140 | 0.152 | 0.266 | 0.385 | 0.422 |

Standard errors in parentheses

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

**Table 4.** Determinants of CBSA-Level Fixed Effects from the 1 Percent Trimmed Long-Difference Regression. *Note:* The dependent variable for each regression is the estimated CBSA fixed effect from the zip code-level long-difference regression reported in Column 3 of Panel C in Table 3. Because the fixed effects in this regression are constrained to have a weighted mean of zero, the constant term is added back to the fixed effects before the regressions are run. All regressions are weighted by the number of tax returns in the CBSA in 2006. The CBSA-level average income variables in the top two rows are deviated from their sample means before running the regressions. The house price variable in the third row is the log change in the CoreLogic CBSA house price from 2001 to 2006 (and is not deviated from its sample mean). See the text for details of how CoreLogic determines this index when there are few available repeat-sales transactions. *Source:* NY Fed Consumer Credit Panel/Equifax, IRS Statistics of Income, and CoreLogic.

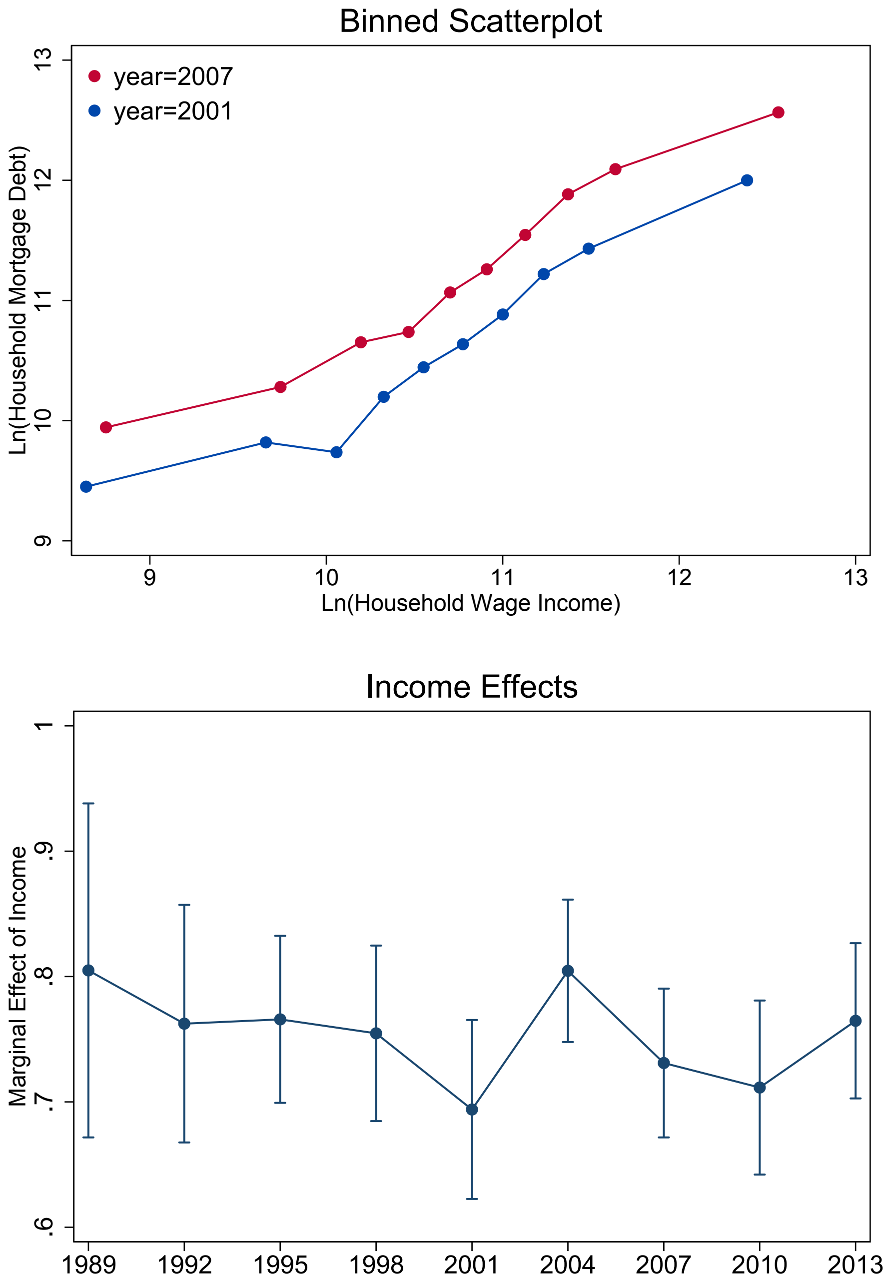



**Figure 9.** The Relationship between Household-Level Mortgage Debt and Wage and Salary Income in the Survey of Consumer Finances. *Note:* The top panel is a binned scatter plot of total household-level mortgage debt and wage and salary income in the SCF, calculated as averages across the five sample implicates in the Combined Extract Data. Both debt and income are expressed as natural logs, and households with no wage and salary income are excluded. The bottom panel depicts income coefficients from a pooled Poisson regression for household debt, in which (the log of) wage and salary income, dummies for the age of the household head (younger than 35, 35–44, 45–54 and 55–64), the number of children, and dummies for nonwhite and marital status are each interacted with yearly dummies. Households with heads 65 and older and households with no wage income are excluded. The reported coefficients are averages of estimates using the five implicates in the Combined Extract Data. Standard errors are calculated as in Rubin (1987) but with no degrees-of-freedom adjustment. *Source:* Combined Extract Data of the SCF.

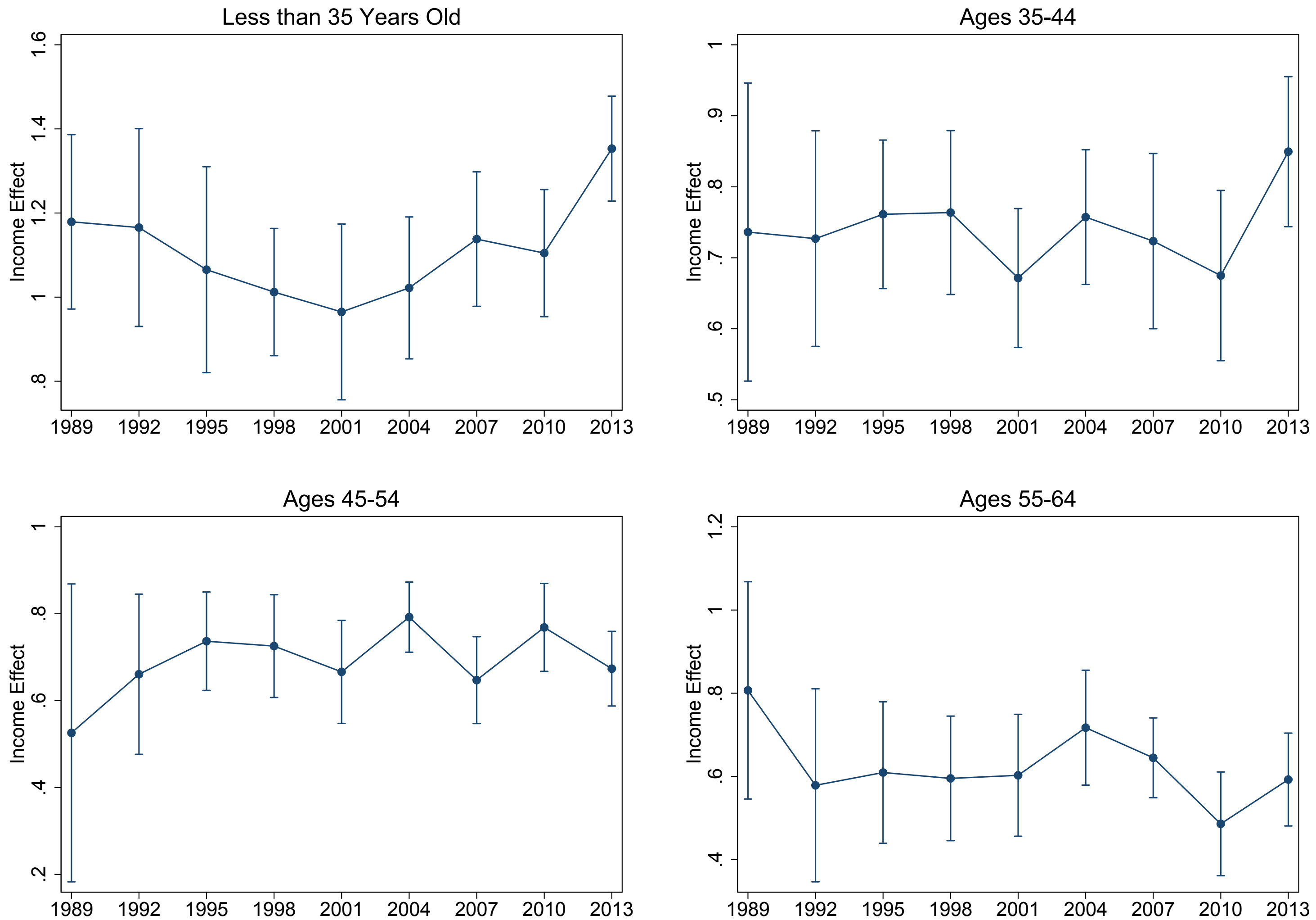



**Figure 10.** The Relationship between Mortgage Debt and Income by Age of Household Head. *Note:* Each of the four panels above is generated by a single Poisson regression for household debt, in which the (the log of) wage and salary income is interacted with dummies for the age group of the household head as well as with yearly dummies. The specification is otherwise identical to the regression that generates the lower panel of Figure 9. See the notes to that table for additional details. *Source:* Combined Extract Data of the Survey of Consumer Finances.

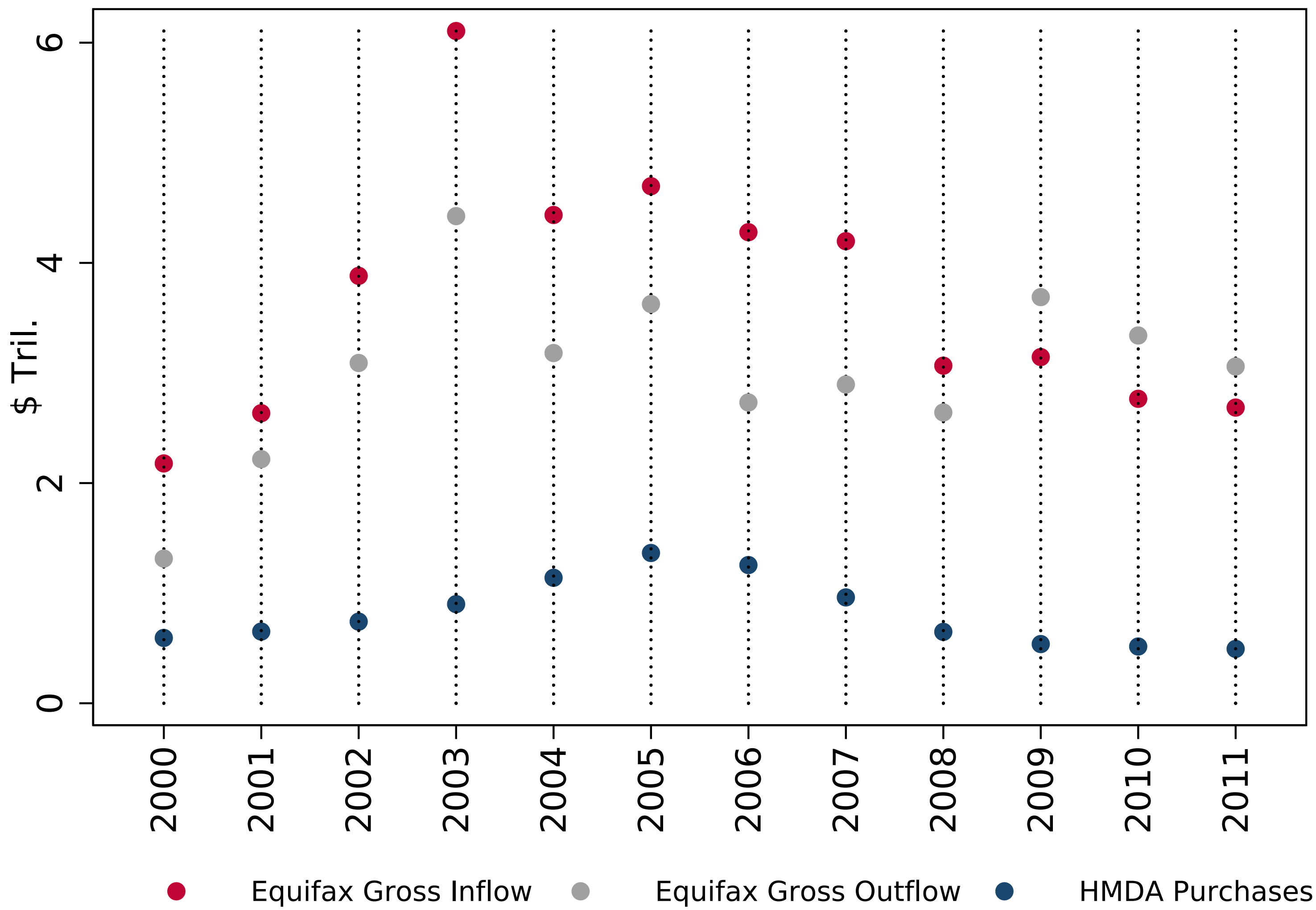
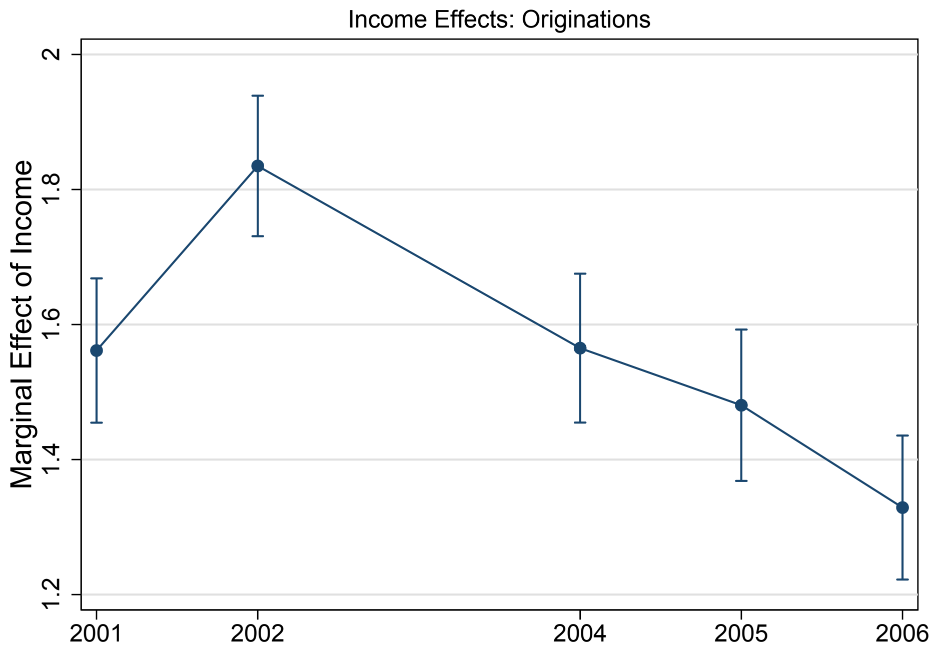



**Figure 11.** Gross Flows of Mortgage Debt. *Note:* This graph plots the gross inflows and outflows of mortgage debt from Equifax, along with the value of purchase mortgage originations from HMDA. The inflows and outflows are calculated using the individual tradelines in Equifax: any origination or increase in balance is counted as an inflow, while any termination or decrease in balance is counted as an outflow. *Source:* NY Fed Consumer Credit Panel/Equifax and the Home Mortgage Disclosure Act.



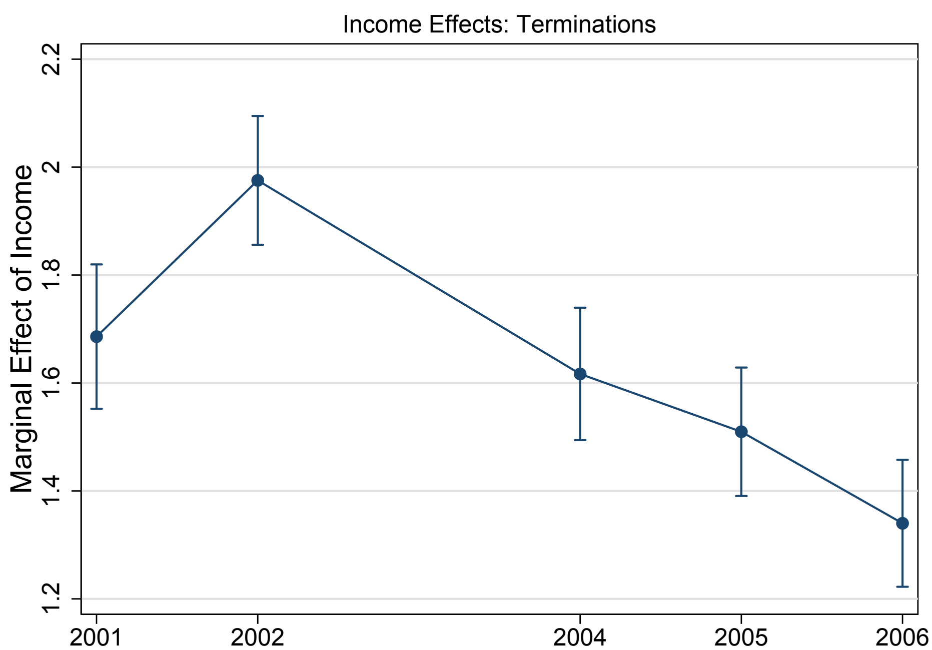




**Figure 12.** Income Effects for Originations and Terminations. *Note: Source:* NY Fed Consumer Credit Panel/Equifax and Internal Revenue Service Statistics of Income.

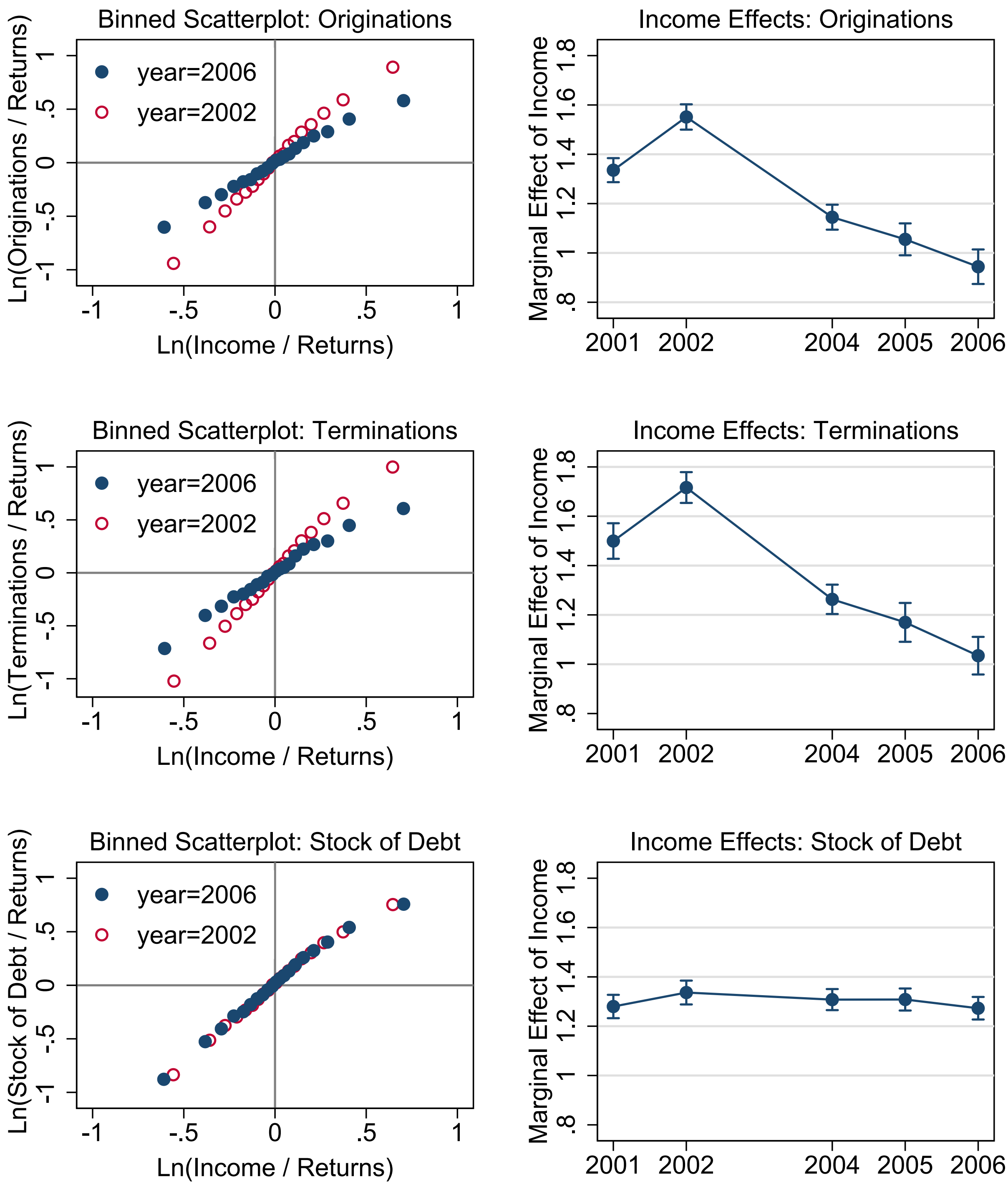



**Figure 13.** The Relationship Between Originations, Terminations, and the Stock of Mortgage Debt in County-Deviated Data. *Note:* The binned scatter plots and estimated income effects above are calculated similarly to those in the bottom row of Figure 7. Here, the scatter plots are deviated from county rather than CBSA means and the data correspond to mortgage originations (top row), mortgage terminations (middle row), and the total stock of mortgage debt (bottom row). The income coefficients depicted in the panels at right are generated by pooled returns-weighted regressions of the relevant variable on income × year interactions and county × year fixed effects. Standard errors are clustered by county (not county-year). *Source:* NY Fed Consumer Credit Panel/Equifax and IRS Statistics of Income.

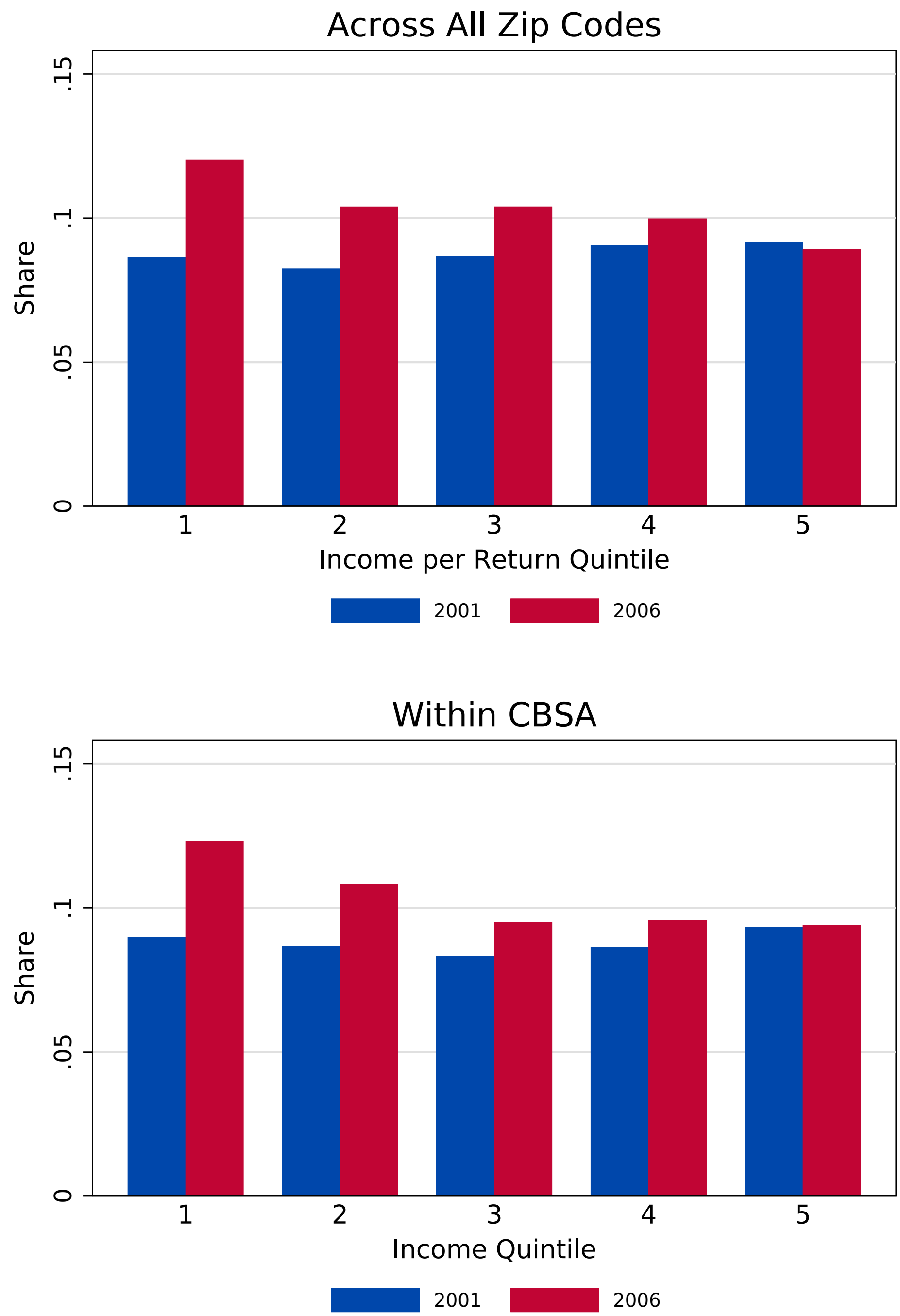



**Figure 14.** Purchase-Mortgage Intensity by Income Quintile of Zip Code: 2001 and 2006. *Note:* The ratio of new purchase mortgages (as measured in HMDA) to outstanding first liens (as measured in the Equifax data) can be thought of as a measure of purchase-mortgage intensity. *Source:* Home Mortgage Disclosure Act (for mortgage originations), NY Fed Consumer Credit Panel/Equifax (for outstanding first liens), and IRS Statistics of Income.

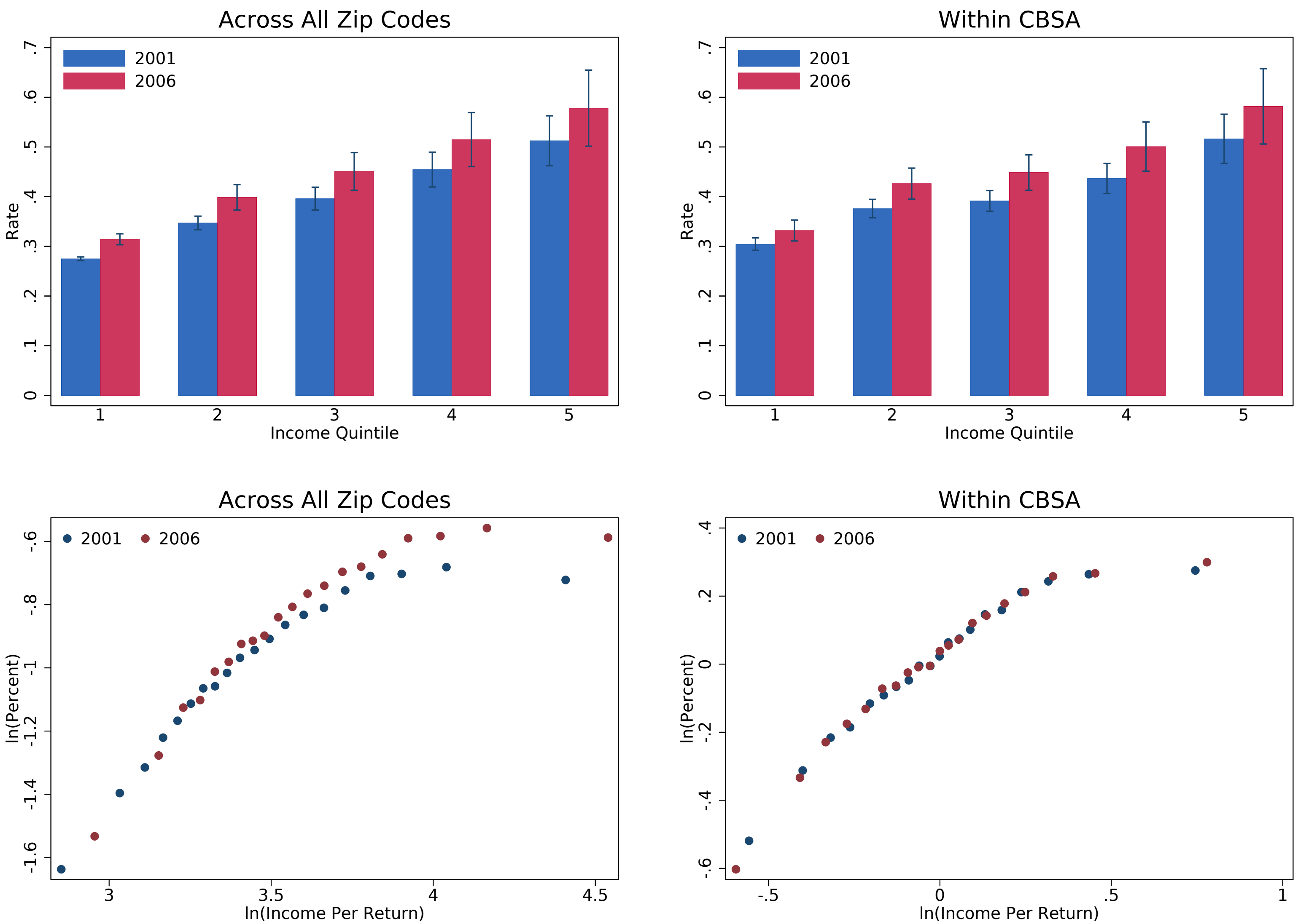



**Figure 15.** Proportion of Mortgaged Households. *Note:* The top two panels show estimates of the proportion of mortgaged households both across all zip codes and within CBSAs. The error bars represent two methods of calculating this proportion, with the height of the bars being the mean of the two. The upper estimate is the number of outstanding first liens in the CCP after adjusting for joint mortgages divided by the number of IRS returns, while the lower estimate is the number of mortgaged "couples" in Equifax: the number of people with a mortgage, adjusting for any joint mortgages. The bottom two panels are binned scatter plots of the natural log of the proportion of mortgaged households on the natural log of income per return, both across all zip codes and within CBSAs. *Source:* NY Fed Consumer Credit Panel/Equifax and IRS Statistics of Income.

| | AGI Growth Rate | 2006 AGI Level | Wage & Salary Growth Rate | 2006 Wage & Salary Level |
|---|---|---|---|---|
| **Model 1** | | | | |
| HMDA Purchase Growth Rate | –.16** | | | |
| | (0.08) | | | |
| **Model 2** | | | | |
| HMDA Purchase Growth Rate | 0.88*** | –0.38*** | | |
| | (0.07) | (0.02) | | |
| **Model 3** | | | | |
| HMDA Growth Rate: Avg. Purchase Mortgage Amount | .29*** | –.04*** | | |
| | (.04) | (.01) | | |
| HMDA Growth Rate: No. of Purchase Mortgages | .60*** | –.34*** | | |
| | (.07) | (.02) | | |
| **Model 4a: Income as AGI** | | | | |
| Equifax Debt Stock / Tax Returns Growth Rate | 0.42*** | –.05*** | | |
| | (0.05) | (0.01) | | |
| **Model 4b: Income as AGI** | | | | |
| Equifax Growth Rate: Mortgage Debt per Mgted. Household | .17*** | –.03*** | | |
| | (.02) | (.01) | | |
| Equifax Growth Rate: Mortgaged Households / Tax Returns | .25*** | –.03*** | | |
| | (.05) | (.01) | | |
| **Model 5a: Income as Wage and Salary** | | | | |
| Equifax Debt Stock / Tax Returns Growth Rate | | | 0.74*** | –.02 |
| | | | (0.05) | (0.01) |
| **Model 5b: Income as Wage and Salary** | | | | |
| Equifax Growth Rate: Mortgage Debt per Mgted. Household | | | .22*** | –.01* |
| | | | (.04) | (.01) |
| Equifax Growth Rate: Mortgaged Households / Tax Returns | | | .51*** | –.01 |
| | | | (.06) | (.01) |

**Table 5.** Decomposing the Negative Correlation Between Growth in Purchase Mortgages and Income Growth: 2002–2006. *Note:* Model 1 replicates the negative sign of the correlation between purchase-mortgage growth and AGI growth reported in Mian and Sufi (2009). Model 2 adds the level term to the regression to show that relationship between purchase-mortgage originations and AGI was never negative in levels. Model 3 follows Adelino, Schoar, and Severino (2015b) by dividing purchase-mortgage originations into the average size of each purchase mortgage and the number of purchase mortgages. This regression shows a significant decline in the slope of the positive relationship between purchase-mortgage originations and income only when considering the number of purchase mortgages, not the average purchase amount. Model 4a considers the per-return stock of mortgage debt, while Model 4b divides this stock into debt per mortgaged household and the proportion of households that have a mortgage. Model 5a and 5b repeat this analysis, using salary and wages rather than AGI. *Source:* NY Fed Consumer Credit Panel/Equifax and IRS Statistics of Income.

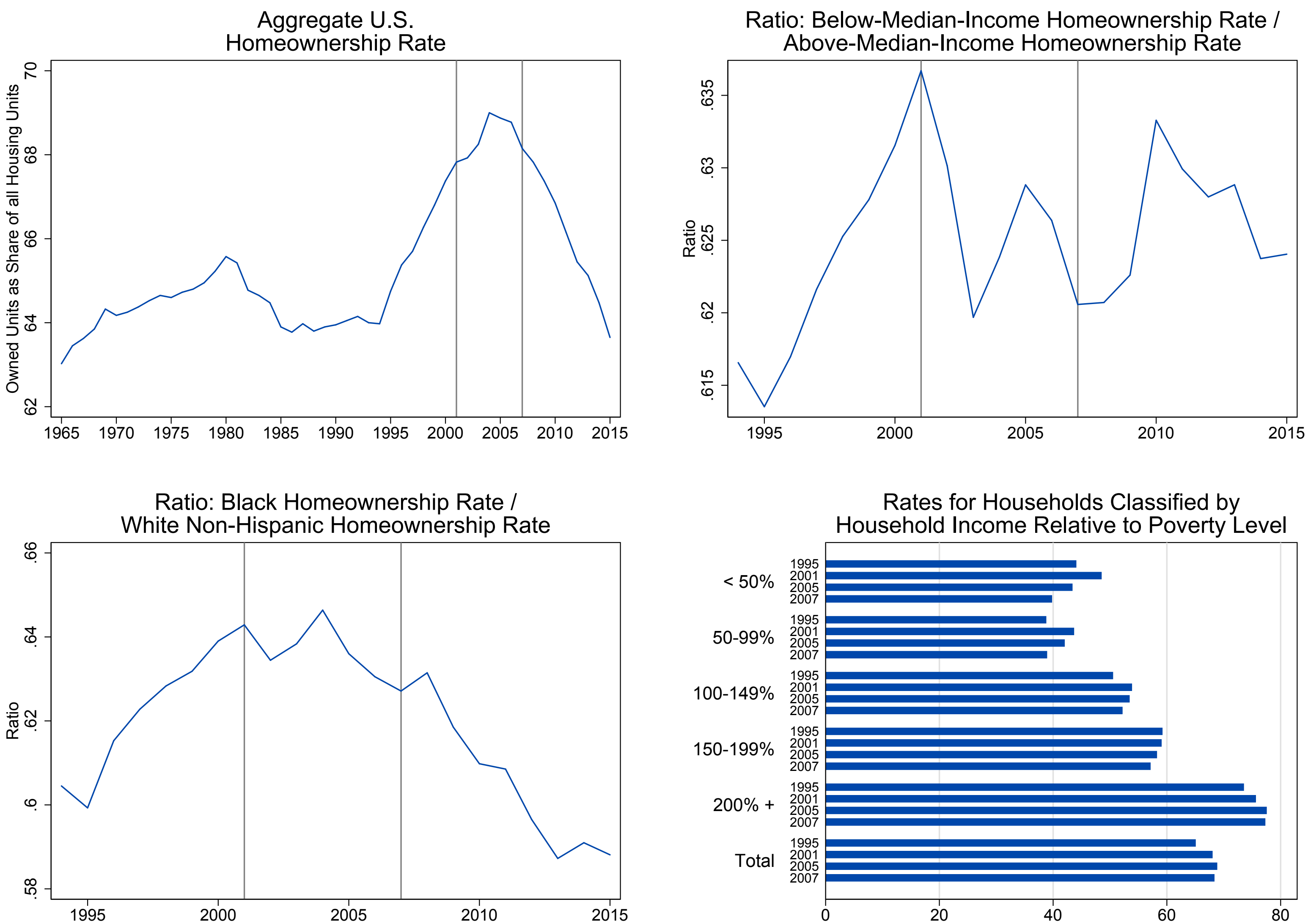



**Figure 16.** Homeownership Statistics Derived from Census Data. *Note:* The top left panel shows the nationwide homeownership rate over time. The grey vertical lines are drawn at 2001 and 2007, the start and end points of the housing boom that is the subject of this paper. The top right panel plots the ratio of the below-median-income homeownership rate to the above-median-income homeownership rate. This ratio did not increase over this time period, which would have happened if people below the median income had disproportionately increased their homeownership rate. The bottom left panel plots the ratio of the homeownership rate of black Americans to that of non-Hispanic white Americans. This also did not increase over the course of the housing boom. The bottom right panel plots the homeownership rates over time for households classified by their income relative to the poverty level, and again shows no increase in the rate for the lowest-income Americans. *Source:* Bureau of the Census.

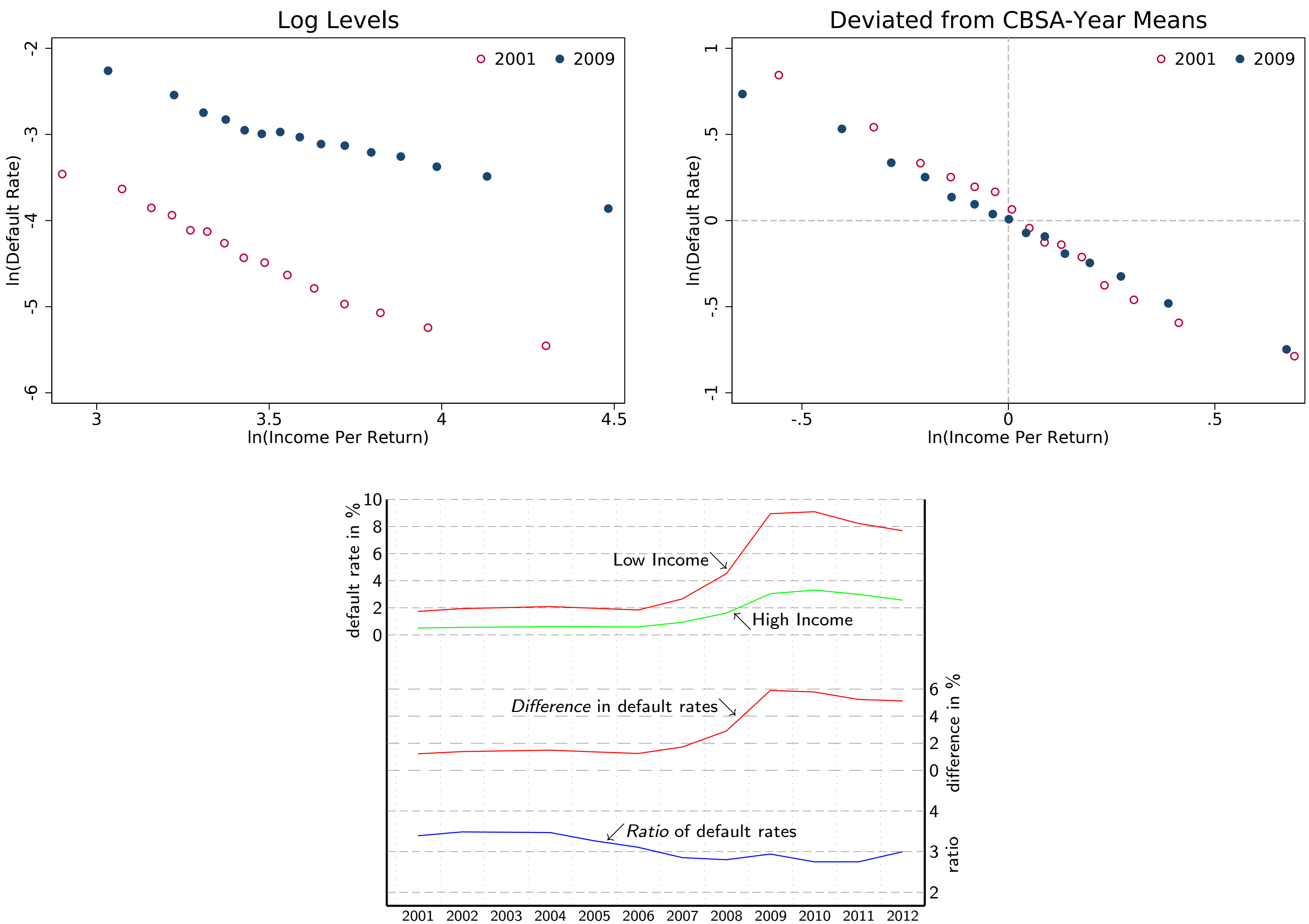
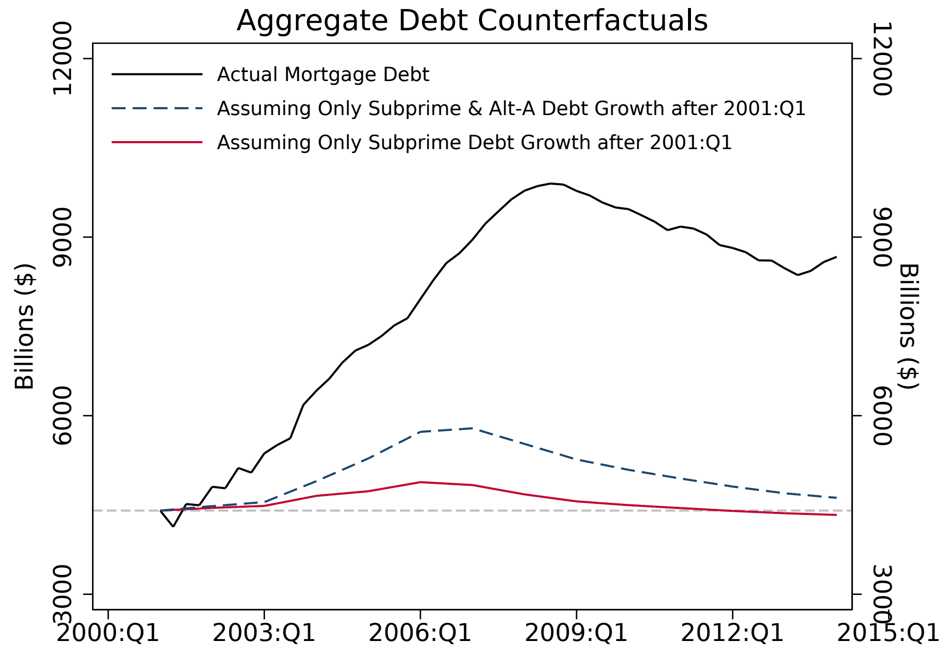



**Figure 17.** Relationships Between Income and Foreclosure Rates at the Zip Code-Level. *Note:* The two panels in the top row are binned scatter plots of foreclosure rates and salary and wage income for zip codes in 2001 and 2009. Both the foreclosure rates and income are expressed as natural logs of per-return values. In the top left panel, the variables are not deviated from CBSA means, while in the the top right panel they are. In the lower panel, the top two lines depict foreclosure rates for zip codes in the highest and lowest income quintiles within CBSA. Income is defined as salary and wage income. The middle line in the lower panel plots the difference in those two default rates, while the bottom line plots their ratio. *Source:* NY Fed Consumer Credit Panel/Equifax and IRS Statistics of Income.



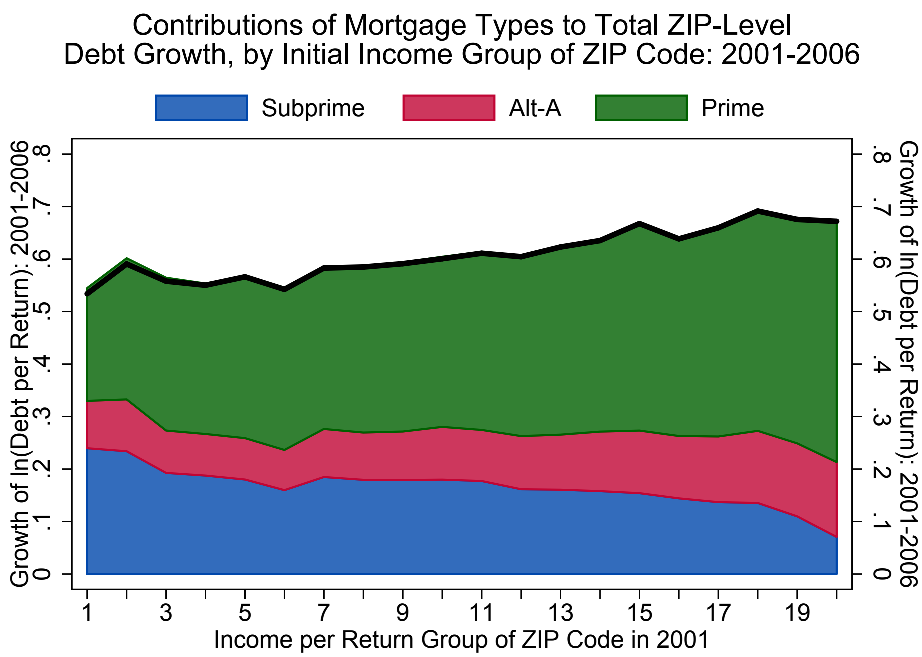




**Figure 18.** The Effect of Growth in Alternative Mortgage Products on Growth of Mortgage Debt. *Note:* In the top panel, the black line depicts U.S. aggregate mortgage debt as measured by the Equifax database. The solid red line shows how aggregate debt would have evolved if the only net additions to debt after 2001:Q1 had been securitized subprime mortgages, as measured by the Securities Industry and Financial Markets Association (SIFMA). The blue dashed line shows the counterfactual if net additions of privately securitized Alt-A debt from the SIFMA database had also been allowed. The bottom panel uses data on subprime and Alt-A mortgage debt from CoreLogic Private Label Securities Database to show the 2001–2006 contributions of prime, subprime, and Alt-A debt for total debt growth among individual ZIP codes, sorted into 20 income-per-return categories. *Source:* NY Fed Consumer Credit Panel/Equifax, Securities Industry and Financial Markets Association, CoreLogic Private Label Securities ABS Database, and IRS Statistics of Income.

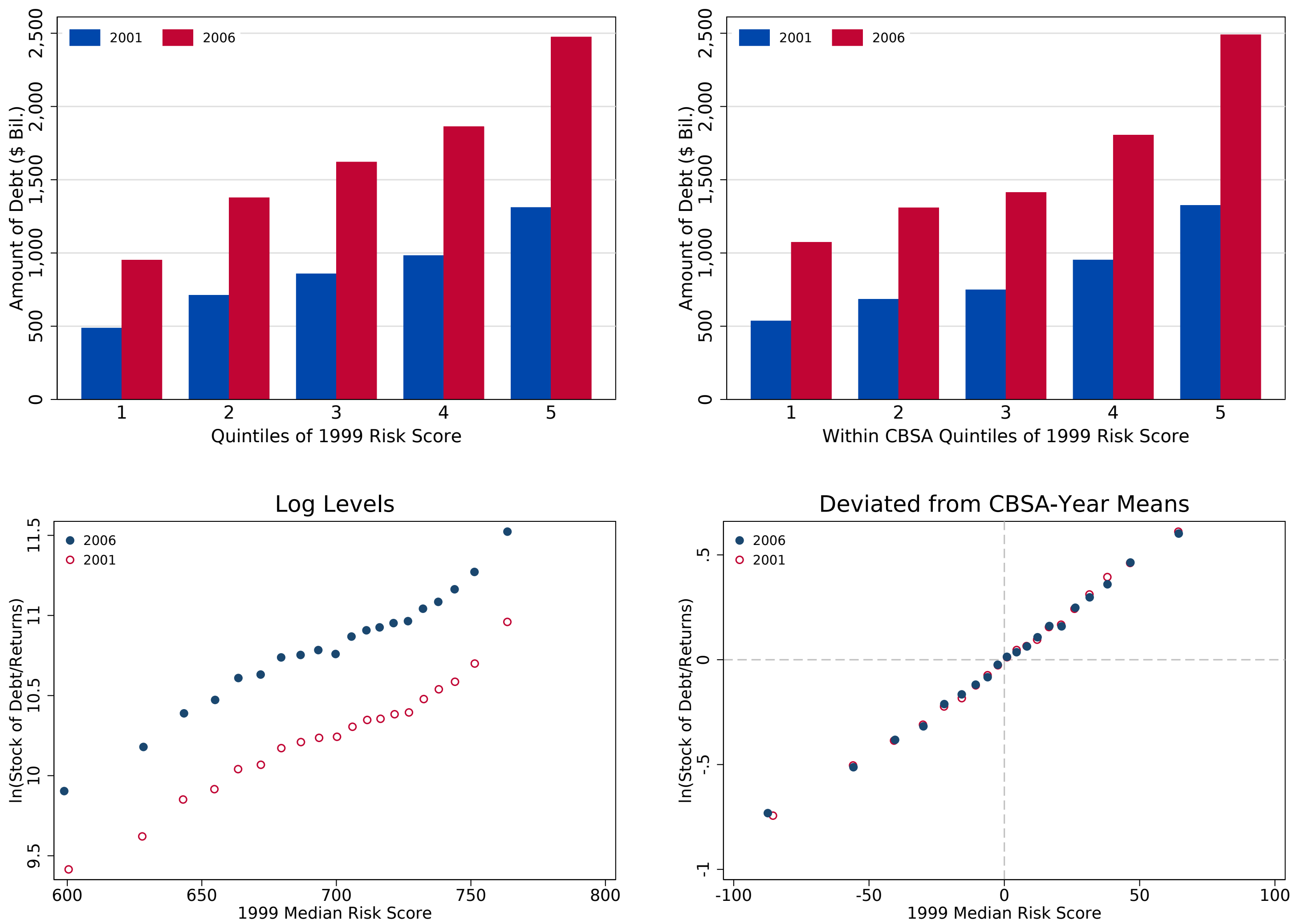



**Figure 19.** Stock of Mortgage Debt by Quintiles of 1999 Median Risk Score. *Note:* The two left panels are across all ZIP codes, while the two right panels are within CBSA. The top two graphs shows the increase in the total dollar value of debt for each quintile of risk score. Quintiles are created by calculating five return-weighted bings of the 1999 median risk score in each ZIP code and year from Equifax. The bottom two panels are binscatters where the 1999 median risk score has been separated into 20 return-weighted bins. The graph plots the average value of the log of the stock of mortgage debt for each bin. *Source:* NY Fed Consumer Credit Panel/Equifax and IRS Statistics of Income.

# A Internet Appendix

## A.1 Additional Comparisons of Aggregate Debt and Income Data

Figure A.1 compares two aggregations of individual-level Equifax mortgage-debt balances from the New York Fed Consumer Credit Panel. The level of aggregation is either the state or the county. The horizontal axes of each panel measure our aggregations of debt for the given geographical unit, while the vertical axes measure aggregates calculated and published by the New York Fed itself. In all panels, the dots lie along 45-degree lines, giving us confidence that we are aggregating up to the ZIP-code level correctly when we construct our main cross-sectional dataset.

Figure A.2 compares aggregates of IRS income data. The blue line is a national-level aggregate of either salary and wage income (top panel) or AGI (bottom panel) that is published by IRS. The red dots are aggregated data from the ZIP-level income dataset we use for cross-sectional analysis. In both panels, the published nationwide aggregate is larger than the ZIP-level aggregate. In large part, this discrepancy stems from the suppression rules that IRS applies to the ZIP-level dataset before they release it. Even so, the two aggregates follow similar time-series patterns.

## A.2 AGI vs. Wages and Salaries and 2006 vs. 2007

The main empirical work in this paper uses salary and wage income as the measure of income, because wage income is more likely to be used by lenders when they evaluate mortgage applications. An alternative choice would be to use AGI (for ZIP-level analysis) and total income (for SCF analysis). Figure A.3 shows that our main results go through even when this alternative choice is made. The two left panels use ZIP-level data from Equifax, with quintiles calculated using AGI per tax return. The two right panels use data from the SCF, with the quintiles calculated using the SCF's measure of total income. The top panels show the same similarity in debt evident in the introductory Figure 2 in both the ZIP-level data and the SCF. The lower panels replicate the finding that, because mortgage debt rises with income in the cross section, equal debt-growth rates imply very large dollar amounts of new debt for the richest borrowers.

The main text also uses 2006 rather than 2007 as the last year of the mortgage boom when performing ZIP-level analysis. (The ending-year issue is not relevant for the SCF.) This choice is necessitated by that spike in tax filing in 2007 that was illustrated in the text by Figure 3. As explained in the data section, this spike is driven by a surge in persons filing for the sole purpose of receiving economic stimulus payments in 2007. A previous appendix graph, Figure A.2, implies that the additional filers had very low incomes, because their tax returns had little effect on 2007 levels of total AGI or of wages and salaries. Further evidence that the extra filers had low incomes appears in Figure A.4. This figure shows that ZIP-level growth in the number of IRS returns filed in 2007 is not only much greater than in other years, but that 2007 growth covaries negatively and monotonically with ZIP-level income. As with the choice of income definition, however, the choice of ending year has little effect on the main results. Figure A.5 shows that using 2007 as the end of the boom for the ZIP-level distributions generates the same patterns seen in earlier figures.

## A.3 Debt Distributions Disaggregated by Lien Types

In the main text, we investigate debt patterns using all types of mortgage debt: first mortgages, second mortgages, and HELOCs. Figure A.6 disaggregates the analysis by lien type. For reference, we include as the upper left panel of this figure the overall debt distribution with respect to income that appeared in the same position of the introductory Figure 2. The top right panel of Figure A.6 shows the distribution of first-mortgage debt. Because the large majority of outstanding debt consists of first liens, it is not surprising that the first-lien distribution remains stable over time. The lower left panel presents distributions of closed-end second mortgages. Here there is a pronounced change in the distribution, with high-income ZIP codes receiving much higher shares of second-mortgage debt in 2006 relative to 2001. The last panel shows distributions of HELOC debt. There is a slight tilt toward higher debt shares among richer quintiles, but this tilt is not as severe as in the previous panel. In any case, none of the panels in Figure A.6 indicate a significant increase in the share of debt held by low-income quintiles. Figure A.7 performs the same analysis using AGI rather than salaries and wages, with similar results.

## A.4 Auxiliary Results Related to Long-Distance Regressions

Figure A.8 and Table A.1 present results related to the long-difference regressions in section 3. Figure A.8 illustrates the trimming of the sample in these regressions, by plotting distributions of income levels in 2001, income growth between 2001 and 2006, and the growth of the stock of mortgage debt from 2001 to 2006. Plots in the left column pertain to all ZIP codes, while plots in the right correspond to ZIP codes located within CBSA boundaries. All distributions are weighted by the number of 2001 returns and include vertical lines marking the first, fifth, 95th, and 99th percentiles. These percentiles are relevant for the alternative trims made in the long-difference regressions. Table A.1 shows how the long-difference regressions are affected by the choice of 2007 as the ending year of the mortgage boom, and is therefore the corollary of Table 3 in the text. As explained in footnote 28, the results in these tables are similar.

## A.5 Data on Non-Agency Mortgage Securities

Figure A.9 compares two measures of aggregate private-label mortgaged-backed securities outstanding, disaggregated by mortgage type. One measure is an aggregate estimate from the Securities Industry and Financial Markets Association (SIFMA), while the other is our summation of individual-level mortgages from the CoreLogic Private Label Securities ABS Database. The top panel shows that CoreLogic's coverage of the total amount of Alt-A mortgages packaged into private-label securities was very good throughout the housing boom, as the SIFMA aggregate and the implied CoreLogic aggregate are very close. The lower panel shows the CoreLogic's coverage of subprime mortgages improved over time. Table A.2 provides a complete breakdown of the components of private-label residential mortgage-backed securities outstanding, according to SIFMA. As noted in the text, securities backed by Alt-A mortgages grew fastest during the boom, although there was significant growth in bonds backed by subprime and prime jumbo mortgages as well.

## A.6 Effect of Subprime and Alt-A on Debt Growth: 2001-07

Figure A.10 replicates the main lessons of Figure 18 for debt growth, using 2007 as the ending year of the mortgage boom. Like Figure 18, which ends the boom in 2006, the appendix figure shows that the use of subprime mortgages grew more in low-income areas. Growth in Alt-A and prime mortgage debt tended to be higher in richer ZIP codes. The end result is that mortgage debt grew at broadly similar rates throughout the income distribution.

## A.7 Yearly Foreclosure Regressions

Figure A.11 plots income coefficients from yearly regressions of foreclosures rates on the log of income. The income coefficients trend up modestly, implying that foreclosures became relatively more prevalent in high-income ZIP codes during the housing bust. This figure complements the binned scatter plots in Figure 17, which also indicate somewhat higher growth in foreclosures in high-income ZIP codes relative to low-income ZIP codes. This relative pattern obtains even though absolute growth in foreclosures was lower in high-income areas.

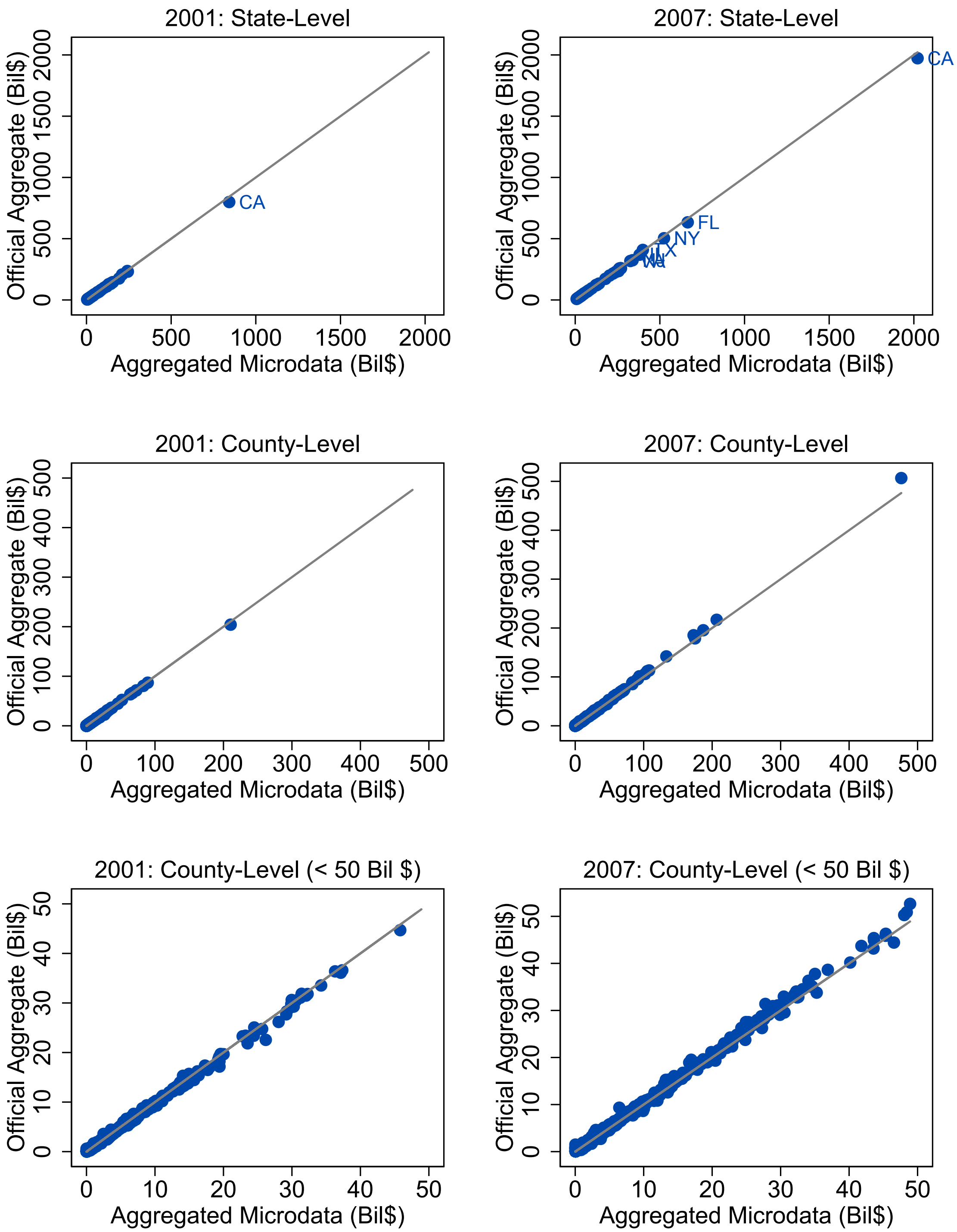



**Figure A.1.** Comparison of Aggregated Mortgage Debt Balances in the New York Fed Consumer Credit Panel. *Note:* Each of the panels above is a comparison of aggregated data from the microlevel records of the NY Fed Consumer Credit Panel. Aggregation along the horizontal axes was performed by the authors, while the vertical axes measure aggregates generated from the same dataset by the Federal Reserve Bank of New York. For the county-level data in the lower two rows, only counties with at least 10,000 consumers possessing credit records in 2010:Q4 are included. *Source:* NY Fed Consumer Credit Panel/Equifax.

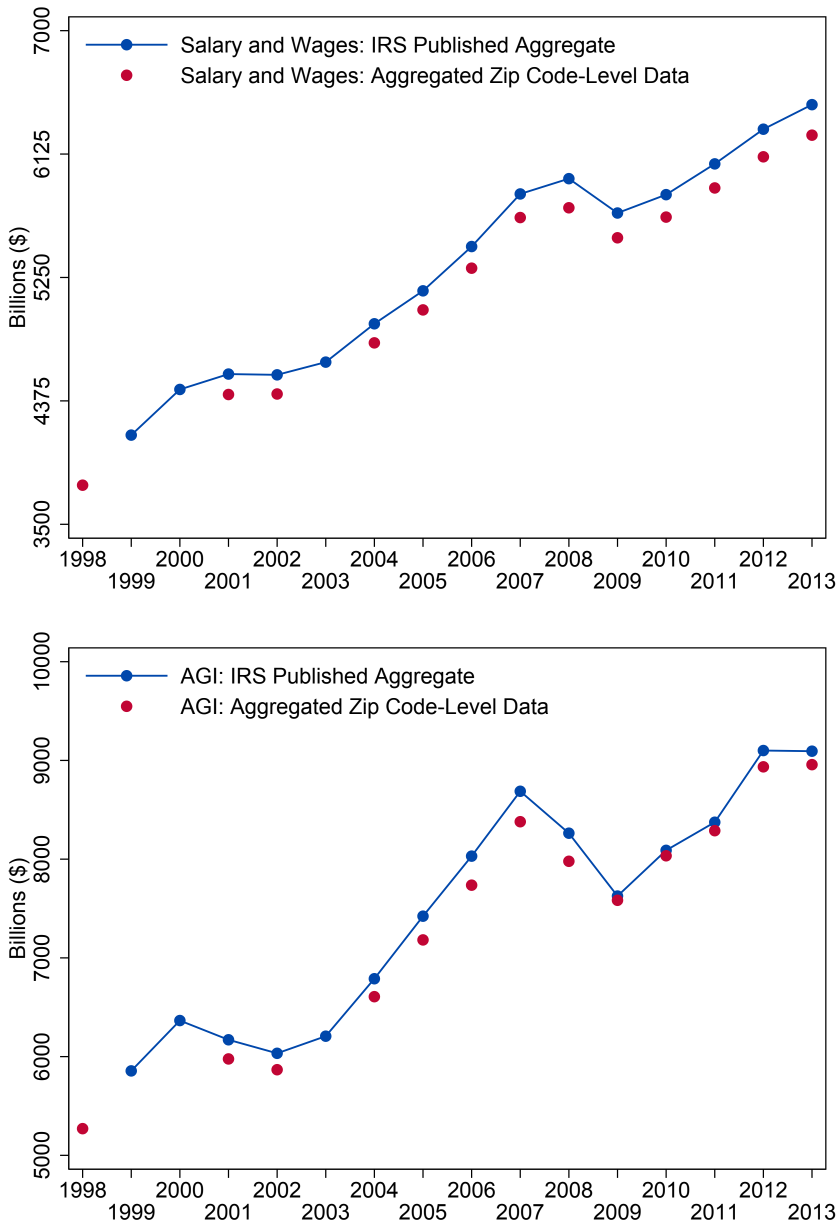



**Figure A.2.** Measures of Aggregate Salary and Wage Income and Adjusted Gross Income. *Note:* In each panel, the blue line depicts the given income aggregate as published by the IRS, and the red dots depict annual aggregates generated from the ZIP-level IRS data. *Source:* Internal Revenue Service, Statistics of Income Historical Table 1 (available at `https://www.irs.gov/uac/SOI-Tax-Stats-Historical-Table-1`).

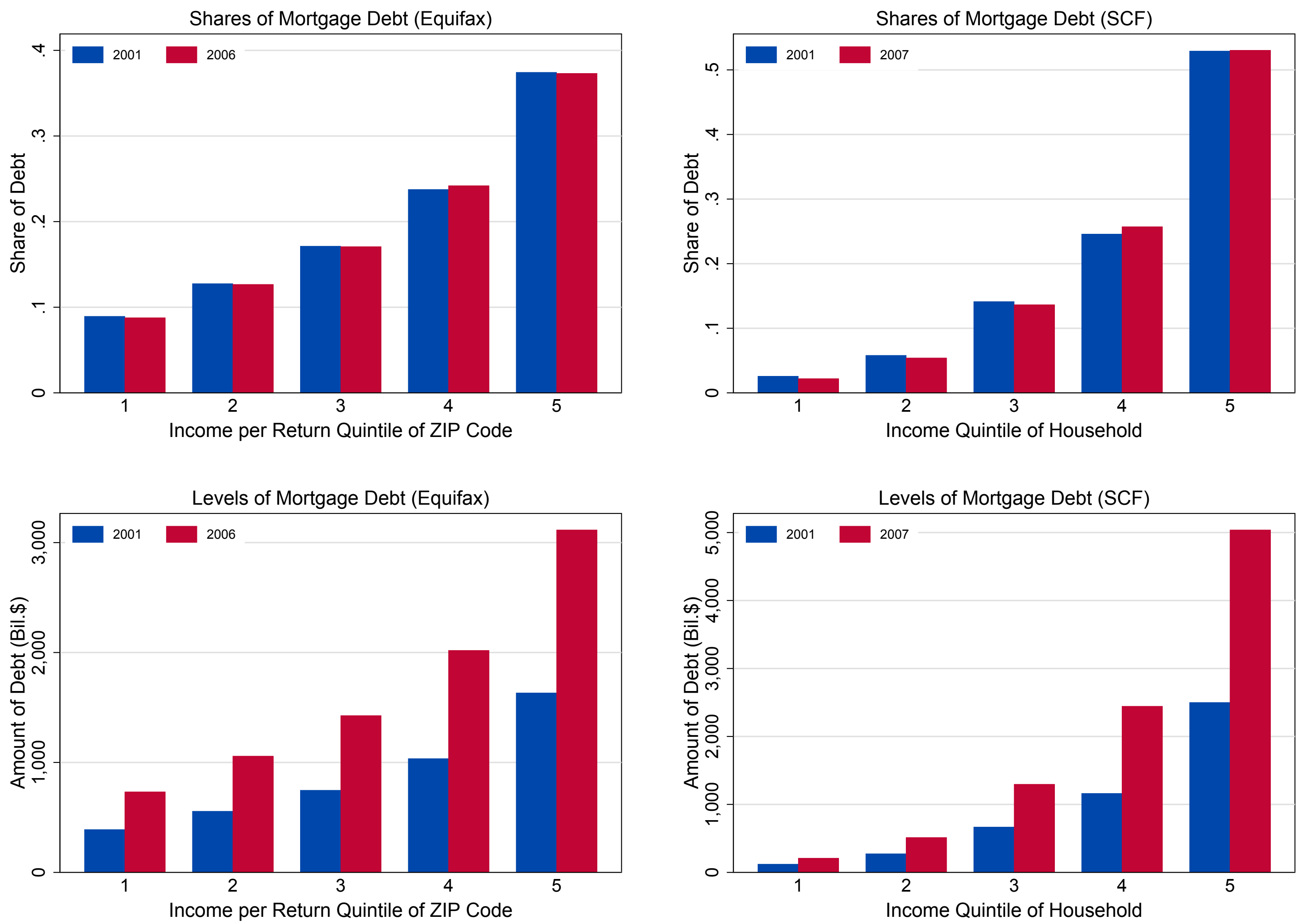



**Figure A.3.** Distributions of Mortgage Debt With Respect to Adjusted Gross Income (for ZIP Codes) and Total Income (for Households). *Note:* The income measure used throughout the main text is salary and wage income. This figure uses AGI as the income measure for ZIP codes in the left panels, and total income from the SCF for households in the right panels. *Source:* NY Fed Consumer Credit Panel/Equifax, IRS Statistics of Income, and Survey of Consumer Finances.

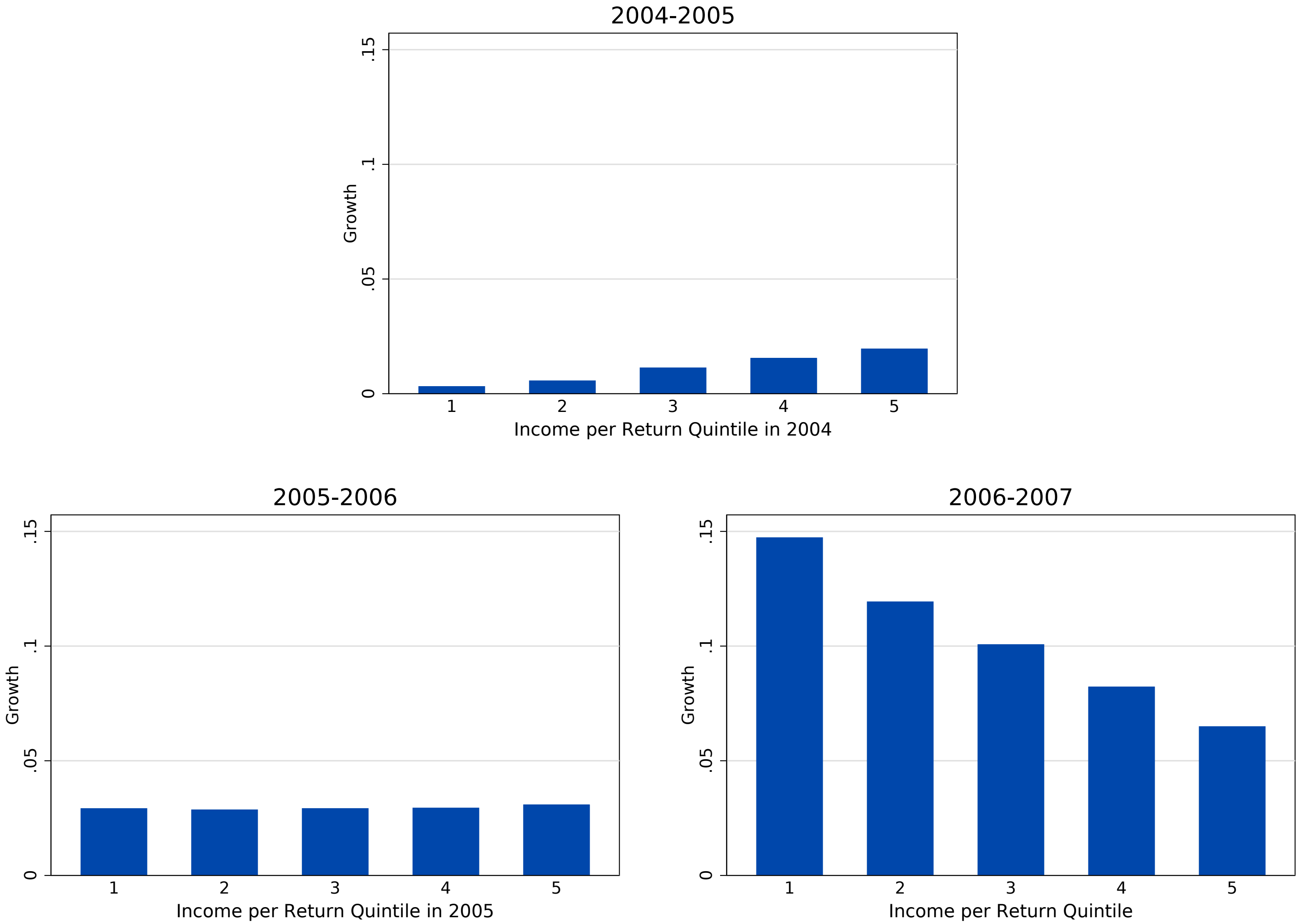



**Figure A.4.** Growth in IRS Returns by Income Quintile. *Note:* The top panel shows the growth in the total number of ZIP-level tax returns between 2004 and 2005, grouped by ZIP-level income in 2004. The bottom panels provide analogous information for returns growth in 2005-2006 and 2006-2007. The bottom right panel shows the strong inverse relationship between ZIP-level returns growth and income between 2006 and 2007 that was generated by a surge of low-income persons who filed solely to take advantage of the 2007 tax stimulus. *Source:* IRS Statistics of Income.

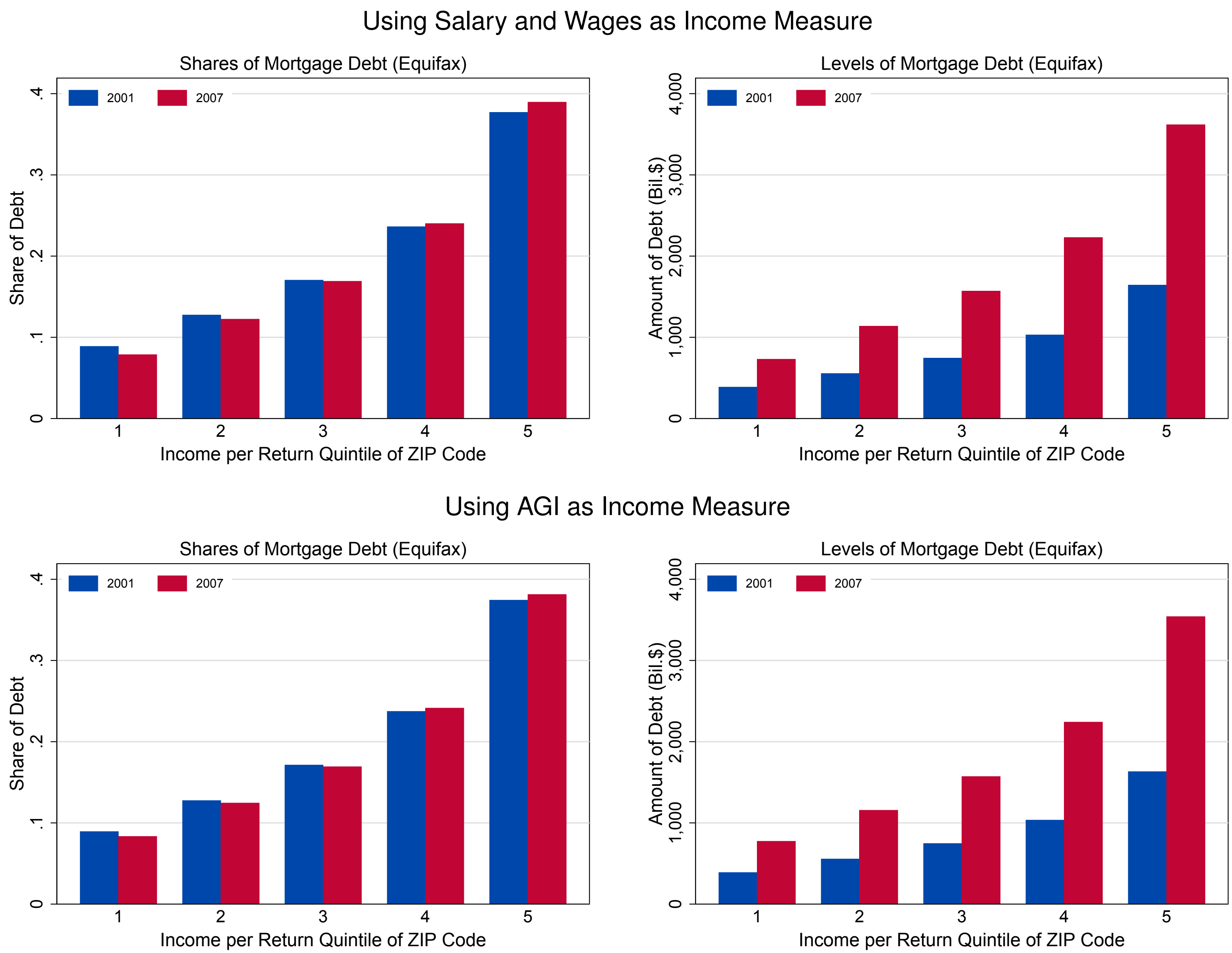



**Figure A.5.** ZIP-Level Distributions of Debt for 2001 and 2007 using Alternative Income Definitions. *Note:* These graphs are analogous to the ZIP-level panels in Figure 2, which depict distributions for 2001 and 2006, rather than 2001 and 2007. The lower panels in this figure also use AGI rather than wage and salary income. *Source:* NY Fed Consumer Credit Panel/Equifax, IRS Statistics of Income, and Survey of Consumer Finances.

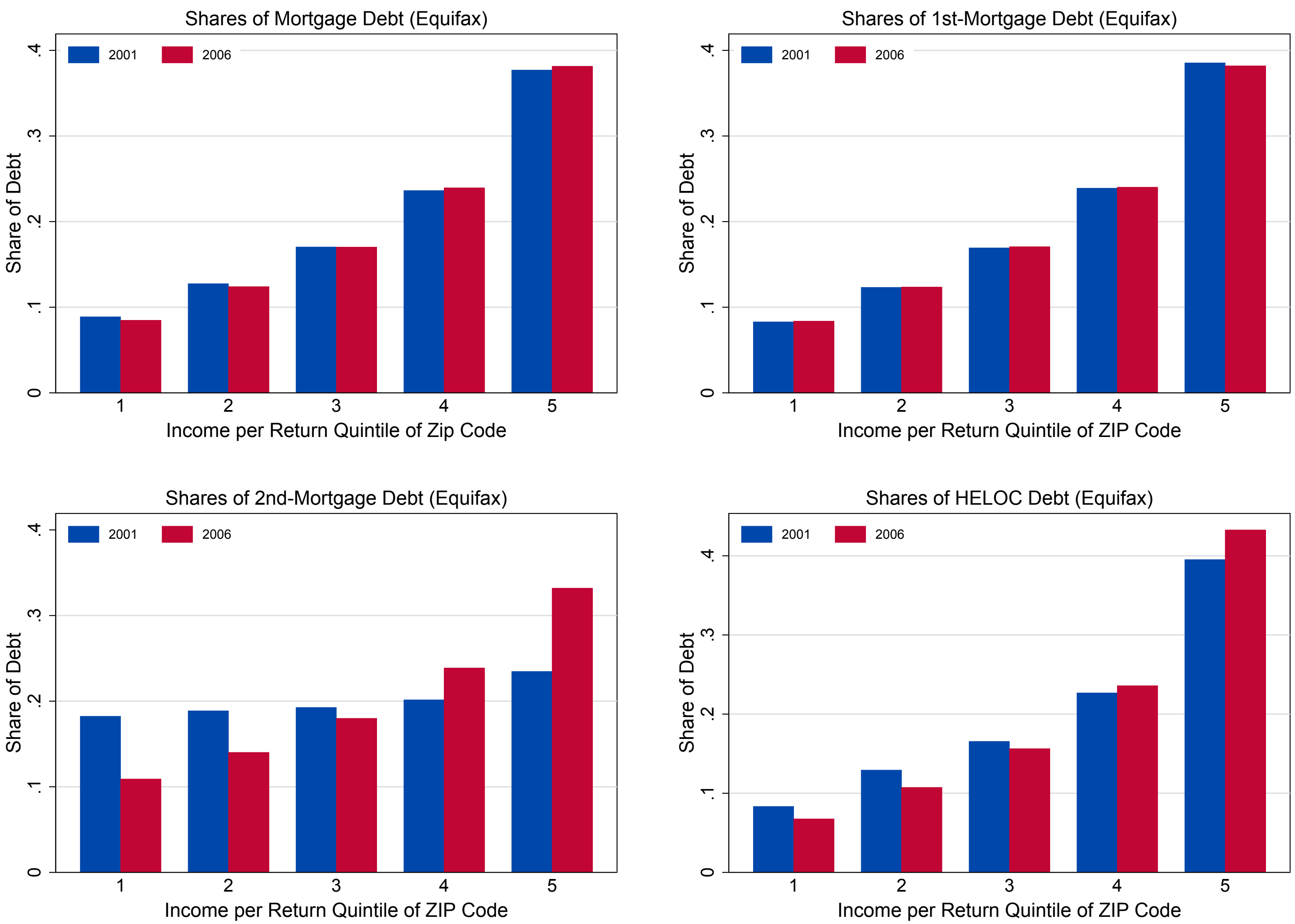



**Figure A.6.** 2001-06 Change in ZIP-Level Distributions of Debt by Mortgage Type, using Salary and Wages as Income Definition. *Note:* First mortgages include all purchase and refinance mortgages that are neither home equity loans nor home equity lines of credit (HELOCs). *Source:* NY Fed Consumer Credit Panel/Equifax and IRS Statistics of Income.

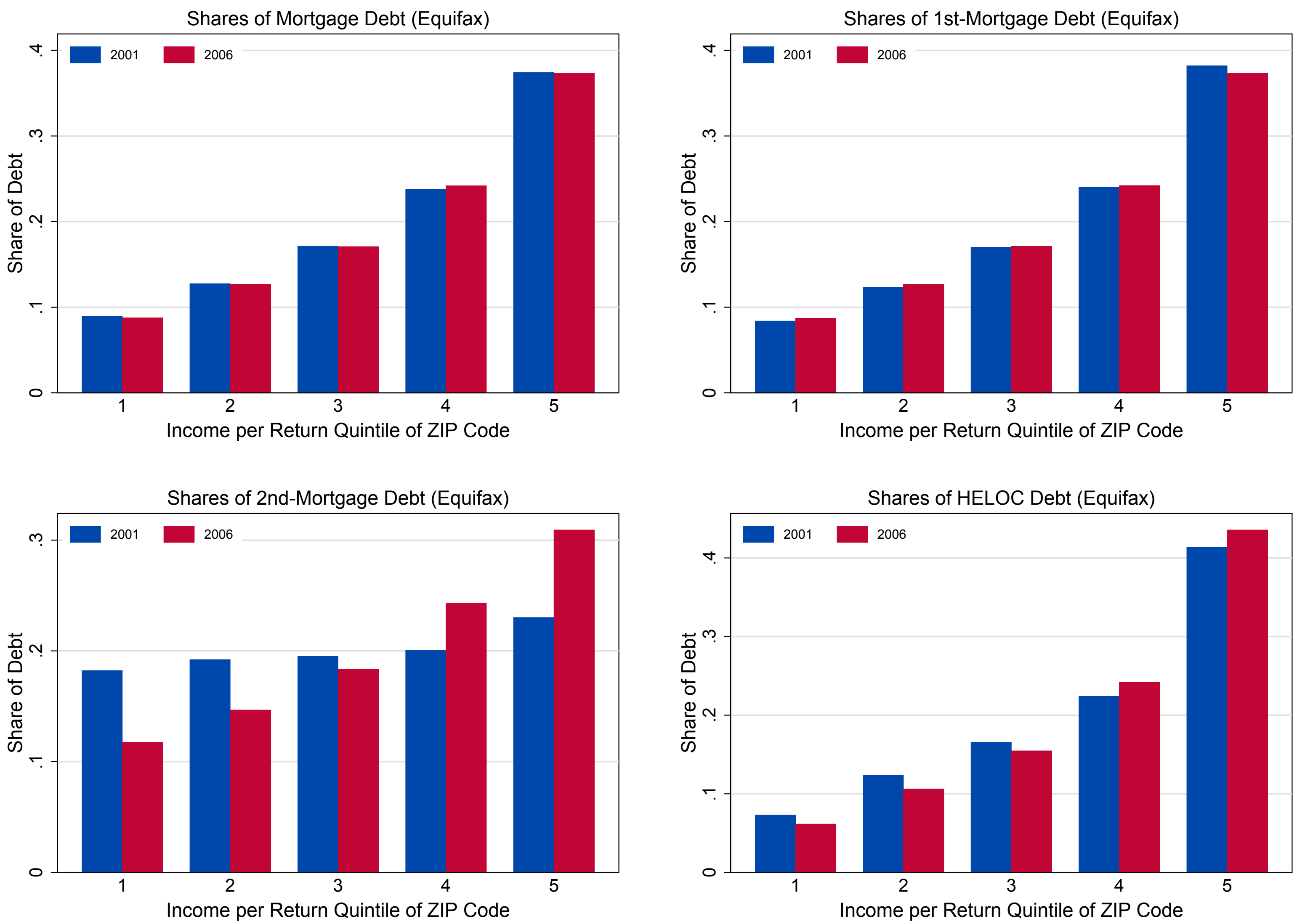



**Figure A.7.** 2001-06 Change in ZIP-Level Distributions of Debt by Mortgage Type, using AGI as Income Definition. *Note:* First mortgages include all purchase and refinance mortgages that are neither home equity loans nor home equity lines of credit (HELOCs). *Source:* NY Fed Consumer Credit Panel/Equifax and IRS Statistics of Income.

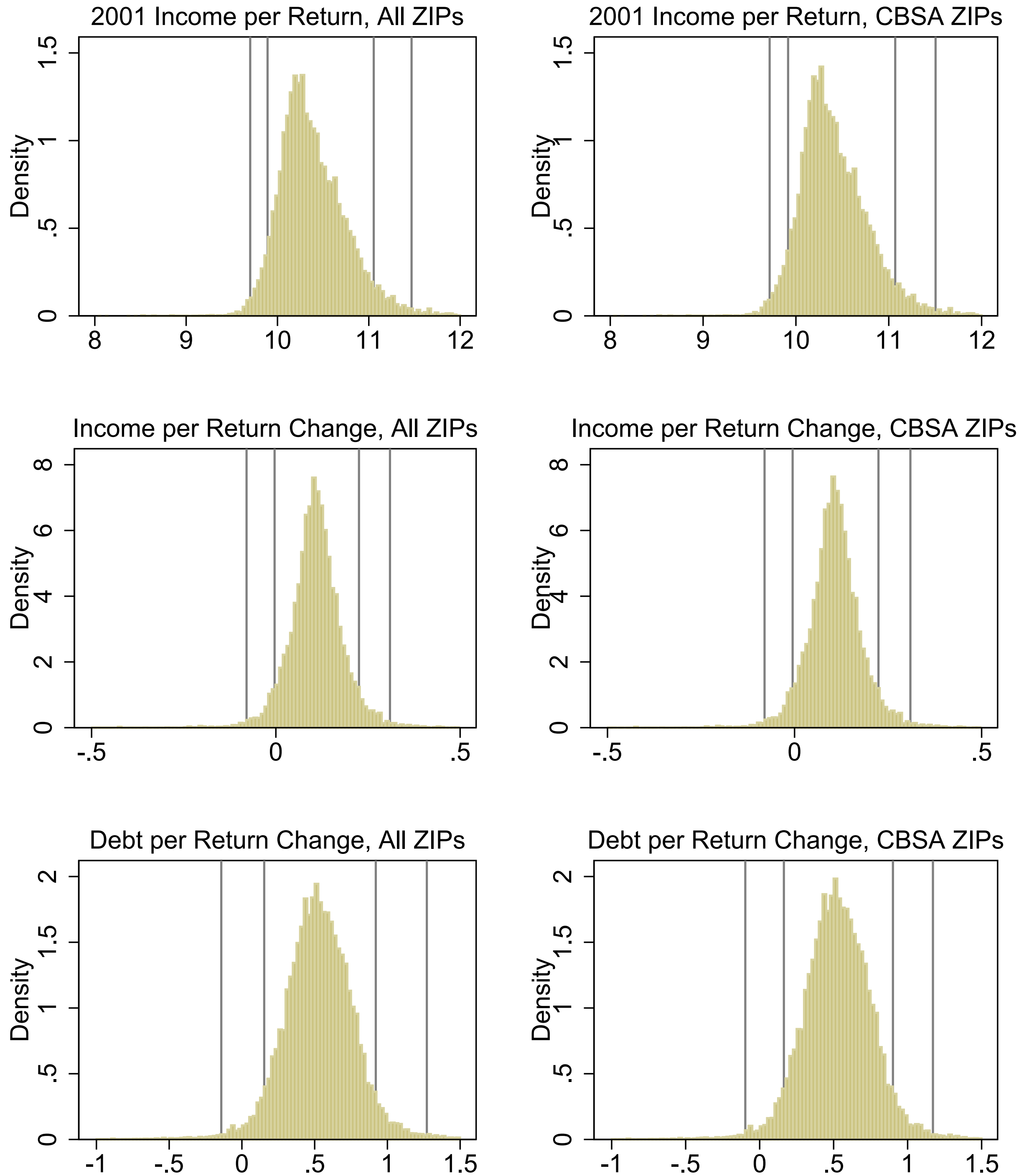



**Figure A.8.** ZIP-Level Densities of 2001 Income, 2001-06 Income Growth, and 2001-06 Mortgage-Debt Growth. *Note:* Growth rates are calculated as log differences. Vertical lines depict 1st, 5th, 95th and 99th percentiles. Panels at right include all ZIP codes, while those at left include only ZIP codes located in CBSAs. All distributions are weighted by the number of 2001 returns. *Source:* NY Fed Consumer Credit Panel/Equifax and IRS Statistics of Income.

| Dependent Variable: 2001-07 ZIP-Level Change in Ln Mortgage Debt per Return | | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Sample Restriction | None | None | 1% Trim | 5% Trim |
| | Panel A: All ZIP Codes | | | |
| 2001-07 Change in Ln Income per Return | 1.071*** | 1.050*** | 1.170*** | 1.031*** |
| | (0.040) | (0.043) | (0.054) | (0.061) |
| 2001 Ln Income per Return Level | | 0.019 | 0.010 | 0.031 |
| | | (0.012) | (0.014) | (0.016) |
| Constant | 0.527*** | 0.527*** | 0.528*** | 0.528*** |
| | (0.008) | (0.008) | (0.008) | (0.008) |
| R-sq. | 0.122 | 0.122 | 0.150 | 0.117 |
| Observations (No. of ZIP Codes) | 35,595 | 35,595 | 27,337 | 18,313 |
| Expected Diff. in Debt Growth: 90th 2001 Income Pctile vs. 10th 2001 Income Pctile | | 0.017 | 0.009 | 0.027 |
| | Panel B: CBSA ZIP Codes without Fixed Effects | | | |
| 2001-07 Change in Ln Income per Return | 1.088*** | 1.059*** | 1.192*** | 1.028*** |
| | (0.043) | (0.046) | (0.057) | (0.064) |
| 2001 Ln Income per Return Level | | 0.027* | 0.010 | 0.032 |
| | | (0.014) | (0.016) | (0.018) |
| Constant | 0.527*** | 0.527*** | 0.528*** | 0.529*** |
| | (0.009) | (0.009) | (0.009) | (0.008) |
| R-sq. | 0.141 | 0.142 | 0.164 | 0.120 |
| Observations (No. of ZIP Codes) | 27,567 | 27,567 | 21,634 | 15,165 |
| Expected Diff. in Debt Growth: 90th 2001 Income Pctile vs. 10th 2001 Income Pctile | | 0.023 | 0.009 | 0.028 |
| | Panel C: CBSA ZIP Codes with CBSA Fixed Effects | | | |
| 2001-07 Change in Ln Income per Return | 0.827*** | 0.858*** | 0.990*** | 0.925*** |
| | (0.064) | (0.060) | (0.062) | (0.066) |
| 2001 Ln Income per Return Level | | −0.027 | −0.057*** | −0.052*** |
| | | (0.015) | (0.016) | (0.015) |
| Constant | 0.527*** | 0.527*** | 0.528*** | 0.529*** |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| R-sq. | 0.429 | 0.429 | 0.553 | 0.580 |
| Observations (No. of ZIP Codes) | 27,567 | 27,567 | 21,634 | 15,165 |
| Expected Diff. in Debt Growth: 90th 2001 Income Pctile vs. 10th 2001 Income Pctile | | −0.023 | −0.049 | −0.045 |

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

**Table A.1.** Long-Difference Regressions using 2001-2007 Differences. *Note:* This table is analogous to Table 3, which uses 2001-2006 differences rather than 2001-2007 differences. *Source:* NY Fed Consumer Credit Panel/Equifax and IRS Statistics of Income.

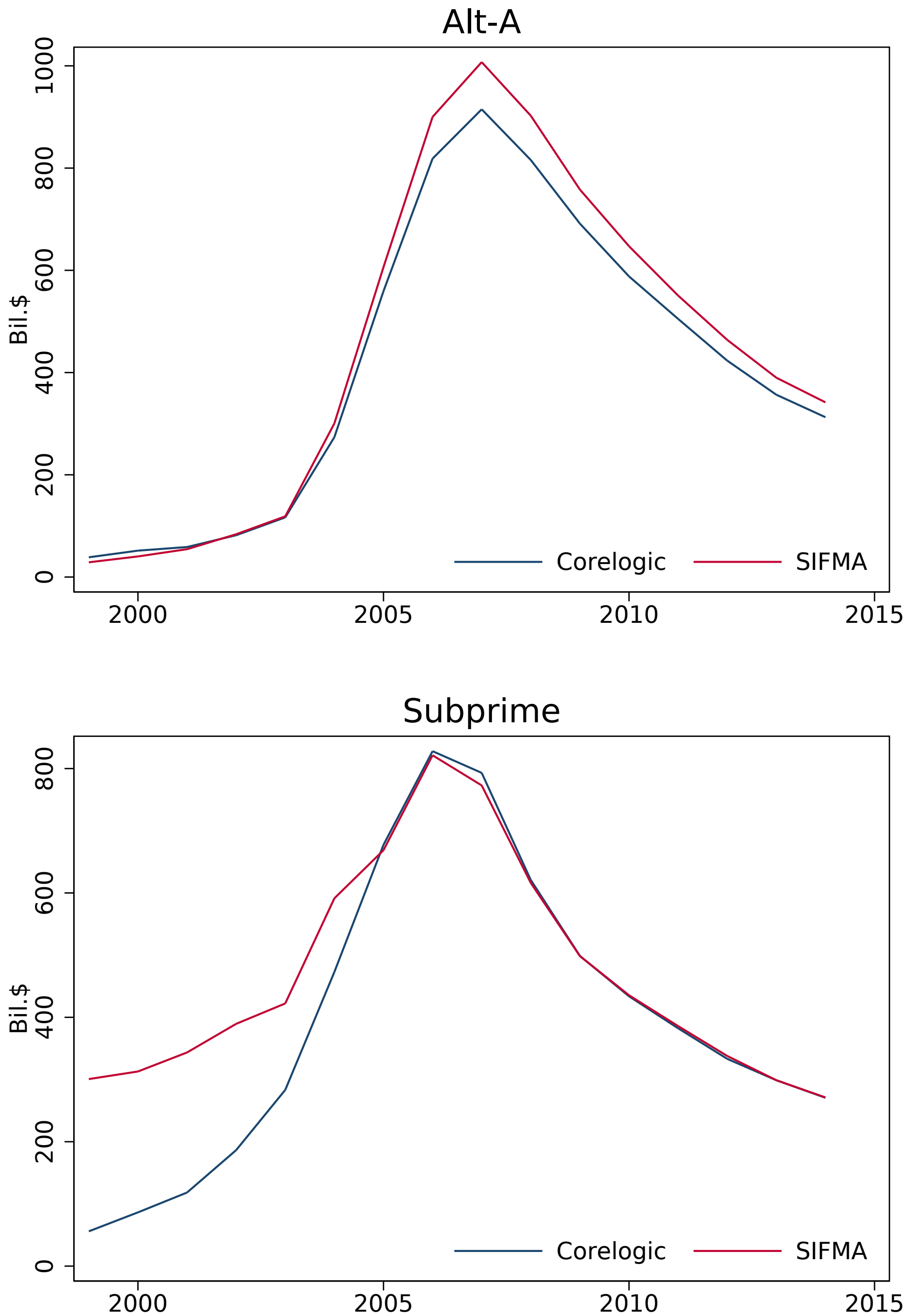



**Figure A.9.** Comparison of Stocks of Mortgage Debt By Type from CoreLogic and SIFMA. *Note:* This figure compares two measures of subprime and Alt-A private-label mortgage-backed securities outstanding. One measure is an aggregate estimate from the Securities Industry and Financial Markets Association (SIFMA). The other is generated by the authors by summing loan-level balances from the CoreLogic Private Label Securities ABS Database. The figure shows that CoreLogic's coverage of Alt-A securities was very good throughout the 2000s, but that its coverage of subprime mortgage-backed securities improved over time. *Source: SIFMA and CoreLogic.*

| Year | Total RMBS | HELOC | Junior Lien | Scratch and Dent | Subprime | Resecur-itization | Alt-A | Jumbo Prime | Manuf. Housing | Other |
|---|---|---|---|---|---|---|---|---|---|---|
| 1980 | 0.7 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.7 |
| 1981 | 0.7 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.7 |
| 1982 | 0.7 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.7 |
| 1983 | 3.7 | 0.0 | 0.0 | 0.0 | 0.0 | 3.1 | 0.0 | 0.0 | 0.0 | 0.7 |
| 1984 | 12.7 | 0.0 | 0.0 | 0.0 | 0.0 | 10.8 | 0.0 | 0.0 | 0.0 | 1.9 |
| 1985 | 28.3 | 0.0 | 0.0 | 0.0 | 0.0 | 24.8 | 0.0 | 0.0 | 0.0 | 3.5 |
| 1986 | 78.1 | 0.0 | 0.0 | 0.0 | 0.0 | 70.2 | 0.0 | 0.0 | 0.0 | 7.9 |
| 1987 | 131.6 | 0.0 | 0.0 | 0.0 | 0.0 | 111.7 | 0.0 | 0.0 | 0.2 | 19.7 |
| 1988 | 185.4 | 0.0 | 0.0 | 0.7 | 0.0 | 147.4 | 0.0 | 1.5 | 2.5 | 33.3 |
| 1989 | 204.0 | 2.0 | 0.0 | 1.2 | 0.2 | 146.0 | 0.0 | 2.2 | 3.4 | 49.0 |
| 1990 | 230.5 | 4.2 | 0.0 | 2.1 | 2.9 | 146.7 | 0.1 | 2.1 | 4.3 | 68.1 |
| 1991 | 271.7 | 6.2 | 0.3 | 3.0 | 8.2 | 141.3 | 0.2 | 5.6 | 4.8 | 102.1 |
| 1992 | 279.9 | 7.1 | 0.3 | 4.0 | 8.4 | 104.8 | 0.2 | 5.1 | 6.7 | 143.1 |
| 1993 | 269.0 | 7.9 | 0.2 | 5.2 | 11.0 | 67.2 | 0.4 | 6.4 | 8.1 | 162.6 |
| 1994 | 278.1 | 8.3 | 0.1 | 6.1 | 16.3 | 45.5 | 0.3 | 7.3 | 10.4 | 183.8 |
| 1995 | 286.7 | 9.1 | 0.1 | 7.3 | 23.7 | 38.2 | 0.5 | 7.0 | 15.6 | 185.2 |
| 1996 | 314.0 | 12.0 | 1.0 | 8.9 | 43.5 | 28.6 | 5.0 | 6.7 | 21.7 | 186.7 |
| 1997 | 383.9 | 12.5 | 2.0 | 11.2 | 72.4 | 26.4 | 11.2 | 7.2 | 28.2 | 212.6 |
| 1998 | 478.8 | 10.6 | 7.4 | 13.8 | 106.8 | 19.6 | 24.9 | 12.6 | 37.1 | 246.0 |
| 1999 | 692.6 | 11.9 | 7.9 | 15.0 | 303.0 | 13.6 | 28.9 | 33.9 | 47.8 | 230.6 |
| 2000 | 743.6 | 13.5 | 8.7 | 16.1 | 315.0 | 13.3 | 40.3 | 59.8 | 52.3 | 224.4 |
| 2001 | 818.7 | 12.4 | 9.5 | 16.8 | 344.8 | 12.3 | 54.5 | 130.1 | 51.7 | 186.5 |
| 2002 | 909.6 | 23.7 | 10.7 | 16.8 | 390.3 | 18.5 | 83.8 | 184.9 | 47.9 | 133.0 |
| 2003 | 1,009.1 | 27.6 | 12.1 | 22.7 | 422.7 | 17.3 | 118.8 | 233.8 | 39.3 | 114.7 |
| 2004 | 1,469.7 | 48.1 | 17.2 | 29.4 | 591.6 | 32.9 | 300.6 | 285.5 | 34.1 | 130.2 |
| 2005 | 2,005.4 | 57.2 | 30.8 | 37.0 | 668.7 | 49.3 | 606.1 | 356.9 | 29.4 | 170.0 |
| 2006 | 2,588.0 | 66.4 | 63.3 | 42.2 | 821.4 | 58.1 | 900.1 | 438.7 | 25.9 | 171.8 |
| 2007 | 2,704.7 | 62.6 | 65.0 | 45.5 | 772.9 | 63.7 | 1,007.3 | 509.2 | 22.8 | 155.6 |
| 2008 | 2,349.8 | 50.0 | 47.4 | 39.0 | 616.2 | 66.4 | 902.6 | 454.2 | 20.5 | 153.5 |
| 2009 | 1,916.1 | 40.4 | 36.5 | 33.5 | 498.6 | 68.5 | 758.2 | 364.5 | 18.2 | 97.8 |
| 2010 | 1,643.9 | 33.3 | 30.4 | 29.8 | 435.7 | 69.8 | 647.3 | 289.3 | 16.6 | 91.5 |
| 2011 | 1,409.6 | 23.9 | 26.0 | 28.4 | 386.4 | 70.5 | 550.8 | 228.1 | 14.9 | 80.5 |
| 2012 | 1,218.9 | 19.9 | 23.1 | 26.3 | 338.3 | 83.9 | 464.0 | 183.2 | 13.3 | 66.8 |
| 2013 | 1,073.8 | 17.5 | 20.5 | 30.5 | 299.4 | 89.1 | 389.7 | 151.9 | 11.4 | 63.7 |
| 2014 | 971.7 | 14.5 | 18.4 | 41.5 | 271.5 | 88.8 | 341.6 | 132.9 | 10.2 | 52.3 |
| 2015 | 875.2 | 11.7 | 16.6 | 54.2 | 242.2 | 87.8 | 290.5 | 117.1 | 9.0 | 51.5 |
| Change: 2001-2006 | 1,769.3 | 54.0 | 53.7 | 25.4 | 476.6 | 45.8 | 845.6 | 308.6 | -25.8 | -14.7 |
| Change: 2001-2007 | 1,886.0 | 50.3 | 55.5 | 28.7 | 428.1 | 51.4 | 952.8 | 379.1 | -28.9 | -30.9 |

**Table A.2.** U.S. Non-Agency Residential Real Estate Securities (RBMS) Outstanding, in Billions. *Source:* Securities Industry and Financial Markets Association (SIFMA). Data are drawn from Tab 2.1 of the spreadsheet titled "U.S. Mortgage-Related Issuance and Outstanding," available at http://www.sifma.org/uploadedFiles/Research/Statistics/StatisticsFiles/SF-US-Mortgage-Related-SIFMA.xls.

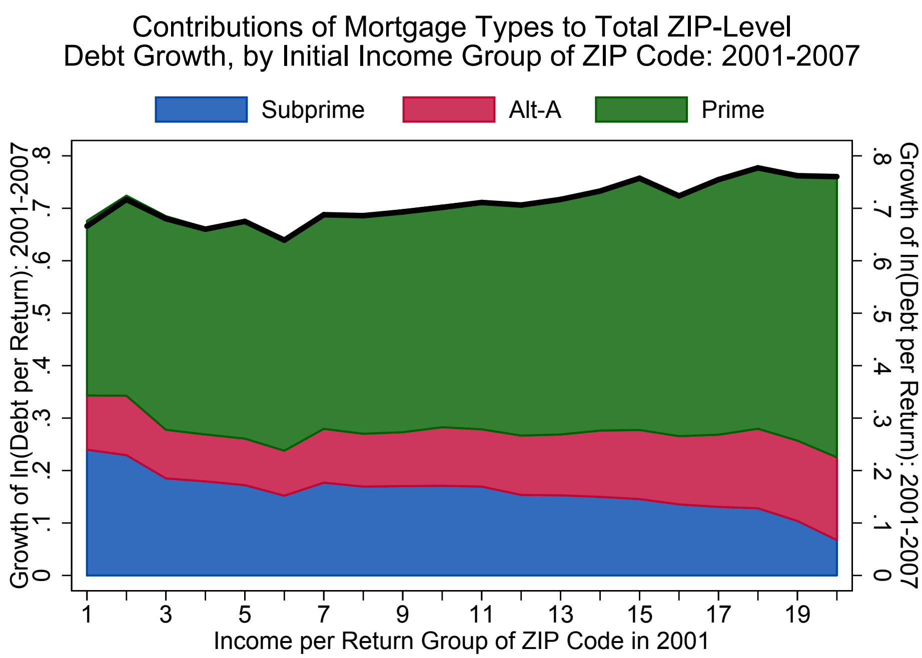



**Figure A.10.** The Effect of Alternative Mortgage Products on Overall Mortgage-Debt Growth at the ZIP-Code Level: 2001-07. *Note:* This figure is analogous to Figure 18, which is based on debt growth from 2001 to 2006 rather thangrowth from 2001 to 2007. *Source:* NY Fed Consumer Credit Panel/Equifax, CoreLogic Private Label Securities ABS Database, and IRS Statistics of Income.

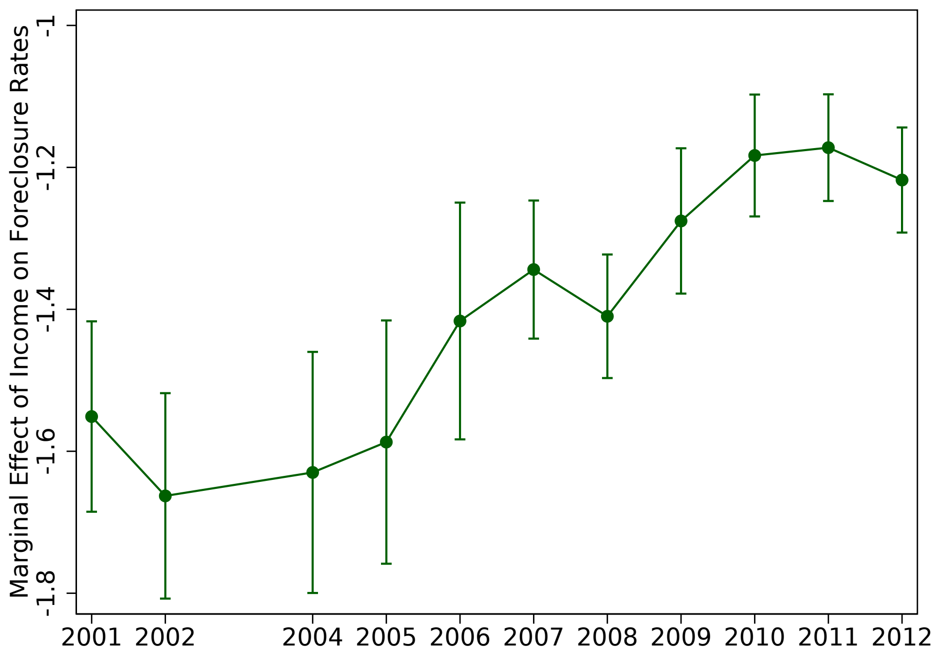



**Figure A.11.** Estimated Income Effects in Foreclosure Regressions. *Note:* This figure depicts yearly coefficients from a regression of ZIP-level foreclosure propensities on income from 2001 through 2012. These foreclosure-income regressions are structured similarly to the debt-income regressions presented in the left panels of Figure 7. *Source:* NY Fed Consumer Credit Panel/Equifax and IRS Statistics of Income.