Borrower Characteristics and Credit Supply Expansion in the U.S. Residential Mortgage Market : Evidence from 2010 - 2015

by

Raghav Saraogi

An honors thesis submitted in partial fulfillment

of the requirements for the degree of

Bachelor of Science

Undergraduate College

Leonard N. Stern School of Business

New York University

May 2018

Professor Marti G. Subrahmanyam
Faculty Adviser

Professor Arpit Gupta
Thesis Adviser

Borrower Characteristics and Credit Supply Expansion in the U.S. Residential Mortgage Market : Evidence from 2010 - 2015

Raghav Saraogi *

May 11, 2018

Abstract

I estimate the relative effects of changes in borrower characteristics and subsequent credit policy changes on the revival in mortgage debt and loan approval rates, post the financial crisis of 2007-2009. Using loan-level data from the Home Mortgage Disclosure Act (HMDA) and IRS tax-return income data, I find that trends in median applicant income closely match those of median IRS income per ZIP code between 2010 and 2015 indicating that the mortgage market has attracted applicants from across the income distribution during this time. The aggregate increase in debt has not been disproportionately higher in high-income areas and approval rates have increased across the distribution of income and credit scores. In contrast with existing literature that suggests lower credit access along credit scores, I find that ZIP codes with lower average credit scores have experienced the highest levels of income growth. I use a Bartik instrument for income to show that this has driven higher approval rates and mortgage origination both along the extensive and intensive margins of mortgage origination.

*NYU Stern Undergraduate Honors Program. I would like to thank my advisor Professor Arpit Gupta (NYU Stern) for his unyielding and generous support in helping me through the research process. I am also grateful to Professor Marti Subrahmanyam (NYU Stern) for his wisdom and the opportunity to participate in this invaluable research opportunity.

1 Introduction

This paper aims to analyze the nature of the revival in the U.S. residential mortgage market, post the financial crisis, in the context of borrower characteristics. The mortgage market in the United States has revived considerably since the aftermath of the 2007-2009 financial crisis. Total mortgage debt has increased from 13.3 trillion dollars in 2013 to 14.3 trillion in 2016, after steadily declining between 2008 and 2013. Between this period, the aggregate approval rate has also increased from 68.5% to 72%. The specific characteristics of this trend in the mortgage market has implications for the broader economy. It is a measure of the transmission of monetary policy, credit quality standards, economic expectations of both lenders and borrowers and other economic agents. However, this paper focuses on two other key aspects of the residential mortgage market; firstly, whether there has been a consistent and systematic change in borrower characteristics and secondly whether greater demand for mortgage debt has economic underpinnings with changes in the real economy. Borrower characteristics have largely remained unchanged since 2010, relative to the average household in their region. Then one may ask, what has justified the higher levels of household debt? The answer lies in higher levels of overall income, which have resulted in a much larger applicant pool in the mortgage market in 2015 as compared with 2010.

This study primarily focuses on three dimensions of the mortgage market: loan approval rates, income and credit scores. I analyze how these have interacted with and influenced each other between the years 2010 and 2015. I begin by examining borrower characteristics such the share of mortgage debt originated for different income groups and credit score groups. I also study applicant incomes relative to the median income in the ZIP code and find that the ratio for each income group has not shifted significantly since 2010. This means that applicant income and median income levels have increased proportionally between 2010 and 2015.

Credit supply expansion can be tested using variation in approval rates between years and between regions. I use income data from IRS tax returns and find that income levels are associated with higher approval rates. Further, I test whether income has become more closely related with mortgage origination along the extensive and intensive margin. Using year interaction effects with income in a linear regression, I find that this is indeed the case - income has a monotonically larger coefficient for both the extensive and intensive margins of mortgage origination between 2010 and 2015. Finally, to specifically test the role of income growth within three-digit ZIP code in mortgage origination, I use a Bartik Instrument for income to empirically show a causal effect of income growth on all dimensions of mortgage origination. I use a Two-Stage Least Squares (TSLS) model with three-digit ZIP code fixed effects to show that differences in income between 2010-2015 at the three-digit ZIP code level drives higher approval rates, and both the extensive and intensive margins of mortgage origination, as well as the number of applications, while controlling for average credit scores, with all coefficients highly statistically significant.

In the broader view, economists find that the average credit score of originated mortgages has increased since the crisis - an indication of lower access to credit. However, I find that in a large sample of single-family fixed-rate mortgages owned by Fannie Mae and Freddie Mac, the average credit score has steadily declined between 2010 and 2015, contradicting the view that credit access has tightened for the marginal or risky borrower. This data is given in Figure 7. Moreover, aggregating differences between ZIP codes using this data, there is a higher proportion of overall value of debt that has been originated in lower credit score regions. This represents partial evidence of credit supply expansion and lending policy changes, along the credit quality dimension. Further, aggregated data shows that areas with lower credit quality have experienced the highest income growth rates. This result is subsequently verified using regression analyses. Therefore, there is a strong signal about credit lending policies from this insight. In line with the view that credit policies are tightening in the mortgage market, this insight supplements this view by hypothesizing that lenders are balancing their credit risk exposure by lending to households with higher income, even though they may have lower credit scores.

The Fed's Senior Loan Officer Survey 2017 reports only marginal evidence of easing standards of residential mortgage loans since 2010. Lenders are still constrained by far reach of the Dodd-Frank Wall Street Reform and Consumer Protection Act (Dodd-Frank) and other regulations emerging from the crisis such as the Home Affordable Modification Program (HAMP), which incentivize lenders, both and large and small, to reduce their exposure to risky loans. After the imminent reversal in lending patterns, mortgage debt has been on the rise once again and if there has been a reversal of the tightening standards that regulations imposed on lenders it is important to understand its drivers and correlation with the real economy. The pattern of lenders adjusting credit risk measures by income levels and income growth implies that mortgage lenders have a positive outlook for the economy and are increasing their exposure to regions they believe will achieve relatively high levels of income growth, despite lower credit scores.

On the other hand, a borrower with a given set of characteristics who would have been denied a mortgage in 2010 probably has better odds of being approved for the same mortgage today. Borrowers in 2016 have better applications on average, with higher incomes and credit scores. Therefore, it wouldn't be difficult to assume that as the real economy has recovered over the past five years or so, more people have increased their ability to borrow and are looking toward homes for a more stable standard of living and/or financial future. The number of applicants has grown from 3.5 million in 2011 to over 4.75 million in 2015, as depicted in Figure 1. The approval rate for 'risky' loans in the 90th percentile of loan amount to applicant income ratio (LTI) has increased marginally since 2010. Simultaneously, there is a steep increase in the number of applications in this percentile, and across the distribution as well. I empirically show that the number of applications are higher in regions with higher income growth and this result is robust within a region as well - suggesting that application growth is directly proportional to income growth.

In the next section, I describe how this research fits into a broader set of related literature. In the following section, I outline data sources and key assumptions made in using these data. In section 4 I present key insights from aggregated data summaries, with a focus on borrower characteristics and the distribution of debt across the income and credit score spectrum. In section 5, I quantify the hypotheses formed in section 4 and provide causal estimates of key results through an instrumental variable two-stage least squares estimation. Finally in section 6, I summarise the key insights, draw policy implications from them and present some questions for further research.

2 Related Literature

Existing literature on the impact of regulation post the housing boom by DeFusco et. al (2017). shows the effect of leverage constraints imposed by Dodd-Frank on credit markets. The Ability-to-Repay (ATR) Rule has affected and eliminated a significant proportion of credit markets and has effectively restricted leverage for a larger share of borrowers by acting on both the demand and supply of credit. Additionally D'Acunto et. al. (2017) show that mortgage lenders reduced credit by 15% to middle-class households due to large banks facing increased costs of origination due to financial regulation after the crisis in 2007-2009. Bordo et. al. (2018) show that small business lending has been constrained in the aftermath of the crisis due to the Dodd-Frank Act.

In addition, there is a body of work that debates the role of distribution of mortgage credit before and during the crisis. Sufi (2016) use HMDA data to illustrate the role of credit supply expansion in redistributing credit disproportionately toward lower-income borrowers. On the other hand, Schoar et. al. (2016) have demonstrated the role of the demand side in the mortgage crisis and they suggest that middle and high-income borrowers increased their share of mortgage defaults in the run-up to the crisis.

In showing that income levels are have become more intimately coupled with mortgage origination post the crisis, I borrow the methodology in Schoar (2016) to show the relative importance of income levels in a regulatory environment that restricts borrowers with lower credit scores and that deters high levels of individual or household leverage. Moreover, in contrast with various literature that emphasizes the regulatory effects of reducing access to credit to the marginal borrower, I find that lower and middle credit score ZIP codes have experienced a rate of mortgage origination relative to the highest average credit score regions, largely based on stronger income growth.

3 Data Sources and Assumptions

There are four large data sets that I plan to turn to for this analysis, all of which are publicly available. First, the Home Mortgage Disclosure Act (HMDA) data provided by the Consumer Finance Protection Bureau (CFPB) is a large collection of home mortgage applications, released annually since 1975, from over 6000 financial institutions in the U.S.. It contains loan-level data of borrower characteristics, income, geographical location, loan decision and other important features for each mortgage application. Specifically, I use the applicant's income, the census tract reported on the application, the loan amount and the credit decision in this analysis. Most importantly, this can then be used to compare mortgage approval rates across a given time period and the income levels tied to these approval rates for a given census tract or county.

Researchers including Mian and Sufi (2017) and Schoar et. al. (2016) have debated the role of fraudulent overstating of income in the subprime crisis. They have focused specifically on the difference between income reported on mortgage applications and its equivalent reported on tax returns. Here too, I will use tax return data provided by the Internal Revenue Service to test for levels of income reported in HMDA. Specifically, I use the "Taxable Income" item to estimate the median taxable income for each ZIP code, for each year between 2010 and 2015. In aggregating this data to three-digit ZIP codes, I used a weighted average of median taxable income in each five-digit ZIP code. To estimate median income in a census tract, I matched each census tract with one five-digit ZIP Code using the census tract with the highest proportion of residents and business within a given ZIP code. Further, to estimate county level measures for income and all other variables, I again used a weighted-average of census tract measures, using the number of mortgage applications per tract as the weights.

Thirdly, two large GSEs: the Federal National Home Mortgage Association (Fannie Mae) and the Federal Home Loan Mortgage Association (Freddie Mac) publish loan-level data for a section of the mortgages they hold. Fannie Mae provides 5 year, 10 year, 15 year, 20 year, 25 year and 30 year fixed rate single-family data. Freddie Mac publishes loan performance data for single-family fixed rate mortgages purchased or guaranteed by the GSE. This subset of agency backed mortgages can be mined to indicated further characteristics of the new or marginal borrowers: including debt-to-income ratios, credit scores, loan-to-value (LTV) and delinquency status. These data can further be compared with HMDA, to reveal how the share of agency mortgages has evolved over time.

Lastly, the U.S. Census Bureau's County Business Patterns (CBP) data is used to construct the instrumental variable: the Bartik instrument. These data provide estimates for the number of payroll employees employed by each industry (by NAICS code) in each county on an annual basis. These data is then used to construct the share of each industry in every county, in terms of the number of employees hired across industries. The Bartik instrument also uses aggregate employee growth rates per industry and this is estimated from the number of payroll employees variable in this data-set.

4 Aggregate Evidence

4.1 Borrower Characteristics

The mortgage market has faced fairly unchanged characteristics from borrowers, when considering the entire mortgage applicant pool since 2010. There have been no systematic trends in borrower characteristics, on average, along any dimension: income, credit score, DTI or LTI. Here, income among applicants refers to their income relative to the median or average income in the ZIP Code or county. Figure 2 plots this for every income group for all six years. The parallel slopes of the best-fit lines suggest that there has been no change in the income of mortgage applicants relative to the median income in the ZIP Code for each IRS income group. This evidence shows that the mortgage market has not necessarily deterred less-wealthier households from taking on more debt post the crisis - a significant result considering high delinquency rates at the beginning of this period, coupled with conservative lending regulation. This evidence them also implies that there isn't a 'self-selection' phenomenon that is causing higher approval rates.

Furthermore, when considering borrower characteristics by credit-quality groups, loan sizes have not been significantly higher since 2010, scaled for borrower income. Figure 3 plots this trend for different levels of FICO scores [1], with no systematic trend. While there is some variability in LTI levels over this period, the average trend represented here shows that there is no noticeable change in LTI being demanded over this period for any of the groups. This provides a partrial measure of the risk and 'ambitiousness' of a mortgage application. Then, it can inferred that borrowers have not increased their expectation and are not looking for a higher level of indebtedness since 2010, at least when it comes to their home. Note that this measure, scales for income as reported on the mortgage application.

[1] Here, the FICO groups are derived by classifying each county based on its median FICO score, into 5 groups with 1 corresponding to the lowest FICO score counties.

To extend this picture, LTI ratios have also not varied significantly by income group. Figure 4 shows that the ratio has increased slightly for all income groups since 2010, after decreasing slightly in 2010. The highest increase has been for the highest income quintile - a total increase of just over 2% over the six-year period.

However, the extensive margin of mortgage origination has increased significantly. In the HMDA data, the number of applications for purchase-only, single family home mortgages increased by 38% between 2010 and 2015. Figure 5 visualizes this trend spatially by county. It is evident from the figure that mortgage applications were on the rise by greater than 14% in a most counties, with 20% of counties experiencing 75% or greater increase in applications since 2010. This suggests that since the high-default period of 2009-2010, when many households were deterred from taking on higher debt, there has been a significant revival of demand along the extensive margin.

There is a body work that shows that post the crisis, mortgage debt has been redistributed regressively. D'Acunto and Rossi (2017) note that mortgage lenders reduced credit to middle-class households by 15% while increasing origination to wealthy households by 21% since 2011. However, in this subset of loans represented in the HMDA data, there is no such indication when considering the cross-sectional distribution of ZIP codes by IRS income. Figure 6 plots the shares of mortgage origination by value. Notably, the wealthiest quintile of ZIP codes have accumulated more than 50% of the debt each year. However, there is little evidence of redistribution of debt towards wealthier ZIP codes. This result is also holds when considering quintiles of applicant income. Therefore, there has not been an increase in approval rates only in high-income ZIP codes. Credit quality is the next obvious dimension along which credit access could be assessed. Paralleling the trend around regressive redistribution along income, Goodman et al (2018). document that credit access remains limited for low credit score borrowers. However, Fannie Mae and Freddie Mac loan data suggest an opposite trend. Figure 7 shows that the credit score across different percentiles in this set of loans has decreased since 2010, which suggests that FHFA sponsored loans have expanded credit access since 2010 to increase origination toward lower credit score borrowers.

Fannie Mae and Freddie Mac loans being issued for lower credit score borrowers is also reflected in the total origination value for different quintiles of FICO scores. Figure 8 shows that there has been lower overall mortgage debt been issued to ZIP codes with the highest median FICO scores. The proportion of debt issued to the second, third and fourth quintiles has increased monotonically since 2012. This parallels the trend of lower credit scores in Fannie Mae and Freddie Mac loans since these are the regions that have experienced higher proportions of debt issuance. In conjunction with this pattern is the key piece of data on incomes in these regions. Figure 8 shows that median income is higher in ZIP Codes where FICO scores are lower. In fact, the highest FICO score quintile of ZIP codes have slightly lower incomes in 2015 than in 2010 on average. There is a strong signal here from the market that it is attaching a higher weight to incomes for lower credit score regions now than in 2010. This also shows that there is a motive for lenders to expand credit in these regions, namely, growing income levels. If it is the case that there is a increased sensitivity to income, given a credit quality level, then it also represents a systemic shift towards assessing a loan application in a more forward-looking perspective by lenders than they have in the past.

5 Microevidence - Regression Analysis

In this section, I test three specific hypotheses. First, I argue that income differences between regions has a positive impact on mortgage origination, including approval rates. Second, I assert that between 2010 and 2015, income is more closely tied to measures of mortgage origination as well as demand for mortgage debt, along the extensive and intensive margins. Finally, I empirically show that holding credit quality fixed, income growth within a given region (a county, three-digit ZIP code or FICO-score group of ZIP Codes), is positively associated with approval rates. Moreover, there is causal evidence through the use of an instrumental variable that suggests that income growth is an important explanation for higher approval rates and credit expansion.

5.1 Cross-sectional Differences in Income

To begin testing the a hypothesis that claims higher income levels meriting higher approval rates in low credit quality regions, I begin by examining cross-sectional variation within county and cross-county differences in two measures of income - IRS income and Applicant Income - as explanatory variables of differences in approval rates. The level of measurement is a census tract - the most granular geographic identifier available in the HMDA data-set. The other covariates include median measures of FICO scores, CLTV ratios and DTI ratios taken at their median for a given ZIP code that was matched with a census tract. Year and county fixed effects control for other unobserved temporal and regional differences. This specification is given by:

$$Q_{i,t} = \beta Ln(y)_{i,t,j} + \gamma CLTV_{i,t} + \rho FICO_{i,t} + \phi DTI_{i,t} + FE_t + FE_{county} + \epsilon_{i,t} \quad (1)$$

Where:

- Q: Approval rate
- y_j: Median Income - IRS or Applicant Income
- $CLTV$: Median Combined Loan-to-Value ratio
- $FICO$: Median FICO (Credit) Score
- DTI: Median Debt-to-Income Ratio (Percentage Points)
- FE_t: Year Fixed Effects
- FE_{county}: County Fixed Effects

In the columns 1 and 2 of Table 1, the explanatory variable of interest in median applicant income. The coefficient is positive and significant both within counties and across counties, as represented by the models with and without county fixed effects respectively. The same result also holds for the coefficient of the median FICO score. Similarly, there is a positive and significant coefficient for median taxable income in the region, both within and across counties. This specification suggests that income is positively associated with approval rates, and perhaps more importantly, even within-county differences in approval rates can be partially explained by either measure of income, while holding credit quality and other loan features fixed.

5.1.1 The Bartik Instrument for Income

To facilitate the marginal exogenous variation of IRS income across counties, I estimate the specification using an instrumental variable. I have used a Bartik instrument for IRS income as constructed in Goldsmith-Pinkham et. al. (2017). The instrument constructs a measure of industry growth for each county. The CBP data from the Census is only available at the county level and therefore all analysis that include the Bartik instrument will be estimated with a county as a single instance of data. The Bartik instrument is computed as follows. Consider income growth $y_{l,t}$,at time t and location l, the Bartik instrument is then given by:

$$\tilde{y_{l,t}} = Z'_{l,t}G_{t,t-1} \tag{2}$$

where $Z'_{l,t}$ is a vector of local industry shares for total number of locations L at time t, $G_{t,t-1}$ is a $K \times 1$ vector of the overall growth in the national median of incomes earned by employees in an industry between years t and $t-1$. Employment share of a given industry is fixed to year t. Therefore the Bartik Instrument is the inner product of these two vectors. This can be used to estimate a causal effect of income growth in a given region on demand for mortgage debt. More precisely, the reduced form using the instrument would be given by:

$$Mtg_{i,j} = \beta_j \tilde{y}_i + \boldsymbol{X_i'}\phi_j + FE_t + \epsilon_i \tag{3}$$

Where:

- $Mtg_{i,j}$: Measure of origination: approval rates, number of loans originated or average loan size.
- y: Median Taxable Income in the region
- $\boldsymbol{X_i}'$: Vector of controls (same as in tract-level regression)
- FE_t: Year Fixed Effects

The results are presented in Table 2. Here as well, IRS income has a positive and significant coefficient for all three measures of mortgage origination, this time, instrumented using the Bartik instrument. The coefficients are small here, note however, that this specification uses the dollar value of income, which is not on a log scale, therefore the marginal effect of 1 dollar change is small. Another interesting insight from these results is the negative coefficient on FICO scores in column 2, for number of approved or originated loans as the response variable. This confirms the remark made from aggregated data in Figure 7, which suggested that regions with lower credit scores share a higher proportion of debt than before. This result confirms this at the county level, using the means of the coefficient for the six years between 2010 and 2015.

5.2 The Convergence of Mortgage Origination and Income Over Time

To estimate the year-to-year differences in sensitivity to income, I further estimate the differences in the income coefficient at the census tract level, with county fixed effects. At this level, I also explore the interaction between IRS income and FICO scores. These specifications use 3 different mortgage origination measures j: approval rates, number of loans originated and average loan size. The year-interaction effects specification is given by:

$$Mtg_{i,j} = \beta_j(Ln(y_i) \times Y_t) + FE_{county} + \epsilon_i \quad (4)$$

These results are presented in Table 4. The coefficients of interest here are primarily IRS Income interacted with the year. The first two columns represent approval rates as the response, columns 3 and 4 use the median loan size [2] and columns 5 and 6 have the number of loans originated as the response. An increasing sensitivity to income would be represented by increasingly positive coefficients for the interaction terms, for every origination measure. The coefficient for Ln(IRS Income)[3] represents the average effect of IRS Income on the origination measure and the interaction terms represent the marginal effect of income levels *relative* to 2010 levels. The interaction terms for approval rate don't have a monotonically increasing coefficient, however they are all positive and significant. On the other hand, the coefficients for both the other origination measures increase sequentially with time. The interaction term coefficient for the year 2012 is not significant for loan size, with county fixed effects (column 3), but notably, the coefficient is positive, large and significant for the following years. The same is true for the interaction terms for years 2011 and 2012 in column 5, with county fixed effects, where the subsequent coefficients are all positive and significant. There are two conclusions that can be drawn from this analysis. First, income was more closely tied to both the extensive and intensive margins of mortgage origination as we progress from 2010 to 2015 within counties. Second, the same was also true comparing counties across the country.

Adapting the line of argument in Schoar (2016) and Mian (2016), and to build on this result further, there are two important implications of the trends highlighted above, at both the census tract level and the causal specification at the county level. Mortgage

[2] the loan sizes in HMDA are in thousands of USD and this analysis maintains that unit of measurement.
[3] Given in the first row of table 4.

market origination measures being more closely tied to income levels suggests that the revival in the volume of mortgage debt is coherent with *real* changes in the economy. This could also mean that higher income levels are allowing more households to take on debt.

5.3 The Impact of Income Growth Within a County and ZIP Code

One of the main reasons that a higher share of debt has been originated in lower FICO score regions progressively over this time period is that income levels in these areas are are increasing. To estimate this change and its impact on approval rates, I estimate the following simple specification on a panel data set aggregated to the three-digit ZIP code level[4]

$$Q_i = \beta Ln(y_i) + \gamma FICO + FE_{three-digitZIP} + \epsilon_i \tag{5}$$

In the above equation, by holding the three digit ZIP code fixed with the $FE_{three-digitZIP}$ fixed effect, the only variation within the ZIP Code is between the years. Therefore, the coefficient of interest β indicates the effect on approval rates of temporal changes in median IRS income in the region. Table 5 presents the results. Column 1 represents the above specification. Column 2 represents:

$$Q_i = \beta Ln(y_i) + FE_{FICOquintile} + \epsilon_i \tag{6}$$

and finally, the third column represents the above model with year fixed effects. Broadly, the results confirm that income growth within ZIP codes predict better approval rates over time. Column 1 represents this result with a positive and significant coefficient for IRS income, holding FICO scores fixed, within a three-digit ZIP code. Extending this insight, even within each quintile of FICO score ZIP codes, higher income levels are asso-

[4]This allows for a cleaner comparison of FICO scores since the smallest geographic identifier in Fannie/Freddie loans was the three-digit ZIP code.

ciated with higher approval rates, both between years and within each year, as given by the positive coefficients in columns 2 and 3. Further, I note that higher income levels also identify a higher number of applicants in a given ZIP code, and therefore, the number of loans originated is also the highee. This result is presented in Table 6.

I show that these results have causal implications through the use of the Bartik instrument, at the county level. I estimate the specification in equation 5 with IRS income instrumented using the Bartik instrument. Table 7 presents the results and confirms that the coefficients of interest are positive and significant. This underscores the causal underpinnings of income growth on increases in approval rates, holding the other most important feature of loans, credit quality, fixed.

5.4 Approval Rate Comparisons for Large Loans

In this section, I further show that credit policies over this time-period have expanded towards originating a larger class of loans by examining approval rates of specific categories of loans. Specifically, I examine approval rates for the 90th percentile of loans by LTI ratio as of 2010. Table 5 presents the results. In this sample of loans, for each year, I picked loans with an LTI ratio of 3 ± 0.1 and simply compared the number of loans and the proportion of these that were approved. The loans represented here represent risky loans because they belong in the 90th percentile of LTI which means these were the largest loans relative to the applicant's income as defined in 2010. Notably, a lot more loans within this LTI range exist in the data, an indication of the large increase in number of overall loans, since this subset still represents the same proportion of all mortgage applications.

6 Conclusion

There are three broad conclusions that can be drawn from this analysis. First, there is a monotonically increasing correlation between mortgage origination and income levels between 2010 and 2015; mortgage origination has not decoupled from income levels in this expansion (and in turn the real economy) the way that it did in the run-up to the mortgage crisis in 2007-2009. Second, average borrower credit scores are lower. This suggests that at least a section of mortgage lenders are extending credit to lower credit score borrowers in 2015 that may have denied in 2010. Third, and most importantly, income level increases have resulted in higher approval rates and volume of mortgage origination, even in the lowest FICO score quintile ZIP codes. Therefore, it is not surprising to see the surge in mortgage applications over this period. Between the growth rate in applications and income levels, the total originated debt volume has increased, even with moderate increases in the aggregate approval rates.

Therefore under specific conditions post the crisis, including regulation, lower levels of securitization and the dominance of large lenders, the market has expanded on the basis of increasing income levels. This result improves our understanding of mortgage markets under unusual conditions, in this case: recovering from a large-scale mortgage crisis. It shows that mortgage lenders are looking for ways to expand credit supply and if they can find specific parameters other than credit scores with which also to lower credit risk, then they will go after it. This study also then implies that as the U.S. economy continues its economic recovery with employment growth and wage increases, its effects are likely to be positively transmitted to the mortgage market. On the borrowers' side, there is positive effect of income shocks on the propensity to borrow, as established by the relation between income levels and the number of applicants in section 5, which may be predictive of higher levels of both demand and ultimately higher levels of mortgage origination along the extensive margin as the economy expands.

References

[1] ADELINO, M., A. SCHOAR AND F. SEVERINO (2016): ”Loan Originations and Defaults in the Mortgage Crisis: The Role of the Middle Class” The Review of Financial Studies vol. 29, n. 7.

[2] A. DEFUSCO, S. JOHNSON, J. MONDRAGON (2017): ”Regulating Household Leverage” DeFusco, Available at SSRN: https://ssrn.com/abstract=3046564

[3] D. GREENWALD (2016): ”The Mortgage Credit Channel of Macroeconomic Transmission”, MIT Sloan Research Paper No. 5184-16. Available at SSRN: https://ssrn.com/abstract=2735491 or http://dx.doi.org/10.2139/ssrn.2735491

[4] A. MIAN, A. SUFI (2016): ”Fraudulent Income Overstatement on Mortgage Applications During the Credit Expansion of 2002 to 2005”. Kreisman Working Papers Series in Housing Law and Policy No. 21. Available at SSRN: https://ssrn.com/abstract=2561366 or http://dx.doi.org/10.2139/ssrn.2561366

[5] A. MIAN, A. SUFI (2016): ”Household Debt and Defaults from 2000 to 2010: The Credit Supply View”. Kreisman Working Papers Series in Housing Law and Policy No. 28. Available at SSRN: https://ssrn.com/abstract=2606683 or http://dx.doi.org/10.2139/ssrn.2606683

[6] P. GOLDSMITH-PINKHAM, I. SORKIN, H. SWIFT (2017): ”Bartik Instruments: What, When, Why and How”. Stanford Institute for Economic Policy Research Working Paper No. 17-052.

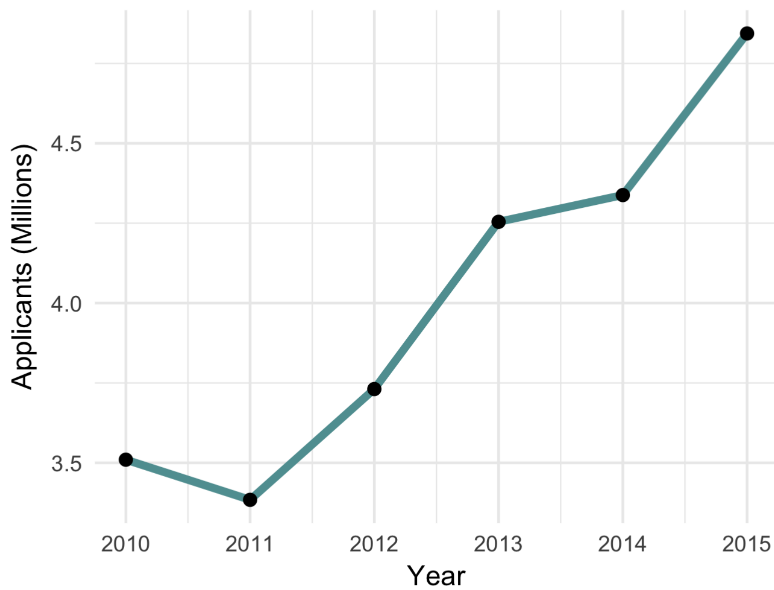
[7] L. GOODMAN (2018): ”Housing Finance at a Glance: A Monthly Chartbook, April 2018”. Urban Institute Research Report Series.

[8] M. BORDO, J. DUCA (2018): "The Impact of the Dodd-Frank Act on Small Business". NBER Working Paper No. 24501. Available at NBER: http://www.nber.org/papers/w24501.

[9] F. D'ACUNTO, A. ROSSI (2017): "Regressive Mortgage Credit Redistribution in the Post-crisis Era". Robert H. Smith School Research Paper No. RHS 2833961. Available at SSRN: https://ssrn.com/abstract=2833961 or http://dx.doi.org/10.2139/ssrn.2833961.

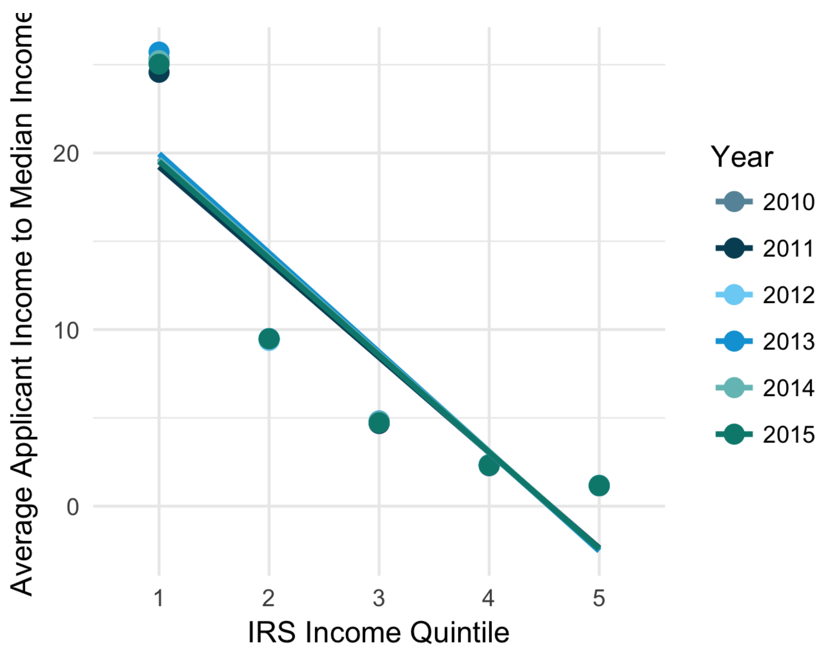
[10] C. FOOTE, P. WILLEN (2016): "Cross-Sectional Patterns of Mortgage Debt during the Housing Boom: Evidence and Implications". Federal Reserve Bank of Boston Research Department Working Paper No. 16-12.

Figure 1: Number of Applicants by Year



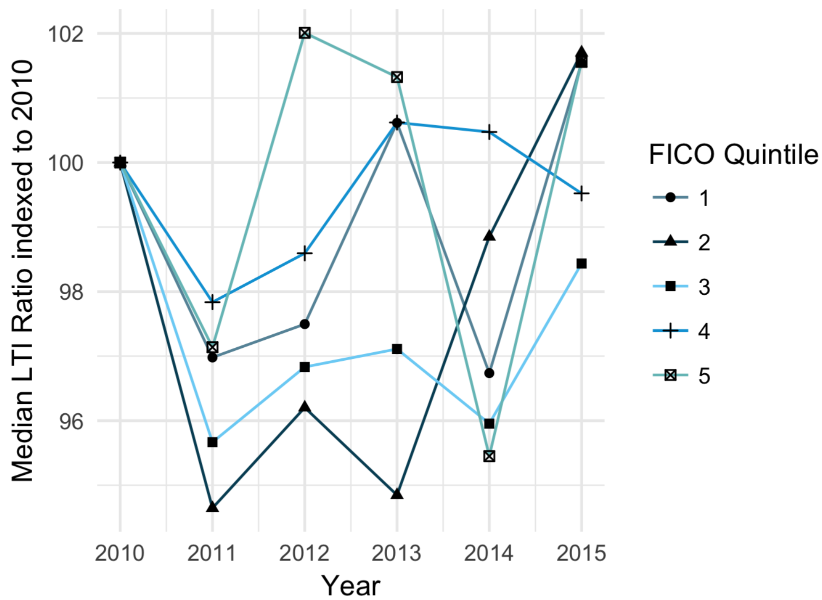
The figure plots the total number of applicants by year in the HMDA data-set. Data Sources: HMDA

Figure 2: Applicant Income - Median Income Ratio by IRS Income Group



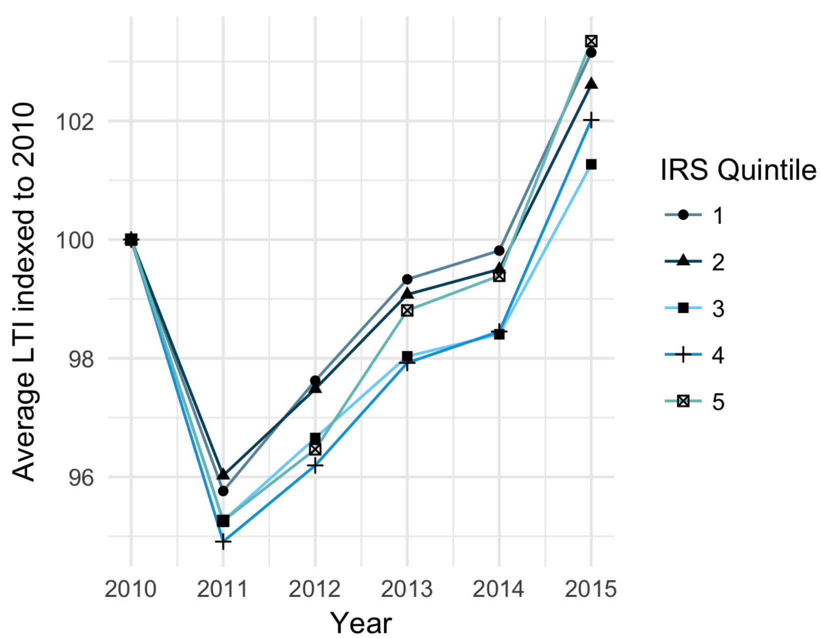
The figure plots mean applicant income to median income in zipcode per IRS income quintile of ZIP codes. Data Source: HMDA, IRS

Figure 3: LTI by FICO Group



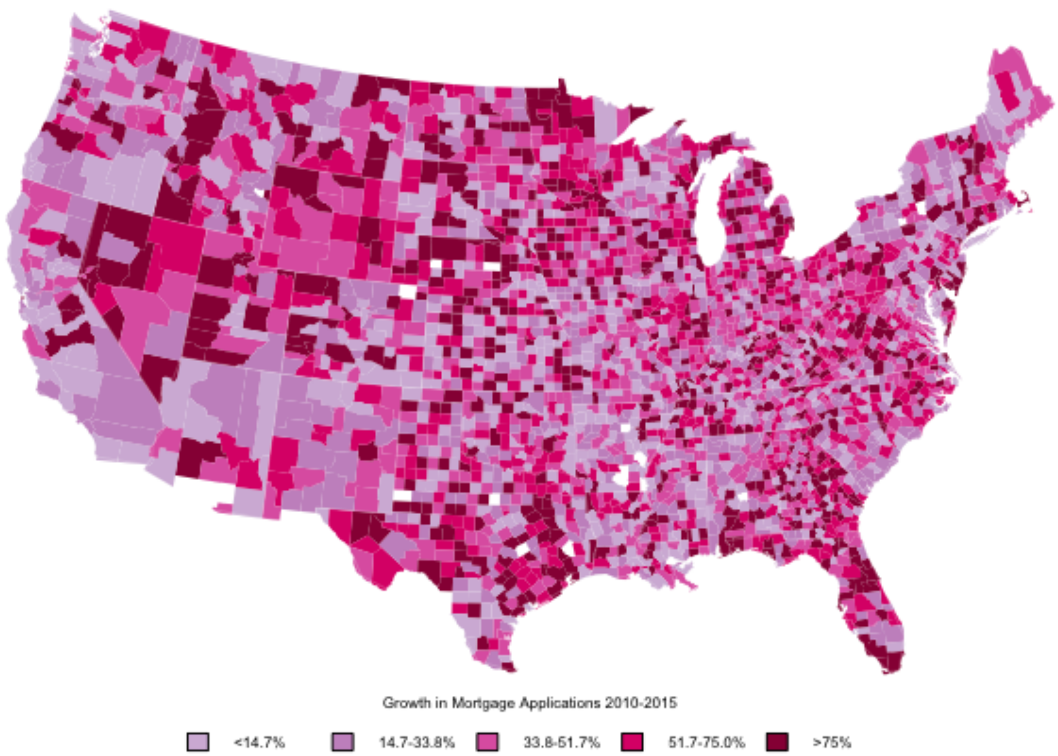
The figure plots mean Loan-to-Income ratio by FICO-score quintiles of Counties income quantiles that loans fall under.

Figure 4: LTI by IRS Income Group



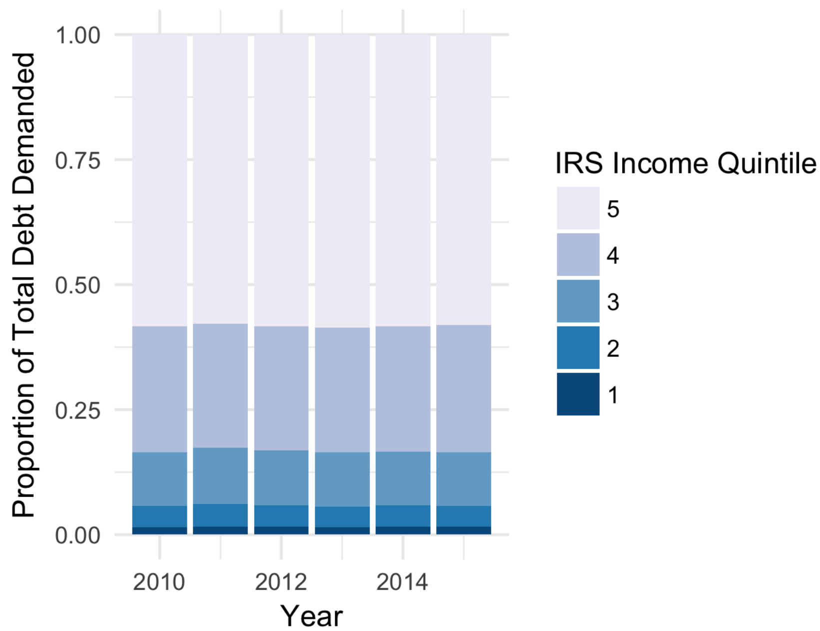
The figure plots mean Loan-to-Income ratio by IRS income quintiles of ZIP codes.

Figure 5: Total Increase in Mortgage Applications 2010-2015 by County



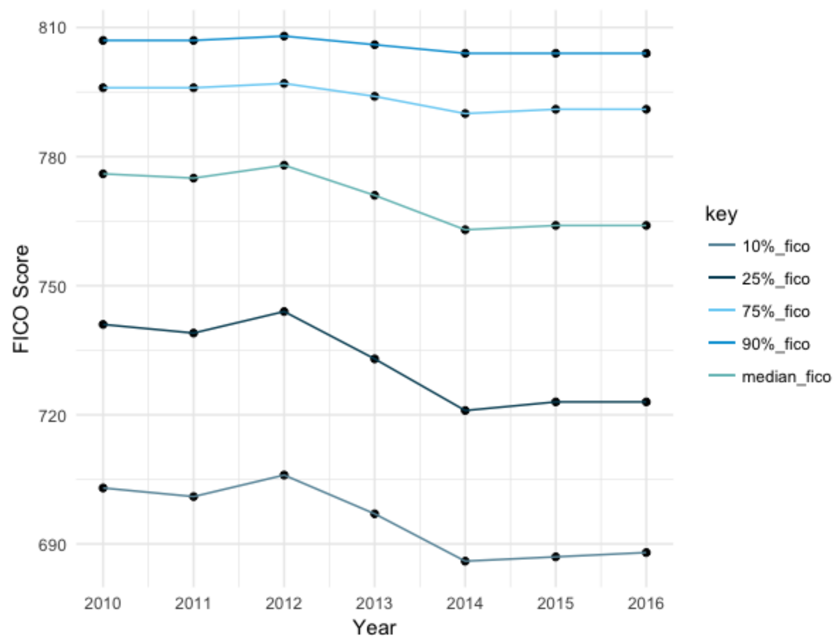
The figure plots the percentage change in total number of applications in the HMDA data per county, by diving the growth rate into 5 equally weighted buckets. Data Source: HMDA

Figure 6: Proportion of Debt Issued by IRS Income Group



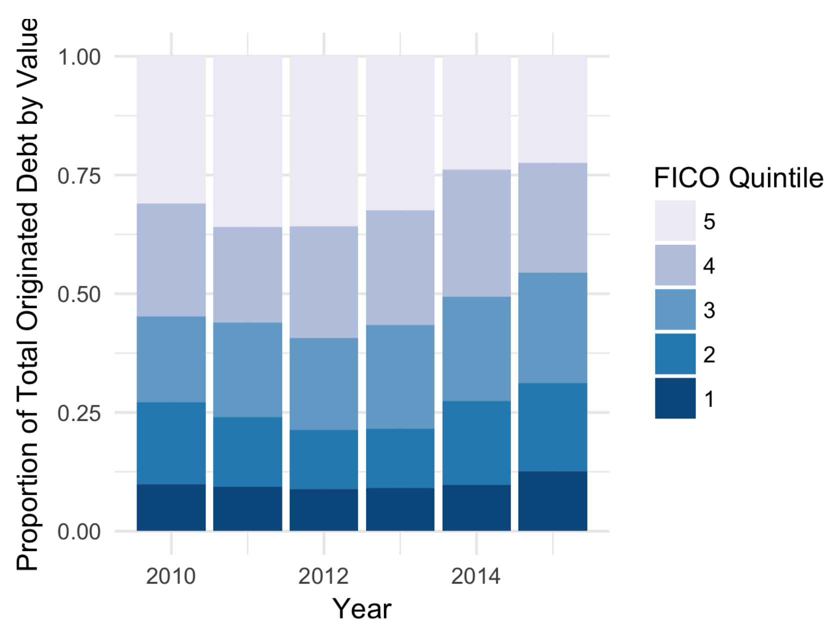
The figure plots the share of the total dollar value of mortgage debt that has been issued to each quintile of ZIP codes by median IRS income, defined within each year. Data Source: HMDA, IRS

Figure 7: FICO Score Percentiles By Year



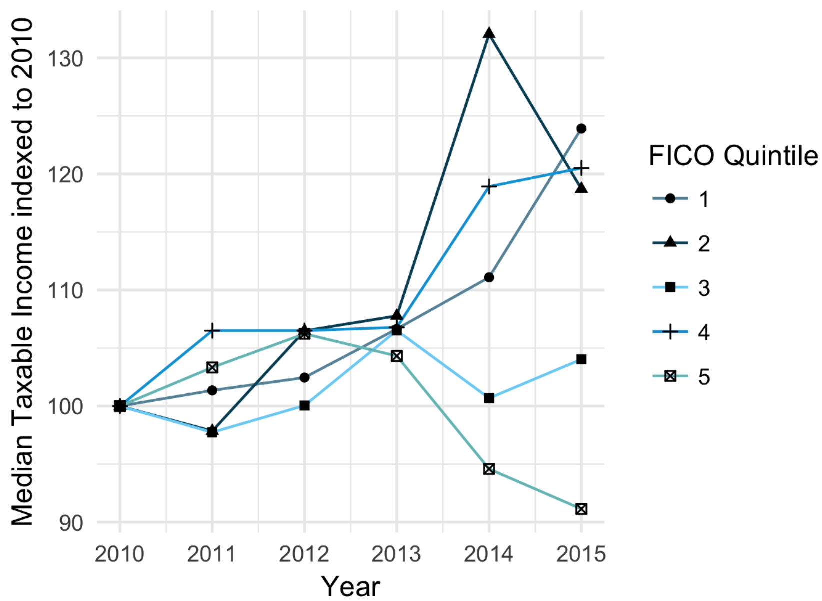
The figure plots different percentiles of FICO scores by year. Data Source: Fannie Mae, Freddie Mac

Figure 8: Proportion of Debt Issued by FICO Score Group



The figure plots the share of the total dollar value of mortgage debt that has been issued to each quintile of ZIP codes by median FICO score, defined within each year. Data Source: HMDA, Fannie Mae, Freddie Mac.

Figure 9: Income by FICO Score Group



The figure plots Median IRS Income, indexed to 2010, within each quintile of ZIP Codes by median FICO score, defined within each year. Data Source: IRS, Fannie Mae, Freddie Mac.

Table 1: This census tract level panel regression estimates the effect of differences in levels of two types of income: median loan applicant income and median IRS income in the region on approval rates, with CLTV, FICO scores and DTI controls, all estimated at the median for each tract. Columns 1 and 2 use applicant income as the main independent variable, whereas columns 3 and 4 use IRS income. The specification is estimated both with and without county fixed effects, however each specification uses year fixed effects.

	Applicant Income		Median Taxable Income	
	(1)	(2)	(3)	(4)
Applicant Income	9.927***	10.320***		
	(0.056)	(0.055)		
Ln IRS Income			3.666***	3.535***
			(0.026)	(0.022)
CLTV	0.191***	0.744***	−0.021	0.380***
	(0.013)	(0.008)	(0.013)	(0.008)
FICO	0.154***	0.433***	0.201***	0.456***
	(0.008)	(0.004)	(0.008)	(0.004)
DTI	−0.017	−0.582***	0.056	−0.531***
	(0.035)	(0.012)	(0.035)	(0.012)
County Fixed effects?	Yes	No	Yes	No
N	432,440	432,440	432,404	432,404
R^2	0.282	0.126	0.263	0.109
Adjusted R^2	0.277	0.126	0.258	0.109
Residual Std. Error	13.539 (df = 429316)	14.889 (df = 432430)	13.719 (df = 429281)	15.031 (df = 432394)

Notes:

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table 2: Here the estimation from Table 1 is repeated with the Bartik Instrument used for IRS income, at the county level. I have note estimated applicant income effects. In this case, there are three mortgage origination measures used as the response: approval rate in column 1, number of originated loans or “approved count“ in column 2 and LTI ratio in column 3, all within-year estimates.

	Approval Rate	Approved Count	LTI Ratio
	(1)	(2)	(3)
IRS Income	0.0002***	0.275***	0.00001***
	(0.00001)	(0.009)	(0.00000)
FICO	0.407***	−127.089***	0.016***
	(0.014)	(8.527)	(0.001)
CLTV	0.521***	93.505***	−0.038***
	(0.051)	(31.733)	(0.002)
DTI	−0.215***	−370.029***	0.066***
	(0.062)	(38.998)	(0.002)
Year Fixed effects?	Yes	Yes	Yes
N	18,455	18,455	18,455
R^2	0.201	0.168	0.361
Adjusted R^2	0.201	0.168	0.361
F Statistic (df = 4; 18445)	1,145.759***	−3,868.203	2,584.230***

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table 3: This specification uses census tract level data to compare the the coefficient of IRS income across years 2010-2015, by using a year interaction effect to illustrate the raw correlation between income in each year and approval rates in columns 1 and 2, median loan size in columns 2 and 3 and number of loans originated in 3 and 4. The specification is estimated with county fixed effects for each response variable, in columns 2, 4 and 6.

	Approval Rate		Median Loan Size		Loans Originated	
	(1)	(2)	(3)	(4)	(5)	(6)
Ln IRS Income	2.671***	2.435***	−0.174	27.190***	7.823***	8.155***
	(0.046)	(0.048)	(0.319)	(0.416)	(0.150)	(0.152)
2011	−11.133***	−10.883***	3.009	−18.104***	−0.980	−1.310
	(0.656)	(0.733)	(4.589)	(6.415)	(2.160)	(2.344)
2012	−16.996***	−17.335***	2.669	−90.205***	−5.847**	−16.914***
	(0.703)	(0.783)	(4.920)	(6.855)	(2.315)	(2.504)
2013	−13.807***	−14.040***	−31.523***	−117.627***	−13.628***	−23.574***
	(0.692)	(0.770)	(4.841)	(6.746)	(2.279)	(2.465)
2014	−13.287***	−13.542***	−59.794***	−146.809***	−11.636***	−21.858***
	(0.693)	(0.771)	(4.848)	(6.755)	(2.282)	(2.468)
2015	−9.516***	−9.487***	−87.233***	−184.616***	−20.496***	−30.554***
	(0.699)	(0.777)	(4.886)	(6.804)	(2.300)	(2.486)
Ln IRS Income × 2011	1.001***	0.979***	−0.769*	1.149*	−0.089	−0.052
	(0.062)	(0.069)	(0.431)	(0.602)	(0.203)	(0.220)
Ln IRS Income × 2012	1.648***	1.676***	−0.165	8.242***	0.186	1.377***
	(0.066)	(0.073)	(0.460)	(0.641)	(0.216)	(0.234)
Ln IRS Income × 2013	1.360***	1.379***	4.259***	12.007***	1.404***	2.500***
	(0.065)	(0.072)	(0.452)	(0.630)	(0.213)	(0.230)
Ln IRS Income × 2014	1.464***	1.485***	7.865***	15.572***	1.360***	2.477***
	(0.065)	(0.072)	(0.452)	(0.630)	(0.213)	(0.230)
Ln IRS Income × 2015	1.200***	1.195***	11.560***	20.118***	2.695***	3.792***
	(0.065)	(0.072)	(0.455)	(0.634)	(0.214)	(0.232)
Constant		39.908***		−118.630***		−48.078***
		(0.506)		(4.427)		(1.618)
County Fixed effects?	Yes	No	Yes	No	Yes	No
N	432,405	432,405	432,405	432,405	432,405	432,405
R^2	0.263	0.068	0.542	0.092	0.208	0.054
Adjusted R^2	0.258	0.068	0.539	0.092	0.202	0.054
Residual Std. Error	13.718 (df = 429280)	15.370 (df = 432393)	95.953 (df = 429280)	134.611 (df = 432393)	45.159 (df = 429280)	49.179 (df = 432393)

Notes:

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table 4: Approval Rates for High LTI ratio Loans

	Year	Approval Rate	Proportion of all loans	LTI	Number of Loans
1	2010	0.721	0.055	3	194,700
2	2011	0.718	0.052	3	175,683
3	2012	0.732	0.052	3	194,557
4	2013	0.730	0.053	3	227,025
5	2014	0.741	0.055	3	239,657
6	2015	0.749	0.056	3	271,418

Table 5: This specification estimates the effect of income growth within a three-digit ZIP code using a panel data set, with approval rates being the response in all columns. In column 1, approval rates are regressed with log IRS income with median FICO scores as a covariate, with three-digit ZIP code fixed effects. Column 2 again uses approval rates but does not use FICO scores as a covariate, and uses a yearly-defined median FICO score quintile. Finally column 3 represents the same specification as column but with year fixed effects.

	(1)	(2)	(3)
Ln IRS Income	0.117^{***}	0.051^{***}	0.050^{***}
	(0.007)	(0.002)	(0.002)
FICO	-0.001^{***}		
	(0.0001)		
3-digit ZIP Fixed effects?	Yes	No	No
FICO Score Quintile Fixed effects?	No	Yes	Yes
Year Fixed effects?	No	No	Yes
N	5,303	5,303	5,303
R^2	0.923	0.315	0.331
Adjusted R^2	0.908	0.314	0.330
Residual Std. Error	0.027 (df = 4417)	0.073 (df = 5297)	0.072 (df = 5292)

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table 6: This specification estimates the effect of income growth within a three-digit ZIP code using a panel data set, on the number of applications and number of loans originated. In column 1, number of applications is regressed with log IRS income with median FICO scores as a covariate, with three-digit ZIP code fixed effects. Column 2 uses the number of loans originated as the response.

	No. of Applications	Loans Originated
	(1)	(2)
Ln IRS Income	5,029.820***	4,092.115***
	(282.972)	(214.065)
FICO	−56.014***	−45.652***
	(2.867)	(2.169)
3-digit ZIP Fixed effects?	Yes	No
N	5,303	5,303
R^2	0.966	0.961
Adjusted R^2	0.959	0.954
Residual Std. Error (df = 4417)	1,085.211	820.950

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table 7: This TSLS specification estimates the effect of income growth within a three-digit ZIP code using a panel data set, with approval rates being the response in all columns, using the Bartik Instrument to arrive at a causal estimate of the effect of income growth on approval rates. In column 1, approval rates are regressed with log IRS income with median FICO scores as a covariate, with three-digit ZIP code fixed effects. Column 2 again uses approval rates but does not use FICO scores as a covariate, and uses a yearly-defined median FICO score quintile. Finally column 3 represents the same specification as column but with year fixed effects.

	(1)	(2)	(3)
Ln IRS Income	3.464***	12.107***	12.107***
	(0.070)	(0.468)	(0.468)
FICO	0.409***		
	(0.011)		
County Fixed effects?	Yes	No	No
FICO Score Quintile Fixed effects?	Yes	Yes	Yes
Year Fixed effects?	No	No	Yes
N	18,449	18,449	18,449
R^2	0.193	0.128	0.128
Adjusted R^2	0.193	0.128	0.128
F Statistic	2,199.058*** (df = 2; 18441)	−6,694.030 (df = 1; 18442)	−6,694.030 (df = 1; 18442)

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.