The Effect of Mortgage Credit Availability on House Prices and Construction: Evidence from a Frontier Estimation Approach *

Elliot Anenberg† Aurel Hizmo‡ Edward Kung§
Raven Molloy¶

November 22, 2015

Abstract

We present new evidence on changes in mortgage credit availability from 2001 to 2014 and estimate its effects on house prices and residential construction. To isolate changes in mortgage credit supply from changes in credit demand, we construct a new measure of supply based on production frontier estimation. This "loan frontier" allows us to examine changes in credit availability for different types of borrowers in different housing markets, dimensions that have yet to be fully explored. Exploiting the disaggregated nature of our measure and national trends in the amount of credit extended to various groups, we construct an instrument for mortgage credit supply. The exogenous variation in the loan frontier can explain 27% of the total variation in price appreciation. We find that a one percent increase in the change in the aggregated loan frontier increases the change in house prices by 0.9 percent and the change in the single-family housing stock by 0.09 percent.

*We thank Leonard Nakamura, Dan Hartley, Tomasz Piskorski, James Vickrey, Giovanni Favara for helpful comments, and Jimmy Kelliher and Jack Popper for excellent research assistance. The analysis and conclusions set forth are those of the authors and do not indicate concurrence by other members of the research staff or the Board of Governors.

†Board of Governors of the Federal Reserve System

‡NYU Stern

§UCLA

¶Board of Governors of the Federal Reserve System

1 Introduction

A common narrative associated with the housing market over the past decade is that changes in mortgage credit supply played an important role in explaining the boom and bust in housing prices and residential investment over this time period. Generally speaking, the easiness or tightness of mortgage credit supply should be reflected in the price of credit– i.e. in interest rates. However, most studies have found it difficult to ascribe much of the recent housing cycle to changes in mortgage interest rates.[1] Indeed, mortgages rates declined to historical lows during the recent recession while house prices plummeted.

Besides the price of credit, another important feature of mortgage credit supply is that not every borrower can borrow as much as he or she wants or needs. For example, some borrowers cannot borrow more than 80 percent of their house value, while others cannot borrow an amount greater than the conforming loan limit. We refer to such limits on the quantity of credit as *mortgage credit availability*.[2] Credit availability can co-move with interest rates, but the two are not necessarily direct functions of each other. Therefore, it is possible that mortgage credit availability moved up and down over the past decade even as real interest rates fell steadily over this period.

Relative to the literature on the effects of interest rates, the existing literature on the effects of mortgage credit availability on housing market outcomes is less developed. Several studies present evidence that certain elements of mortgage credit availability loosened during the 2000s (see Mian and Sufi (2009), Demyanyk and Van Hemert (2011), Nadauld and Sherlund (2013), Keys et al. (2010)), but these studies do not characterize changes in more comprehensive measures of mortgage credit availability or their general effects

[1] See, for example, Glaeser et al. (2010), which argues that interest rates cannot explain much of the boom in prices, or Adelino et al. (2012) which estimates a small elasticity of house prices with respect to interest rates.

[2] Geanakoplos (2010) provides a theoretical framework to understand why collateral constraints, in addition to the interest rate, plays an important role in equilibrium mortgage credit supply. We use the generic term *credit availability* rather than downpayment or collateral constraints because borrowing can be constrained even if downpayment is allowed to be zero.

on housing market outcomes. One important obstacle to this line of research is that, as noted by Glaeser et al. (2010) and Li and Goodman (2014), there are few direct measures of credit availability that convincingly disentangle mortgage credit supply from mortgage credit demand. Some recent papers that estimate an elasticity of house prices with respect to mortgage credit supply by instrumenting for imperfect measures of mortgage credit supply are generally unable to disentangle the effects of mortgage interest rates from credit availability (e.g. Favara and Imbs (2015) and Maggio and Kermani (2015)). Therefore, more evidence is needed to assess whether mortgage credit availability played a prominent role in explaining the dynamics of the housing market over the past decade.

To illustrate the difficulty in measuring mortgage credit supply separately from demand, consider a commonly used measure of credit availability: the approval rate on loan applications. The approval rate does not isolate supply from demand because a change in the approval rate could result from a change in the distribution of applicants, rather than a change in credit availability. The approval rate is also difficult to interpret because borrowers are unlikely to apply for loans for which they are likely to be rejected. Another commonly used measure of availability, the median borrower credit score of new originations, has similar limitations. Indeed, Figure 1 shows that if approval rates or median credit scores were used as measures of credit availability, then one would think that credit tightened or stayed the same between 2000 to 2006, an impression that runs counter to much of the narrative evidence that credit supply expanded during this period. In the Appendix, we discuss the attractive features and limitations of some other measures of credit availability that have been developed.

The second challenge facing any analysis of mortgage credit and house prices is the simultaneity between these two variables, which arises due to the unique nature of housing as collateral for loans. The use of the home as collateral means that expectations of future house price growth are an important determinant of credit availability, and these expectations are likely to be affected by past house price growth. Thus, areas with strong house price appreciation in the past may also have greater credit availability. At the same time, credit availability influences the demand for housing, which in turn can affect house prices and house price expectations. Consequently, any study which seeks to estimate the effect of credit availability on house prices needs a source of variation in credit market conditions that is exogenous to local house price movements.

In this paper, we will tackle both of these challenges by proposing a new measure of mortgage credit availability. Not only does this new measure isolate changes in credit supply from changes in credit demand, but it also provides a convenient way to construct a new instrumental variable for credit availability that is independent from local housing market conditions. We use this measure to first describe mortgage credit availability conditions, and then to estimate the elasticity of house price growth and housing stock growth with respect to changes in mortgage credit availability.

Our method is motivated by the literature on estimating production frontiers. In the mortgage context, the outputs of the production process are the characteristics of the loan that a lender is willing to underwrite, such as loan amount and the required downpayment, and the inputs are borrower characteristics such as credit score and income. We interpret changes in the loan frontier as changes to mortgage credit availability. The rationale is that for any given set of borrower characteristics, it is likely that at least a small number of potential borrowers would demand the maximum amount of credit available to them. Therefore, by focusing on the frontier, we are capturing changes in lender policy, not borrower demand.

We follow the approach of Cazals et al. (2002) to estimate the frontier non-parametrically using data on mortgage originations. Given the available data, the estimated frontier may be interpreted as the maximum loan amount that a borrower could obtain, given her FICO score, downpayment amount, income, metropolitan area, and year of origination.[3] Our estimated loan frontiers suggest that increases in credit availability during the boom (2001-2006) were

[3]The loan amount is a "combined" loan amount in that it includes the balance of simultaneous second liens at the time of origination.

fairly similar across borrower types and metropolitan areas, whereas the contraction in credit (2007-2014) was much sharper for low-score and low-income borrowers.[4]

A key advantage of the loan frontier is that it can be estimated for different borrower types, locations, and time periods, so we have substantial cross-sectional and time-series variation in credit availability. We exploit this variation to create a Bartik-style instrument (Bartik (1991)) for the loan frontier that is plausibly exogenous to local house prices and construction activity. Identification is driven by the fact that shocks to the national credit market that are exogenous to local housing market conditions can still have different effects on different metropolitan areas, depending on the distribution of potential borrowers in the area.

Using our instrument, we find that a one percent increase in the change in the aggregated loan frontier[5] increases the change in house prices by 0.9 percentage points and the change in the single-family housing stock by 0.09 percentage points for a metropolitan area with a mean housing supply elasticity as computed by Saiz (2010). Our estimates are comparable when we add a control for the average interest rate among loans near the frontier, which supports our interpretation that our estimated elasticity is with respect to changes in mortgage credit availability and is not driven by changes to the price of credit. Our instrumented loan frontier can explain about 54% of the variation in price appreciation that cannot be explained by national trends. Overall the exogenous variation in the loan frontier can explain 27% of the total variation in price appreciation.

A number of other papers have tried to empirically assess the role of credit supply in housing markets, using regulatory shocks to achieve identification.[6]

[4]Adelino et al. (2015) and Bhutta (2015) also find evidence consistent with this result. Ferreira and Gyourko (2015) additionally show that the subsequent foreclosure crisis was widespread as well, suggesting that riskier credit was extended to both prime and subprime borrowers during the boom.

[5]The aggregated loan frontier can be interpreted as the maximum loan amount available to the average borrower in a metro-year.

[6]In a unique approach, Fuster and Zafar (2015) use survey evidence to estimate the elasticity of willingness to pay for a house with respect to changes in the downpayment

Adelino et al. (2012) and Kung (2015) measure the effect of exogenous changes to conforming loan limits on house prices. Favara and Imbs (2015) and Maggio and Kermani (2015) use regulatory shocks to instrument for different measures of credit supply, such as the volume of mortgage originations and the loan-to-income ratio.[7] Consistent with our results, each of these papers demonstrates—to varying degrees of magnitude—a positive causal effect of credit supply on house prices.

We make two main contributions to this literature. First, we argue that our estimate provides a cleaner measure of the elasticity of house prices with respect to mortgage credit availability. For example, Favara and Imbs (2015) use deregulation shocks to instrument for the volume of mortgage originations, but it is possible that the deregulation shocks increase the volume of mortgage originations solely through its effect on the interest rate rather than by increasing mortgage credit availability. Second, our measure of credit availability is easily understood in terms of lender policy. Suppose, for example, that lenders lower the credit score requirement on jumbo loans. This would show up as an increase in the loan frontier of low credit-score borrowers, and it would be easy to use our measure to assess this policy's effect on house prices. It would be harder to use price elasticity estimates from a more indirect measure of credit supply, like the volume of originations, because one would first have to predict how the lender policy affects the volume of originations, and then apply the elasticity estimate. Another advantage of our approach is that our instrument does not rely on one-time events to achieve identification, and therefore is more easy to extrapolate to other time periods or policy changes.

2 Methodology

Consider a mortgage origination process in which a borrower of characteristics $x \in \mathbb{R}^p$ (i.e. credit score, income) obtains a loan of characteristics $y \in \mathbb{R}^q$

requirement.

[7]See also Rajan and Ramcharan (2014) for related work on credit availability and land prices in the 1920s, and Gropp et al. (2014) for evidence that changes in credit availability may be the real cause of recent household deleveraging.

(i.e. loan amount, required downpayment). The set of possible mortgage originations is given by:

$$\Psi = \left\{ (x, y) \in \mathbb{R}^{p+q} |\text{ Borrower } x \text{ can obtain loan } y \right\}$$

We assume an ordinal ranking for x and y such that that if $(x, y) \in \Psi$ then $x' \geq x$ and $y' \leq y$ implies $(x', y') \in \Psi$, where the inequality is taken element by element. In words, a borrower with better characteristics can always obtain all the loans available to a borrower with worse characteristics. Similarly, if a borrower could obtain a loan with good characteristics, then the same borrower could also obtain a loan with worse characteristics.[8]

Formulated in this way, the mortgage origination process is equivalent to a production process with free disposal in which the borrower characteristics are inputs and the loan characteristics are outputs. The econometric problem is to estimate Ψ from a random sample of mortgage originations, $\{x_i, y_i\}_{i=1}^n$. Cazals et al. (2002) (henceforth CFS) describe a robust non-parametric approach to this problem, which we adopt in this paper.

To illustrate the CFS method, we begin with the case of a single output $y \in \mathbb{R}$ and multiple inputs $x \in \mathbb{R}^p$. We note that the possibility set Ψ can equivalently be described by the efficient output frontier $\varphi(x)$, defined as:

$$\varphi(x) = \sup \{ y | (x, y) \in \Psi \}$$

Suppose the data, $\{x_i, y_i\}_{i=1}^n$, are drawn from the joint distribution (X, Y). Let us define the expected maximum output function of order m, $\varphi_m(x)$, as:

$$\varphi_m(x) = E[\max\{Y_1, \ldots, Y_m\} | X \leq x]$$

Intuitively, $\varphi_m(x)$ is the highest expected level of output that would be observed with inputs less than x, out of m draws.

Following CFS, let us construct the empirical analog to $\varphi_m(x)$. First, let

[8]For loan characteristics where smaller is better (i.e. the required downpayment), we can simply redefine y as measuring the negative of that characteristic. We can also do this with borrower characteristics where smaller is better, such as other debt holdings.

us construct :

$$\hat{S}_{c,n}(y|x) = \frac{\frac{1}{n}\sum_{i=1}^{n} I[y_i \leq y, x_i \leq x]}{\frac{1}{n}\sum_{i=1}^{n} 1[x_i \leq x]}$$

which is the empirical analog of $P(Y \leq y|X \leq x)$. Noting that:

$$P(\max\{Y_1, \ldots, Y_m\} \leq y|X \leq x) = P(Y \leq y|X \leq x)^m$$

we can compute the empirical analog of $\varphi_m(x)$ by the following procedure. First, let $n(x)$ be the number of observations with $x_i \leq x$. Then, denote by y_j^x the jth smallest value of y_i that is observed with $x_i \leq x$; i.e. for $x_i \leq x$ we have $y_1^x < y_2^x < \ldots < y_{n(x)}^x$. Then, we compute:

$$\hat{\varphi}_{m,n}(x) = \hat{S}_{c,n}(y_1^x|x)^m y_1^x + \sum_{j=2}^{n(x)} \left[\hat{S}_{c,n}\left(y_j^x|x\right)^m - \hat{S}_{c,n}\left(y_{j-1}^x|x\right)^m\right] y_j^x$$

CFS establish the asymptotic properties of the estimator, but the key point to note is that $\hat{\varphi}_{m,n}(x)$ is a $\sqrt{n}$-consistent estimator for $\varphi_m(x)$. Therefore, as m and n grow large, $\hat{\varphi}_{m,n}(x)$ approaches $\varphi(x)$, the efficient output frontier. The reason to use a finite m is that choosing a smaller m makes the estimator robust to outliers that may actually fall outside the possibility set (e.g. due to measurement error) while still maintaining the interpretation as an expected minimum out of m draws.[9] $\hat{\varphi}_{m,n}(x)$ is therefore a consistent estimator of the maximum level of output that inputs x could achieve.

To extend the method to multiple outputs, one simply notes that there is no special distinction between inputs and outputs other than in the assumption: $(x, y) \in \Psi$ implies $(x', y') \in \Psi$ if $x' \geq x$ and $y' \leq y$. If one were to take the negative of an output, it would be interpreted as an input according to the above definition. Therefore, one can estimate the efficient frontier for a single output as a function of the all the inputs and other outputs, simply by recasting the other outputs as negative inputs. To illustrate, let the outputs

[9] One source of measurement error in our sample will be that we merge our loan origination data to HMDA data to obtain borrower income, and the merge is not perfect.

be $y^* = (y, z)$ where y is a scalar and z is a vector, and let the inputs be x. Then, the estimator $\hat{\varphi}_{m,n}(-z, x)$ is a consistent estimator for the maximum level of y obtainable when inputs are less than x and non-y outputs are greater than z.

2.1 Discussion and Example

To illustrate further how this frontier can be interpreted, consider an application where the output is loan amount and the inputs are the borrower's credit score. $\hat{\varphi}_{m,n}(x)$ is therefore interpreted as the highest loan amount that a borrower with credit score x could obtain.

Of course, in practice, loan amount is not the only output and credit score is not the only input. The required downpayment and income are relevant outputs and inputs, respectively. Not all inputs and outputs may be observed well in the data. Therefore, it is useful to discuss the interpretation of the frontier when required downpayment and income are ignored in the computation of $\hat{\varphi}_{m,n}(x)$. In this case, $\hat{\varphi}_{m,n}(x)$ measures the maximum loan amount that could be obtained by a borrower with credit score x, *irrespective of downpayment and income*. So if the maximum loan amount obtainable is increasing in income, then $\hat{\varphi}_{m,n}(x)$ is not representative of the average borrower; rather it measures the maximum loan amount obtainable by borrowers with relatively high incomes.[10] In our implementation of the methodology described in the subsequent sections, there will be some omitted inputs due to data limitations (e.g. household wealth). We will instrument for the frontier and conduct robustness checks to make sure that our estimated elasticities of house prices and construction with respect to credit availability are not driven by changes in unobserved characteristics.

For some applications, it will be useful to aggregate the frontier. Suppose we know the distribution of characteristics over the population of potential borrowers, $F(x)$. We can then compute the expected maximum output over

[10]Moreover, if income is correlated with credit score, then $\hat{\varphi}_{m,n}(x)$ is not a structural estimate of the effect of credit score on the maximum borrowing amount; rather it conflates the effects of both income and credit score.

the population of potential borrowers as:

$$\hat{\psi} = \int \hat{\varphi}_{m,n}(x)\, dF(x)\,.$$

$\hat{\psi}$ is therefore an aggregate measure of mortgage credit availability. In the example with loan amount as output and credit score as inputs, it has the clear interpretation as the maximum loan amount obtainable by individuals with average credit score (but possibly high levels for other omitted inputs).

To aid the reader's understanding of the methodology, consider the example in Figure 2, which shows a frontier using loan amount as the output and the borrower's FICO score as the input. The dots represent individual mortgage originations and the solid line is an estimate of the frontier for $m = 1000$. The data sources will be described below.[11] Note that the frontier is not literally the outer envelope of the data. A higher choice of m would result in fewer observations that lie beyond the frontier. $m = 1$ would produce a frontier at FICO level x that is equal to the sample mean of all loan amounts with $FICO < x$.[12]

Also note that the frontier is rather smooth and monotonic. This result occurs because our method assumes that a borrower with a given FICO is as least as credit worthy as an otherwise-similar borrower with a lower FICO, and so the estimate of the frontier at a given FICO uses the data on originations made to lower FICO borrowers. That said, for finite m, our methodology does not impose monotonicity on the frontier, and the frontier could have declined at high FICO scores if there were a sufficiently large number of originations with lower loan amounts at such a FICO score.

This example shows a frontier in only two dimensions. In the analysis below, we will focus on four dimensions: credit score, borrower income, downpayment and loan amount. One important issue worth clarifying is that higher

[11]The example shown here uses data from the Chicago metro area for the year 2012. Borrower income and downpayment are restricted to be less than 150k and 30k, respectively.

[12]In results that are available upon request, we computed the frontier for a sample of GSE loans only and we find that indeed our estimate of the frontier picks up the conforming loan limit.

house prices do not necessarily imply a higher frontier. Since we condition the frontier on a borrower's downpayment, the required downpayment would be larger in higher-priced areas if lending standards are the same. In other words, because variation in house prices affects downpayment size as well as loan amount, it implies movement along the frontier, not shifts in the frontier.

The loan frontier does not capture all aspects of credit supply, most notably the mortgage rate. Incorporating the mortgage rate is complicated because the interest-related costs to the borrower depends on other terms of the loan, such as the amount of origination fees paid (e.g. "points"), the length to maturity, and how the loan ammortizes, and not all of these terms are observable in our data. There is also no clear ordinal ranking for some of these terms. In addition, the loan frontier does not capture potential constraints at the extensive margin, such as how many loans lenders are willing to originate at the frontier. Nonetheless, we find that our measure of credit availability has large effects on housing market outcomes.

3 Data

In applying the CFS methodology to mortgages, we combine two sources of loan-level data. The first source is McDash Analytics, which collects data from a large number of mortgage servicers, including 19 of the 20 largest servicers. Since 2005, McDash has covered roughly 65 to 75 percent of agency loans (i.e. loans subsequently purchased by the GSEs or the FHA), and 20 to 40 percent of loans held on banks' portfolios.[13] McDash covered fewer servicers in the first half of the 2000s. However, the proportions of GSE, FHA, and portfolio loans in the McDash data are fairly similar to the comparable proportions in the aggregate market, so we are reasonably confident that changes in McDash's coverage of these three segments of the market will not influence our results.

The second dataset that we use is compiled by CoreLogic and covers loans that were subsequently sold into non-agency mortgage-backed securities. This

[13] We determine market coverage by comparing total loan volumes for each market segment to aggregate loan volumes published by Inside Mortgage Finance.

dataset has covered more than 90 percent of these loans since 2000. Consequently, when we combine these two data sources, we obtain a dataset that provides a comprehensive picture of all of the major segments of the residential mortgage market since 2000.[14]

Our combined dataset includes many variables of interest related to the mortgage origination including the loan amount, the loan-to-value (LTV) ratio, the borrower's credit score, and the zip code of the property associated with the mortgage loan. To obtain the borrower's income, we merge our loan level data with the confidential version of the Home Mortgage Disclosure Act (HMDA) data.[15] Our match rate ranges between 90-95 percent, depending on the year. The Appendix describes the matching algorithm in detail. The loans that cannot be matched to HMDA are excluded from the rest of the analysis. A number of researchers have found that reported incomes in HMDA appears to be overstated in 2005 and 2006, but not from 1995 to 2004 or from 2007 to 2011 (e.g. Avery et al. (2012), Blackburn and Vermilyea (2012)). In Section 5.2 below we show that our results are similar when we focus on loans with fully-documented income, for which income overstatement is less likely.[16]

Another advantage of merging the loan level data with HMDA is that it allows us to obtain the junior liens that are associated with each first lien at the date of origination.[17] Therefore, we are able to obtain the "combined" LTV and the combined loan amount for each origination. We will use this combined loan amount in the analysis that follows, but we will refer to it simply as the loan amount. For the years 2004-2010 – the years when junior liens are most prevalent, the HMDA data have a flag that connects first liens to junior liens. For the other years, we match first liens to junior liens using an algorithm that we describe in the Appendix.

In this paper we compute the frontier using the loan amount and down-

[14] Although the BlackKnight dataset also includes some non-agency securitized loans, we exclude these loans to avoid double-counting.

[15] For more information on the HMDA data, see Bhutta and Ringo (2014).

[16] In our data, loan originations are classified as fully documented (41%), limited/no documentation (15%), or unknown documentation (44%).

[17] We exclude junior liens taken out after the purchase origination date, such as HELOCS. For more information on second liens, see Lee et al. (2012).

payment level as outputs, and the borrower's FICO score and income as the inputs. We measure the loan amounts, downpayment levels, and incomes in real terms by converting the nominal levels into to 2014 dollars using the price index for personal consumption expenditures. We compute the frontier separately for the 100 largest (by average population between 2001 and 2013) metropolitan areas.

We focus exclusively on purchase originations because we are interested in the extension of new credit to households. After dropping a small number of loans with LTVs>120 and loans with appraisal amounts below $10000 or above $5 million, we are left with a sample of *17 million loans* originated between 2001 and 2014 that we use to compute our frontier.

4 The Loan Amount Frontier

In this section, we apply the methodology developed in Section 2 and the data introduced in Section 3 to construct the "loan amount frontier", which is the maximum loan amount that borrowers are able to obtain in a particular period given their FICO score, income, and downpayment amount.[18] We set the parameter m – defined in Section 2 – equal to 1000.[19] We assign FICO scores, downpayments, and incomes to equally-sized bins and estimate the frontier for each bin in each year and each metropolitan area.[20] We limit the sample to the largest 100 metropolitan areas (as ranked by average population between 2001 and 2013) because cell sizes become too small to reliably estimate a frontier in metropolitan areas with fewer mortgage originations. In particular, using weights that we describe below, we aggregate the loan amount frontiers by

[18] Although we think of downpayment amount as an output, we can still calculate the loan amount frontier as conditional on a given downpayment.

[19] When computing the loan frontier at a given fico, income and downpayment, we first drop the min(5,0.0001*$nobs_j$) largest loan balances (where j indexes the CBSA) to minimize the influence of any measurement error. This drop does not have any effect on the asymptotic properties of the frontier.

[20] We use a FICO grid of 480 to 840 with bins of length 20; income bins of $10,000 from $40,000 to $180,000 with additional bins for $200,000, $250,000 and $1,000,000; and a downpayment grid of $0 to $300,000 with bins of length $10,000.

metropolitan area and bootstrap the aggregated frontiers using 100 repetitions to compute standard errors around our estimates. The 95 percent confidence intervals associated with the aggregate frontier for the 10th, 50th, 100th, 150th, and 200th largest CBSAs are shown in Figure 3. Our estimates of the frontier are fairly precise up until around the 100th largest CBSA, but beyond the 100th largest CBSA, it seems that our dataset does not have enough loans to precisely characterize changes in credit availability given our choice of time frequency for the loan frontier (yearly) and our choice of bin size for the input variables.

Returning to the disaggregated loan frontiers, Figure 4 shows the distribution of the data around the estimated frontier using a histogram, averaged across all borrower types, metro areas, and years. There is a clear discontinuity in the distribution near the estimated frontier. This suggests that there are indeed strict constraints on loan amount given fico, income, and down-payment, and that our estimation approach is doing a good job of identifying them.[21]

Table 1 presents some basic facts about the variance of the loan frontier. The average loan frontier is \$281k and the standard deviation is \$197k. One half of the variance in the frontier can be explained by fixed effects for each FICO bin, illustrating that credit supply is strongly affected by a borrower's credit score. Income is also an important determinant of credit supply, accounting for an additional 13 percent of the variation in the frontier. Metropolitan area fixed effects explain 10 percent of the variation. These differences could reflect differences in the market structure of banks or geographic variation in the types of lenders.[22] Overall, the dimensions of credit that we consider account for 80 percent of the variation in the frontier, with 20 percent reflecting idiosyncratic variation within these categories.

The top-left panel of Figure 5 shows a contour plot of the loan amount

[21] We also plotted the distribution for $m = 500$ and $m = 2000$ and the results looked very similar.

[22] Some of the metro variation could also be due to persistent differences in economic conditions that are not captured by income. For this reason, in the subsequent analysis we focus on the time series variation in the frontier rather than the cross sectional variation.

frontier by FICO and downpayment for a select metro area (Boston) for the year 2004, holding income fixed at $150,000. Not surprisingly, the frontiers indicate that lenders are generally willing to extend larger loans to borrowers with better credit scores and higher downpayments. The dark blue areas of the frontiers indicate that borrowers with very low credit scores were essentially unable to obtain a loan at all in 2004. The top-right panel of Figure 5 shows the frontier in 2012. Credit tightened substantially for lower FICO borrowers between 2012 and 2004.

The bottom-left panel of Figure 5 shows the contour plots by FICO and income, holding downpayment fixed at $50,000 for the year 2004. The frontier generally rises with income, suggesting that lenders are willing to supply more credit to higher income borrowers, even holding constant credit score and downpayment amount. This result supports our claim that the loan frontier is determined by credit supply rather than credit demand. If the frontiers were driven by demand, one might expect higher income borrowers to be associated with lower frontiers conditional on downpayment, as higher income households tend to be wealthier and would not want to lever themselves as much as poorer borrowers, all else equal.

To more completely describe the loan amount frontiers across years and the dimensions of credit that we consider (credit score, income, downpayment and location), we aggregate the frontiers across all dimensions but one, and then examine how the frontier changes along the remaining dimension of credit. Income, credit score, and downpayment are weighted according to the joint distribution of these three variables across all observations in our sample, and metropolitan areas are weighted by population.[23] For example, to asses the importance of credit score we calculate the average frontier for each FICO bin across all downpayments, incomes, and metropolitan areas.

[23]Because our weights are constant over time and across locations, the aggregated frontiers are not a function of changes in observed borrower characteristics over time or differences across locations. An alternative weighting scheme would weight income and credit score according to their shares in the aggregate population, rather than their shares only among mortgage borrowers. However, doing so puts a lot of weight on cells with low credit scores and low incomes, and these cells are imprecisely measured because they contain few mortgage originations.

As shown by Figure 6, consistent with the contour plots the frontier is higher for higher credit scores. Changes over time are striking. From 2001 to 2005 the frontier expanded by 30 to 40 percent for all credit scores above 560. The result that credit did not expand by more for low-score borrowers is consistent with Ferreira and Gyourko (2015), Bhutta (2015), and Adelino et al. (2015). We do this result implies that the growth of the market for private-label MBS did not expand credit to low-score borrowers. Rather, it seems likely that a variety of factors increased mortgage credit availability to high-score borrowers as well during this time period, with the result that credit expanded substantially for all groups.

During the financial crisis, the estimated loan frontiers contracted for all credit scores, but by much larger amounts for borrowers at the lower end of the distribution. Whereas decreases between 2005 and 2011 were in the range of 20 to 25 percent for borrowers with a credit score above 640, the frontier fell by nearly 45 percent for borrowers with a credit score around 620 and by nearly 75 percent for borrowers with scores around 600. For borrowers with even lower scores, the frontier fell to zero, indicating that borrowers with these scores were no longer able to obtain credit.

Turning to income, Figure 7 shows the relationship between income and the frontier in 2001, 2004 and 2013.[24] The frontier shifted up by roughly 35 percent at all incomes above $60,000 from 2001 to 2004, indicating that standards eased by similar amounts for borrowers at all income levels. The frontier shifted back down during the financial crisis,and this shift was larger for lower incomes. For borrowers with incomes betwen $60,000 and $250,000, the 2013 frontier was fairly close to its 2001 level. For borrowers with incomes below $60,000, standards in 2013 were somewhat tighter than in 2001.

Figure 8 shows the relationship between downpayment and the frontier in 2001, 2004, 2008 and 2013.[25] The loan amount frontier is increasing and con-

[24]We examine 2004 to reflect the peak of the housing boom rather than 2005 or 2006 because, as noted above, income misreporting was common in 2005 and 2006.

[25]The frontiers shown in the figure may seem low because they are lower than the maximum loan size allowed by the GSEs. This result can be explained by the fact that the figure shows the average of maximum loan sizes across a range of borrower characteristics, and so

cave in downpayment, illustrating that borrowers can lever themselves more by increasing their downpayment at low downpayment levels, whereas at high downpayment levels larger downpayments are not generally associated with larger loan amounts. The frontier shifted up from 2001 to 2005. This upward shift was larger for lower downpayment amounts, indicating that maximum loan-to-value ratios rose more for smaller loans. Maximum loan sizes decreased substantially in the first few years of the housing market contraction. This decrease was similar across all loan amounts except for very small downpayment amounts, where it contracted by more as the availability of 100 percent LTV loans largely disappeared. Over the next four years the frontier shifted down further, returning to a roughly similar level as in 2001.

Figure 9 depicts geographic variation over time in the frontier by aggregating the frontier by metropolitan area and year, and plotting percentiles of the distribution across metropolitan areas in each year. A large majority of metro areas experienced a boom and bust in mortgage credit availability over our sample period. The 90th percentile of growth between 2001 and 2005 (the peak year for most cities) was about 60 percent while the 10th percentile was 20 percent. Despite the large contraction of credit availability in the latter half of our sample, credit availability in 2013 was still more available than it was in 2001 for a majority of the metro areas in our sample.

In summary, the loan amount frontiers are consistent with a number of standard predictions about mortgage credit availability: credit score, income and downpayment are important factors influencing the amount of credit that a borrower can obtain, with more credit available to borrowers with higher scores, higher incomes and larger downpayments. Holding these factors constant, credit availability expanded during the first half of the 2000s and contracted during the financial crisis. Our measure also provides some new insights into credit availability. For example, increases in credit availability during the boom were fairly similar across borrower types, whereas the contraction in credit was much sharper for low-score and low-income borrowers.

puts some weight on maximum loan sizes available to borrowers that do not qualify for the maximum GSE-backed loan.

On net, for low-score and low-income borrowers credit was more difficult to obtain in 2013 than in 2001, while for high-score and high-income borrowers the reverse is true. Another noteworthy result is that there are differences in credit availability growth across metropolitan areas, even for borrowers with the same credit scores, incomes, and downpayments. It is this variation that we will use below to study the effects of mortgage credit availability on housing market outcomes.

5 The Effect of Credit Availability on House Prices and Construction

The dissagregated nature of the loan frontier can offer new insights on the relationship between credit supply, house prices, and residential construction activity.

First, we investigate the sensitivity of house prices and construction activity to credit availability using the loan frontier. Existing measures of credit availability are unsatisfactory for this purpose because they either (i) capture credit demand in addition to credit supply, or (ii) lack variation across locations over time, making it difficult to conduct an empirical analysis that controls for a variety of other factors.

To this end, we estimate regressions of the following form:

$$\Delta y_{jt} = \gamma \Delta F_{jt} + \beta \Delta X_{jt} + \alpha_j + \delta_t + \epsilon_{jt}. \tag{1}$$

All the variables enter in changes. Δy_{jt} is either the change in the log quality adjusted house price level or log single family housing stock in metro j at year t.[26] F_{jt} is the loan frontier aggregated up to the metro-year level, as described in Section 4. α_j and δ_t capture a set of metro area and year fixed effects,

[26]The housing stock estimates for each metropolitan area are created from the stock in the 2000 Census, the stock in the 2013 ACS, annual building permits from the Census' building permits data, and the equation $stock_{jt} = stock_{jt-1} + permits_{jt-1} - depreciation_j$. CBSA-specific depreciation estimates are imputed from the difference between the 2013 stock and the sum of the 2000 stock and cumulative building permits from 2000 to 2012.

respectively. To control for time-varying CBSA-level fundamentals that may affect both housing market activity and credit availability, we include CBSA-by-year log-income, employment, and delinquency rate in the controls X_{jt}.[27] We estimate (1) using the estimated loan frontiers for the 100 largest metro areas in our data. Standard errors are clustered at the metro area.

Table 2 shows the results for both house price and housing stock growth. In columns 2 and 4, we also interact the change in the loan frontier with the measure of housing supply elasticity developed by Saiz (2010), as the effect of credit availability on prices and construction should depend on the slope of the housing supply curve. The results reveal that the change in the loan frontier is significantly positively related to both price growth and housing stock growth. For a metro area with the mean supply elasticity, a one percent increase in the change in the loan frontier increases the change in prices by 0.47 percent and housing stock by .017 percent.[28] The relationship is stronger for prices in inelastic metros but is not as sensitive to supply elasticity for changes in the housing stock.

One issue with interpreting these results is that credit availability may be endogenous. First, omitted inputs from our loan frontier may be correlated with house prices. For example, suppose the distribution of household wealth (unobserved to us) in a particular metro area increases. The resulting increase in wealth might increase demand for housing, which would tend to increase local prices, and increase the loan frontier to the extent that lenders are willing to extend more credit to borrowers with higher household wealth. Second, as discussed in Section 1, there may be simultaneity bias if house prices and credit availability are jointly determined.

To address these potential endogeneity issues, we exploit the disaggregated nature of our frontier measure to create an instrument for credit availability in

[27] The house prices come from Zillow. The employment rate and income measures come from the Bureau of Economic Analysis. The delinquency rate is computed using our loan level data described in Section 3. The BEA data are not yet available for 2014, so data associated with 2014 are dropped from the regression equation (1).

[28] Recall that the loan frontier in the regression is an average across borrower types, and so a one percent increase in the frontier does not necessarily imply that credit expanded everywhere in the borrower distribution by one percent.

the spirit of Bartik (1991). The main identification idea is to use the fact that shocks to the national credit markets are exogenous to the local conditions in any one particular metro area, but can still have differential effects across metro areas, because different metro areas have different population distributions. For example, suppose that there is a national shock (such as regulatory changes or the financial crisis of 2007) that reduces the willingness of banks to lend to risky borrowers. The impact of such a change will be greater in metros where there are a large number of people with low credit scores.

To construct our instrument for a particular CBSA, we first estimate changes to the national loan frontier for each combination of income, FICO score and down payment. This is done by taking the population weighted average of the changes in the corresponding frontiers for all CBSAs except for the CBSA in question. Next, we integrate the changes in the national frontiers using the local distributions of income, FICO and downpayment of the CBSA we are constructing the instrument for. In other words, the instrument, Z_{jt}, for metro j at time t is constructed as the weighted average of changes in the loan frontiers in *other* metro areas, weighted by the population shares in metro j:

$$Z_{jt} = \sum_k s_j^k \sum_{i \neq j} \omega_i \Delta F_{it}^k \tag{2}$$

where k is a FICO/income/downpayment bin, and s_j^k is the share of individuals in bin k in metro j, averaged across time periods in our data. ω_i is the overall population share of metro area i (excluding metro j), and F_{it}^k is the loan frontier in metro i time t for bin k.

We need two features of the data for our instrument to have power in the first stage. The main requirement is that there are differential trends in the national measures credit availability across different borrower types. Such differential trends can be seen in Figures 6-9, and were likely driven by a variety of changes in the national mortgage market including the expansion and subsequent collapse of the market for private-label mortgage-backed securities, changes in long-term interest rates, and changes in government policies regarding GSE and FHA-backed mortgages. The second requirement is that there is cross-sectional variation in the distribution of borrowers across metro areas, i.e. that not all the metro areas have the same types borrowers living in them. As expected, this requirement also holds up in the data. Table 3 shows the first-stage results and indeed we see that the instrument is strongly positively correlated with our measure of credit availability.

The second stage results of the IV procedure are displayed in Table 4. The IV estimates are larger in magnitude to the OLS estimates. This could be because the instrument is isolating variation in the frontier that we observe across many metro areas, which should help address any attenuation that arises due to measurement error in changes in the local frontier. For a metro area with the mean supply elasticity, a one percent increase in the change in the average loan frontier increases the change in prices by 0.9 percent and housing stock by 0.1 percent.

To investigate whether the relationship between credit availability and housing market activity is different at different points in the housing cycle, in Table 5, we interact all the regressors in regression (1) with a dummy for whether the year is 2007 or later. We do not find evidence of a significant difference in the effect of credit availability on prices and housing construction during the boom versus the bust.

Returning to our baseline IV estimates in Table 4, we now explore the extent to which the exogenous changes to credit availability we identify can explain variation in house prices. Since our exogenous variation comes from differential effects that national shocks have on local markets, the shocks to the frontier cannot directly explain national movements in house prices. These national trends explain about 50% of the variation in local house prices. Our instrumented loan frontier can explain about 54% of the remaining variance that cannot be explained by national trends. Overall the exogenous variation in the loan frontier can explain 27% of the total variation in price appreciation.

Taken together, our evidence strongly suggests that easier mortgage credit has a significant positive effect on both house prices and construction activity.

5.1 Differential effects of credit availability across borrower types

In the previous section, we were able to construct the Bartik-style instrumental variable because of the disaggregated nature of the loan frontier measure. Another advantage of the disaggregation is that it allows us to investigate the differential effects of credit availability across borrower types. Existing measures of credit availability cannot address this issue because they do not characterize the heterogeneity in credit conditions across borrowers. To this end, we estimate regressions of the following form:

$$\Delta y_{jt} = \gamma_1 \Delta F_{jt}^{FICO<680} + \gamma_2 \Delta F_{jt}^{FICO>=680} + \beta \Delta X_{jt} + \alpha_j + \delta_t + \epsilon_{jt} \qquad (3)$$

where we disaggregate the loan frontier into a frontier for low credit score borrowers (FICO<680) and a frontier for high credit score borrowers (FICO≥680). Within these two categories, income/FICO/downpayment are aggregated as before.

The results for house prices and housing stock are presented in Table 6. In our preferred specification (Column 6), we find that only changes in the loan frontier for high-FICO borrowers is significantly related to changes in prices and housing stock. The effect of the loan frontier for low-FICO borrowers tends to be small and statistically insignificant.

In Table 7, we repeat this exercise by splitting the population on income, at $100,000. Similar to before, we find that house price changes and housing stock changes are more strongly correlated with credit availability for high-income households than for low-income households.

These results are interesting in light of the recent commentary pointing to tight mortgage credit for riskier borrowers as an explanation for the lackluster housing recovery in recent years. Our results suggest that mortgage credit has indeed been especially tight for riskier borrowers in recent years, but the historical correlation between credit to these borrower types and house prices seems weak. Instead, it seems that the availability of credit for less-risky borrowers exhibits a stronger correlation with prices and construction. The results also seem to rule out the expansion of mortgage credit to subprime borrowers as the primary driver of house price appreciation during the boom.

6 Robustness

In this section, we show that our estimates of the elasticity of house price and housing stock growth with respect to changes in the frontier are qualitatively robust to (i) alternative choices of weights used to compute the instrument (ii) using an approach that attempts to control for unobserved borrower characteristics (iii) alternative choices of the parameter m used to implement the method of Cazals et al. (2002) (iv) using full-documentation loans only where our borrower income measure is likely to be the most accurate and (v) including a measure of interest rates as an additional explanatory variable.

First, we test the robustness of our IV results to the choice of weights, s_j^k, used to compute the instrument as shown in equation (2). Table 8 shows our main results when s_j^k is defined as the share of individuals in bin k in metro j in 2001, rather than the share of individuals in bin k in metro j averaged across time periods in our data. By fixing the weights using the data at the beginning of our sample period, we address potential concerns regarding households sorting over our sample period in a way that is affected by credit availability or housing market outcomes. The estimated elasticities of house price growth and housing stock growth with respect to credit availability are somewhat higher than in our baseline specification.

One potential issue with our regression estimates is that there is an omitted, unobserved borrower characteristic that is both increasing the frontier and house prices in such a way that our instrument is not purging the frontier of this variation. In particular, our instrument will be valid only if metro by year specific shocks to the distribution of unobservables (that also independently affect house prices) are not correlated across metro areas.[29] To address this concern, we construct the frontier using the borrower's residualized inter-

[29] In addition, shocks to the distribution of unobservables that are correlated across metro areas would be captured by our fixed effects if the shocks are spread across all borrower types.

est rate at the time of origination as an additional input.[30] The motivation for this approach is that one might expect that, conditional on observable characteristics, lower interest rates are available to borrowers with better unobserved characteristics. Then, the interest rate residual can be used as a proxy for borrower unobserved characteristics. We find that the frontier tends to be increasing in the negative of the residual, which is consistent with this interpretation. Table 9 shows the results of this exercise. For the results in the first two columns, we include the frontier associated with borrowers of low unobservable type only; for the results in the next two columns, we include the frontier associated with borrowers of high unobservable type only; and for the results in the last two columns, we aggregate over the unobserved borrower type using equal weights for low and high types. In all specifications, the estimated elasticities of house price growth and housing stock growth with respect to credit availability are comparable to the ones in our baseline specification, suggesting that changes in the distribution of borrower unobservables are not driving our main estimation results.

Next, we test the robustness of our main results to our choice of m, which as explained above, is the number of draws one takes from the sample when computing the expected maximum loan amount. Table 10 shows results for $m = 500$ and $m = 2000$. The results do not appear to be very sensitive to our choice of m.

Next, we re-estimate the frontier, dropping all loan originations that are not flagged as fully documented. As discussed above, the motivation for this specification is that researchers have found that reported incomes in HMDA appears to be overstated, particularly in 2005 and 2006. By focusing on loans with full-documentation, we are focusing on a sample for which income overstatement is less likely. The right-most columns Table 8 shows that our results are similar to our main results using this subsample of the data.

Finally, in Table 11 we include the median interest rate among loan origi-

[30] In practice, we obtain the residual by regressing the interest rate on fico, ltv, income, origination amount, ARM dummy, loan type dummies, 30-year term dummy, cbsa fixed effects, and interaction terms. The regressions are run separately for each year.

nations within $10,000 of the estimated frontier as an additional explanatory variable in our main specification.[31] We do this to test whether our main estimate is being driven by the relationship between mortgage credit availability and housing market outcomes, or whether our estimate is also affected by the price of credit, which may be correlated with mortgage credit availability and is omitted from our main specification. After controlling for year and CBSA fixed effects, mortgage rates at the frontier have little effect on growth in prices or the housing stock. The estimate on the instrumented frontier is not significantly changed. This result supports our interpretation of our main result as an elasticity of changes in housing prices with respect to changes in mortgage credit availability.

7 Conclusion

We construct a new nonparametric measure of mortgage credit availability and argue that it reflects changes in credit supply rather than changes in demand for credit. Our estimation strategy allows us to examine changes in credit availability for different types of borrowers and in different housing markets, dimensions that have not yet been explored. In a panel of 100 metropolitan areas, we study how changes in our measure for credit availability affect housing construction and house prices. We exploit the disaggregated nature of our estimator to construct Bartik-type instruments to address potential endogeneity concerns. Our elasticity estimates imply that changes in credit availability can explain about 30 percent of the recent boom in house prices and 40 percent of the bust.

That credit availability has such a large effect on the housing market has interesting implications for monetary policy. As Geanakoplos (2014) notes, the Federal Reserve typically affects the housing sector through policies that influence the riskless interest rate. Our results suggest that the Federal Reserve could increase its influence on the housing sector, and perhaps ultimately on

[31] The median interest rate is computed for each bin on inputs, and then the median interest rate is aggregated using the same weights that we use to aggregate the frontier.

the broader economy and financial stability, by intervening to affect leverage requirements and other aspects of mortgage credit availability.

References

Adelino, Manuel, Antoinette Schoar, and Felipe Severino, "Credit supply and house prices: evidence from mortgage market segmentation," Technical Report, National Bureau of Economic Research 2012.

_ , _ , and _ , "Changes in Buyer Composition and the Expansion of Credit During the Boom," Technical Report, National Bureau of Economic Research 2015.

Avery, Robbert, Glen Canner, Ken Brevoort, and Neil Bhutta, "The 2011 Home Mortgage Disclosure Act Data," *Federal Reserve Bulletin*, 2012, *100* (4).

Bartik, Timothy J, "Who benefits from state and local economic development policies?," *Books from Upjohn Press*, 1991.

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

_ and Daniel R Ringo, "The 2013 Home Mortgage Disclosure Act Data," *Federal Reserve Bulletin*, 2014, *100* (4).

Blackburn, McKinley L and Todd Vermilyea, "The prevalence and impact of misstated incomes on mortgage loan applications," *Journal of Housing Economics*, 2012, *21* (2), 151–168.

Cazals, Catherine, Jean-Pierre Florens, and Leopold Simar, "Nonparametric frontier estimation: a robust approach," *Journal of Econometrics*, 2002, *106*, 1–25.

Demyanyk, Yuliya and Otto Van Hemert, "Understanding the Subprime Mortgage Crisis," *Review of Financial Studies*, 2011, *24* (6), 1848–1880.

Favara, Giovanni and Jean Imbs, "Credit Supply and the Price of Housing," *American Economic Review*, 2015, *105* (3), 958–92.

Ferreira, Fernando and Joseph Gyourko, "A New Look at the US Foreclosure Crisis: Panel Data Evidence of Prime and Subprime Borrowers from 1997 to 2012," Technical Report, National Bureau of Economic Research 2015.

Fuster, Andreas and Basit Zafar, "The Sensitivity of Housing Demand to Financing Conditions: Evidence from a Survey," Working Paper, Federal Reserve Bank of New York 2015.

Geanakoplos, John, "The leverage cycle," in "NBER Macroeconomics Annual 2009, Volume 24," University of Chicago Press, 2010, pp. 1–65.

_ , "Leverage, default, and forgiveness: lessons from the American and European crises," *Journal of Macroeconomics*, 2014, *39*, 313–333.

Glaeser, Edward L., Joshua D. Gottlieb, and Joseph Gyourko, "Can Cheap Credit Explain the Housing Boom?," NBER Working Paper No. 16230 2010.

Gropp, Reint, John Krainer, and Elizabeth Laderman, "Did consumers want less debt? Consumer credit demand versus supply in the wake of the 2008-2009 financial crisis," 2014.

Keys, Benjamin J, Tanmoy Mukherjee, Amit Seru, and Vikrant Vig, "Did Securitization Lead to Lax Screening? Evidence from Subprime Loans*," *The Quarterly journal of economics*, 2010, *125* (1), 307–362.

Kung, Edward, "The Effect of Credit Availability on House Prices: Evidence from the Economic Stimulus Act of 2008," Research Report, UCLA 2015.

Lee, Donghoon, Christopher J Mayer, and Joseph Tracy, "A new look at second liens," Technical Report, National Bureau of Economic Research 2012.

Li, Wei and Laurie Goodman, "Measuring Mortgage Credit Availability Using Ex-Ante Probability of Default," Research Report, The Urban Institute 2014.

Maggio, Marco Di and Amir Kermani, "Credit Induced Boom and Bust," Research Report 2015.

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

Nadauld, Taylor D. and Shane M. Sherlund, "The impact of securitization on the expansion of subprime credit," *Journal of Financial Economics*, 2013, *107* (2), 454 – 476.

Rajan, Raghuram G and Rodney Ramcharan, "The Anatomy of a Credit Crisis: The Boom and Bust in Farm Land Prices in the United States in the 1920s," *American Economic Review, Forthcoming*, 2014.

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

A Other measures of credit availability

We now briefly discuss some other measures of credit availability that have been developed, and consider some of their attractive features as well as their limitations.

One measure is the Federal Reserve's Senior Loan Officer Opinion Survey (SLOOS). This is a survey of senior loan officers at 60 of the nation's largest commercial banks in which respondents are asked about whether their institution "tightened" standards on mortgage lending during the previous quarter. While the SLOOS indicator theoretically isolates supply from demand, it relies on the respondent's judgment and interpretation of the survey question, and so is too qualitative to be useful for many applications.

Another measure is the Mortgage Banker Association's Mortgage Credit Availability Index (MCAI). The MCAI is based on the characteristics of loan programs offered by investors that buy residential mortgages. While the MCAI is also theoretically a direct measure of supply, it is fairly limited in its coverage (i.e. it is influenced by the product type of loans that investors will accept, and it does not cover loans held in bank portfolios). The MCAI therefore may not give a complete picture of aggregate mortgage market conditions.

Finally, researchers at the American Enterprise Institute and the Urban Institute publish measures of the riskiness of newly originated mortgages. Conceptually, the riskiness of new mortgages is supposed to reflect underwriting standards. However, the riskiness of new mortgages may also be affected by the types of mortgages demanded, so riskiness does not isolate supply from demand. Moreover, measuring the riskiness of new mortgages requires strong assumptions about how past defaults can be used to predict future defaults. By contrast, our loan frontier measure is purely data driven and does not require any assumptions about loan performance.

B Details on the HMDA to McDash/CoreLogic Merge

The HMDA data are first restricted to first lien, purchase mortgages to be comparable with the McDash/CoreLogic sample.[32] Each HMDA loan is assigned a unique id ("hmdaid"). HMDA reports the census tract of the property whereas McDash/CoreLogic reports the zip code so the first step is to convert census tracts in HMDA into zip codes. We do this using the HUD-USPS Zip Crosswalk files and the Missouri Census Data center crosswalk for years in which the HUD-USPS Zip Crosswalk files are unavailable. This is a one-to-many merge, as census tracts can be contained in multiple zip codes, and so a single hmdaid may appear multiple times in the data after this merge.

Each McDash/CoreLogic loan is assigned a unique id ('mcdashid"). We then match mcdashid to all records in HMDA that have the same loan amount[33], the same zip code, and have origination dates within 45 days of each other. Flexibility on origination dates is permitted because some origination dates are missing in McDash/CoreLogic and must be imputed using the closing date of the loan. There could also be recording errors. In the case that a single hmdaid matches to more than one mcdashid, all potential matches for a particular hmdaid are sorted on difference in origination date, difference in occupancy status, and difference in loan type (e.g. FHA, GSE), in that order. Only the best potential match by this sort criteria is kept; the rest are dropped. This ensures that a single hmdaid does not match to more than one mcdashid. Then, in the case of where a mcdashid matches to more than one hmdaid, matches are again sorted on difference in origination date, difference in occupancy status, and difference in loan type, in that order. The first record in the sort is kept as a match.

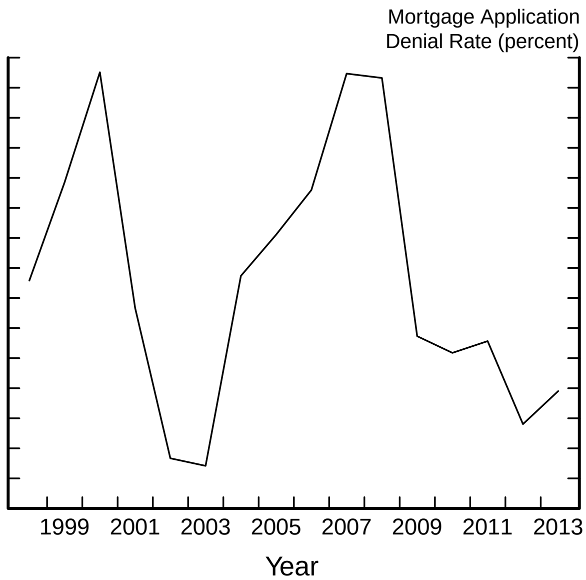
[32] For the years 2001-2003, there is not a first lien flag. For these years, some junior liens are identified by finding loans that have the exact same borrower characteristics (income, sex, race, ethnicity), census tract, occupancy status, origination date, and selecting the loan origination where the loan amount is a small fraction of the larger loan amount.

[33] The loan amount in the McDash/CoreLogic data is first rounded to the nearest 1000 because all loan amounts in HMDA are rounded to the nearest 1000.

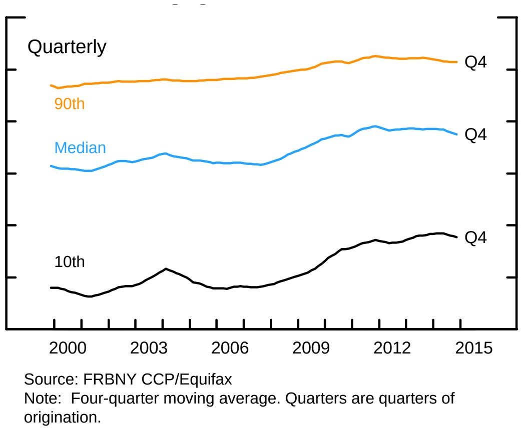
In the case where a mcdashid does not match to any hmdaid, we then do a second round of matching that follows the same procedure as the above paragraph, except we permit zip codes to match on only the first 4 digits of the zip code. Flexibility in the match on zip code is permitted because some error is introduced when translating census tracts to zip codes. There could also be recording errors. All hmdaids and lpsids that are matched in the first round are excluded from the second round.

The next step is to collect all junior liens associated with each first lien mortgage origination at the time of origination. For the years 2004-2010, this is trivial because the HMDA data have a flag connecting first liens to junior liens. For the remaining years, we following the following procedure. For each first lien mortgage origination, we have all the borrower characteristics and property characteristics available in HMDA from the match described above. Therefore, we can match each first lien purchase origination with all junior lien purchase originations in HMDA that have the exact same census tract, origination date, occupancy status, and borrower characteristics (income, race, ethnicity, sex). A match between a first lien and junior lien where the junior lien loan amount is greater than the first lien loan amount, or where the combined LTV $>$ 120 is dropped. In practice, we find that there are very few instances where a single junior lien matches to multiple first lien originations. In addition, the data between 2004-2010, where we can directly link first liens to junior liens, suggests that this algorithm is a good one. In particular, in almost all cases, the first lien and second lien have the same census tract, origination date, occupancy status, and borrower characteristics. The share of originations that can be linked to a junior lien for the years 2001-2014 are: 4.1, 5.7, 7.2, 12.9, 22.7, 25.8, 13, 2, 0.4, 0.3, 1, 0.9, 0.8, 1.4 percent respectively.

Figure 1: Alternative Measures of Credit Availability



(a) Mortgage Application Denial Rate



(b) Credit Score Distribution on New Owner-Occupied Purchase Mortgages

Figure 2: Frontier Example

The dots represent individual mortgage originations and the solid line is an estimate of the frontier using the methodology described in Section 2 for $m = 1000$. The loan frontier, shown on the y-axis, is in thousands of dollars.

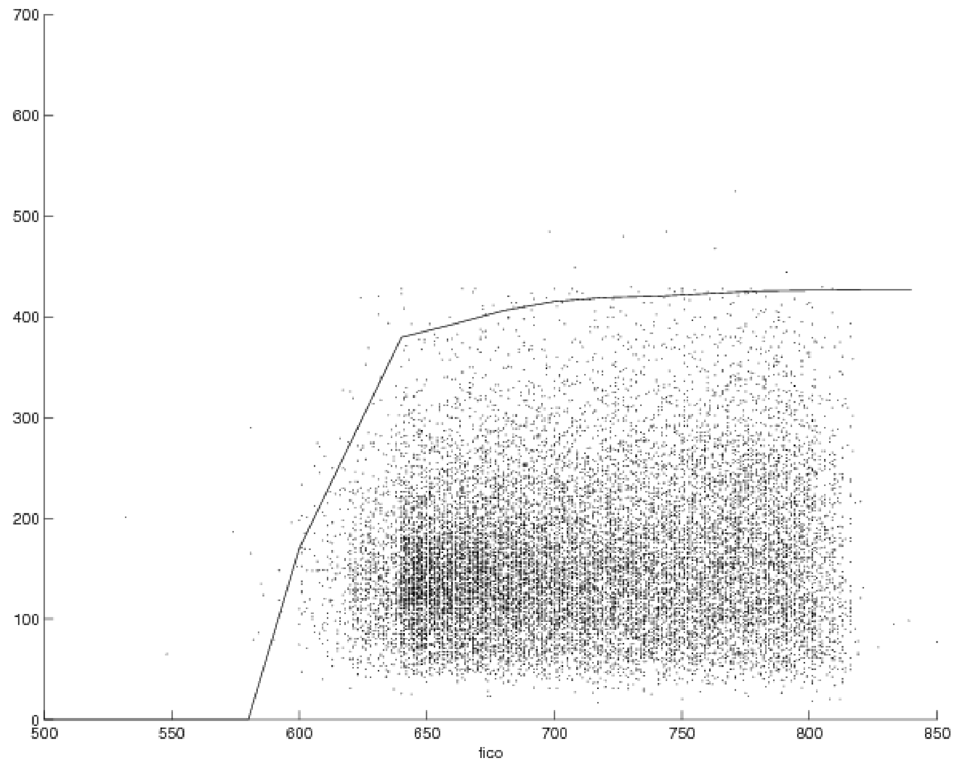


Figure 3: Confidence Intervals for Select Aggregate Loan Frontiers

Dotted lines are 95 percent confidence intervals around the aggregate loan frontier for select metro areas. Standard errors are computed through bootstrap with 100 repetitions. Aggregate loan frontier is the area under the loan amount frontier for each year and city given the choice of weights for each FICO, income, and downpayment described in Section 4. Frontiers for each metro area are normalized to one in 2001.

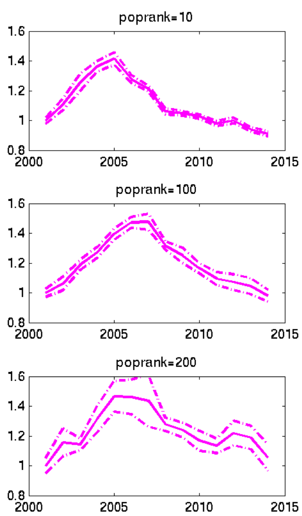


Figure 4: Distribution of Mortgage Originations Around the Loan Frontier

For each borrower type/year/metro area, we compute the share of observations within $5000 intervals around the estimated frontier. The figure plots the histogram when we take the simple average of these shares across all borrower types, years, and metros areas. Following our frontier estimation methodology, the sample for a borrower type of x includes all originations among borrowers with types $\leq x$.

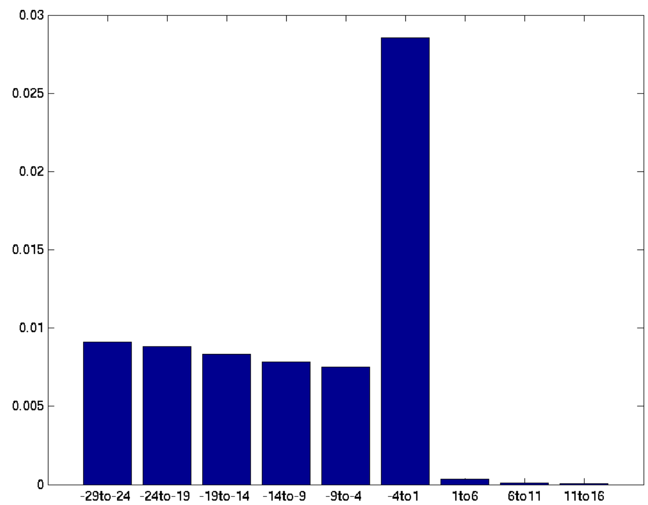


Figure 5: Boston Loan Frontiers

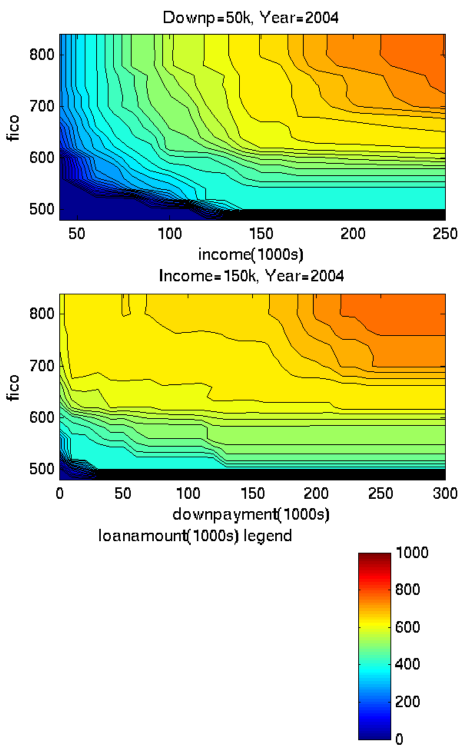


Table 1: Analysis of Variance for Loan Frontier

The average loan frontier is $281k and the standard deviation is $197k.

	Dependent Variable: Loan Frontier				
	(1)	(2)	(3)	(4)	(5)
Rsquared	0.49	0.5	0.63	0.7	0.8
FICO F.E.	x	x	x	x	x
Downp F.E.		x	x	x	x
Income F.E.			x	x	x
Year F.E.				x	x
MSA F.E.					x

Figure 6: Aggregate Loan Frontiers by FICO

The loan frontier is aggregated over metro areas, incomes, and downpayments using the weights described in Section 4. The loan frontier is in thousands of 2014 dollars.

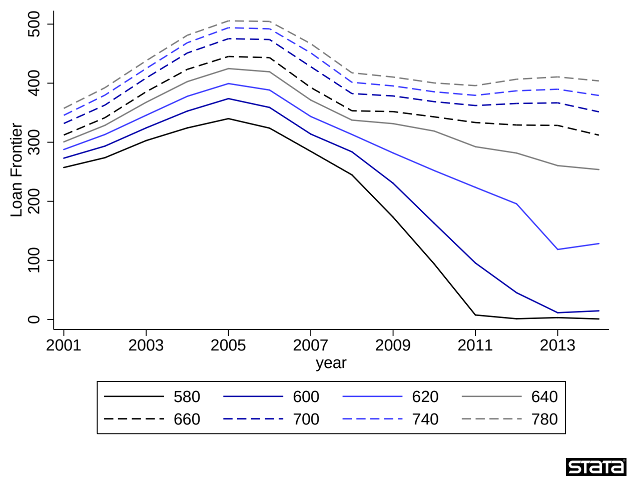


Figure 7: Aggregate Loan Frontiers by Income

The loan frontier is aggregated over metro areas, FICO scores, and downpayments using the weights described in Section 4. The loan frontier is in thousands of 2014 dollars.

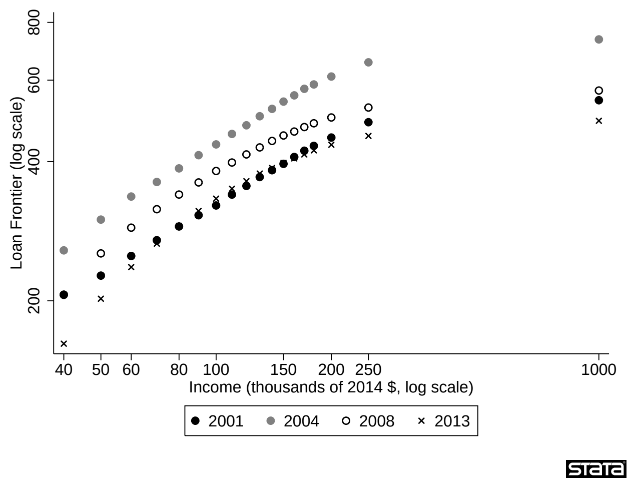


Figure 8: Aggregate Loan Frontiers by Downpayment

The loan frontier is aggregated over metro areas, incomes, and FICO scores using the weights described in Section 4. The loan frontier is in thousands of 2014 dollars.

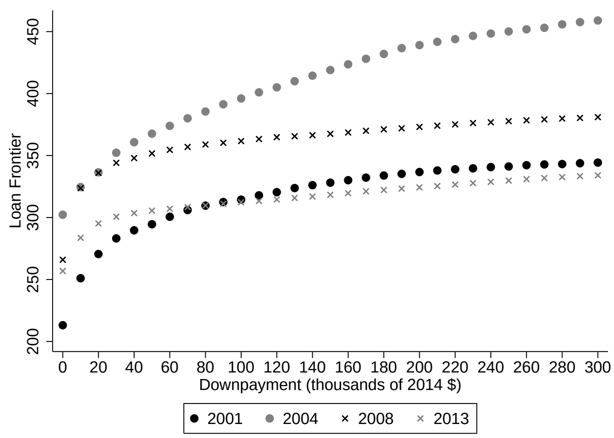


Figure 9: Aggregate Loan Frontiers by Metro Area

The loan frontier is aggregated over downpayments, incomes, and FICO scores using the weights described in Section 4. “pX” denotes the Xth percentile of the loan frontier across metro areas within each year.

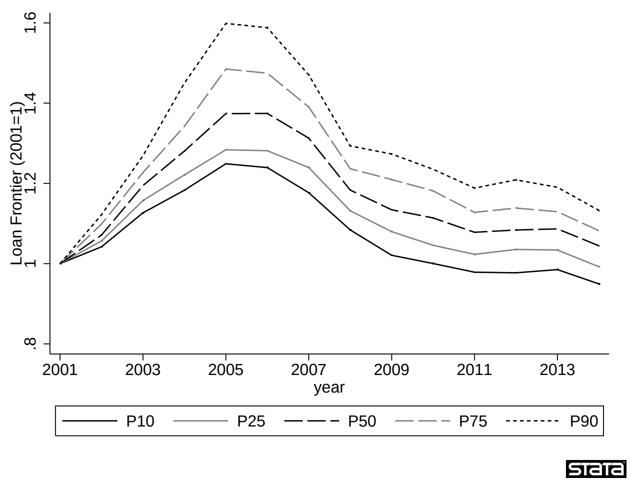


Table 2: The OLS effects of the loan frontiers on housing stock and prices for single family units

Dep. variable:	$\Delta lnPrice$		$\Delta lnHstock$	
	(1)	(2)	(3)	(4)
$\Delta lnFrontier$	0.559***	0.471***	0.018***	0.017***
	(0.087)	(0.092)	(0.006)	(0.006)
$Inelastic \times \Delta lnFrontier$		0.207***		0.006
		(0.049)		(0.005)
ΔLog Delinquency Rate	-0.127***	-0.109***	0.004*	0.005*
	(0.014)	(0.015)	(0.002)	(0.002)
ΔLog Income	0.116	0.074	-0.005	-0.005
	(0.091)	(0.086)	(0.016)	(0.016)
ΔLog Employment	0.966***	1.021***	0.217***	0.220***
	(0.255)	(0.245)	(0.043)	(0.045)
Observations	1120	1060	1200	1140
R^2 overall	0.594	0.608	0.181	0.180

Note - All the variables in this regression are in log differences. The $\Delta lnFrontier$ is the change in the log of the loan frontier for people with $480 < FICO \leq 840$ weighted by the joint distribution of FICO, Income and Downpayment shares of the particular CBSA. The sample consists of annual data from 2001 to 2013 for 100 metropolitan areas (CBSAs). All specifications include CBSA and year level fixed effects. The clustered robust standard errors are given in parentheses.

* statistical significance at the 90% level

** statistical significance at the 95% level

*** statistical significance at the 99% level

Table 3: The first stage effects of the instrument on loan frontiers

Dep. variable:	$\Delta lnFrontier$		$Inelastic \times \Delta lnFrontier$
	(1)	(2)	(3)
$\Delta lnInstrument$	0.556***	0.529***	0.314**
	(0.122)	(0.132)	(0.140)
$Inelastic \times \Delta lnInstrument$		0.061***	0.850***
		(0.019)	(0.048)
ΔLog Delinquency Rate	-0.070***	-0.066***	-0.045***
	(0.005)	(0.005)	(0.008)
ΔLog Income	0.143	0.149	0.081
	(0.095)	(0.092)	(0.082)
ΔLog Employment	0.467**	0.453**	0.055
	(0.201)	(0.197)	(0.180)
Observations	1200	1140	1140
R^2 overall	0.335	0.335	0.734

Note - All the variables in this regression are in log differences. The sample consists of annual data from 2001 to 2013 for 100 metropolitan areas (CBSAs). All specifications include CBSA and year level fixed effects. The clustered robust standard errors are given in parentheses.

* statistical significance at the 90% level

** statistical significance at the 95% level

*** statistical significance at the 99% level

Table 4: The IV effects of the loan frontiers on housing stock and prices for single family units

Dep. variable:	$\Delta lnPrice$		$\Delta lnHstock$	
	(1)	(2)	(3)	(4)
$\Delta lnFrontier$	1.304***	0.900***	0.089**	0.104**
	(0.340)	(0.341)	(0.041)	(0.049)
$Inelastic \times \Delta lnFrontier$		0.105**		-0.005
		(0.053)		(0.006)
ΔLog Delinquency Rate	-0.067**	-0.088***	0.009**	0.010**
	(0.029)	(0.024)	(0.004)	(0.005)
ΔLog Income	-0.005	0.020	-0.018	-0.020
	(0.083)	(0.091)	(0.013)	(0.014)
ΔLog Employment	0.696***	0.864***	0.188***	0.186***
	(0.214)	(0.234)	(0.028)	(0.032)
Observations	1120	1060	1200	1140
R^2 overall	0.472	0.571	0.053	0.006

Note - All the variables in this regression are in log differences. The instrument is a Bartik type instrument that translates shocks to the national frontier to CBSAs where such borrowers are located. The sample consists of annual data from 2001 to 2013 for 100 metropolitan areas (CBSAs). All specifications include CBSA and year level fixed effects. The clustered robust standard errors are given in parentheses.

* statistical significance at the 90% level

** statistical significance at the 95% level

*** statistical significance at the 99% level

Table 5: The IV effects of the loan frontiers interacted with the crisis period dummy

Dep. variable:	$\Delta lnPrice$		$\Delta lnHstock$	
	(1)	(2)	(3)	(4)
$\Delta lnFrontier$	1.490***	1.366**	0.023	0.010
	(0.537)	(0.530)	(0.063)	(0.071)
$Inelastic \times \Delta lnFrontier$		0.048		-0.002
		(0.051)		(0.009)
$I_{t\geq 2007} \times \Delta lnFrontier$	-0.212	-0.117	0.077	0.070
	(0.672)	(0.759)	(0.074)	(0.082)
$I_{t\geq 2007} \times Inelastic \times \Delta lnFrontier$		-0.028		0.013
		(0.112)		(0.013)
ΔLog Delinquency Rate	-0.097**	-0.100**	0.002	0.000
	(0.049)	(0.046)	(0.007)	(0.007)
ΔLog Income	-0.070	-0.031	-0.026	-0.021
	(0.179)	(0.172)	(0.048)	(0.052)
ΔLog Employment	0.165	0.220	0.304***	0.308***
	(0.215)	(0.231)	(0.064)	(0.064)
$I_{t\geq 2007} \times \Delta lnDeliquency$	0.047	0.049	0.008	0.008
	(0.058)	(0.057)	(0.006)	(0.006)
$I_{t\geq 2007} \times \Delta lnIncome$	0.061	0.034	0.041	0.040
	(0.234)	(0.232)	(0.046)	(0.050)
$I_{t\geq 2007} \times \Delta lnEmployment$	1.046**	1.005*	-0.287***	-0.281***
	(0.462)	(0.530)	(0.075)	(0.065)
Observations	1120	1060	1200	1140
R^2 overall	0.470	0.599	0.136	0.227

Note - All the variables in this regression are in log differences. The instrument is a Bartik type instrument that translates shocks to the national frontier to CBSAs where such borrowers are located. The sample consists of annual data from 2001 to 2013 for 100 metropolitan areas (CBSAs). All specifications include CBSA and year level fixed effects. The clustered robust standard errors are given in parentheses.

* statistical significance at the 90% level

** statistical significance at the 95% level

*** statistical significance at the 99% level

Table 6: The effects of the loan frontiers on metropolitan area house prices and housing stock

Dep. Variable	$\Delta lnPrice$			$\Delta lnHstock$		
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta lnFrontier_{FICO<680}$	0.149***		0.048	0.005		0.001
	(0.041)		(0.034)	(0.004)		(0.003)
$\Delta lnFrontier_{FICO\geq 680}$		0.485***	0.452***		0.017***	0.016***
		(0.090)	(0.090)		(0.005)	(0.004)
$Inelastic \times \Delta lnFrontier_{FICO<680}$	0.120***		0.034**	0.005		0.003
	(0.023)		(0.017)	(0.003)		(0.004)
$Inelastic \times \Delta lnFrontier_{FICO\geq 680}$		0.195***	0.159***		0.004	0.001
		(0.052)	(0.051)		(0.004)	(0.004)
ΔLog Delinquency Rate	-0.146***	-0.109***	-0.108***	0.004	0.005*	0.005*
	(0.014)	(0.015)	(0.015)	(0.002)	(0.002)	(0.002)
ΔLog Income	0.162	0.073	0.075	-0.002	-0.005	-0.005
	(0.102)	(0.097)	(0.093)	(0.017)	(0.017)	(0.016)
ΔLog Employment	1.137***	1.009***	1.011***	0.226***	0.220***	0.221***
	(0.283)	(0.266)	(0.256)	(0.047)	(0.046)	(0.046)
Observations	1060	1060	1060	1140	1140	1140
R^2 overall	0.556	0.608	0.611	0.174	0.178	0.180

Note - All the variables in this regression are in log differences. The sample consists of annual data from 2001 to 2013 for 100 metropolitan areas (CBSAs). All specifications include CBSA and year level fixed effects. The clustered robust standard errors are given in parentheses.

* statistical significance at the 90% level

** statistical significance at the 95% level

*** statistical significance at the 99% level

Table 7: The effects of the loan frontiers on metropolitan area house prices and housing stock

Dep. Variable	$\Delta lnPrice$			$\Delta lnHstock$		
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta lnFrontier_{Income\leq 100k}$	0.144*		0.120**	0.009**		0.006*
	(0.077)		(0.051)	(0.004)		(0.004)
$\Delta lnFrontier_{Income>100k}$		0.359***	0.319***		0.015***	0.012***
		(0.070)	(0.062)		(0.004)	(0.004)
$Inelastic \times \Delta lnFrontier_{Income\leq 100k}$	0.132***		-0.025	0.006		0.004
	(0.043)		(0.032)	(0.004)		(0.004)
$Inelastic \times \Delta lnFrontier_{Income>100k}$		0.196***	0.209***		0.005	0.003
		(0.047)	(0.050)		(0.004)	(0.003)
ΔLog Delinquency Rate	-0.148***	-0.119***	-0.114***	0.004*	0.004*	0.005**
	(0.014)	(0.014)	(0.015)	(0.002)	(0.002)	(0.002)
ΔLog Income	0.197*	0.098	0.104	-0.001	-0.004	-0.004
	(0.108)	(0.093)	(0.087)	(0.017)	(0.016)	(0.016)
ΔLog Employment	1.147***	0.985***	0.970***	0.225***	0.218***	0.219***
	(0.292)	(0.260)	(0.246)	(0.046)	(0.045)	(0.044)
Observations	1060	1060	1060	1140	1140	1140
R^2 overall	0.545	0.604	0.608	0.177	0.180	0.183

Note - All the variables in this regression are in log differences. The sample consists of annual data from 2001 to 2013 for 100 metropolitan areas (CBSAs). All specifications include CBSA and year level fixed effects. The clustered robust standard errors are given in parentheses.

* statistical significance at the 90% level

** statistical significance at the 95% level

*** statistical significance at the 99% level

Table 8: Robustness of the IV effects of the loan frontiers on housing stock and prices

	Presample weights from 2001		Only Full Doc. Loans	
Dep. variable:	$\Delta lnPrice$	$\Delta lnHstock$	$\Delta lnPrice$	$\Delta lnHstock$
	(1)	(2)	(3)	(4)
$\Delta lnFrontier$	1.928***	0.142**	1.203**	0.135**
	(0.647)	(0.066)	(0.609)	(0.068)
ΔLog Delinquency Rate	-0.027	0.013**	-0.090**	0.011**
	(0.050)	(0.006)	(0.043)	(0.005)
ΔLog Income	-0.098	-0.025*	0.037	-0.028*
	(0.136)	(0.014)	(0.124)	(0.016)
ΔLog Employment	0.545*	0.173***	0.542	0.147***
	(0.278)	(0.028)	(0.368)	(0.039)
Observations	1120	1200	1120	1200
R^2 overall	0.616	0.548	0.608	0.369

Note - In specifications (1) and (2), the frontier measure and the instrument are constructed using presample population weights of year 2001. In specifications (3) and (4), the frontier measure and the instrument are constructed using only full documentation loans. The instrument is a Bartik type instrument that translates shocks to the national frontier to CBSAs where such borrowers are located. The sample consists of annual data from 2001 to 2013 for 100 metropolitan areas (CBSAs). All specifications include CBSA and year level fixed effects. The clustered robust standard errors are given in parentheses.

* statistical significance at the 90% level

** statistical significance at the 95% level

* * * statistical significance at the 99% level

Table 9: Robustness of the IV effects of the loan frontiers on housing stock and prices

	Low Unobservable Type		High Unobservable Type		Controlling for Unobs. Type	
Dep. variable:	$\Delta lnPrice$	$\Delta lnHstock$	$\Delta lnPrice$	$\Delta lnHstock$	$\Delta lnPrice$	$\Delta lnHstock$
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta lnFrontier$	1.721***	0.110**	1.300***	0.089**	1.474***	0.098**
	(0.571)	(0.046)	(0.339)	(0.041)	(0.418)	(0.043)
ΔLog Delinquency Rate	-0.055	0.010**	-0.068**	0.009**	-0.063*	0.009**
	(0.041)	(0.004)	(0.029)	(0.004)	(0.033)	(0.004)
ΔLog Income	-0.028	-0.019	-0.004	-0.017	-0.014	-0.018
	(0.140)	(0.012)	(0.084)	(0.013)	(0.100)	(0.012)
ΔLog Employment	0.074	0.149***	0.691***	0.188***	0.437	0.172***
	(0.408)	(0.032)	(0.215)	(0.028)	(0.272)	(0.028)
Observations	1120	1200	1120	1200	1120	1200
R^2 overall	0.562	0.606	0.735	0.618	0.380	0.062

Note - Here we use a measure of unobserved borrower type, which is inferred from the unexplained part of the interest rate a borrower pays once we control for fico, ltv, income, origination amount, ARM dummy, loan type dummies, 30-year term dummy, cbsa fixed effects, and interaction terms, as an additional input. In specifications (1) and (2), we use the frontier measure for borrowers with low unobservable type. In specifications (3) and (4), we use the frontier measure for borrowers with high unobservable type. In specifications (5) and (6), we aggregate the frontier over low and high unobserved types using equal weights for each unobserved type. The instrument is a Bartik type instrument that translates shocks to the national frontier to CBSAs where such borrowers are located. The sample consists of annual data from 2001 to 2013 for 100 metropolitan areas (CBSAs). All specifications include CBSA and year level fixed effects. The clustered robust standard errors are given in parentheses.

* statistical significance at the 90% level

** statistical significance at the 95% level

*** statistical significance at the 99% level

Table 10: Robustness of the IV effects of the loan frontiers on housing stock and prices

	$m = 500$		$m = 2000$	
Dep. variable:	$\Delta lnPrice$	$\Delta lnHstock$	$\Delta lnPrice$	$\Delta lnHstock$
	(1)	(2)	(3)	(4)
$\Delta lnFrontier$	1.235***	0.085**	1.343***	0.091**
	(0.309)	(0.039)	(0.362)	(0.042)
ΔLog Delinquency Rate	-0.077***	0.009**	-0.061*	0.010**
	(0.025)	(0.004)	(0.031)	(0.005)
ΔLog Income	0.011	-0.017	-0.015	-0.018
	(0.079)	(0.013)	(0.088)	(0.013)
ΔLog Employment	0.702***	0.189***	0.688***	0.188***
	(0.216)	(0.028)	(0.215)	(0.028)
Observations	1120	1200	1120	1200
R^2 overall	0.753	0.623	0.719	0.614

Note - In specifications (1) and (2), the frontier measure and the instrument are constructed using $m = 500$. In specifications (3) and (4), the frontier measure and the instrument are constructed using $m = 2000$. The instrument is a Bartik type instrument that translates shocks to the national frontier to CBSAs where such borrowers are located. The sample consists of annual data from 2001 to 2013 for 100 metropolitan areas (CBSAs). All specifications include CBSA and year level fixed effects. The clustered robust standard errors are given in parentheses.

* statistical significance at the 90% level

** statistical significance at the 95% level

* * * statistical significance at the 99% level

Table 11: The IV effects of the loan frontiers on housing stock and prices for single family units

Dep. variable:	$\Delta lnPrice$		$\Delta lnHstock$	
	(1)	(2)	(3)	(4)
$\Delta lnFrontier$	1.350***	1.204***	0.089**	0.091*
	(0.384)	(0.363)	(0.044)	(0.047)
$Inelastic \times \Delta lnFrontier$		0.055**		-0.001
		(0.027)		(0.004)
$\Delta lnRate_{Frontier}$	0.041	0.038	-0.000	-0.000
	(0.055)	(0.054)	(0.005)	(0.005)
ΔLog Delinquency Rate	-0.066**	-0.071**	0.009**	0.009**
	(0.030)	(0.029)	(0.005)	(0.005)
ΔLog Income	-0.015	0.022	-0.018	-0.018
	(0.086)	(0.085)	(0.014)	(0.014)
ΔLog Employment	0.666***	0.738***	0.188***	0.187***
	(0.227)	(0.235)	(0.028)	(0.027)
Observations	1120	1120	1200	1200
R^2 overall	0.457	0.507	0.053	0.047

Note - All the variables in this regression are in log differences. The $Rate_{Frontier}$ variable defined as the median interest rate of loans within $10000 of the loan frontier. The instrument is a Bartik type instrument that translates shocks to the national frontier to CBSAs where such borrowers are located. The sample consists of annual data from 2001 to 2013 for 100 metropolitan areas (CBSAs). All specifications include CBSA and year level fixed effects. The clustered robust standard errors are given in parentheses.

* statistical significance at the 90% level

** statistical significance at the 95% level

*** statistical significance at the 99% level