

**Finance and Economics Discussion Series
Divisions of Research & Statistics and Monetary Affairs
Federal Reserve Board, Washington, D.C.**

**Measuring Mortgage Credit Availability: A Frontier Estimation
Approach**

Elliot Anenberg, Aurel Hizmo, Edward Kung, and Raven Molloy

2017-101

Please cite this paper as:

Anenberg, Elliot, Aurel Hizmo, Edward Kung, and Raven Molloy (2017). "Measuring Mortgage Credit Availability: A Frontier Estimation Approach," Finance and Economics Discussion Series 2017-101. Washington: Board of Governors of the Federal Reserve System, <https://doi.org/10.17016/FEDS.2017.101>.

NOTE: Staff working papers in the Finance and Economics Discussion Series (FEDS) are preliminary materials circulated to stimulate discussion and critical comment. 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. References in publications to the Finance and Economics Discussion Series (other than acknowledgement) should be cleared with the author(s) to protect the tentative character of these papers.

Measuring Mortgage Credit Availability: A Frontier Estimation Approach*

Elliot Anenberg[†] Aurel Hizmo[†] Edward Kung[‡] Raven Molloy[†]

March 28, 2017

Abstract

We construct a new measure of mortgage credit availability that describes the maximum amount obtainable by a borrower of given characteristics. We estimate this “loan frontier” using mortgage originations data from 2001 to 2014 and show that it reflects a binding borrowing constraint. Our estimates reveal that the expansion of mortgage credit during the housing boom was substantial for all borrowers, not only for low-score or low-income borrowers. The contraction was most pronounced for low-score borrowers. Using variation in the frontier across metropolitan areas over time, we show that borrowing constraints played an important role in the recent housing cycle.

*We thank Erik Hurst, Edward Glaeser, Adam Guren, Leonard Nakamura, Dan Hartley, Tomasz Piskorski, James Vickery, Giovanni Favara, John Krainer, and numerous conference and seminar participants 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

[‡]UCLA

1 Introduction

Many economic models emphasize the role of borrowing constraints in the real economy. The housing market is a prime example of a market where borrowing constraints are economically important. Availability of collateralized borrowing through mortgages has direct effects on household portfolio choice, housing and asset prices, homeownership rates, defaults and transmission of monetary policy.¹ Recently, policymakers have been keenly interested in measuring and monitoring borrowing constraints because of the role they played in the housing boom and bust in the 2000s, and in the current economic recovery.

Despite their known importance, there are few direct measures of mortgage borrowing constraints. Instead, most studies use measures such as the volume of originations, the median credit score of new borrowers, or approval rates, which are influenced by a variety of factors besides mortgage credit availability. In this paper, we propose a new, *direct* measure of mortgage borrowing constraints. The measure aims to estimate the maximum mortgage size that banks are willing to extend to a borrower, conditional on the borrower’s observable characteristics, including the value of the collateral. In other words, we estimate the quantity of credit available to borrowers at the individual level, conditional on observables.

Our methodology is motivated by the literature on production frontier estimation, which asks: “Given a vector of inputs, what is the maximal set of outputs that could be obtained?” Under certain conditions, we can think of the possibility set of mortgage originations as analogous to a production possibility frontier, with borrower characteristics as inputs and contract terms as outputs. Our measure, which we call the *loan frontier*, therefore answers: “Given a vector of borrower characteristics, what is the maximum mortgage amount that the borrower can obtain?” As long as there are at least some borrowers who borrow the maximum because their demand for credit exceeds the amount of credit that lenders are willing to extend, the loan frontier can be estimated from observed mortgage originations using data envelopment techniques.

¹Some examples include Cocco (2005) and Chetty et al. (2016) who study the effects on portfolio choice, Ortalo-Magne and Rady (2006) and Favilukis et al. (2015) who study the effects on equilibrium prices and transactions, Corbae and Quintin (2015) who focus on foreclosures in the housing bust, Gete and Reher (2016) who study the effects on homeownership, and Iacoviello and Neri (2010) and Keys et al. (2014) who explore the role of housing in transmitting monetary policy. For an overview of the literature see Davis and Van Nieuwerburgh (2015).

Cazals et al. (2002) describe a non-parametric frontier estimation method which is robust to outliers. We apply this method to estimate the loan frontier using observed U.S. mortgage originations data from 2001 to 2014. Given our data, the estimated loan frontier measures the maximum loan amount that a borrower could obtain (across all contract types, interest rate offers, and lenders), conditional on her credit score, downpayment, and income.² The loan frontier is also computed separately by metropolitan statistical area (MSA) and by year. In the end, we computed the loan frontier for over 10,000 borrower types, defined as combinations of credit score, downpayment, and income, for the 100 largest MSAs over the fourteen year period from 2001 to 2014.

We conduct two initial exercises to validate that the loan frontier measures borrowing constraints. First, we show that there is a mass point in the distribution of loan originations exactly at the estimated frontier. This is true within most borrower types, cities, and years. Because there is no reason for unconstrained borrowing demand to be concentrated at this point, we take the bunching of loans at the frontier as evidence that the frontier represents a binding borrowing constraint. The bunching of loans at the frontier also supports our earlier assertion that there are at least some borrowers who borrow the maximum that they are able. Second, we show that, at the aggregate level, changes to the loan frontier are correlated with two other measures of credit availability that use different data sources and methodologies: the Federal Reserve’s Senior Loan Officer Opinion Survey, and the Mortgage Banker Association’s Mortgage Credit Availability Index.

One issue with the estimated loan frontier is that it does not account for unobserved heterogeneity in borrowing constraints. Unobserved heterogeneity is a concern because lenders likely observe more information about borrowers than is available in our data, and thus the frontier will measure borrowing constraints for borrowers with the “best” (in terms of credit availability) unobservables. Changes to the frontier for these borrowers may therefore not be representative of changes in borrowing constraints for the rest of the population. To address this concern, we impose some parametric assumptions on the distribution of unconstrained borrowing demand and borrowing constraints that allow us to identify and estimate a full

²The loan amount is a “combined” loan amount in that it includes the balance of simultaneous second liens at the time of origination.

distribution of borrowing constraints. Estimates of the parametric model show that the average borrowing constraint in the population is highly correlated with the non-parametric loan frontier. This result is robust to various assumptions about the correlation between unobserved borrowing demand and unobserved borrowing constraints. Thus, the nonparametric loan frontier seems to capture movements in underwriting standards that affect a wide range of borrowers, despite the fact that only a relatively small fraction of loans are originated at the estimated frontier.

The estimated frontiers show that mortgage credit availability expanded during the first half of the 2000s and contracted significantly during the financial crisis. Increases in credit availability during the boom were fairly similar across borrower types regardless of their income, credit score, or downpayment. In contrast, the decline in credit availability during the financial crisis was much sharper for borrowers with low credit scores, and somewhat sharper for borrowers with lower incomes. Interestingly, the contraction was not significantly larger for borrowers with low downpayments relative to borrowers with high downpayments, perhaps due to the expansion of low downpayment mortgage programs through the FHA. On net, mortgage credit availability was lower for low credit-score borrowers in 2014 than in 2001, while the opposite was true for high credit-score borrowers. In addition, we document substantial differences in the expansion and contraction of credit availability across metro areas, even conditional on fixed borrower attributes.

Our new methodology for measuring borrowing constraints is useful for a range of purposes. For one, it can help policymakers and other market observers monitor mortgage credit availability to better assess financial conditions and risks to financial stability.³ One benefit of our loan frontier over alternative measures is that it can be constructed for narrowly-defined types of borrowers and locations. Another benefit is that the non-parametric estimation approach of our measure is transparent, easily reproducible, and uses data that are currently more accessible than the data required to compute other measures.

Beyond monitoring mortgage credit conditions, the loan amount frontier can be a useful

³Many empirical papers including Mian and Sufi (2009a), Mian and Sufi (2011), and Keys et al. (2010) have emphasized the role of mortgage finance in amplifying housing boom and bust cycles. Geanakoplos (2010, 2014) emphasizes the importance of monitoring credit conditions for monetary policy and for managing the leverage cycle.

input into other economic analysis. To provide one application, we exploit the variation in the loan frontier across metropolitan areas to show that mortgage credit availability played an important role in the recent housing market boom and bust. To this end, we regress changes in metropolitan area house prices and housing stocks on changes in the loan frontier. We control for the potential endogeneity between borrowing constraints and housing market outcomes by constructing a shift-share instrumental variable (Bartik (1991)) for the loan frontier—a strategy that would be impossible without disaggregated data on frontiers by borrower type and location. We find that a 1% increase in the loan frontier predicts a 0.9% increase in house prices and a 0.09% increase in the housing stock. Our estimates do not change much when controlling for mortgage interest rates, suggesting that the effects of borrowing constraints on the housing market are separate from the effects of the price of credit. Our findings are consistent with existing studies that have found it difficult to ascribe much of the recent housing cycle to changes in mortgage interest rates, but find much larger effects for broader measures of credit supply.⁴ Because the weight of the evidence suggests that borrowing constraints are an important determinant of housing decisions, even controlling for interest rates, it is important for economists and policymakers to be able to measure these constraints.

The paper is organized as follows. Section 2 describes the frontier estimation methodology as it applies to mortgage originations data. Section 3 describes the data. In Section 4 we compute the frontier and show its patterns across individual borrowers and over time. We also present evidence that the frontier accurately measures a binding borrowing constraint. In Section 5, we use a parametric model to deal with unobserved heterogeneity, and show that the resulting measure of the borrowing constraints facing the average borrower is highly correlated with the frontier. In Section 6, we study the effect of the frontier on house prices and the housing stock using a shift-share IV strategy. Section 7 concludes and describes some ways in which the loan frontier could be used in future research.

⁴See Adelino et al. (2012), Glaeser et al. (2012), Favara and Imbs (2015); Maggio and Kermani (2015), Favilukis et al. (2015). In addition, Mian and Sufi (2009b), Demyanyk and Van Hemert (2011), Nadauld and Sherlund (2013), Keys et al. (2010), Haughwout et al. (2011), Ben-David (2011) provide evidence that certain elements of mortgage credit availability loosened during the 2000s, suggestive of a relationship between mortgage availability and the housing boom.

2 The Frontier Estimation Methodology

Consider a mortgage origination process in which borrowers of observed characteristics $x \in \mathbb{R}^p$ (i.e. credit score, income) obtain loans of observed characteristics $y \in \mathbb{R}^q$ (i.e. loan amount, required downpayment). The set of *all possible* mortgage originations is given by:

$$\Psi = \{(x, y) \in \mathbb{R}^{p+q} \mid \text{Borrower } x \text{ can obtain loan } y\} \quad (1)$$

We assume an ordinal ranking for x and y :

Assumption 1. 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.

x and y are therefore ordered in such a way that increases to x expand the possibility set while increases to y shrink it. One could think of x as borrower attributes which reduce the riskiness of the loan, and y as mortgage terms that increase the riskiness of the loan.⁵

The econometric problem is to estimate Ψ from a sample of mortgage originations $\{x_i, y_i\}_{i=1}^n$. Clearly, if $(x, y) \notin \Psi$, then $P(y_i \geq y \mid x_i = x) = 0$. We also assume the converse:

Assumption 2. If (x, y) is in the interior of Ψ , then $P(y_i \geq y \mid x_i = x) > 0$.

In words, there is always positive demand for the riskiest loans available. We note that this is not the same as assuming that all possible loans have positive demand. It is possible that some loans in the possibility set are very safe for the lenders, but unattractive to borrowers, and therefore not demanded at all.

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. Assumption 2 guarantees that some borrowers are actively constrained by the boundary of the possibility set Ψ . In the production context, Assumption 2 is equivalent to assuming that efficient production units are represented in the data.

⁵For borrower attributes that increase the riskiness of the loan, or mortgage terms that reduce it, we can simply define x and y as the negative of that attribute.

Cazals et al. (2002) (henceforth CFS) describe a robust, non-parametric method for estimating the efficient output frontier, which we adopt in this paper. To illustrate the CFS method, we begin with the case of a single output $y \in \mathbb{R}$ (i.e. loan amount) and multiple inputs $x \in \mathbb{R}^p$. The efficient output frontier is given by:

$$\varphi(x) = \sup \{y | (x, y) \in \Psi\} \quad (2)$$

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

$$\varphi_m(x) = E [\max \{Y_1, \dots, Y_m\} | X_1, \dots, X_m \leq x] \quad (3)$$

Intuitively, $\varphi_m(x)$ is the expected highest loan amount that would be observed with borrowers of characteristics less than x , out of m draws.

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

$$\hat{S}_{c,n}(y|x) = \frac{\frac{1}{n} \sum_{i=1}^n I [y_i \leq y, x_i \leq x]}{\frac{1}{n} \sum_{i=1}^n I [x_i \leq x]} \quad (4)$$

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

$$P (\max \{Y_1, \dots, Y_m\} \leq y | X_1, \dots, X_m \leq x) = P (Y \leq y | X \leq x)^m \quad (5)$$

we can compute the empirical analog of $\varphi_m(x)$ by the following procedure. Let $n(x)$ be the number of observations with $x_i \leq x$. Then, denote y_j^x as the j th smallest value of y_i conditional on $x_i \leq x$. We compute:

$$\hat{\varphi}_{m,n}(x) = \hat{S}_{c,n}(y_1^x|x)^m y_1^x + \sum_{j=2}^{n(x)} [\hat{S}_{c,n}(y_j^x|x)^m - \hat{S}_{c,n}(y_{j-1}^x|x)^m] y_j^x \quad (6)$$

as the estimator for $\varphi_m(x)$.

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. Choosing a finite m makes the estimator robust to outliers that may actually fall outside the possibility set (i.e. due to measurement error) while still maintaining the interpretation as an expected maximum out of m draws. $\hat{\varphi}_{m,n}(x)$ is therefore a robust, consistent estimator of the maximum borrowing amount that borrowers with characteristics x can achieve.

To extend the method to multiple outputs, one simply notes that there is no special distinction between inputs and outputs other than in their ordering. If one were to take the negative of an output as an input instead, then Assumptions 1 and 2 would continue to hold.⁶ Therefore, we can estimate the efficient frontier for a single output as a function of all the inputs and of the other outputs, simply by recasting the other outputs as negative inputs. In practice, we will use loan amount as the output, and other available contract terms, such as the downpayment, as inputs.

2.1 Example and discussion

To illustrate the frontier and its interpretation, consider an application where the output is loan amount and the input is the borrower's credit score. $\hat{\varphi}_{m,n}(x)$ is therefore an estimate of the highest loan amount that a borrower with credit score x could obtain. Figure 1 shows the frontier calculated using data from the Chicago metropolitan statistical area in 2012. The dots represent individual mortgage originations and the solid line is an estimate of the frontier with $m = 1000$. The figure shows that the frontier is generally increasing and concave in credit score.

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 that is equal to the sample mean of loan amounts for borrowers with $creditscore \leq x$. Generally speaking, however, the frontier will not be very sensitive to m when m is already high, because the methodology will tend to pick out the location where there is some bunching in the data, as we will show later.

⁶These are statistical statements. Economically, the distinction remains that contract terms (outputs) are chosen while characteristics (inputs) are fixed. However, Assumption 2 guarantees that for each output dimension, the constraint along that dimension will be binding for all chosen levels of other output dimensions (i.e. limits to the loan amount will be binding conditioning on all levels of chosen downpayment).

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 the population of potential borrowers as:

$$\hat{\psi} = \int \hat{\varphi}_{m,n}(x) f(x) dx \quad (7)$$

$\hat{\psi}$ is an aggregate measure of mortgage credit availability, defined as the maximum borrowing amount faced by the average borrower in the population. Other methods of aggregation may also be considered, depending on the application.

In practice, loan amount is not the only output and credit score is not the only input. However, not all possible inputs and outputs may be observed in the data. Therefore, it is important to discuss the interpretation of the frontier in the presence of unobservables. We discuss unobserved heterogeneity in much more detail in Section 5, but for now we will simply give some intuition. Consider the interpretation of the estimated frontier in Figure 1 when output is loan amount, but the true inputs are credit score and income. If we only observe the credit score, then $\hat{\varphi}_{m,n}(x)$ measures the maximum loan amount that could be obtained by a borrower with credit score x , *irrespective of the borrower's income*. Assuming that borrowing limits are increasing in income, then the frontier is not representative of the average borrower, but rather those with the highest incomes conditional on credit score.

In general, the frontier will measure the borrowing limit for borrowers with the most extreme unobservables, and will therefore be greater than the average borrowing constraint in the population. Even though the level of the frontier is not representative of the average borrower, changes in the frontier can still be if the factors that influence mortgage credit supply are similar across borrower types. In Section 5, we will consider an identification strategy based on shape restrictions on the unobservables and show that, in practice, changes in the average borrowing constraints in the population are indeed highly correlated with changes in the frontier.

In the main analysis below, we will focus on four characteristics of borrowers and loans: credit score, borrower income, downpayment, and loan amount. In addition, we will estimate the frontier separately by metropolitan area and year.

Before proceeding, there are two important issues worth clarifying. First, changes to the estimated loan frontier are not mechanically related to house price growth.⁷ Since we condition the frontier on downpayment, an increase in metro-area house prices that is not accompanied by an increase in borrowing constraints would be reflected in a movement of borrowers along and within the frontier, rather than a change in the frontier itself. Consider, for example, a borrower who in year t can borrow up to \$200,000 with zero downpayment. This borrower can purchase a home worth up to \$200,000 in year t . Now, suppose the same home is worth \$220,000 in year $t + 1$, but that the level of borrowing constraints did not change. Because borrowing constraints do not change, we still would not see in the data any borrowers with zero downpayment borrowing \$220,000 to purchase this home, and we therefore would not estimate an increase in the frontier for borrowers with zero downpayment. Instead, what we are likely to see in the data is that some other type of borrower is purchasing this home in year $t + 1$ —perhaps a borrower with larger downpayment or a borrower with a higher credit score.

The second issue to clarify is that we exclude the mortgage rate as an output in our implementation below. Thus, the mortgage rate is an unobserved output and our frontier has the interpretation as the maximum loan amount obtainable by borrowers who are willing and able to pay relatively high interest rates. We made this decision for two reasons. First, it is difficult to compare mortgage rates across contracts and we do not observe everything that would affect the true cost of a mortgage, such as points paid. Second, the previously cited literature as well as our results in Section 6 suggest that the quantity of mortgage credit available is more important for understanding some key aspects of housing market dynamics than the price of that credit.⁸ In fact, market observers and policymakers have focused on the perceived tightness of mortgage credit supply as a key headwind on the housing and broader economic recovery in the aftermath of the financial crisis, even as mortgage rates remained historically low. A loan frontier that included the mortgage rate as an output might show that credit supply was relatively loose following the financial crisis because mortgage rates

⁷If lending standards are endogenously affected by collateral values or by expectations of house price growth, then the two would be correlated (though not mechanically).

⁸Furthermore, previous literature has had more success measuring the price of mortgage credit than mortgage availability, so it is the measurement of mortgage availability that we view as the important contribution of our research.

declined to historically low levels over this time period. Thus, it would not measure the component of credit supply that is currently capturing the attention of many policymakers and researchers (see Gete and Reher (2016); Laufer and Paciorek (2016); Bernanke (2012)). That said, our methodology is general enough to accommodate the mortgage rate as an output, given available data, and the mortgage rate would be appropriate to include for certain applications.

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.⁹ 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.

The second dataset that we use is compiled by CoreLogic and covers loans that were subsequently sold into non-agency mortgage-backed securities. This 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 2000.¹⁰

Our combined dataset includes many variables of interest related to the mortgage origination process 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 using an algorithm described in the Appendix. We are able to match 90 to 98 percent of all loans in the McDash and CoreLogic dataset, de-

⁹We determine market coverage by comparing total loan volumes for each market segment to aggregate loan volumes published by Inside Mortgage Finance.

¹⁰Although the McDash dataset also includes some non-agency securitized loans, we exclude these loans to avoid double-counting.

pending on the year. We also match junior liens with first liens using information on date of origination and property location, as described in the Appendix.¹¹ 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, although we will refer to it simply as the loan amount.

Given the available data, we compute the frontier using the loan amount as the output, and the borrower’s credit score, income and downpayment as the inputs. We measure the loan amounts, downpayments, and incomes in real terms by converting the nominal levels into 2014 dollars using the price index for personal consumption expenditures. We compute the frontier separately for the 100 most populous metropolitan areas, and for mortgage originations on single-family properties only.¹² 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 loan-to-value ratios >120 and loans with appraisal amounts below \$10,000 or above \$5 million, we are left with a sample of 14 million loans originated between 2001 and 2014 that we use to compute our frontiers.

4 The Loan Frontier

In this section, we report on the estimated loan frontiers using the methodology described in Section 2 and the data introduced in Section 3. As a reminder, the loan frontier can be interpreted as the maximum loan amount that borrowers are able to obtain in a particular metro/year, given their credit score (measured as FICO score), income, and downpayment amount. We set $m = 1,000$. We discretize the distributions of FICO scores, downpayments, and incomes and estimate the frontier for each bin in each year and each metropolitan area.¹³ We limit the sample to the largest 100 metropolitan areas because cell sizes become too small to reliably estimate a frontier in metropolitan areas with fewer mortgage originations.

¹¹We exclude junior liens taken out after the purchase origination date, such as HELOCS. For more information on second liens, see Lee et al. (2012).

¹²We distinguish between single-family and condo because underwriting standards could depend on property type, and we choose to focus on the single-family housing market in this paper.

¹³We 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. Metro areas are defined using core-based statistical area definitions.

Figure 2 illustrates the loan frontier for Boston, in 2004 and 2012. The left panels show the frontiers for 2004 and the right panels show the frontiers for 2012. The top panels show the contour plots by FICO and income, holding downpayment fixed at \$50,000. The bottom panels show the contour plots by FICO and downpayment, holding income fixed at \$150,000. Unsurprisingly, the frontiers indicate that lenders are willing to extend larger loans to borrowers with better credit scores, higher incomes, and higher downpayments. The contour plots also reveal complementarity between credit score, income, and downpayment in determining maximum borrowing amounts. Generally, to obtain the largest loan amounts, one must have high income, high credit score, and high downpayment.

We now conduct two exercises to validate that the loan frontier is measuring borrowing constraints. First, Figure 3 shows the distribution of borrowing amounts, as a function of distance to the frontier calculated for that borrower.¹⁴ The histogram shows a clear mass of loans that are within -\$4,000 and +\$1,000 of the estimated loan frontier. This bunching of loan amounts at the frontier suggests the existence of borrowing constraints based on FICO, income, and downpayment that are indeed binding, and that the loan frontier accurately identifies these constraints.

The bunching in Figure 3 is not being driven by a lot of bunching among just a few borrower types. Rather, there is bunching across a wide range of borrower types. To illustrate, Figure 4 shows the distribution of borrowing amounts around the estimated frontier for different borrowers in different MSAs and years, conditional on downpayment less than \$50k and income between \$80k and \$120k. The plot shows that, even for borrowers of similar income and downpayment, FICO significantly shifts the location of the frontier, and moreover that there is bunching around the frontier even within borrower types, MSAs, and years. Figure 4 reinforces our assertion that the loan frontier estimates are not mechanically related to house prices. Unless housing markets are very segmented within MSAs for borrowers of different credit score—even when they have similar income and wealth—we would not expect to see discontinuities in the loan amounts for borrowers with different credit scores, unless there are binding borrowing constraints.

Overall, using the Kleven and Waseem (2013) procedure for detecting bunching, we find

¹⁴The histograms for alternative choices of m (i.e. $m = 500$ and $m = 2000$) look very similar.

that statistically significant bunching is detected in 75 percent of borrower type/metro/year bins. We describe this exercise in more detail in the Appendix.

The second piece of evidence that supports the frontier as reflecting borrowing constraints is that the aggregate loan frontier is correlated with two other aggregate measures of mortgage availability: the Federal Reserve’s Senior Loan Officer Opinion Survey (SLOOS) and the Mortgage Banker Association’s Mortgage Credit Availability Index (MCAI). The SLOOS is a qualitative survey of senior loan officers at banks, and it asks whether the bank tightened or loosened underwriting standards for residential mortgages during the previous quarter. The MCAI is an index computed from the underwriting standards of loan programs offered by select investors. It roughly has the interpretation of a risk-weighted count of loan programs offered by investors. Like our frontier measure, these alternative measures of mortgage availability aim to measure changes in lender policy.¹⁵

To aggregate the loan frontier, we compute the weighted mean of the loan frontier across metro areas and borrower bins for each year. Downpayment bins are assigned equal weight, income and FICO scores are weighted according to the joint distribution of these two variables across all observations in our sample, and metro areas are weighted by population. Figure 5(a) plots changes in the aggregate loan frontier against the net fraction of banks reporting having tightened standards for residential mortgages in the SLOOS. The two measures are negatively correlated, indicating that years when more banks tightened lending standards were also years when our loan frontier contracted. Figure 5(b) plots the aggregate loan frontier and the MCAI over time, and shows that they have evolved similarly from 2004 to 2014.¹⁶

We now turn to documenting some basic facts about the loan frontier from 2001 to 2014. Table 1 summarizes some basic facts about the variance of the multidimensional loan frontier. The average loan frontier is \$283k (averaged across metro areas, years, and bins) and the standard deviation is \$199k. 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

¹⁵To be clear: none of these measures, including the loan frontier, are necessarily exogenous to changes in housing market and economic conditions. The measures aim to measure changes to lender policy regardless of the source of the change.

¹⁶The swings in the MCAI have a much larger magnitude, but this difference is difficult to interpret as the loan frontier and the MCAI do not have comparable units.

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 geographic variation in the market structure of banks, types of lenders, or persistent differences in economic conditions that are not captured by borrower income.

Figure 6 shows that the aggregate loan frontiers are fairly precisely estimated. The figure shows the estimated loan frontiers for various metro areas, along with 95% confidence intervals, which we computed using 100 bootstrapped repetitions. Confidence intervals are very tight, generally on the order of $\pm 5\%$ for the 100 largest metro areas that form our estimation sample. Beyond the 100th largest metro area, confidence intervals become larger, which reinforces our decision to restrict our analysis to the 100 largest metro areas.

Figure 7 plots how the loan frontier has changed over time for borrowers of differing credit score. The changes over time are striking. From 2001 to 2005, the frontier expanded by 30 to 45 percent for all credit scores above 560. During the financial crisis, the loan frontier 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 credit score above 640, the frontier fell by nearly 45 percent for borrowers with credit score around 620, and by nearly 75 percent for borrowers with credit scores around 600. For borrowers with even lower credit scores, the frontier fell to zero, indicating that borrowers with these scores were no longer able to obtain mortgage credit.

Turning to income, Figure 8 plots the evolution of the loan frontier for various income groups. The frontier expanded by 35 to 50 percent at all incomes above \$40,000 from 2001 to 2004, with larger increases for borrowers with higher incomes. The frontier also contracted a little more for higher-income borrowers during the financial crisis. On net, credit availability in 2014 was roughly the same as its 2001 level for borrowers with incomes above \$50,000. For borrowers with incomes below \$50,000, standards in 2014 were somewhat tighter than in 2001.

Figure 9 shows the loan frontiers for borrowers with various downpayment amounts. Conditional on downpayment, the loan amount frontier expanded substantially from 2001 to

2005, illustrating the conventional wisdom that lenders reduced downpayment requirements during this period. Maximum loan sizes decreased substantially in the first few years of the housing market contraction, and then flattened out during the last few years of our sample.

Figure 10 shows how the frontier varied across locations. To better compare changes over time, we normalize the value of the frontier to equal 1 in 2001 for each metropolitan area. The figure shows that changes in the frontier were much more pronounced in some locations than others. For example, in Las Vegas, the frontier expanded by more than 60 percent from 2001 to 2005, whereas in Dallas, it only expanded by 20 percent during the same period. Similarly, the contraction in credit was much more pronounced in Las Vegas than Dallas. Some areas like Detroit experienced more tightening during the bust than easing during the boom, with the net result that the frontier in 2014 was lower than in 2001. Other metro areas experienced the opposite: i.e. in Washington D.C., the contraction in credit after 2005 was smaller than the expansion from 2001 to 2005.

Finally, the solid black line in Figure 10 depicts the overall evolution of the aggregate frontier in our sample. On average, credit expanded by 45 percent from 2001 to 2006, contracted sharply from 2006 to 2008, and then continued to shrink from 2008 to 2014 (although at a more modest pace). On net, for the average potential borrower, mortgage credit was about as available in 2014 as it was in 2001.

In summary, the loan frontiers are consistent with a number of standard predictions about mortgage credit availability: borrowing ability is increasing in income, downpayment, and credit score. Holding these factors constant, availability expanded during the first half of the 2000s and contracted significantly during the financial crisis. The loan frontier also provides some new insights into mortgage credit availability. Increases in credit availability during the boom were fairly similar across borrower types, but the contraction was much sharper for low credit score and very low income borrowers.¹⁷ On net, mortgage credit availability was lower for low-score and very low-income borrowers in 2014 than it 2001, while the opposite is true for other borrowers. Another noteworthy result is that there are

¹⁷Adelino et al. (forthcoming) and Bhutta (2015) also find evidence consistent with this result. Also consistent with our evidence that the credit expansion during the boom was not limited to subprime borrowers, Ferreira and Gyourko (2015) find that the foreclosure crisis was widespread among prime and subprime loans.

differences in credit availability growth across metro areas, even for borrowers with the same credit scores, incomes, and downpayments. Thus, differential changes to credit availability across metro areas are not driven solely by compositional changes in the types of borrowers demanding mortgages.

5 Unobserved Heterogeneity

When there are relevant borrower characteristics that are observed by the lender but are unobserved in the data, the loan frontier measures the borrowing limit for borrowers with the “best” unobservables (from the perspective of borrowing ability). If there is a lot of unobserved heterogeneity, and if changes in the constraints of the borrowers with the best unobservables are not representative of changes in the constraints of more typical borrowers, then the loan frontier may not describe a borrowing constraint that is of much economic interest on its own because it only applies to a small fraction of borrowers. Indeed, Figure 3 shows that a small fraction of borrowers originate near the frontier, suggesting a role for unobserved heterogeneity in practice.

In this section, we consider a parametric estimation approach that allows us to identify the full distribution of borrowing constraints for borrowers with a particular set of observable characteristics. The estimated model provides a good fit to the data and shows that the average borrowing constraint in the population of mortgage borrowers turns out to be highly correlated with the loan frontier that we estimate in the previous section. These results suggest that, in practice, changes in the loan frontier are representative of changes of the borrowing constraints of typical borrowers. We prefer the loan frontier as a headline measure because it is more transparently computed and relies on fewer assumptions.

We begin by proposing a very general model of mortgage originations. Let borrowers be indexed by i and let them be characterized by observed characteristics $x_i \in \mathbb{R}^p$. Borrowers have indirect utility over borrowing amount l given by $V_i(l)$. Each borrower faces a maximum borrowing amount c_i .¹⁸ Neither V_i nor c_i are fully determined by x_i , so there may be

¹⁸For simplicity, we currently abstract away from contract terms other than the loan amount. The model is general enough to accommodate the choice of multidimensional contracts. A fuller specification of the model is available from the authors on request.

unobserved heterogeneity in both preferences and constraints. Unobserved heterogeneity in preferences may be driven by variation in housing demand based on family size, or variation in risk tolerance and preference for leverage. Unobserved heterogeneity in constraints may be driven by information about the borrowers that lenders observe but that we do not.

The borrower’s problem is therefore:

$$l_i = \arg \max_l V_i(l) \text{ s.t. } l \leq c_i \quad (8)$$

We additionally define the *unconstrained borrowing demand* for borrower i as:

$$d_i = \arg \max_l V_i(l) \quad (9)$$

We assume that V_i is continuous and single-peaked, so that borrowers have satiation points for mortgage borrowing and d_i exists. We can easily see that:

$$l_i = \min \{c_i, d_i\} \quad (10)$$

That is, the loan amount actually demanded is the minimum between the unconstrained demand and the borrowing constraint. We want to make statements about the distribution of c_i using only data on (l_i, x_i) .

Under the notation of Section 2, the possibility set can be written:

$$\Psi = \{(x, l) : P(c_i \geq l | x_i = x) > 0\} \quad (11)$$

And thus the loan frontier estimates:

$$\varphi(x) = \sup \{l : P(c_i \geq l | x_i = x) > 0\} \quad (12)$$

Through equation (12), we see that the loan frontier estimates the upper bound of the support of the distribution of borrowing constraint c_i , conditional on observable characteristics x_i . As discussed earlier, this is the same as saying the frontier estimates the borrowing constraints of borrowers with the “best” unobservables.

While we have argued that the frontier accurately estimates a real borrowing constraint, we have thus far not claimed that it estimates borrowing constraints for the average borrower. But the researcher is probably more interested in estimating *average* borrowing constraints for borrowers of type x : $E[c_i|x_i = x]$. Without further assumptions, the full distribution of c_i is not identified.¹⁹ However, we are able to identify the distribution of c_i if we are willing to make some additional assumptions.

In particular, let us assume that $c_i, d_i|x_i$ are bivariate log-normal with means μ_c, μ_d , variances σ_c^2, σ_d^2 , and correlation ρ . Basu and Ghosh (1978) show that the shape parameters are identified, up to a switch in the identity of c and d , from the distribution of $l_i = \min\{c_i, d_i\}$. In order to separate c from d , we will assume that $\sigma_d^2 > \sigma_c^2$. That is, variance in the unobserved heterogeneity in unconstrained demand is larger than the variance in unobserved heterogeneity in constraints. This seems to be a reasonable assumption, given that we are already conditioning on the most important variables that lenders would use to determine underwriting standards.

To illustrate how these shape restrictions allow us to separately identify the two distributions, consider Figure 11, which illustrates the distributions of unconstrained demand, borrowing constraints, and observed mortgage originations.²⁰ The figure shows that the distribution of originated loan amounts closely follows the distribution of unconstrained demand on the left tail, while it follows the distribution of constraints more closely on the right tail. Intuitively, borrowers with small loan amounts are unlikely to be constrained, and the distribution of small loans more closely reflects the distribution of unconstrained demand. Borrowers with large loan amounts are more likely to be constrained, and the distribution of large loan amounts will more closely reflect the distribution of constraints. Figure 11 also illustrates the intuition for how we may identify changes to constraints over time separately from changes to demand. If the left tail of the loan distribution remains the same from one period to the next, while the right tail of the distribution changes, we can reasonably attribute these changes to changes in the distribution of constraints.

¹⁹To see this, note that one could rationalize any data set (l_i, x_i) either by writing $c_i = l_i$ and $d_i > l_i$, or $d_i = l_i$ and $c_i > l_i$.

²⁰The means and variances of borrowing constraints and unconstrained demand are set so that the distribution of mortgage originations is close to the observed distribution of loans in Chicago in 2003.

We implement this approach on our entire dataset. To avoid having to separately estimate means and variances for every bin of borrower characteristics, we instead specify:

$$c_{ijt} = \delta_{jt}^c + \alpha_{1t}^c \log(\text{fico}_{ijt}) + \alpha_{2t}^c \log(\text{income}_{ijt}) + \alpha_{3t}^c \log(1 + \text{downp}_{ijt}) + \epsilon_{ijt} \quad (13)$$

$$d_{ijt} = \delta_{jt}^d + \alpha_{1t}^d \log(\text{fico}_{ijt}) + \alpha_{2t}^d \log(\text{income}_{ijt}) + \alpha_{3t}^d \log(1 + \text{downp}_{ijt}) + \xi_{ijt} \quad (14)$$

where δ_{jt} denotes a separate dummy variable for each metro j and year t , and $\epsilon_{ijt}, \xi_{ijt}$ are iid bivariate normal with zero mean and variances $\sigma_{c,t}^2, \sigma_{d,t}^2$. Note that all of the parameters in (13) and (14) are allowed to vary by year. The variances are assumed to be constant across metro areas and borrower types, but we allow them to vary by year. We estimate the parameters in equations (13) and (14) separately for each year using the entire sample of loan originations in that year. We can write the likelihood function in closed form, and estimate the parameters by maximum likelihood.²¹ Although ρ is formally identified, in Monte Carlo simulations we found that it was difficult to estimate precisely in practice, so we set $\rho = 0$. However, given that unobserved demand and unobserved supply could be positively correlated due to, for example, a relevant borrower characteristic that is unobserved to us but is observed to the lender, we also repeated this exercise with $\rho = 0.5$ and $\rho = 0.75$ and the results were very similar.

We start by examining whether the parametric model can provide a reasonable fit to the data. Figure 12 shows that the estimated model does a good job of fitting the empirical distributions of loan amounts in each year. We also verified the model fit for the distribution of loan amounts at the MSA level.

Figure 13 plots our estimates of $E[c_i]$, $E[d_i]$, and the aggregate loan frontier over time. To compute these estimates, we calculate the average of c_{ijt} and d_{ijt} across all borrowers in the sample for each year. To be comparable, the aggregate loan frontier is also computed as the average loan frontier across these same borrowers.²² The results show that average

²¹We modify the likelihood function to account for right-censoring at the loan frontier that we observed in Figure 3. However, our results are largely unchanged when we use an uncensored likelihood and do not give the estimation procedure any information about the loan frontiers. The results are also unchanged if we exclude all the loans near the frontier from our analysis suggesting that the estimates are not driven by what is happening exactly at the frontier.

²²The aggregate loan frontier in Figure 13 is slightly different from the aggregate frontier presented in Figure 10 because the weights are different.

borrowing constraints are highly correlated with the non-parametric loan frontier, though lower in levels. This result helps to alleviate concerns that changes in the frontier are not representative of changes to borrowing constraints for typical borrowers. In fact, it appears that the loan frontier is very informative about movements in credit availability for borrowers with average levels of unobservables. Interestingly, the average level of unconstrained demand is not nearly as volatile or as correlated with the frontier as the average level of borrowing constraints. Intuitively, this results from the fact that the left tail of the loan distribution is more stable over time, and less correlated with the frontier, than the right tail of the loan distribution.

The estimation also produces an estimate of the the share of borrowers who are bound by their constraints in each year; that is, the share of borrowers for whom $d_i \geq c_i$. This is a feature of the data that we do not explicitly target in estimation. Figure 14 shows this predicted share by year. As expected, the share is negatively correlated with the frontier, suggesting that when lending constraints are looser, the share of constrained borrowers is lower. The model predicts that 60 to 70 percent of borrowers take out the maximum obtainable loan amount given their FICO, income, and downpayment. We are not aware of any rigorous attempts to measure the share of constrained borrowers, but our estimates are similar in magnitude to other indirect and ad-hoc measures. Using the Survey of Consumer Expectations, Fuster and Zafar (2015) show that 42% of respondents would increase their demanded house value if downpayment requirements decreased from 20% to 5%. Applying institutional mortgage rules to the NLSY, Barakova et al. (2014) estimate that 58% of homeowners in 2003 and 72% in 2007 borrowed the maximum amount allowable.

6 Application: The effect of mortgage availability on house prices and construction

We close the paper with an application that illustrates how the loan frontier can be useful in analysis that goes beyond a description of credit availability conditions. In particular, we use the frontier to measure the sensitivity of the price and quantity of housing to credit

availability. We estimate regressions of the following form:

$$\Delta y_{jt} = \gamma \Delta F_{jt} + \beta \Delta X_{jt} + \alpha_j + \delta_t + \epsilon_{jt}. \quad (15)$$

Δy_{jt} is either the change in the log quality-adjusted house price or the change in an estimate of the log single-family housing stock in metro j at year t . 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, respectively. To control for time-varying metro-level factors that may affect both housing market activity and credit availability, we include changes in metro-by-year log-income, employment, and delinquency rate in X_{jt} .

The data for these regressions come from a number of sources. House prices come from the Zillow’s metro-area House Value Indexes. Housing stock estimates 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 $stock_{jt} = stock_{jt-1} + permits_{jt-1} - depreciation_j$. Metro-specific depreciation rates are imputed from the difference between the 2013 stock and the 2000 stock plus cumulative building permits from 2000 to 2012. Employment rate and income measures come from the BEA. Delinquency rate is computed using our loan level data described in Section 3.

Table 2 shows the results for both house price and housing stock growth. Standard errors are clustered at the metro level. In columns 2 and 4, we interact the change in the loan frontier with the measure of housing supply elasticity developed by Saiz (2010) to test whether the effect of credit availability on prices and construction depends 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 housing supply elasticity, a one percent increase in the loan frontier is associated with 0.53 percentage point higher house price growth and .018 percentage point higher housing stock growth. The relationship is stronger for prices in inelastic metros but we do not find that the relationship is weaker for construction in inelastic metros.

One issue with interpreting these results is that credit availability may be endogenous to local housing market conditions so that $cov(\epsilon_{jt}, \Delta F_{jt}) \neq 0$. For one reason, omitted variables

affecting both the loan frontier and the housing market may create a spurious correlation. Also, house prices and credit availability may be jointly determined in equilibrium, leading to a simultaneity bias.

To address these potential endogeneity issues, we exploit the disaggregated nature of the loan frontier to create an instrument for credit availability in 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 low credit score borrowers in particular. The impact of such a change on lending will be greater in metros where there are a large number of people with low credit scores. Our strategy is to estimate how local housing market outcomes respond to national changes to credit market conditions that affect a larger vs. smaller share of their borrower populations.

To construct our instrument for a given metropolitan area, we first estimate changes in 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 metros except for the metro in question. Next, we integrate the changes in the national frontiers using the local distributions of income, FICO and downpayment of the metro we are constructing the instrument for. Specifically, the instrument, Z_{jt} , for metro j at time t is equal to:

$$Z_{jt} = \sum_k s_j^k \sum_{i \neq j} \omega_i \Delta F_{it}^k \quad (16)$$

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.²³ ω_i is the overall population share of

²³Our methodology for creating this instrument differs slightly from standard practice because we use shares that are derived from the average over our entire sample period rather than shares from the initial period or shares in year t-1. We do not use shares from year t-1 because changes in the types of borrowers who obtain credit could be endogenous to current and future (expected) local housing market conditions, so we think it is critical to use shares that are fixed over time. We do not use initial shares because borrower types changed substantially during our sample period, so predicting credit supply based on 2001 borrower types could weaken the predictive power of the instrument. Nevertheless, we show in the Appendix that the results are similar when we use 2001 shares.

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 of credit availability across different borrower types. Such differential trends can be seen in Figures 7-10, 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; this holds in the data as not all metro areas have the same types of borrowers living in them. The technical condition for the instrument to be valid is $cov(\varepsilon_{j,t}, Z_{jt}) = 0$. That is, changes in unobserved local fundamentals, excluding metro and year fixed effects, should not affect national trends in credit availability to different borrower types. For example, if household wealth increases in a specific metro and year so that ε_{jt} increases, this may affect local lending conditions, but it should not affect what happens to lending in other markets ΔF_{it}^k .

Table 3 shows the first-stage results; the instrument is strongly positively correlated with the local loan frontier. The second stage results of the IV procedure are displayed in Table 4. The qualitative results from the OLS continue to hold, though the magnitudes of the coefficients are somewhat larger. This amplification could be because the instrument is isolating variation in the frontier that is common across many metro areas, which might reduce any attenuation bias that would arise due to measurement error in the local frontiers. For a metro area with the mean supply elasticity, a one percentage point larger change in the loan frontier for the average borrower leads to 0.9 percentage point higher house price growth and 0.1 percentage point larger growth in the housing stock. The price effect is stronger for more inelastic areas but, as with the OLS regressions, the housing stock effect is not significantly related to supply elasticity.²⁴

Recall that Figure 3 shows that the mass of originations around the frontier is a relatively

²⁴The results are qualitatively and quantitatively robust to alternative specifications including: 1) alternative choices of m , 2) alternative definitions of weights s_j^k for constructing the instrument, 3) using only full-doc loans to address income misreporting, and 4) controlling for unobserved borrower heterogeneity by using the residual of an interest rate regression. We describe these robustness checks and present their results in the Appendix.

small share of all originations. Then, a natural question that arises is why a constraint which is binding for so few borrowers has material effects on the housing market. The likely explanation is that movements in the frontier are correlated with movements in constraints faced by other borrowers, as the evidence in Section 5 suggests.

Table 5 shows results when we control for the median mortgage interest rate by metro-year as an additional regressor. The coefficients on the loan frontier are hardly changed from Table 4, suggesting that credit availability, as measured by the loan frontier, has an additional effect on the housing market that is not captured fully by variation in interest rates. As discussed above, this is consistent with the large body of literature that finds small house price elasticities with respect to interest rates, but larger elasticities with respect to broader measures of credit supply.

To give our estimated elasticities some context, we calculate the contribution of changes in mortgage credit supply to the boom and bust in house prices and residential construction. Because our IV strategy identifies the causal effect of credit supply on housing market outcomes, we can multiply the change in aggregate credit supply by our estimated coefficients to obtain the contribution of the frontier to aggregate changes in prices and quantities. The national aggregate loan frontier increased by 45 percent from 2001 to 2006, then contracted by 26 percent from 2006 to 2011. Based on the coefficients in Table 4, this cycle accounts for 68 percent of the growth in aggregate house prices from 2001 to 2006, and 81 percent of the subsequent house price decrease. At the same time, the expansion in credit accounts for 49 percent of the increase in the single-family housing stock. The contraction in credit from 2006 to 2011 implies a 3 percent decline in the housing stock. Of course, because the housing stock is durable and population growth puts continual upward pressure on housing demand, the housing stock rarely contracts. But we can still compare the predicted contraction in the stock to growth rates of the stock. The 5-year growth rate of the aggregate housing stock stepped down from 7 percent in the 2001-2006 period to 3 percent in the 2006-2011 period, a deceleration of 4 percentage points. Thus, the contraction in credit can account for roughly 72 percent of the slowdown in housing stock growth between these two periods.

To be sure, aggregate changes in the loan frontier are not themselves exogenous, as a wide variety of factors may have influenced aggregate credit conditions during the boom and

bust, including endogenously determined housing market conditions. We therefore do not view the above analysis as revealing the contribution of *exogenous* credit supply shocks on house prices and construction over the cycle. Rather, the results above shed light on the magnitude of the role that credit conditions played over this time, regardless of the reason for the changes in credit supply.

7 Conclusion

We construct a new nonparametric measure of mortgage borrowing constraints that can be interpreted as the maximum mortgage amount obtainable by borrowers of given characteristics. The frontier estimation approach allows us to monitor changes in credit availability for different types of borrowers and in different housing markets, providing a detailed picture of mortgage availability that requires only data on mortgage originations to compute. We show that there is bunching in loan originations at the frontier and that the frontier is correlated with alternative measures of credit availability. To illustrate the usefulness of the loan frontier, we exploit changes in the frontier over time and across locations to show that credit availability played a significant role in house price and housing stock movements over the recent housing cycle.

We now opine on some challenges and opportunities going forward. As to challenges, the model in Section 5 reveals the difficulty in identifying the full distribution of borrowing constraints without making some parametric assumptions about the distributions of borrowing constraints and unconstrained demand. In order to identify these distributions without making such assumptions, future work could exploit variables that are known to affect one distribution and not the other. Another strategy might be to focus on groups of borrowers *a priori* known to be either constrained or unconstrained. As to opportunities, the loan frontier could be useful as an empirical input into structural models of the housing market with heterogeneous borrowing constraints, which would allow for analysis of the effects of policies that affect credit supply. In addition, the frontier could be used to explore the factors that affect mortgage credit availability and the effects of credit on household decision-making and economic activity.

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 – , “Loan Originations and Defaults in the Mortgage Crisis: The Role of the Middle Class,” *Review of Financial Studies*, forthcoming.
- Avery, Robert, Glen Canner, Ken Brevoort, and Neil Bhutta**, “The 2011 Home Mortgage Disclosure Act Data,” *Federal Reserve Bulletin*, 2012, 100 (4).
- Barakova, Irina, Paul S. Calem, and Susan M. Wachter**, “Borrowing constraints during the housing bubble,” *Journal of Housing Economics*, 2014, 24, 4–20.
- Bartik, Timothy J**, “Who benefits from state and local economic development policies?,” *Books from Upjohn Press*, 1991.
- Basu, A.P. and J.K. Ghosh**, “Identifiability of the Multinormal and Other Distributions under Competing Risks Model,” *Journal of Multivariate Analysis*, 1978, 8, 413–429.
- Ben-David, Itzhak**, “Financial Constraints and Inflated Home Prices during the Real Estate Boom,” *American Economic Journal: Applied Economics*, 2011, 3 (3), 55–87.
- Bernanke, Ben**, “Challenges in Housing and Mortgage Markets,” 2012.
- Bhutta, Neil**, “The ins and outs of mortgage debt during the housing boom and bust,” *Journal of Monetary Economics*, 2015.
- 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.

- Chetty, Raj, Sandor Laszlo, and Adam Szeidl**, “The Effect of Housing on Portfolio Choice,” *The Journal of Finance*, 2016, *Forthcoming*.
- Cocco, Joao F.**, “Portfolio Choice in the Presence of Housing,” *Review of Financial Studies*, 2005, *18* (2), 535–567.
- Corbae, Dean and Erwan Quintin**, “Leverage and the Foreclosure Crisis,” *Journal of Political Economy*, 2015, *123* (1), 1–65.
- Davis, Morris and Stijn Van Nieuwerburgh**, “Housing, Finance, and the Macroeconomy,” *Handbook of Regional and Urban Economics*, April 2015, *5*.
- 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.
- Favilukis, Jack, Sydney C. Ludvigson, and Stijn Van Nieuwerburgh**, “The Macroeconomic Effects of Housing Wealth, Housing Finance, and Limited Risk-Sharing in General Equilibrium,” *Journal of Political Economy*, 2015, *Forthcoming*.
- 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**, “Solving the Present Crisis and Managing the Leverage Cycle,” *Federal Reserve Bank of New York Economic Policy Review*, 2010.
- , “Leverage, default, and forgiveness: lessons from the American and European crises,” *Journal of Macroeconomics*, 2014, *39*, 313–333.

- Gete, Pedro and Michael Reher**, “Systemic Banks, Mortgage Supply and Housing Rents,” *Georgetown university working paper*, 2016.
- Glaeser, Edward L, Joshua D Gottlieb, and Joseph Gyourko**, “Can cheap credit explain the housing boom?,” in “Housing and the Financial Crisis,” University of Chicago Press, 2012, pp. 301–359.
- Haughwout, Andrew, Donghoon Lee, Joseph S Tracy, and Wilbert Van der Klaauw**, “Real estate investors, the leverage cycle, and the housing market crisis,” *FRB of New York Staff Report*, 2011, (514).
- Iacoviello, Matteo and Stefano Neri**, “Housing Market Spillovers: Evidence from an Estimated DSGE Model,” *American Economic Journal: Macroeconomics*, April 2010, 2 (2), 125–64.
- 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.
- Keys, Benjamin, Tomasz Piskorski, Amit Seru, and Vincent W. Yao**, “Mortgage Rates, Household Balance Sheets, and the Real Economy,” 2014.
- Kleven, Henrik J and Mazhar Waseem**, “Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from Pakistan,” *The Quarterly Journal of Economics*, 2013, p. qjt004.
- Laufer, Steven and Andrew Paciorek**, “The Effects of Mortgage Credit Availability: Evidence from Minimum Credit Score Lending Rules,” *FEDS working paper*, 2016.
- Lee, Donghoon, Christopher J Mayer, and Joseph Tracy**, “A new look at second liens,” Technical Report, National Bureau of Economic Research 2012.
- 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 U.S. Mortgage Default Crisis,” *Quarterly Journal of Economics*, 2009, *124* (4), 1449–1496.

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

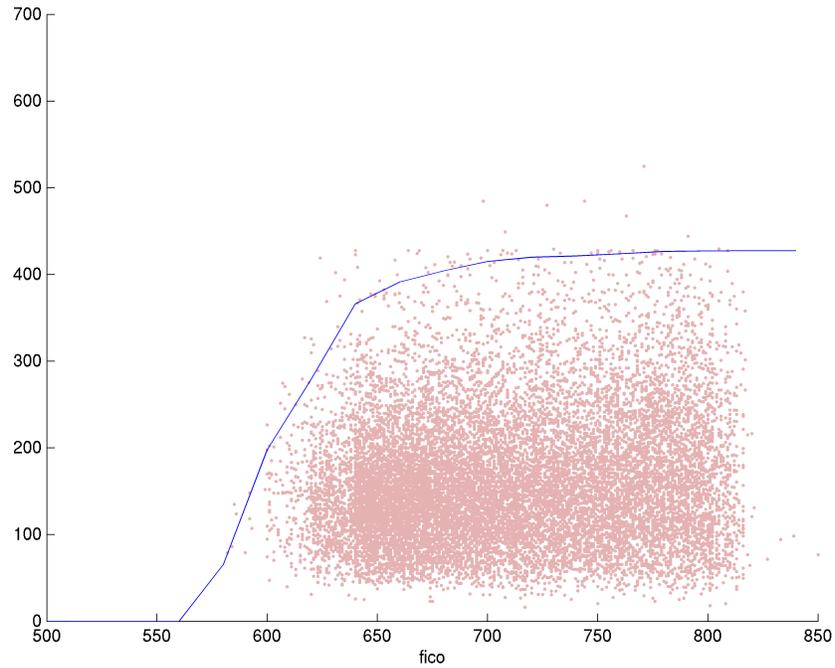
– **and** –, “House Prices, Home Equity-Based Borrowing, and the U.S. Household Leverage Crisis,” *American Economic Review*, 2011, *101* (5), 2132–2156.

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.

Ortalo-Magne, Francois and Sven Rady, “Housing Market Dynamics: On the Contribution of Income Shocks and Credit Constraints,” *The Review of Economic Studies*, 2006, *73* (2), 459–485.

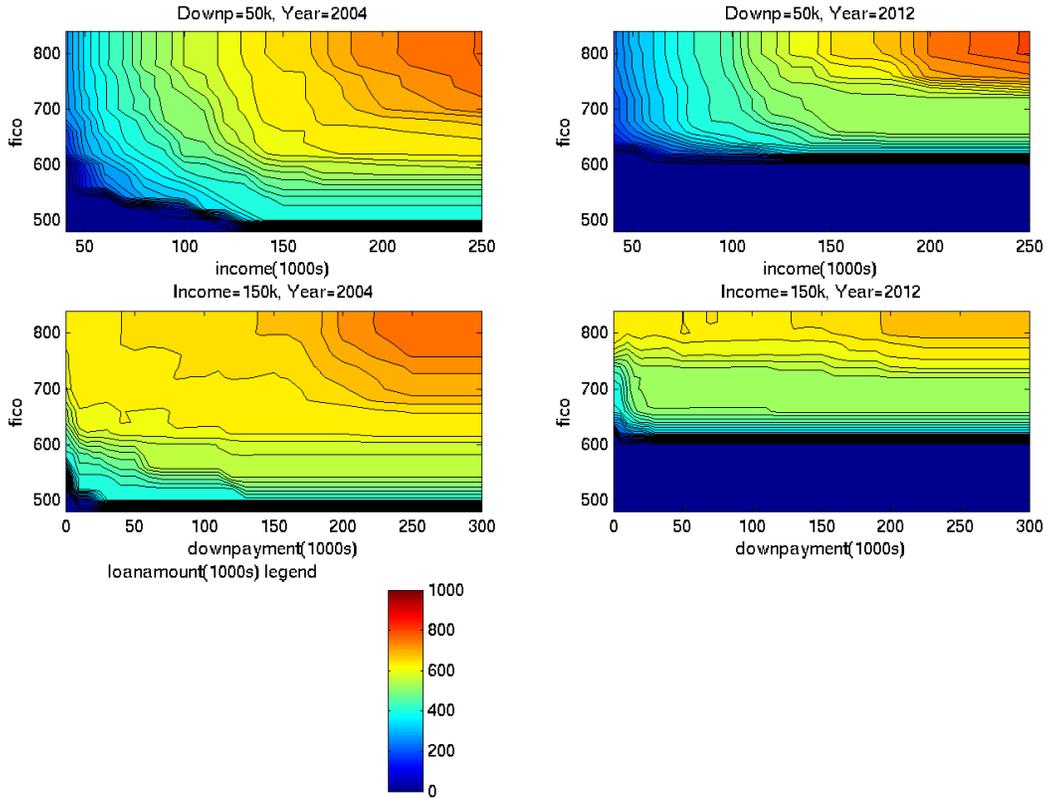
Saiz, Albert, “The geographic determinants of housing supply,” *The Quarterly Journal of Economics*, 2010, *125* (3), 1253–1296.

Figure 1: Example of Loan Frontier



Note: This is loan frontier for Chicago, 2012, computed with loan amount as output and FICO score as input. Each dot represents a mortgage origination in the data, and the solid line is the estimate of the loan frontier with $m = 1,000$. The loan frontier is reported in thousands of dollars. Source: Staff calculations based on data provided by McDash Analytics, LLC, a wholly owned subsidiary of Black Knight; and Corelogic.

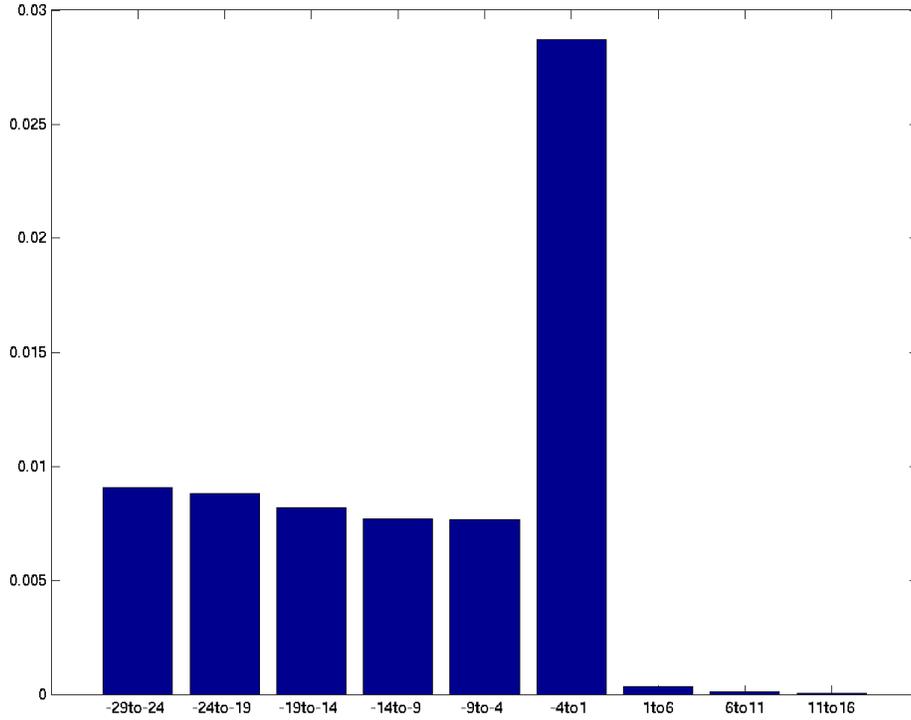
Figure 2: Boston Loan Frontiers, 2004 and 2012



Note: This figure shows contour plots of the frontier computed for the Boston metro area, in 2004 and 2012. The left panels show 2004 and the right panels show 2012. The top panels show the contours with respect to credit score and income, for a downpayment fixed at \$50k. The bottom panels show the contours with respect to credit score and downpayment, for borrower income fixed at \$150k.

Source: Staff calculations based on data provided by McDash Analytics, LLC, a wholly owned subsidiary of Black Knight; and Corelogic.

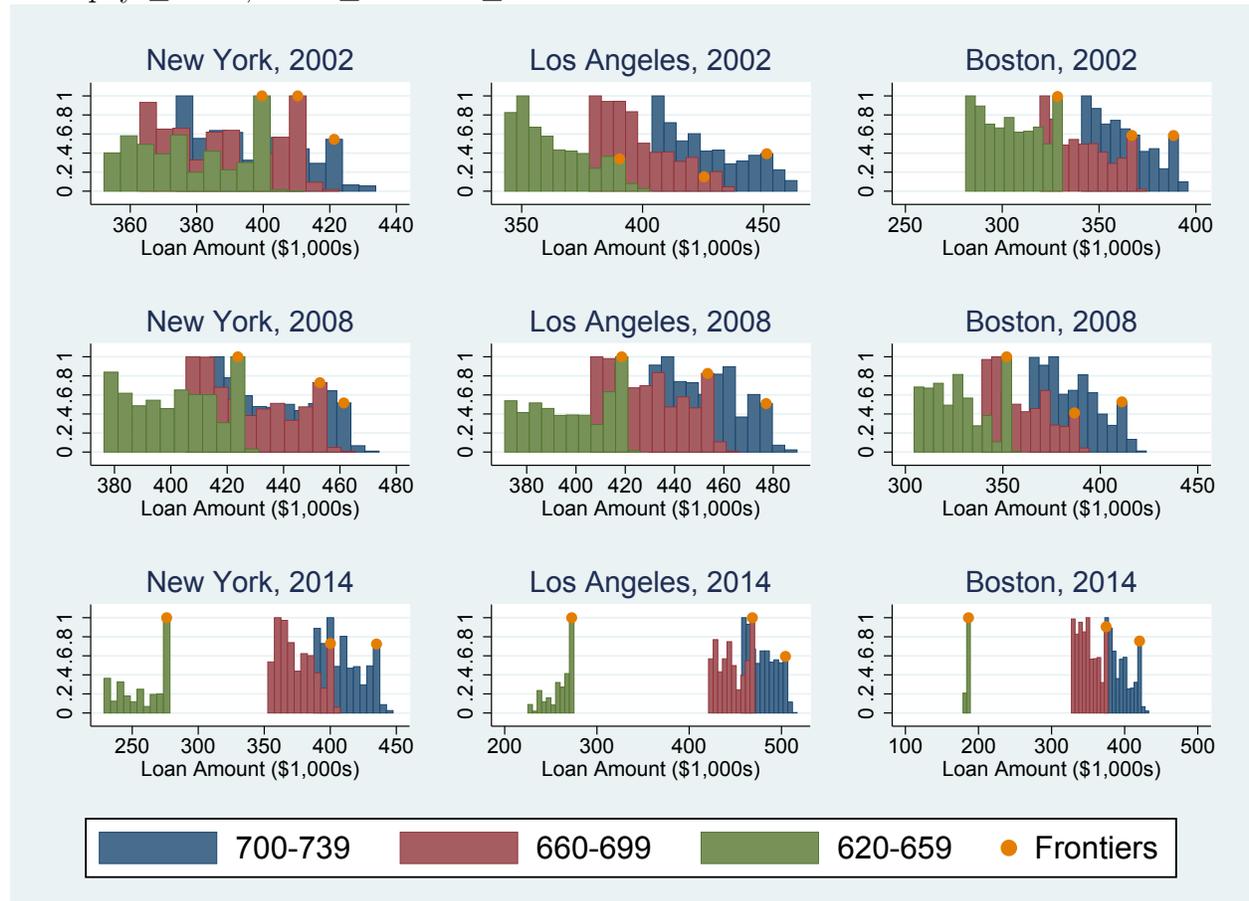
Figure 3: Distribution of Mortgage Originations Around the Loan Frontier



Note: For each borrower type/year/metro area, we compute the share of observations within \$5,000 intervals around the estimated frontier for that borrower. The figure plots the histogram when we take the simple average of these shares across all borrower types, years, and metro areas.

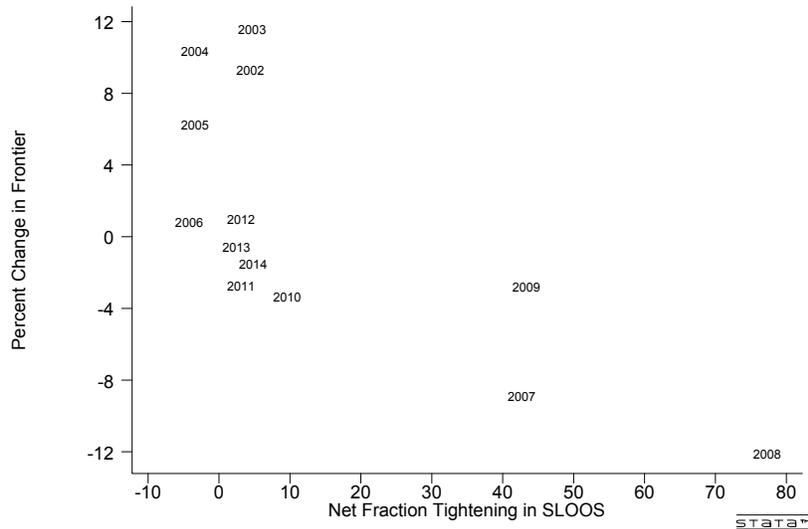
Source: Staff calculations based on data provided by McDash Analytics, LLC, a wholly owned subsidiary of Black Knight; and Corelogic.

Figure 4: Distribution of Mortgage Originations Around the Loan Frontier by Credit Score, Downpay \leq \$50k, $\$80k \leq$ Income \leq \$120k

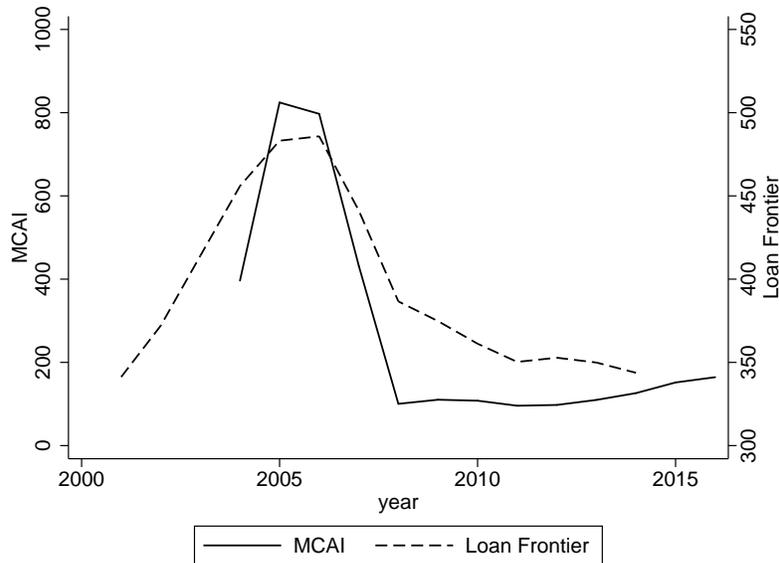


Note: For borrowers with downpayment less than \$50,000, income between \$80,000 and \$120,000, and various credit scores, we first compute the share of observations within \$5,000 intervals around the estimated frontier for those borrowers. A simple average is then computed for each distance bin. We then plot the share against the distance bin, shifted by the average estimated loan frontier for each group. The results show that different borrower types within MSAs and years have different estimated loan frontiers, but that there is bunching around the frontier within each group. Source: Staff calculations based on data provided by: McDash Analytics, LLC, a wholly owned subsidiary of Black Knight; and Corelogic.

Figure 5: Correlation Between the Loan Frontier and the SLOOS / MCAI
 (a) Senior Loan Officer Opinion Survey



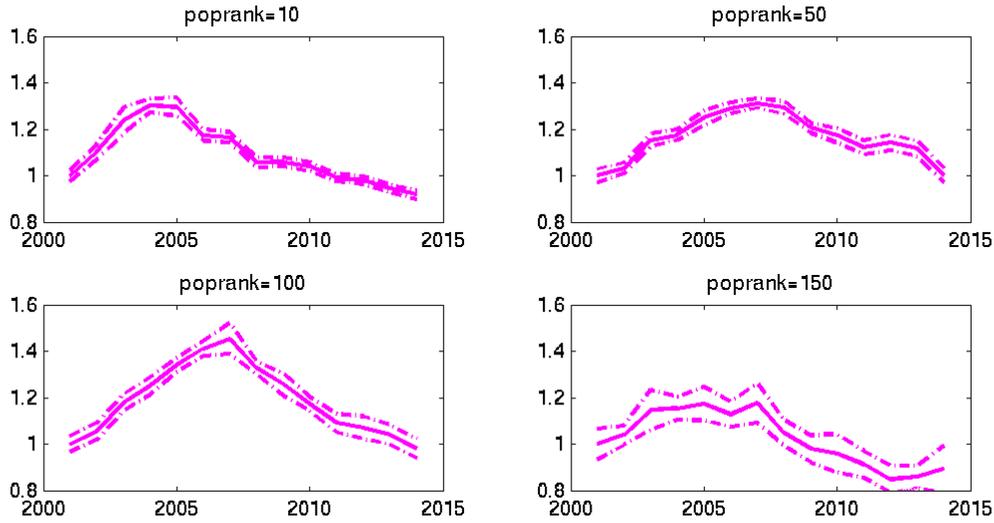
(b) Mortgage Credit Availability Index



Note: Panel (a) shows the correlation between the net fraction of banks reporting a tightening of standards for residential mortgages in the Senior Loan Officer Opinion Survey (SLOOS) and changes in the aggregate loan frontier. SLOOS responses are reported separately for prime, non-traditional and subprime loans. To obtain aggregate SLOOS responses for each year, we average three categories using equal weights. Also, we average quarterly responses to obtain annual estimates. Panel (b) shows the loan frontier along with the Mortgage Credit Availability Index (MCAI) produced by the Mortgage Bankers' Association. The MCAI is a function of the number of loan programs offered by large investors and the risk characteristics that define the types of loans that these programs will accept. The loan frontier is aggregated over metro areas, incomes, and down-payments using the weights described in Section 4.

Source: Staff calculations based on data provided by: McDash Analytics, LLC, a wholly owned subsidiary of Black Knight; and Corelogic.

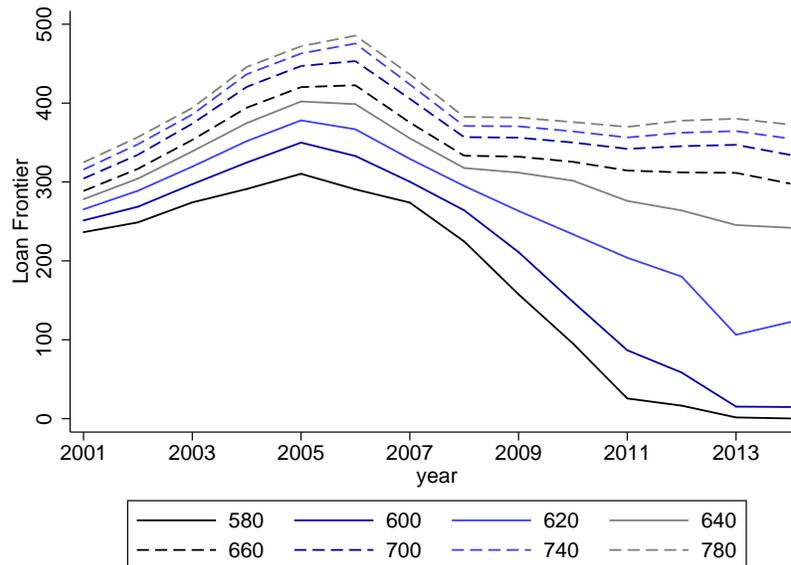
Figure 6: Confidence Intervals for Aggregate Loan Frontiers



Note: This figure shows 95 percent confidence intervals, in dotted lines, of the aggregate loan frontier, the solid line, for select MSAs. Poprank is the population rank of the MSA. Confidence intervals are computed using 100 bootstrap repetitions.

Source: Staff calculations based on data provided by: McDash Analytics, LLC, a wholly owned subsidiary of Black Knight; and Corelogic.

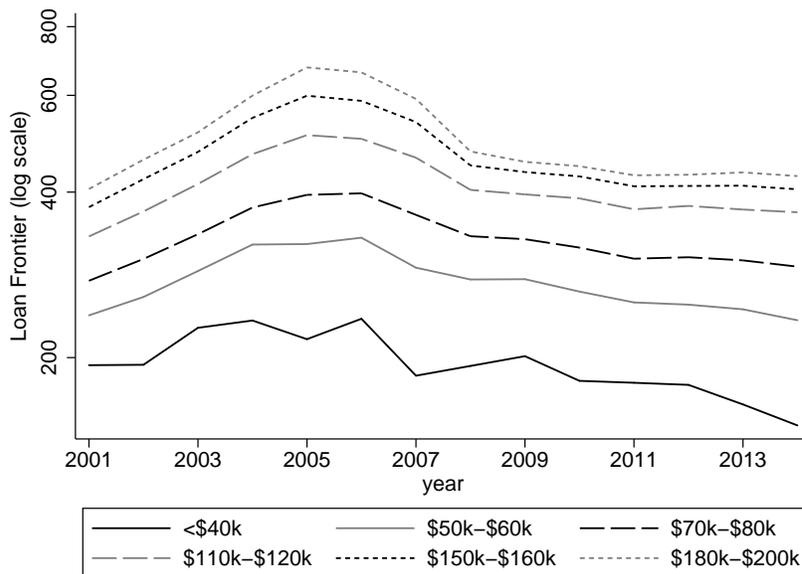
Figure 7: Aggregate Loan Frontiers by FICO



Note: 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.

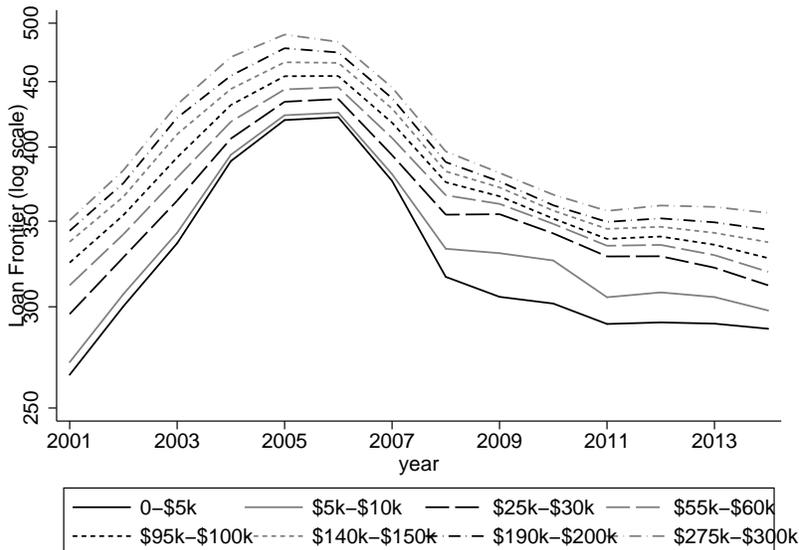
Source: Staff calculations based on data provided by: McDash Analytics, LLC, a wholly owned subsidiary of Black Knight; and Corelogic.

Figure 8: Aggregate Loan Frontiers by Income



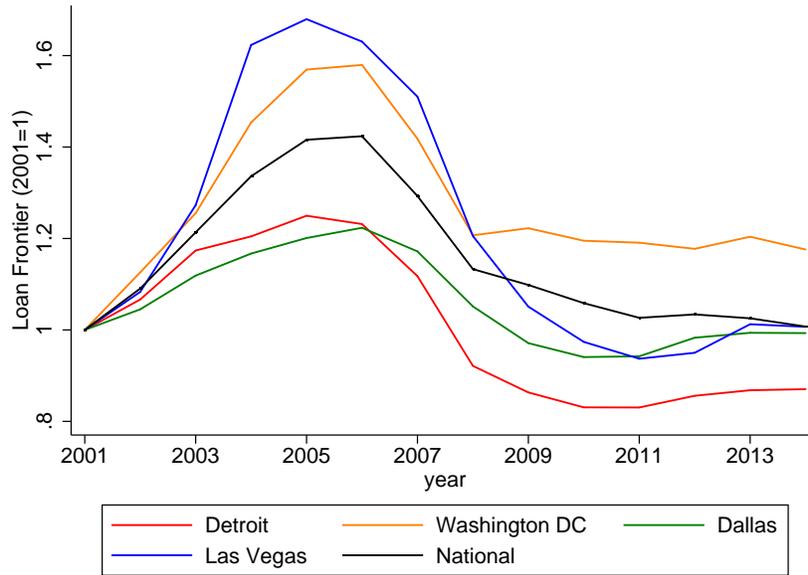
Note: The loan frontier is aggregated over metro areas, FICO scores, and downpayments using the weights described in Section 4. The loan frontier is in thousands of 2014 dollars.
 Source: Staff calculations based on data provided by: McDash Analytics, LLC, a wholly owned subsidiary of Black Knight; and Corelogic.

Figure 9: Aggregate Loan Frontiers by Downpayment



Note: 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.
 Source: Staff calculations based on data provided by McDash Analytics, LLC, a wholly owned subsidiary of Black Knight; and Corelogic.

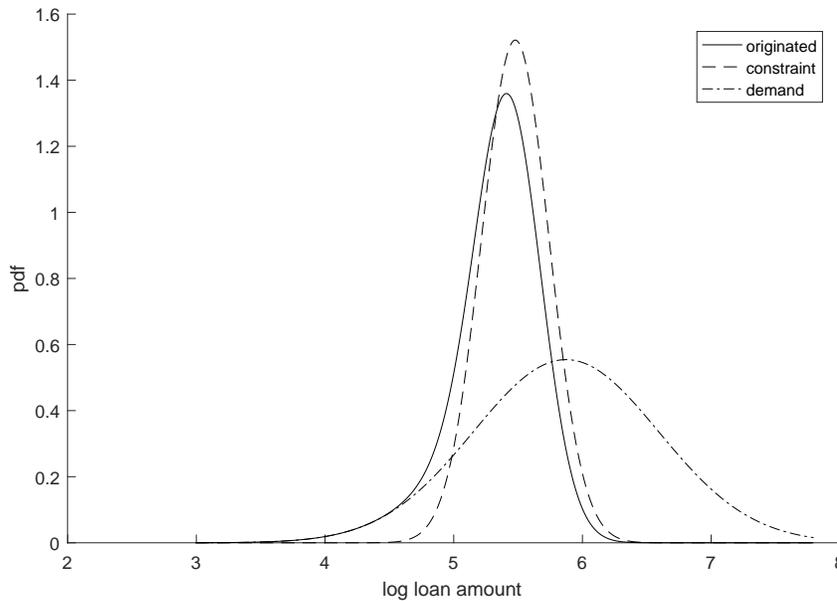
Figure 10: Aggregate Loan Frontiers by Metro Area



Note: The loan frontier is aggregated over downpayments, incomes, and FICO scores using the weights described in Section 4. The solid black line also aggregates over metro areas using population weights. The loan frontier is in thousands of 2014 dollars.

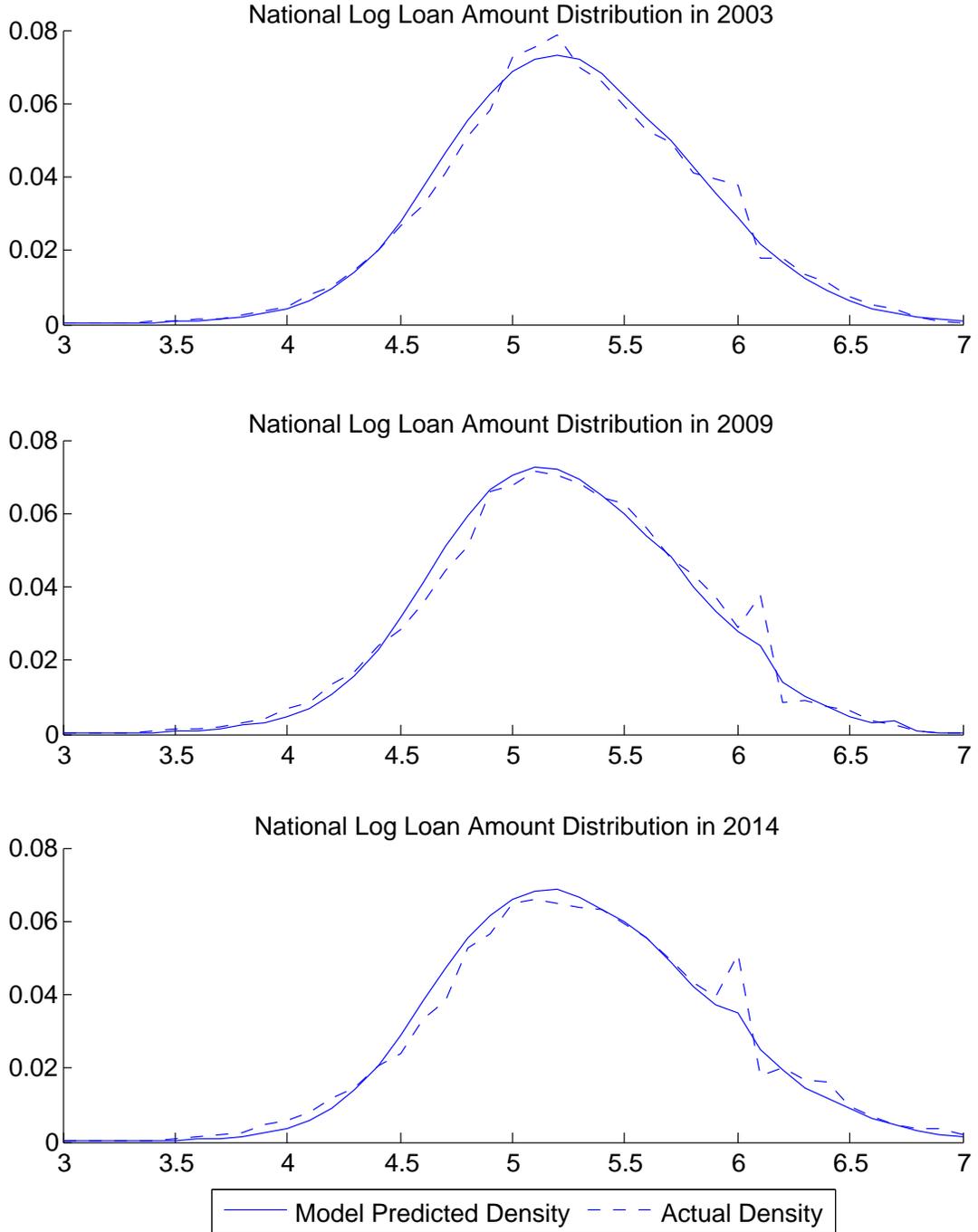
Source: Staff calculations based on data provided by McDash Analytics, LLC, a wholly owned subsidiary of Black Knight; and Corelogic.

Figure 11: Distribution of the Minimum of a Bivariate Normal



Note: Constraints and unconstrained demand are jointly log-normal with parameters described in Section 5. Originated loan amount is the minimum of constraint and unconstrained demand.

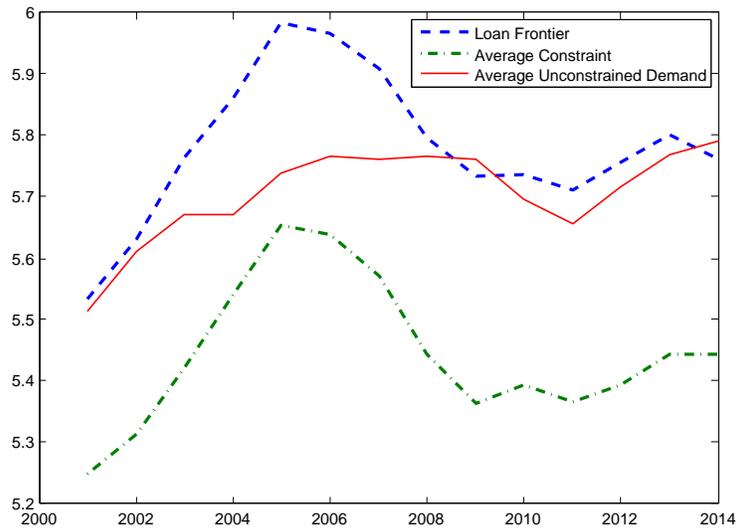
Figure 12: Model Fit of Loan Amount Distributions, by Year



Note: The figure compares the empirical distribution of mortgage originations with the simulated distribution from the estimates of the parametric model in Section 5.

Source: Staff calculations based on data provided by McDash Analytics, LLC, a wholly owned subsidiary of Black Knight; and Corelogic.

Figure 13: Estimates for Average Constraint, Average Unconstrained Demand, and Loan Frontier, 2001-2014



Note: The national average borrowing constraint and average unconstrained borrowing demand are estimated using the parametric model described in Section 5. The loan frontier is estimated non-parametrically as in Section 4.

Source: Staff calculations based on data provided by McDash Analytics, LLC, a wholly owned subsidiary of Black Knight; and Corelogic.

Figure 14: Estimates for the Share of Constrained Borrowers, 2001-2014



Note: The share of constrained borrowers is constructed by simulating the parametric model in Section 5 and computing the share of borrowers for whom the unconstrained borrowing demand is higher than their borrowing constraint.

Source: Staff calculations based on data provided by McDash Analytics, LLC, a wholly owned subsidiary of Black Knight; and Corelogic.

Table 1: Analysis of Variance for Loan Frontier

	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

Note: The average loan frontier is \$283k and the standard deviation is \$199k.

Table 2: The OLS Effect of the Loan Frontier on House Prices and Housing Stock

Dep. variable:	$\Delta \ln Price$		$\Delta \ln H stock$	
	(1)	(2)	(3)	(4)
$\Delta \ln Frontier$	0.582*** (0.083)	0.532*** (0.087)	0.018*** (0.006)	0.018*** (0.006)
$Inelastic \times \Delta \ln Frontier$		0.155*** (0.039)		0.006 (0.004)
$\Delta \text{Log Delinquency Rate}$	-0.122*** (0.013)	-0.108*** (0.015)	0.005** (0.002)	0.006** (0.003)
$\Delta \text{Log Income}$	0.077 (0.088)	0.034 (0.081)	-0.013 (0.017)	-0.013 (0.017)
$\Delta \text{Log Employment}$	0.983*** (0.243)	1.031*** (0.235)	0.216*** (0.044)	0.221*** (0.046)
Observations	1217	1152	1217	1152
R^2 overall	0.598	0.611	0.163	0.164

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. All specifications include metro area and year fixed effects. The clustered robust standard errors are given in parentheses. *, **, *** indicate statistical significance at the 90%, 95%, and 99% level respectively.

Table 3: First Stage Effects of the Instrument on Loan Frontiers

Dep. variable:	$\Delta \ln Frontier$		$Inelastic \times \Delta \ln Frontier$
	(1)	(2)	(3)
$\Delta \ln Instrument$	0.567*** (0.110)	0.532*** (0.118)	0.537*** (0.168)
$Inelastic \times \Delta \ln Instrument$		0.065*** (0.018)	0.841*** (0.028)
$\Delta \text{Log Delinquency Rate}$	-0.073*** (0.005)	-0.067*** (0.005)	-0.034*** (0.008)
$\Delta \text{Log Income}$	0.203** (0.096)	0.208** (0.089)	0.128 (0.091)
$\Delta \text{Log Employment}$	0.389** (0.196)	0.373** (0.184)	0.164 (0.170)
F-test of excluded Instruments	26.07	27.15	525.38
Underidentification test (p-values)	0.000	0.000	0.000
Observations	1217	1152	1152
R^2 overall	0.336	0.338	0.729

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. All specifications include metro area and year fixed effects. The clustered robust standard errors are given in parentheses. *, **, *** indicate statistical significance at the 90%, 95%, and 99% level respectively.

Table 4: The IV Effects of Loan Frontiers on House Prices and Housing Stock

Dep. variable:	$\Delta \ln Price$		$\Delta \ln Hstock$	
	(1)	(2)	(3)	(4)
$\Delta \ln Frontier$	1.205*** (0.301)	0.889*** (0.338)	0.088** (0.035)	0.095** (0.043)
$Inelastic \times \Delta \ln Frontier$		0.081* (0.047)		-0.003 (0.006)
$\Delta \text{Log Delinquency Rate}$	-0.071*** (0.025)	-0.089*** (0.023)	0.010** (0.004)	0.011** (0.005)
$\Delta \text{Log Income}$	-0.074 (0.094)	-0.038 (0.107)	-0.030* (0.017)	-0.031* (0.019)
$\Delta \text{Log Employment}$	0.782*** (0.200)	0.913*** (0.226)	0.194*** (0.030)	0.197*** (0.034)
Observations	1217	1152	1217	1152
R^2 overall	0.508	0.582	0.035	0.021

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. All specifications include metro area and year fixed effects. The clustered robust standard errors are given in parentheses. *, **, *** indicate statistical significance at the 90%, 95%, and 99% level respectively.

Table 5: The IV Effects of the Loan Frontier Directly Controlling for Interest Rates

Dep. variable:	$\Delta \ln Price$		$\Delta \ln Hstock$	
	(1)	(2)	(3)	(4)
$\Delta \ln Frontier$	1.209*** (0.294)	0.898*** (0.335)	0.085** (0.039)	0.089* (0.047)
$Inelastic \times \Delta \ln Frontier$		0.079* (0.046)		-0.002 (0.006)
$\Delta \ln MedianRate$	-0.036 (0.115)	-0.118 (0.108)	0.013 (0.012)	0.010 (0.013)
$\Delta \text{Log Delinquency Rate}$	-0.071*** (0.024)	-0.088*** (0.023)	0.010** (0.004)	0.010** (0.005)
$\Delta \text{Log Income}$	-0.074 (0.094)	-0.036 (0.106)	-0.028 (0.021)	-0.026 (0.023)
$\Delta \text{Log Employment}$	0.778*** (0.195)	0.900*** (0.221)	0.215*** (0.032)	0.217*** (0.036)
Observations	1217	1152	1120	1060
R^2 overall	0.507	0.581	0.066	0.066

Note: All the variables in this regression are in log differences. *MedianRate* is the median interest rate of all purchase loans in a metro-year. The sample consists of annual data from 2001 to 2013 for 100 metropolitan areas. All specifications include metro area and year fixed effects. The clustered robust standard errors are given in parentheses. *, **, *** indicate statistical significance at the 90%, 95%, and 99% level respectively.

A Online Appendix—Not for publication

A.1 Details of 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.²⁵ 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 initial 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 amount²⁶, 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.

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

²⁵For 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.

²⁶The 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.

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 mcdashids 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. We follow 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. 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.

A.2 Detail on Detecting Bunching at the Frontier Across Bins

For each of our fico, downpayment, income, year, msa bins that we compute frontiers for (“frontier bins”), we first calculate the share of observations within a certain distance of the frontier. We use twelve distance bins of length 5k, beginning at -49k (i.e. 44k-49k less than the frontier). Let s_{jb} denote the share of observations for frontier bin b within distance j of the frontier. Let j be the midpoint of the interval (e.g. for the interval $[-4k,1k]$, $j = -1.5$). We then estimate the following regression:

$$s_{jb} = \alpha_0 + \alpha_1 j + \alpha_2 j^2 + \alpha_3 I[j = -1.5k] + \alpha_4 I[j > -1.5k] + \epsilon_{jb} \quad (17)$$

separately by group. $\alpha_3 > 0$ and $\alpha_4 < 0$ would be suggestive of bunching because it implies that the bin just before the frontier and the bins just after the frontier have more and less

mass, respectively, relative to what a flexible function of j would suggest.²⁷ We define groups by first combining our 31 FICO frontier bins, 19 downpayment frontier bins, 18 income frontier bins, 14 year bins, and 100 frontier msas into 4 FICO bins (500-550, 550-600, etc), 6 downpayment bins (0-50k, 50k-100k,etc), 5 income bins (0-70k, 70k-110k, 110k-150k, etc), 14 year bins (i.e. years are not further grouped) and 10 city bins (cities are divided into bins according to their population rank). Each unique fico/downpayment/income/city/year bin combination constitutes a group, so we have 16800 groups ($4*6*5*14*10$). We find that 75 percent of groups have α_3 statistically significantly greater than zero and α_4 statistically significantly less than zero at the ten percent level, indicating that bunching is fairly widespread across frontier bins.

A.3 Robustness Results for Section 6

In this section, we show that our estimates in Table 4 are both qualitatively and quantitatively robust to (i) alternative choices of m when computing the frontier, (ii) alternative choices of weights s_j^k in computing the instrument, (iii) using only full-documentation loans to reduce the bias associated with income misreporting, and (iv) computing the frontier conditioning on unobserved borrower heterogeneity, defined as the residual from an interest rate regression.

First, we test the robustness of our main results to our choice of m , which as explained in the text, is the number of draws one takes from the sample when computing the expected maximum loan amount. Table 6 shows results for $m = 500$ and $m = 2,000$. The results do not appear to be sensitive to our choice of m .

Second, we test the robustness of our main result to the choice of weights, s_j^k , used to compute the instrument as in equation (16). Columns 1 and 2 of Table 7 show the regression results when s_j^k is defined as the share of individuals in bin k in metro j in 2001, rather than 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

²⁷We also tried including higher order j terms, and the results were very similar.

respect to the frontier are comparable to those in the baseline specification.

Third, we re-estimate the frontier, dropping all loan originations that are not flagged as fully documented.²⁸ The motivation for this specification is that researchers have found that reported incomes in HMDA appear to be overstated, particularly in 2005 and 2006 (e.g. Avery et al. (2012), Blackburn and Vermilyea (2012)). By focusing on loans with full documentation, we are focusing on a sample for which income overstatement is less likely. Columns 3 and 4 of Table 7 show that our results are similar when using this subsample of the data.

Finally, we consider the possibility of omitted variables. As discussed in Section 5, unobserved heterogeneity may be a concern if changes to the frontier are not correlated with changes to borrowing constraints faced by typical borrowers. In the IV regression, 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 interest 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 larger for metro/year/borrower type bins where the residual is more negative, which is consistent with this interpretation. To keep the analysis tractable, we categorize borrowers into two types: high types who have residual interest rates below average, and low types who have residual interest rates above average. Columns 5 and 6 of Table 7 report the results when we aggregate over the unobserved borrower type using equal weights for low and high types. The estimated elasticities of house price growth and housing stock growth with respect to the frontier are comparable to the ones in our baseline specification, suggesting that changes in the distribution of borrower unobservables are not

²⁸In our data, 41% of loan originations are classified as fully documented, 15% are limited/no documentation, and 44% are of unknown documentation.

²⁹Shocks 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.

³⁰In particular, we obtain the residual by regressing the interest rate at origination on FICO, LTV, income, origination amount, ARM dummy, loan type dummies, 30-year-term dummy, metro fixed effects, and interaction terms. The regressions are run separately for each year.

driving the estimation results.

Table 6: Robustness with respect to choice of m

Dep. variable:	$m = 500$		$m = 2000$	
	$\Delta \ln Price$	$\Delta \ln Hstock$	$\Delta \ln Price$	$\Delta \ln Hstock$
	(1)	(2)	(3)	(4)
$\Delta \ln Frontier$	1.119*** (0.264)	0.082*** (0.032)	1.291*** (0.338)	0.094** (0.038)
$\Delta \text{Log Delinquency Rate}$	-0.076*** (0.022)	0.010** (0.004)	-0.066** (0.027)	0.011** (0.004)
$\Delta \text{Log Income}$	-0.047 (0.084)	-0.029* (0.016)	-0.099 (0.104)	-0.032* (0.017)
$\Delta \text{Log Employment}$	0.798*** (0.196)	0.195*** (0.030)	0.765*** (0.205)	0.193*** (0.029)
Observations	1217	1217	1217	1217
R^2 overall	0.775	0.627	0.720	0.598

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. All specifications include metro area and year fixed effects. The clustered robust standard errors are given in parentheses. *, **, *** indicate statistical significance at the 90%, 95%, and 99% level respectively.

Table 7: Robustness with respect to alternate specifications

Dep. variable:	Presample weights from 2001		Only Full Doc. Loans		Controlling for Unobs. Type	
	$\Delta \ln Price$	$\Delta \ln Hstock$	$\Delta \ln Price$	$\Delta \ln Hstock$	$\Delta \ln Price$	$\Delta \ln Hstock$
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta \ln Frontier$	1.029*	0.140**	0.919**	0.130**	1.723***	0.105**
	(0.592)	(0.065)	(0.432)	(0.057)	(0.503)	(0.047)
$\Delta \text{Log Delinquency Rate}$	-0.090*	0.014**	-0.106***	0.012**	-0.033	0.011**
	(0.047)	(0.006)	(0.031)	(0.005)	(0.039)	(0.005)
$\Delta \text{Log Income}$	0.009	-0.038**	0.019	-0.037**	-0.246	-0.037*
	(0.123)	(0.017)	(0.097)	(0.015)	(0.161)	(0.019)
$\Delta \text{Log Employment}$	0.889***	0.184***	0.736***	0.161***	0.470*	0.180***
	(0.226)	(0.028)	(0.269)	(0.031)	(0.270)	(0.028)
Observations	1217	1217	1217	1217	1217	1217
R^2 overall	0.747	0.497	0.719	0.406	0.624	0.577

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. All specifications include metro area and year fixed effects. The clustered robust standard errors are given in parentheses. *, **, *** indicate statistical significance at the 90%, 95%, and 99% level respectively.