Forced Asset Sales and the Concentration of Outstanding Debt: Evidence from the Mortgage Market*

Giovanni Favara
Federal Reserve Board

Mariassunta Giannetti
Stockholm School of Economics
CEPR, and ECGI

August 2015

Abstract

We provide evidence that lenders differ in their ex post incentives to internalize price-default externalities associated with the liquidation of collateralized debt. Using the mortgage market as a laboratory, we conjecture that lenders with a large share of outstanding mortgages on their balance sheets internalize the negative spillovers associated with the liquidation of defaulting mortgages and are thus less inclined to foreclose. We find that zip codes with higher concentration of outstanding mortgages experience fewer foreclosures, more renegotiations of delinquent mortgages, and smaller house prices declines. These results are not driven by prior local economic conditions, mortgage securitization or unobservable lender characteristics.

Keywords: House Prices; Foreclosures; Bank Concentration

JEL Codes: G01; G21; R31; R38

*We thank Manuel Adelino, Gene Amromin, Elliot Anenberg, Eric Engstrom, Julian Franks, Paul Kalem, Anastasia Kartasheva, Benjamin Keys, Elena Loutskina, Steven Ongena, Karen Pence, Anthony Pennington-Cross, Uday Rajan, Avri Ravid, Amit Seru, Steve Sharpe, Shane Sherlund, Greg Udell, Vikrant Vig, and seminar participants at the Western Finance Association, the NBER Financing Housing Capital Conference, the London Business School Finance Summer Symposium, the New York University Stern School of Business, the University of British Columbia Sauder School of Business, the Federal Reserve Bank of Chicago, the Board of Governors of the Federal Reserve System, Boston University, Georgia State University, University of California, Irvine, the University of Lausanne and EPFL, the University of Amsterdam, the Goethe University in Frankfurt, the IDC Annual Conference in Financial Economics Research in Herzliya, the EFA in Lugano, the FIRS Conference in Quebec City, the SED Conference in Seoul, and the Summer Meeting of the Econometric Society at the University of Southern California. We also thank Mihir Gandhi for outstanding research assistance. Giannetti acknowledges financial support from the Jan Wallander and Tom Hedelius Foundation and the Bank of Sweden Tercentenary Foundation. This paper represents the views of the authors and not those of the Federal Reserve System or its Board of Governors. Favara: giovanni.favara@frb.gov; Giannetti: mariassunta.giannetti@hhs.se

Price dislocations associated with forced asset sales may generate externalities that feedback on asset values and cause price-default spirals, especially in illiquid markets with collateralized lending (Kiyotaki and Moore, 1997; Gromb and Vayanos, 2002; Brunnermeier and Pedersen, 2009). As asset price spirals may impair the balance sheets of other market participants, an important question is why lenders do not take actions to avoid collateral liquidation (Shleifer and Vishny, 2011). Surprisingly, the theoretical and empirical literature on this question is scant.

The purpose of this paper is to provide evidence that lenders differ in their ex post incentives to internalize price-default externalities, and that this heterogeneity depends on the share of collateralized debt in their portfolios. Our conjecture is that lenders holding a large share of the outstanding collateralized debt internalize the feedback effects of liquidation decisions on collateral values and may be inclined to renegotiate their debt to avoid price-default spirals. Using data on foreclosures and house prices during the 2007-2010 U.S. housing crisis, we find evidence that such incentives are at work and are economically significant.

The recent real estate crisis is an ideal laboratory for testing this conjecture for three reasons. First, mortgages, the standard debt contracts in the housing market, entitle lenders to seize the houses and sell them through a foreclosure process if borrowers default. Second, as the housing market is illiquid, foreclosures may generate price discounts that tend to spillover to non-distressed neighboring houses (Campbell, Giglio and Pathak, 2011; Harding, Rosenblatt, and Yao, 2009; Anenberg and Kung, 2013; Hartley, 2014). Third, the recent crisis has seen an unprecedented increase in foreclosures and decline in house prices, with feedback loops between foreclosures and prices contributing to the severity of the crisis. For instance, it has been shown that foreclosures led to a generalized decline in house prices (Mian, Sufi and Trebbi, 2015), which in turn caused additional foreclosures, as borrowers moved into negative equity positions (Elul, Souleles, Chomsisengphet, Glennon, and Hunt, 2010), triggering further price declines (Guren and McQuade, 2013).

We begin the analysis with a stylized model of the housing market in which negative income shocks force distressed homeowners to default on their debt obligations, and foreclosures trigger a decline in house prices, as debt liquidation creates an imbalance of housing demand and supply. In our simple setting, the initial decline in house prices is amplified as non-distressed homeowners that move into negative equity positions find it optimal to default. When mortgages are held by many (atomistic) lenders, each lender places little weight on the effects of its foreclosure decisions on local house prices, and so defaults are followed by further defaults. In contrast, when lenders hold a large share of outstanding mortgages on their balance sheets they internalize the adverse effects of their liquidation decisions on house prices, and have stronger incentives to renegotiate defaulting loans. Fewer liquidations mitigate the adverse impact of the initial decline in house prices, leading to fewer defaults.[1]

To test this theoretical prediction, we focus on the 2007-2010 U.S. housing crisis and perform two sets of tests. First, we use zip code level data on foreclosures, house prices and the concentration of outstanding mortgages. Zip codes are the smallest geographical areas for which we can measure the concentration of outstanding mortgages on lenders' balance sheets, and arguably the largest areas within which foreclosures are likely to generate negative externalities on house prices.[2] Second, for a subset of lenders with available data on mortgage performance, we use loan level data to test whether the same lender's incentives to foreclose on defaulting mortgages depend on the proportion of the zip code's outstanding mortgages on its balance sheet.

In the zip code level analysis, we construct an index of local concentration of mortgages on lenders' balance sheets using data on mortgages retained by the four biggest holders in a zip code. Lenders with a large share of retained mortgages in a zip code are expected to avoid foreclosures in order to minimize the negative impact on house prices and the consequent

[1] Although the foreclosure externality in this setting works through the borrowers' incentives to strategically default, other mechanisms may lead lenders to internalize the consequences of their foreclosure decisions. Negative spillovers could, for example, arise if the negative equity positions of borrowers constrain their ability to refinance their mortgages or if the fall in prices impairs the balance sheet of lenders, for instance through the holdings of previously repossessed properties.

[2] See Mian, Sufi and Trebbi (2015) for empirical evidence on foreclosure externalities at the zip code level.

losses due to foreclosure externalities. We find empirical support for this conjecture. All else equal, an increase in the index of outstanding mortgage concentration from the bottom to the top decile of the distribution reduces the foreclosure rate by approximately 25 percent in the average zip code. These results are obtained controlling for the delinquency rate in the zip code, proxies for borrower creditworthiness, and standard controls for zip code housing, income, and demographic characteristics. Furthermore, these results are obtained controlling for unobserved factors that uniformly affect zip codes within a county, such as income shocks, local market conditions and ex ante lender competition.

A causal interpretation of these results relies on the assumption that given our set of controls the share of loans retained on a lender's balance sheet is orthogonal to the credit quality of mortgages originated. However, there might be selection effects with safer loans retained and riskier loans securitized. In addition, lenders' decision to keep mortgages on their balance sheets may be jointly determined by unobservable lender and local market characteristics. We address these concerns in several steps. First, we control for the fraction of loans that are 90 or more days delinquent throughout the analysis. Mortgage delinquency is the single most important predictor of foreclosures and is likely to absorb most of the differences in ex ante loan quality and ex post economic and financial distress across zip codes.[3] Second, we provide evidence that the concentration of outstanding mortgages is uncorrelated with several observable zip code characteristics. Third, we show that our main results hold when we instrument the share of loans retained on a lender's balance sheet with changes in local lending conditions due to exogenous banks' mergers and with demographic characteristics of local markets that are likely to influence banks' incentives to retain mortgages, but not real estate market dynamics. Last but not least, we provide evidence consistent with a causal mechanism running from the concentration of outstanding mortgages to foreclosure rates and house prices by exploiting the cross-sectional implications of our hypothesis.

For instance, we find that the concentration of outstanding mortgages reduces foreclosure

[3]Supporting this conjecture, common predictors of foreclosures that we consider in our analysis lose explanatory power once we control for the delinquency rate in the zip code.

rates to a larger extent in zip codes with higher delinquency rates, reassuring that our findings are not driven by ex ante differences in the quality of borrowers. We also study how the results vary across jurisdictions with different foreclosure procedures. We expect that any lender, regardless of its outstanding mortgages in the neighborhood, has weaker incentives to foreclose in states with costly foreclosure procedures. Consistent with this idea, we find that the concentration of outstanding mortgages is associated with fewer foreclosures only in non-judicial states where foreclosure costs are lower. We also find that in zip codes with a higher concentration of outstanding mortgages more delinquent mortgages are renegotiated. These results strengthen the interpretation that lenders with a large fraction of the outstanding mortgages differ in their ex post incentives to resolve distress and not in their ex ante ability to screen borrowers.

We next explore the role of securitization. While securitization may strengthen lenders' incentives to foreclose because it decreases the concentration of outstanding mortgages, it may also lead to renegotiation frictions for reasons that are orthogonal to the one we propose. For instance, dispersed ownership brought about by securitization or agency problems between servicers of securitized loans and investors may impede mortgage renegotiation (Piskorski, Seru and Vig, 2010; Agarwal et al., 2011). To separate the role played by securitization, all our specifications control for the share of loans securitized in each zip code and the share of mortgages securitized with GSEs. In some specifications, we also decompose the index of concentration of outstanding mortgages to separate the fraction of loans retained by the four biggest holders in a zip code from the fraction of loans securitized in the same zip code. The outcomes of these tests reassure us that the effects of outstanding mortgage concentration are distinct from those related to securitization.

We then address the concern that our findings are driven by lenders' unobserved characteristics, such as organizational structures and renegotiation capabilities. For example, lenders' ability to collect soft information on borrowers' quality may affect their ex ante incentives to hold mortgages on their balance sheets and their ex post incentives to renegotiate defaulting mortgages. Using loan level data, we perform within-lender tests and evaluate whether the propensity of the same lender to foreclose on a defaulting mortgage varies with the share of mortgages that the lender has retained on its balance sheet in a zip code. Consistent with our zip code level evidence, we find that lenders are less willing to foreclose on delinquent loans in areas where they retained a higher fraction of outstanding mortgages. Conversely, the share of loans that lenders hold on their balance sheets is not statistically related to the foreclosure probability of securitized delinquent mortgages.

In a final step, we explore whether areas with higher concentration of outstanding mortgages also experience lower house price declines as the causal mechanism of our hypothesis implies. We find that a move in the concentration index from the bottom to the top decile of the distribution leads to 6 percent lower rate of house price declines in the average zip code.

Our paper is most closely related to empirical research on forced sales of real assets (Shleifer and Vishny, 1992, 2011). This literature focuses on the negative externalities associated with asset sales in economic downturns. For example, Benmelech and Bergman (2011) document that during recessions, a firm's bankruptcy reduces collateral values of other industry participants imposing negative externalities on their non-bankrupt competitors. Asquith, Gertner, and Scharfstein (1994) find that lenders avoid liquidation and prefer to renegotiate troubled loans when industry conditions deteriorate. We depart from this literature by studying how lenders' incentives to avoid externalities due to liquidation depend on lenders' market structure. To our knowledge, this is the first paper to explore the role of market structure on lenders' liquidation incentives and asset prices.

As our analysis focuses on the housing market, our paper also contributes to the literature on the recent housing crisis. A number of papers explore how differences in the local mortgage markets are associated with the intensity of the crisis (e.g., Mian and Sufi, 2009 and 2011; Mayer, Pence and Sherlund, 2009; Keys, et al., 2010; Purnanandam, 2011) and whether securitization has exacerbated the intensity of the crisis by inhibiting the renegotiation of delinquent loans (Agarwal, et al., 2011; Piskorski, Seru and Vig, 2010; Adelino, Gerardi and Willen, 2013). Our paper focuses, instead, on the incentives of lenders to foreclose portfolio loans, relates such incentives to the share of outstanding mortgages they retained in a neighborhood, and studies the implications for foreclosure rates and house prices.

A related strand of literature studies the costs and benefits of ex post loan renegotiations. While renegotiations may prevent foreclosures and limit deadweight losses for borrowers and lenders, such policies may strengthen borrowers' incentives to default strategically (Agarwal et al., 2014; Mayer at al., 2014). Our stylized model suggests that loan renegotiations limit the losses of a lender with a large share of outstanding mortgages even in presence of strategic defaults.

The paper is also related to the literature that explores the effects of banks' loan concentration on bank-firm relationships (Berger, Miller, Petersen, Rajan, and Stein, 2005), the loan supply (Garmaise and Moskowitz, 2006), and the transmission of monetary policy to mortgage rates (Scharfstein and Sunderam, 2015). All these papers study the effects of market power concentration on loan origination and contract terms. We focus, instead, on the role of concentration of outstanding mortgages on lenders' ex post incentives. By showing that a market with dispersed lenders is more prone to liquidation externalities, we also provide an alternative interpretation to the view that competition in the credit market erodes financial stability because it distorts lenders' risk taking decisions by lowering their profit margins (Keely, 1990).

The rest of the paper is organized as follows. Section 1 describes the theoretical model and summarizes the theoretical predictions on the relationship between the concentration of outstanding mortgages on lenders' balance sheets, foreclosures and house prices. Section 2 describes the data and our empirical strategy. Section 3 presents our main empirical results on foreclosure rates and Section 4 on house prices. Section 5 concludes.

1 Theory and Testable Implications

In this section, we develop a simple model to illustrate the relationship between foreclosures, house prices and the concentration of outstanding mortgages. In the model, foreclosures generate an imbalance of housing supply and demand and cause a decline in the equilibrium prices. As prices decline, borrowers that would otherwise have been able to repay their mortgages default strategically, because the value of their homes falls below the value of their mortgages. We discuss other negative spillovers of foreclosures that may be internalized by concentrated lenders and show that lenders holding a large share of the outstanding mortgages internalize the negative spillovers of foreclosures on house prices and are thus less inclined to foreclose. In the following sections, we bring this conjecture to the data.

1.1 The model

1.1.1 Assumptions

There are two dates and two groups of agents of mass 1, households (indexed by i) and lenders. At $t = 0$, some households enter the period with one unit of housing endowment, $h_{0i} = 1$, and an outstanding mortgage payment, B. At $t = 1$, households enjoy utility from consumption, $c_i \geq 0$, and housing, $h_i \in \{0, 1\}$:

$$U_i = c_i + \gamma_i h_i,$$

where γ_i is uniformly distributed, $\gamma_i \sim \mathcal{U}\ [0, \overline{\gamma}]$, and captures heterogeneity in utility from home ownership. Households with endowment $h_{0i} = 1$ have the highest utility from housing services. Aggregate housing supply is fixed at $\overline{H} < \overline{\gamma}$.

At $t = 1$, households receive a random income, w_i, which is independently distributed from γ_i. With probability q, everyone receives w. With probability $1 - q$, a fraction e of households suffer a negative income shock and receives θw, with $0 < \theta < 1$. We assume that income shocks are observable even though not verifiable. We also assume that distressed households are unable to repay B:

$$w > B > \theta w, \tag{1}$$

and that lenders may partially recover B by selling the houses of these households at a price p (to be derived below).

Under these assumptions, household i's budget constraint at $t = 1$ depends on the realization of the income shock, the repayment or default on the mortgage debt, and whether the lender forecloses in case of default:

$$w_i = \begin{cases} c_i + B + p(h_{1i} - h_{0i}) & \text{no default} \\ c_i + ph_{1i} & \text{default \& foreclosure} \end{cases}.$$

1.1.2 Equilibrium housing prices and strategic defaults

In absence of shocks, the unit housing demand is pinned down by the following condition:

$$\gamma_i \geq p,$$

which relates the utility value of owning to the price of housing. Since γ_i is uniformly distributed, the equilibrium price is determined by equating aggregate demand and supply:

$$p = \overline{\gamma} - \overline{H} > B.$$

At this price, all households repay B and, under our assumption on the initial distribution of housing, they hold on to their houses.[4]

In contrast, when some households are hit by a negative income shock, they cannot afford to repay B (by (1)). If lenders foreclose on these distressed households, a fraction e of them

[4] If the repayment obligation were larger than the equilibrium price, $B > p$, households would default as they have the option to surrender their houses to the lender. The condition $\overline{\gamma} - \overline{H} > B$ rules out this possibility.

is excluded from the housing market. The market clearing condition becomes:

$$(1-e)\left(\overline{\gamma}-p\right)=\overline{H},$$

and the equilibrium price is:

$$p^L=\overline{\gamma}-\frac{\overline{H}}{1-e}.$$

It follows immediately that p^L is strictly lower than p, because a fraction e of households with high utility from owning cannot participate in the market, reducing aggregate demand.

An equilibrium in which lenders foreclose on distressed households implies that $p^L < B$, otherwise distressed households would prefer to sell their houses and pay back their mortgage payment. This equilibrium also implies that non distressed households always default strategically because they can purchase a house at a price lower than B, even though they can afford to repay B.[5] Therefore, in equilibrium, it must be that:

$$\theta w<\overline{\gamma}-\frac{\overline{H}}{1-e}\leqslant w,$$

meaning that households that suffer a negative income shock are unable to participate in the housing market (the first inequality), while non distressed households default strategically.[6] The above discussion can be summarized in the following Lemma.

Lemma *If lenders foreclose on distressed households, house prices fall and non distressed borrowers find it optimal to default strategically.*

It is important to note that in this setting atomistic lenders always find it optimal to foreclose because the highest payment a distressed borrower can promise is θw, but the

[5] The same result would be obtained if other investors purchased the houses and provided housing services to households that have strategically defaulted.

[6] This is the only equilibrium with foreclosure and strategic defaults. The condition $p^L=\overline{\gamma}-\frac{\overline{H}}{1-e}<w$ is implied by $p^L<B<w$. An equilibrium in which $p^L>w$ does not exist because no households would be able to purchase a house, causing the house price to fall. Similarly, it cannot be that $p^L<\theta w$. If this were the case, at least as many households as in the state of the world in which no income shock occurs would be able to purchase a house, driving the equilibrium house price above θw.

equilibrium price that prevails under foreclosure is $p^L > \theta w$.

1.1.3 Lenders' shares of outstanding mortgages and foreclosure decisions

We now consider the case in which one lender holds a large share, ξ, of the outstanding mortgages in the market (ξ-lender), and the remaining share $(1-\xi)$ is dispersed among many atomistic lenders.[7]

If the ξ-lender were to renegotiate the mortgage payment of distressed households, while the other atomistic lenders foreclosed on defaulting borrowers, the aggregate housing demand would be:

$$(1-\xi)\,(1-e)\,(\overline{\gamma}-p^{L'})+\xi(\overline{\gamma}-p^{L'}),$$

and the equilibrium price

$$p^{L'}=\overline{\gamma}-\frac{\overline{H}}{(1-\xi)\,(1-e)+\xi},$$

which is strictly larger than p^L, and increasing in ξ. For values of ξ close to 1, the equilibrium price could be such that $p^{L'} \geq B$, and no default (strategic and non-strategic) would take place. For lower values of ξ, the equilibrium house price may fall below B and any borrowers, including non distressed ones, find it optimal to default.

Under the assumption that income shocks are observable, the $\xi-$lender can offer to reduce the mortgage payment of non distressed households to $B' = p^{L'} < B$, and these households would find it optimal to accept the offer.[8] In this case, the $\xi-$lender is willing to renegotiate with, rather than foreclose on, distressed households if

$$(1-e)p^{L'}+e\theta w > p^L, \tag{2}$$

[7] The trust of the results we present hereafter continues to hold if we allow for several ξ-lenders and mixed strategies over foreclosure and renegotiation decisions.

[8] The assumption that at least the lender with a high share of outstanding mortgages can distinguish households that suffer a negative income shock is crucial (see Ghent (2011) for some supporting evidence). If this was not possible, intact households could strategically ask for a loan modification. The model can, however, be modified to allow banks to imperfectly distinguish between intact and distressed households.

where the left hand side of (2) is the total return from renegotiation. Using the equilibrium prices $p^{L'}$ and p^L, this condition can be rewritten as

$$\frac{\xi}{(1-\xi)(1-e)+\xi}\frac{\overline{H}}{1-e} > \overline{\gamma} - \frac{\overline{H}}{(1-\xi)(1-e)+\xi} - \theta w_1$$

which is more likely to hold as ξ increases.

Also, if this condition holds, not only there are fewer foreclosures and smaller house price declines, but aggregate mortgage losses are also lower. The reason is that all lenders obtain a higher average repayment, including dispersed lenders who are able to foreclose houses at a higher equilibrium price.[9]

The following proposition summarizes this discussion

Proposition *There are fewer foreclosures and smaller declines in house prices in areas where lenders hold a large share of the outstanding mortgages.*

1.1.4 Discussion

In the model, foreclosures trigger more defaults and exacerbate house price declines because of the non-distressed borrowers' incentives to default strategically. Although the notion that a generalized fall in house prices may lead to strategic defaults is supported by empirical evidence (Elul, Souleles, Chomsisengphet, Glennon, and Hunt, 2010; Guiso, Sapienza and Zingales, 2013; Li and Zhao, 2015), the concentration of outstanding mortgages could mitigate the adverse effects of foreclosures even if the externality operated through different mechanisms. For instance, a given income shock may be amplified if foreclosures lead to price declines that deteriorate the balance sheets of all households and cause a reduction in real activity, which in turn may lead to further defaults. Also, foreclosures and price declines may impair the balance sheets of lenders if lenders have direct exposures to local real es-

[9] When a lender renegotiates a defaulting loan, other lenders have stronger incentives to foreclose because the equilibrium house price is higher. Thus, a lender with a large share of outstanding mortgages cannot prevent strategic defaults, but it can mitigate the effects of negative income shocks on foreclosures and house prices. These mitigating effects would be larger if defaults were costly for borrowers.

tate through, for instance, the holdings of previously repossessed properties. Alternatively, externalities may arise because foreclosures reduce the amenity value of neighboring houses (Fisher, Lambie-Hanson, and Willen, 2014) or because the price declines due to foreclosures prevent borrowers from refinancing. In all these cases, we would expect that lenders with a large shares of outstanding mortgages have stronger incentives to avoid foreclosures.

Our empirical analysis aims to capture any of these mechanisms. Inspired by the equilibrium relations of our stylized model, we estimate reduced form equations relating foreclosures and house prices changes to the share of the outstanding mortgages that lenders retain in a neighborhood.

2 Empirical Challenges

The main testable hypothesis of our analysis is that lenders that retain a large share of the outstanding mortgages on their balance sheets are more likely to internalize the adverse effects of foreclosures on house prices and are therefore less inclined to foreclose.

To assess the validity of this hypothesis, we use data on local housing markets. We start by computing the share of mortgages outstanding in a neighborhood that are on a lender's balance sheet. This requires information on the identity of the lender originating mortgages in a given neighborhood and information on whether these mortgages are retained or securitized. This information can only be obtained from the Home Mortgage Disclosure Act (HMDA) data, the largest source of primary U.S. mortgage originations, covering over 90 percent of the mortgage activity of commercial banks, thrifts, credit unions, and mortgage companies (see, e.g., Mian and Sufi, 2009; Loutskina and Strahan, 2011, Favara and Imbs, 2015). From HMDA we use data on lender's identity, location of the property purchased with a mortgage, and mortgage disposition to compute zip code level measures of outstanding mortgage concentration.[10]

[10]Since HMDA reports census tract information of the property location, we match census tracts to zip codes using the crosswalk provided by the U.S. Department of Housing and Urban Development and the

Unfortunately, HMDA does not provide information on loan performance, such as defaults and foreclosures. Since this information is crucial for our analysis, we obtain foreclosure data from RealtyTrac.com, a leading online marketplace for foreclosure properties, covering over 92 percent of the U.S. housing units. RealtyTrac data, however, cannot be matched with mortgage originations in HMDA, preventing us to test our main hypothesis with loan level data. RealtyTrac information on the location of the foreclosed properties can still be used to compute the number of foreclosures in a given zip code, which enables us to conduct most of our analysis with zip code level data. Section 2.1 provides precise details on variables' construction.

To ascertain that lenders' propensity to foreclose on defaulting mortgages does not depend on unobservable lenders' characteristics (such as organizational structures and renegotiation capabilities), we also perform a loan level analysis by merging HMDA information on mortgage originations with information on mortgage performance from the Lender Processing Services (LPS) Applied Analytics. LPS covers roughly 60 percent of the mortgage market in the U.S. and contains information on mortgage defaults and foreclosure starts obtained from the largest mortgage servicers.

The merged HMDA-LPS data enable us to test whether the same lender has a stronger propensity to foreclose in zip codes where it has retained a smaller share of outstanding mortgages. Unfortunately, the merged dataset has some limitations arising from some specific features of LPS. First, LPS provides a less comprehensive coverage of the U.S. mortgage market and of portfolio loans in particular, as mortgages serviced by third party servicers are usually securitized mortgages. Second, LPS records accurately when foreclosures start, but not when they are completed. This introduces some measurement error because a delinquent borrower may become current for several reasons, including a loan modification. The HMDA-LPS data may therefore record too many foreclosures, and even count renegotiations as foreclosures. While this measurement error is likely to inflate the standard errors of our

number of residential addresses as weights.

estimates, we cannot envisage any reason why it should be correlated with the concentration of outstanding mortgages.

2.1 Main variables and descriptive statistics

In the zip code level analysis, our sample consists of approximately 6,000 zip codes in 694 urban counties, as illustrated in Figure 1. We focus on urban areas because house price dynamics, borrower characteristics, and mortgage lending decisions have different determinants in rural, often poor, areas. In addition, it is well known that HMDA coverage is limited and not representative in rural areas (Avery, Brevoort, Canner, 2007).

Table 1 reports definitions, data sources, and summary statistics for the main outcome and control variables used in the analysis. Outcome variables are computed between 2007 and 2010 to measure zip code performance during the U.S. housing crisis.[11] In contrast, all controls (with the exception of the delinquency rate) are measured during the period preceding the crisis, i.e. 2004-2006.

The main outcome variable in our analysis is the foreclosure rate, which is measured using RealtyTrac.com's records on properties that receive a notice of sale, divided by the number of single family owner occupied housing units.[12] As shown in Table 1, the foreclosure rate has large cross-sectional variation.

The second outcome variable is the loan modification rate. We use this variable to study whether lenders are more inclined to renegotiate defaulting loans in areas with higher

[11] Our results are invariant if we consider only the period between 2007 and 2009. We include 2010 because foreclosure completions may have been delayed until 2010 in jurisdictions that require the use of a judge to complete a foreclosure procedure (Mian, Sufi and Trebbi, 2014). The fact that our results are robust to the exclusion of 2010 suggests that the Home Affordable Modification Program (HAMP), introduced at the end of 2009 to help financially struggling homeowners avoid foreclosure by modifying loans, is unlikely to influence our results.

[12] RealtyTrac.com collects information from the moment a foreclosure procedure begins, through a notice of default, until the defaulting property is sold at a public auction through a notice of sale. We use only notice of sales to identify foreclosures, because a defaulting borrower can always reinstate loan payments after a foreclosure's start and before its completion. We also exclude real estate owned (REO) properties, i.e., properties that have been repossessed by lenders, because almost all REOs occur after a notice of sale. None of our results depends on the exclusion of REOs from the total number of foreclosures. Zip code measures of the single family housing stock come from the 2000 Census.

concentration of outstanding mortgages. We compute the number of loan modifications in a zip code using LPS data and the algorithm of Adelino, Gerardi and Willen (2013). The algorithm identifies a loan modification whenever a mortgage is in default and there is a change in its contractual terms, such as a reduction in interest rates, a term extension or a change in the outstanding mortgage balance. For our purposes, we compute the share of portfolio and securitized loans that have been modified in a given zip code relative to all defaulting loans in the same zip code.

The third outcome variable is the change in house prices. If foreclosures generate negative externalities that reduce the value of nearby homes, changes in our measure of outstanding mortgage concentration should also correlate with zip code changes in house prices. We obtain data on house prices from CoreLogic, which provides quality-adjusted house price indexes for existing single-family properties. Between 2007 and 2010, some zip codes experienced house prices depreciations of over 50 percent, others witnessed house price depreciations of 2 percent or less.

To measure the local concentration of outstanding mortgages on lenders' balance sheets, we construct an index, called $Top4$, which is computed from HMDA as the number of mortgages retained by the four biggest holders in a zip code between 2004 and 2006, divided by the total number of mortgages originated in that zip code during the same period:

$$Top4_{z,04-06} \equiv \frac{MR_{1,z} + MR_{2,z} + MR_{3,z} + MR_{4,z}}{TotalOriginations_z}. \tag{3}$$

In equation (3), $MR_{i,z}$ is the number of mortgages retained by the lender ranked i in zip code z over the 2004-2006 period and $TotalOriginations_z$ is the total number of loans originated in zip code z by all lenders over the same period.[13] The numerator of the $Top4$ index uses mortgages retained because we want to measure the credit risk exposure of lenders to the local market. A mortgage is classified as retained, if it is not sold within a year to a GSE or a non-affiliated institution. Since the process of securitization takes on average two to three

[13]Results are similar to the ones we present if we use the volume rather than the number of mortgages.

months (Rosen (2010)), we consider only mortgages originated in the first three quarters of the year, as mortgages issued at the end of the year may be securitized at the beginning of the following year and thus improperly classified as retained.[14] The $Top4$ index is computed over the three-year interval, 2004-2006, because we want to measure concentration of mortgage holdings in terms of the stock of retained mortgages just before the U.S. foreclosure crisis.[15]

As shown in Figure 2, there is substantial nationwide variation in the concentration of outstanding mortgages. In some zip codes, the top 4 holders retain less than 5 percent of the outstanding loans, in others over 30 percent. The $Top4$ index has also considerable dispersion *within* counties. Figure 3 displays the distribution of the $Top4$ index within three of the largest counties (Los Angeles, Dallas, and Cook) and a relatively smaller one (Philadelphia) and shows that the range of variation of the $Top4$ index within these counties is not dissimilar from the one we observe nationwide.

By construction, the proportion of securitized mortgages in each zip code is negatively correlated with the concentration of mortgages held on lenders' balance sheets. To isolate variation in the $Top4$ index due to securitization, we decompose the index as follows:

$$Top4_{z,04-06} = Top4\_ret_{z,04-06} \times \left(1 - \sec_{z,04-06}\right),$$

where

$$Top4\_ret_{z,04-06} \equiv \frac{MR_{1,z} + MR_{2,z} + MR_{3,z} + MR_{4,z}}{Loans\,\mathrm{Re}\,tained_z}$$

captures variation in $Top4$ due to the share of loans retained by the four biggest holders in a zip code, and

[14] The coefficient of correlation between the index of concentration computed with and without the mortgages originated in the last quarter of each year is approximately 0.9.

[15] We consider only mortgages originated for the purchase of single-family, owner-occupied houses, because common house price indexes, including Corelogic, which we use to study the house price implications of lenders' incentives to foreclose, consider only this type of properties. We also exclude loans for refinancing and home improvement.

$$\mathrm{sec}_{z,04-06} = \frac{Loans\, Securitized_z}{TotalOriginations_z}$$

measures the share of loans securitized in the same zip code. As shown in Figure 4, the nation-wide distribution of $Top4\_ret$ is less skewed than $Top4$, but equally dispersed across zip codes.

2.2 Empirical Framework

To test our hypothesis, we estimate the following cross-sectional regressions on zip code level data (z):

$$y_{z,07-10} = \alpha\ Top4_{z,04-06} + \beta\ 90^{+}Delinquencies_{z,07-10} + \gamma\ X_{z,l,04-06} + \delta_C + \epsilon_{z,07-10},$$

where the dependent variable is the foreclosure rate, the loan modification rate, or the logarithmic change in house prices, all measured between 2007 and 2010, and $Top4_{z,04-06}$ is our measure of local concentration of mortgage holdings between 2004 and 2006.

Our empirical strategy consists in exploring the effect of $Top4$ conditionally on the 90 plus days delinquency rate, $90^{+}Delinquencies_{z,07-10}$, during the 2007-2010 period. The delinquency rate is the single most important determinant of foreclosures and helps us to address the main concern that the $Top4$ index may be correlated with unobserved factors that affect both foreclosures and house price dynamics. For instance, a negative correlation between the $Top4$ index and foreclosure rates could arise because lenders have kept on their balance sheet mortgages originated to more creditworthy borrowers. In this case, the ex ante loan quality, rather than differences in ex post lenders' incentives to foreclose, could bias our empirical analysis. Similarly, a positive correlation between $Top4$ and house prices could reflect heterogenous income shocks across zip codes during the 2007-2010 period that happen to be correlated with our concentration index. The contemporaneous zip code delinquency rate is likely to absorb such confounding effects.

Our regression framework also includes county or MSA fixed effects, δ_C, to control for other unobserved factors that uniformly affects zip codes within the same geographical area, such as state lending or foreclosure laws and economic conditions specific to a given county or MSA. In addition, since households are likely to approach different lenders within a county or MSA to obtain a mortage, the inclusion of county or MSA fixed effects allows us to hold constant ex ante competition in the mortgage market and the contract terms faced by borrowers. Finally, the matrix of controls, $X_{z,04-06}$, includes observable zip code characteristics, measured between 2004 and 2006, that are likely to predict our outcome variables. We include characteristics of the mortgage market, such as the proportion of private label and GSE securitized mortgages, and various proxies of borrower's credit quality and leverage, including the median income, the fraction of subprime borrowers, as well as the total debt per capita in each zip code. While these control variables are predetermined, none are truly exogenous. Their inclusion is an attempt to ensure that the $Top4$ index has explanatory power, correcting for the usual determinants of foreclosures and house prices.

We provide evidence consistent with a casual mechanism of our main findings through a series of exercises. First, we show that the $Top4$ index is uncorrelated with prior observable zip code level characteristics. Second, we use propensity scores to show that our main results hold in a sample of zip codes that are very similar along observable dimensions. Third, we instrument the $Top4$ index with changes in local lending conditions due to exogenous banks' mergers and with demographic characteristics of local markets that are likely to influence banks' incentives to retain mortgages and to satisfy the exclusion restriction. Finally, we show cross-sectional differences in the effect of $Top4$ across zip codes that are consistent with a casual effect running from the $Top4$ index on foreclosure rates and house prices. We discuss these tests after describing the main results.

3 Foreclosures and outstanding mortgage concentration

Table 2 reports our main results, with the number of foreclosures per homeowner as the dependent variable.[16] A negative correlation between the $Top4$ index and foreclosure rates is consistent with the hypothesis that the local concentration of mortgage holdings mitigates the incentives to foreclose. Across all specifications estimated in Table 2, there is a statistically significant negative correlation between the $Top4$ index and foreclosure rates. Importantly, the magnitude of this correlation is invariant if we control for unobserved geographical heterogeneity by including MSA fixed effects (columns 1 and 3) or county fixed effects (columns 2 and 4). The stability of the estimated coefficients in the specifications with MSA and county fixed effects suggests that unobservable factors correlated with neighborhood effects are unlikely to bias our findings. While the inclusion of the 90+ delinquency rate reduces the size of the estimated coefficients, the economic significance remains sizable. In column 4, a two-standard-deviation increase in the $Top4$ index is associated with a reduction in foreclosure rate of 25 percent in the average zip code.[17]

3.1 Endogeneity concerns: Determinants of Top4, matched samples and IV regressions

A casual interpretation of the results in Table 2 requires that the concentration of mortgages on lenders' balance sheets is uncorrelated with zip code characteristics, given our set of controls. However, lenders' decisions to retain loans are unlikely to be random and the concentration of outstanding mortgages may reflect local market characteristics. A prominent concern is that our measure of outstanding mortgage concentration correlates with

[16] In untabulated results, we standardize the number of foreclosures with the number of mortgages that are more than 90 days delinquent. The results are qualitatively similar to the ones we report hereafter.

[17] We use a two-standard-deviation change because the distribution of *Top4* is highly skewed, as shown in Figure 2. A two-standard-deviation change corresponds, approximately, to a move from the 10th to the 90th percentile of the *Top4* distribution.

demographic characteristics and the credit quality of loans originated in a zip code.

We start addressing this concern in Table 3 by reporting within-county correlations between the $Top4$ index and zip code level characteristics. We test whether the $Top4$ index is significantly related to observable measures of a neighborhood's credit quality, mortgage credit and house prices. In columns 1 to 4, we consider the median income, the share of minority population, the fraction of subprime borrowers, and the total debt per capita, respectively. We find that none of these observable characteristics is significantly related to the $Top4$ index. In column 5 and 6, we correlate the $Top4$ index to the fraction of mortgages originated between 2004 and 2006, and the change in house prices during the same time period. We find no indication that the $Top4$ index is correlated with local mortgage and housing market dynamics within a county. If local credit and house price dynamics reflect heterogeneity in banking competition, these results also point to a lack of within county correlation between the $Top4$ index and lending competition, suggesting that ex ante contract terms are also unlike to differ across zip codes.

The first column in Table 4 provides corroborating evidence that observable zip code characteristics do not drive our findings. Using a propensity score matching methodology, we match zip codes with $Top4$ index above the median with zip codes with $Top4$ index below the median. To perform the matching, we estimate propensity scores using all control variables in column 4 of Table 2. This procedure ensures that the estimation sample consists of zip codes that are similar in terms of demographic, credit quality and market characteristics. It is reassuring that the parameter estimates in column 1 of Table 4 are if anything larger than those of the baseline regressions in Table 2.

While this evidence suggests that the estimated effect of the $Top4$ index on foreclosures is unlikely to be spuriously driven by factors related to observable characteristics, the $Top4$ index may still reflect *unobservable* market characteristics and differences in the *unobserved* quality of mortgages. To start tackling this more challenging problem, we exploit arguably exogenous variation in the $Top4$ index caused by differences in the concentration of retained mortgages due to bank mergers and by differences in the lenders' propensity to securitize (retain) mortgages due to changes in the local supply of deposits.

Specifically, building on Garmaise and Moskovitz (1996), we use a zip code level measure of mergers between large non-failing commercial banks as an instrument for $Top4$. We focus on mergers between large non-failing banks of at least $1 billion in assets, because the nature and size of the lenders make it unlikely that economic conditions in a given zip code drive the mergers. This allows us to capture changes in the $Top4$ index that are orthogonal to current and prior neighborhoods' performance. Our zip code measure of bank mergers is based on mergers that occur between 2004 and 2006 and exploits information on the location of bank operations using the zip code address of survivor and non-survivor banks' branches.[18]

The second instrument for the $Top4$ index is based on the insight of Loutskina and Strahan (2009) and Gilje, Loutskina and Strahan (2013) that mortgage retention is more likely in areas with more bank deposits, especially for riskier and less liquid loans. Becker (2007) shows that the proportion of population in a county that is aged 65 and above affects positively bank deposits, but is uncorrelated with local economic outcomes. Building on Becker's results, we use the fraction of seniors as an instrument for local banks' propensity to retain mortgage loans.[19] For this instrument to violate the exclusion restriction, it would have to be that markets with a higher fraction of seniors received differential credit and income shocks during the 2004-2006 period that predict ex-post foreclosure rates. In the Appendix, Table A.1 reports within-MSA correlations showing that this does not appear to be the case. We find that local mortgage origination and economic activity do not vary systematically with the proportion of local population aged 65 and above. Instead, we find

[18] Information on bank mergers is obtained from the Mergers and Acquisitions Database of Banks and Bank Holding Companies at the Federal Reserve of Chicago. Information on banks' assets and the location of banks' branches comes from the FDIC's Summary of Deposits. There are 110 distinct non survivors commercial banks involved in a merger or acquisition between 2004 and 2006 (43 in 2004, 23 in 2005 and 44 in 2006). These non survivor banks have branches in 9,760 zip codes of 1,950 counties in 50 U.S. states. In our estimation sample of 5,547 zip codes, the median number of mergers per zip code is 2.6.

[19] Counties are the smallest geographical area with available data on the proportion of population aged 65 and above. For this reason, our IV regressions use MSA instead of county fixed effects. We measure the proportion of seniors in a county in 2004 using Census Bureau projections based on the 2000 census.

that the fraction of mortgages retained during the boom years significantly increases in zip codes of counties with more seniors. These correlations suggest that the fraction of seniors is relevant to explain variation in $Top4$ and is also likely to satisfy the exclusion restriction in our context.

Table 4 reports the IV estimates. Columns 2 to 4 provide the first-stage estimates. As expected, the concentration of mortgages on lenders' balance sheets is larger in zip codes that experienced more bank mergers and that have a higher fraction of seniors in the population. The two instruments have significant statistical power when used individually (columns 2 and 3) or jointly (column 4). As reported in the last two rows, both instruments pass the F-test of weak instruments at conventional significance level and can jointly explain up to one fifth of the variation in the $Top4$ index. Columns 5 and 6 report the second stage estimates, with and without control variables. In both columns, the estimates for the coefficient of interest are negative and significant. The estimated coefficients are also large. A one standard deviation change in $Top4$ explains half of the standard deviation of the foreclosure rates (it explained 10% of the standard deviation of the dependent variable in the OLS regression in column 4 of Table 2). While we base most of our inference on the more conservative OLS estimates, the increase in the magnitude of the IV coefficients suggests that if anything omitted variables bias the effect of the $Top4$ index downward.

3.2 Securitization and outstanding mortgage concentration

By construction, our $Top4$ index is correlated with the proportion of securitized mortgages. Securitization reduces the concentration of outstanding mortgages and thus strengthens lenders' incentives to foreclose for reasons consistent with the mechanism proposed in this paper. For instance, the lack of incentives of atomistic lenders to renegotiate defaulting loans could explain why securitized mortgages are handled by third-party servicers and why pooling and servicing agreements include restrictions that inhibit loan renegotiations (Piskorski, Seru and Vig, 2010; Agarwal et al., 2011). However, securitization may also introduce renegotiation frictions because it generates dispersed ownership of mortgage claims.[20]

Given the role played by securitization, our main specifications in Tables 2 and 4 control for the fraction of mortgages securitized in each zip code. Table 5 presents the results for other specifications that help disentangle the effect of the geographic concentration of mortgages held on lenders' balance sheets from the one of securitization. In column 1, we control separately for the share of mortgages securitized between 2004 and 2006 with GSEs and non-GSEs. This breakdown is important as the two categories of securitized mortgages differ in many respects. For example, GSE mortgages are usually originated with stricter underwriting standards (Agarwal, Chang Yavas, 2012) and carry no default risk for investors of mortgage backed securities. In addition, servicers' duties and obligations in private-label-securitized mortgages differ from agency-securitized mortgages, as GSEs use reputational and financial incentives to improve servicers' performance (Levitin and Twomey, 2011). As shown, the breakdown of GSE and non-GSE securitized loans does not affect the negative correlation between $Top4$ and foreclosure rates.

In column 2, we control also for the fraction of mortgages originated to subprime borrowers, i.e. mortgages with a spread of 3 percentage points above the rate of comparable maturity Treasury securities. This additional control is meant to capture the likely deterioration in the quality of mortgages securitized during the period leading to the 2007-2010 foreclosure crisis. Once again, the effect of $Top4$ on foreclosure rates is invariant.

To provide additional evidence that the relationship between $Top4$ and foreclosure rates is independent from the role of securitization, we use the decomposition of the $Top4$ index described in Section 2.1, which separates the fraction of loans retained by the four biggest holders in a zip code ($Top4\_ret$) from the fraction of loans securitized in the same zip code. Clearly, any effects of the $Top4\_ret$ index on foreclosure rates cannot arise from securitization or any alternative mechanisms through which securitization may affect foreclosure decisions.

The estimates in column 3 of Table 5 show that the negative effect of $Top4$ on foreclo-

[20] Adelino, Gerardi and Willen (2013) provide evidence that securitization is unlikely to be the main reason why lenders are reluctant to renegotiate delinquent mortgages.

sures is uniquely driven by variation in $Top4\_ret$. The effect is not only statistically, but also economically significant: a one-standard-deviation increase in $Top4\_ret$ decreases the number of foreclosures by over 33 percent in relative terms.

3.3 Loan modifications

So far we have shown that foreclosure rates are negatively associated with the concentration of outstanding mortgages in a neighborhood. To strengthen the interpretation of our results, we obtain data on loan modifications. We investigate whether the lower incidence of foreclosures in zip codes with higher concentration of mortgage holdings is attained through lenders' willingness to renegotiate defaulting loans.

To measure mortgage renegotiations, we use LPS data and rely on the algorithm of Adelino, Gerardi and Willen (2013), which identifies a loan modification in LPS whenever the interest rate, maturity or the outstanding balance of a mortgage change, conditional on the mortgage being delinquent. We keep track of the number of modifications of portfolio and securitized loans during the 2007-2010 period, and compute modification rates as the number of loans modified in a given zip code relative to all defaulting loans in the same zip code.[21]

Table 6 presents ordinary least squares and instrumental variable estimates. Consistent with the interpretation of the results presented so far, the $Top4$ index is positively correlated with modification rates only for portfolio loans (columns 1 and 3). The $Top4$ index appears to have a negative effect on the likelihood that securitized loans are renegotiated (columns 2 and 4). This is consistent with the implication of our model that lenders with a low fraction of outstanding mortgages have stronger incentives to foreclose when other lenders renegotiate and real estate prices are consequently higher.

The difference in the estimated coefficients in columns 1 (3) and 2 (4) suggests that the renegotiation of portfolio loans in zip codes with higher outstanding mortgages concentration

[21] This variable may understate lenders' willingness to avoid foreclosures, as lenders may also wait for the borrower to become current again.

is unlikely to depend on omitted borrower or zip code characteristics. If this were the case, modification rates for both securitized and portfolio loans would respond similarly to changes in the $Top4$ index.

3.4 Differences across zip codes

This subsection provides additional cross-sectional evidence consistent with a casual mechanism running from the concentration of outstanding mortgages on lenders' balance sheets to foreclosure rates.

To start with, a causal interpretation of our findings implies that the $Top4$ index is associated with lower foreclosure rates in zip codes with higher delinquency rates. A stronger negative correlation in areas with *lower* delinquency rates would otherwise suggest that our findings are driven by ex ante differences in borrowers' quality. In column 1 of Table 7, we interact $Top4$ with a dummy variable that takes a value equal to one if the zip code has a delinquency rate in the upper tercile of the distribution in our sample, and zero otherwise. Consistent with a causal effect, we find that the negative effect of $Top4$ on foreclosure rates is stronger in the subsample of zip codes with more mortgage delinquencies.

In column 2, we distinguish between zip codes with different house price dynamics. As shown in Mian and Sufi (2011), areas with inelastic housing supply experienced the largest house price boom between 2004 and 2006 and suffered the largest decline in prices when house prices reversed in 2007. To ensure that the effect of the $Top4$ index on foreclosure rates is not due to unobservable dynamics of local house prices, we test whether the relationship between $Top4$ and foreclosure rates is significantly different in zip codes of MSAs with housing supply elasticity in the top tercile of its distribution. We measure housing supply elasticity with the Saiz (2010) index, which quantifies restrictions to the supply of new housing due to geographical constraints. The estimates in column 2 show that our main findings are unrelated to factors that explain booms and busts in house prices.

Since our theory implies that only lenders that have retained a sufficiently large proportion of mortgages in a neighborhood have incentives to renegotiate, we explore nonlinearities in the relationship between the concentration of outstanding mortgages and foreclosure rates. We expect the negative effect of $Top4$ on foreclosure rates to be stronger in zip codes with higher concentration. The estimates in column 3 confirm that the negative effect of the $Top4$ index on foreclosure rates is stronger in areas with higher outstanding mortgage concentration (top tercile of the $Top4$ index distribution) than in lower outstanding mortgage concentration areas (bottom tercile).

In column 4 we compare zip codes in states with different foreclosure laws. In the United States, some states require that a foreclosed sale takes place through the court (judicial foreclosure states), while other states give lenders the automatic right to sell the property of the defaulting borrower (power-of-sale states). As discussed in Pence (2006), judicial procedures impose higher costs and lengthier foreclosure timelines on lenders. Accordingly, any lenders' incentives to foreclose should be weaker in judicial foreclosure states regardless of their share of outstanding mortgages. We thus expect a stronger effect of $Top4$ on foreclosure rates in the subsample of zip codes located in power-of-sale states.[22] Consistent with our conjectures, in column 4, the $Top4$ index has a muted effect on foreclosure rates in the subsamples of zip codes where foreclosure procedures are more costly.[23] This is an important result as lenders that keep mortgages on their balance sheets should have stronger incentives to originate mortgages of better quality in judicial foreclosure states than in power of sale states. The reason is that lenders' payoffs in case of borrowers' defaults is likely to be lower in judicial states. Thus, any endogeneity problem related to the ex ante quality of mortgages should bias the results against our findings.

In column 5, we expand the set of controls to include a measure of mortgage fraud. Mian

[22] We obtain the list of states where lenders must receive a judge's approval to foreclose (judicial foreclosure states) from Rao and Walsh (2009).

[23] Results are qualitatively similar if we use measure the overall cost of a foreclosure procedure using the estimated number of days required to accomplish a foreclosure. Results are also similar if we consider the sample of zip codes in counties that abut judicial and non-judicial states, which should share similar unobservable economic conditions given the high degree of social and economic integration among adjacent counties.

and Sufi (2015) show that mortgage fraud, defined as overstatement of income on mortgage applications, was highest in zip codes that experienced the strongest mortgage credit growth during the housing boom and the largest incidence of defaults in the subsequent bust. Fraudulent income overstatement could therefore pick up other zip code characteristics that are not already accounted for by our set of controls and that may be correlated with the $Top4$ index. Following Mian and Sufi (2015), we measure mortgage fraud between 2004 and 2006 as the growth in zip code income reported on HMDA home-purchase mortgage applications minus the average zip code IRS-reported income growth. Controlling for mortgage fraud affects neither the sign nor the size of the $Top4$ coefficient.[24]

Finally, we check that the effect of $Top4$ on foreclosure rates is not driven by heterogeneity in the size of zip codes. We report results based on weighted least squares, with weight given by either the population in a zip code (column 6) or the fraction of mortgages originated between 2004 and 2006 (column 7). Heterogeneity in population size or mortgage origination may affect the correlation between $Top4$ and foreclosure rates for several reasons. For instance, zip codes with higher population or more originations may experience more foreclosures. Alternatively, banks may set up renegotiation facilities to deal with more defaulting customers in more populated zip codes. We do not find any such effects: the weighted least square estimates in column 6 and 7 are not any different from the benchmark OLS regression results reported in Table 2.

3.5 Lender characteristics

Another important concern is that our findings may depend on lender characteristics that correlate with the $Top4$ index. For example, lenders' ability to collect soft information on borrowers' quality may affect their incentives to hold mortgages on their balance sheets. As a result, the negative correlation between $Top4$ and foreclosure rates may not be arise because

[24]Mortgage fraud has a negative and insignficant coefficient in the regression. Interestingly, mortgage fraud explains significantly foreclosure rates if we exclude the 90+ delinquency rate from our control set. This result confirms that that the delinquency rate summarizes well a large array of ex-ante mortgage and market characteristics affecting foreclosures.

of lenders' incentives to mitigate foreclosure externalities ex post, but rather it may reflect lenders' ability to select borrowers of better quality ex ante. We think this is unlikely as both large and small (community) banks tend to be among the top four holders of mortgages in a typical zip code. Nevertheless, to address this concern rigorously we use loan level data and run regressions with lenders fixed effects. Such regressions allow us to check whether a given lender's propensity to foreclose is lower in zip codes where the lender holds a higher share of outstanding mortgages on its balance sheet, holding lenders' characteristics constant.

For this purpose, we merge HMDA information on mortgage originations with LPS information on mortgage performance.[25] Specifically, we focus on loans originated between 2004 and 2006, and keep track of their performance between 2007 and 2010. We classify a loan as delinquent if LPS reports a delinquent status for the loan at least once between 2007 and 2010, and as foreclosed, if LPS records that a lender has started a foreclosure procedure on the loan at least once during the same period.

We estimate the following linear probability model:

$$\Pr\left(For \mid Delinquency\right)_{i,l,z,07-10} = \alpha Ret_{l,z,04-06} + \beta X_{i,z,04-06} + \delta_l + \epsilon_{i,l,z,07-10},$$

where $\Pr\left(For \mid Delinquency\right)_{i,l,z,07-10}$ denotes the probability that loan i originated and retained by lender l in zip code z is foreclosed during the period 2007–2010 conditional on being 90 or more days delinquent. The main variable of interest is:

$$Ret_{l,z,04-06} \equiv \frac{MR_{l,z}}{TotalOriginations_z},$$

which measures lender l's share of loans retained in zip code z during the 2004-2006 period.[26]

Our hypothesis is that $\alpha < 0$. That is, conditional on default, the probability that a

[25] The proprietary nature of LPS requires that lender identifiers are replaced with randomly generated identifiers. This implies that the HMDA-LPS dataset cannot be merged with other lenders information (for instance, from the Call Reports).

[26] We continue to consider only mortgages originated in the first three quarters because HMDA records whether the loan has been securitized at the end of the year. Since the process of securitization takes on average two months, loans issued in the last quarter could be inappropriately classified as retained.

mortgage is foreclosed is negatively related to the lender's share of outstanding mortgages in zip code z. The regression model is estimated holding constant a vector, $X_{i,z,04-06}$, of loan level controls at origination, such as the borrower's credit score, the loan to value ratio, the debt to income ratio, the loan subprime status, and the borrower's ethnicity. Importantly, the regression model includes lender fixed effects, δ_l , to ensure that unobserved lenders' characteristics do not drive the relationship between Ret and the probability that a delinquent portfolio loan, originated in the same zip code, is foreclosed.

The results are reported in Table 8. The estimates in column 1 confirm that conditional on delinquency, the probability that loan i in zip code z is foreclosed is negatively correlated with Ret. In column 2, we look for non-linear effects. We find that the negative effect of Ret on the probability of foreclosure is stronger for lenders in the top tercile of the Ret distribution. These lenders are 6 percentage points less likely to foreclose a given loan than other lenders. This result reinforces the interpretation that lenders' decisions to foreclose depend on the share of mortgages they retained in the local housing market.

In column 3, we include several zip code level controls to ensure that our results are not due to observable differences across markets. We find that the inclusion of these controls changes only marginally the magnitude of the estimated effect of $Ret_{l,z,04-06}$ on the foreclosure probability. Finally, in column 4, we consider the probability of foreclosure for securitized mortgages. As expected, we find that the proportion of loans that the lender originating the securitized mortgage has retained on its balance sheet is not related to the probability of foreclosing securitized mortgages.

4 House prices and outstanding mortgage concentration

In this section, we study the implications of lenders' incentives to foreclose on house prices. If an increase in $Top4$ is associated with lower foreclosure rates and foreclosures adversely affect local house prices, house prices changes should be positively correlated with the concentration of outstanding mortgages in a neighborhood. Table 9 tests this prediction on zip code level data, with the same empirical framework used in the analysis of foreclosures in Section 3.

The positive and significant coefficient of $Top4$ confirms that during the 2007-2010 period house prices declined less in zip codes with a higher $Top4$ index. In column 2, where we include the full set of zip code controls, we estimate that a two-standard-deviation increase in the $Top4$ index is associated with 6 percent lower house price depreciation in the average zip code. This effect is virtually unchanged in column 3, when we use the propensity score methodology that, as in Table 4, ensures we compare zip codes that are similar along observable dimensions. In addition, the economic effect remains the same when we instrument the $Top4$ index with banks' mergers and the proportion of population with age above 65 (column 4).

The last three columns report the results of three additional cross-sectional exercises. Column 5 checks that the relationship between $Top4$ and house prices is independent from the role of securitization. As in Table 5, we decompose the $Top4$ index into the fraction of loans retained by the four biggest holders, $Top4\_ret$, and the fraction of loans securitized. We continue to find that a higher concentration of outstanding mortgages is associated with a lower decrease in house prices during the 2007-2010 period.

In column 6, we check that the positive effect of $Top4$ on house prices is stronger in zip codes with higher concentration. The estimates confirm that the effect of the $Top4$ index on house prices is indeed non-linear: it is strongest in zip codes that are in the top tercile of the $Top4$ distribution. Finally, in column 7, we exploit the variation in $Top4$ and house prices across states with different foreclosure procedures. As argued in Section 3.4, any lender should have weaker incentives to foreclose if the foreclosure process requires a judicial intervention. Our estimates provide further support for this prediction, as the relationship between $Top4$ and the change in house prices is weaker in jurisdictions with costly foreclosure procedures.

5 Conclusion

We show that in markets with low outstanding mortgage concentration, lenders exhibit an excessive propensity to foreclose because they do not internalize the effects of foreclosures on house prices. We provide evidence supporting this mechanism using cross-sectional differences in foreclosures, renegotiations of delinquent mortgages, house prices and the concentration of outstanding mortgages across zip codes during the recent U.S. housing market crisis. We find that markets with high concentration of outstanding mortgages experience fewer foreclosures and smaller house price declines.

These findings have important policy implications. When negative shocks limit borrowers' ability to repay, measures favoring the consolidation of impaired mortgage lenders with similar geographical exposure may increase the concentration of outstanding mortgages. Our findings suggest that these measures may reduce lenders' aggregate losses because they tend to strengthen their incentives to renegotiate defaulting loans. Similar effects may be achieved with the creation of bad banks that collect troubled loans at times of crises.

The mechanism highlighted in this paper has bearings beyond the context of the housing market. It has implications for the price volatility of any collateralized market with dispersed lending structure. Exploring other areas in which lenders with a high share of the outstanding claims internalize the externalities created by liquidation decisions is an exciting avenue for future research.

References

[1] Adelino, M., K. Gerardi, and P. S. Willen. 2013. Why Don't Lenders Renegotiate More Home Mortgages? The Effect of Securitization, *Journal of Monetary Economics* 60, 835-853.

[2] Agarwal, S., G. Amromin, I. Ben-David, S. Chomsisengphet, and D. Evanoff, 2011. The Role of Securitization in Mortgage Renegotiation, *Journal of Financial Economics* 102,

559-578.

[3] Agarwal, S., Y. Chang, and A. Yavas. 2012. Adverse Selection in Mortgage Securitization, *Journal of Financial Economics* 105, 640–660

[4] Agarwal, S., G., Amromin, I. Ben-David, S. Chomsisengphet, T. Piskorski and A. Seru, 2013. Policy Intervention in Debt Renegotiation: Evidence from Home Affordability Modification Program. Federal Reserve Bank of Chicago, W.P 2013-27.

[5] Anenberg, E., and E. Kung, 2014. Estimates of the Size and Source of Price Declines due to Nearby Foreclosures, *American Economic Review* 104, 2527-51.

[6] Asquith, P., R. Gertner, and D. S., Scharfstein. 1994. Anatomy of Financial Distress: An Examination of Junk Bond Issuers, *Quarterly Journal of Economics*, 109(3): 625–58.

[7] Avery, R. B., K. P. Brevoort, and G. B. Canner, 2007. Opportunities and Issues in using HMDA Data, *Journal of Real Estate Research*, 29, 351-380.

[8] Becker, B., 2007. Geographical Segmentation of US Capital Markets. *Journal of Financial Economics* 85, 151–78.

[9] Benmelech, E. and N. Bergman, 2011. Bankruptcy and the Collateral Channel, *Journal of Finance* 66(2): 337-378.

[10] Berger, A. N., N. H. Miller, M.A. Petersen, R. G. Rajan, and J.C. Stein, 2005. Does Function Follow Organizational Form? Evidence from the Lending Practices of Large and Small Banks, *Journal of Financial Economics* 76, 237–269.

[11] Brunnermeier, M. K., and L. H. Pedersen, 2009. Market Liquidity and Funding Liquidity, *Review of Financial Studies*, 22, 2201–38.

[12] Campbell, J. Y., S. Giglio, and P. Pathak, 2011. Forced Sales and House Prices, *American Economic Review* 101, 2108-2131.

[13] Elul, R., N. Souleles, S. Chomsisengphet, D. Glennon, and R. Hunt, 2010. What "Triggers " Mortgage Default, *American Economic Review* 100, 490-494.

[14] Favara, G. and J. Imbs, 2015. Credit Supply and the Price of Housing, *American Economic Review* 105, 958-992.

[15] Fisher, L. M, L. Lambie-Hanson, and P. S. Willen, 2014. The Role of Proximity in Foreclosure Externalities. Evidence from Condominiums, NBER Working Paper 20080.

[16] Garmaise, M., and T. Moskowitz, 2006. Bank mergers and crime: The real and social effects of credit market competition, *Journal of Finance* 61, 495–538.

[17] Gilje, E., E. Loutskina, and P. E. Strahan, 2013. Exporting Liquidity: Branch Banking and Financial Integration, NBER WP. 19403

[18] Ghent, A. C., 2011. Securitization and Mortgage Renegotiation: Evidence from the Great Depression, *Review of Financial Studies* 24, 1814–1847.

[19] Gromb D. and D. Vayanos, 2002. Equilibrium and Welfare in Markets with Financially Constrained Arbitrageurs. *Journal of Financial Economics* 66,: 361–407.

[20] Guiso, L., P. Sapienza, and L. Zingales, 2013. The Determinants of Attitudes towards Strategic Default on Mortgages, *Journal of Finance* 68, 1473–1515.

[21] Guren, A. and T. McQuade, 2013. How Do Foreclosures Exacerbate Housing Downturns?, Working Paper, Harvard University.

[22] Harding, J. P., E.Rosenblatt, and V. W. Yao, 2008. The Contagion Effect of Foreclosed Properties, *Journal of Urban Economics* 66, 164-178.

[23] Hartley, D., 2014. The Effect of Foreclosures on Nearby Housing Prices: Supply or Dis-Amenity?", *Regional Science and Urban Economics* 49, 108–117.

[24] Jiang, W., A. A. Nelson, and E. Vytlacil, 2014. Securitization and Loan Performance: Ex Ante and Ex Post Relations in the Mortgage Market, *Review of Financial Studies* 27, 454–483.

[25] Keeley, Michael, 1990. Deposit Insurance, Risk and Market Power in Banking, *American Economic Review* 80, 1183-1200.

[26] Keys, B., T. Mukherjee, A. Seru, and V. Vig, 2010. Did Securitization Lead to Lax Screening? Evidence from Subprime Loans 2001-2006, *Quarterly Journal of Economics* 125, 307-62.

[27] Kiyotaki, N. and J. Moore, 1997. Credit Cycles, *Journal of Political Economy* 105, 211–248.

[28] Levitin, A., and T. Twomey, 2011. Mortgage Servicing, *Yale Journal on Regulation* 28, 2-90.

[29] Li, X. and X. S. Zhao, 2015. Strategic Defaults Induced by Loan Modifications, Working Paper, Office of the Currency Comptroller.

[30] Loutskina, E., and P. E. Strahan, 2009. Securitization and the Declining Impact of Bank Finance on Loan Supply: Evidence from Mortgage Acceptance Rates, *Journal of Finance* 64, 861-889.

[31] Loutskina, E., and P. E. Strahan, 2011. Informed and Uninformed Investment in Housing: The Downside of Diversification, *Review of Financial Studies* 24, 1447–1480.

[32] Mayer, C., K. Pence, and S. M. Sherlund, 2009. The Rise in Mortgage Defaults, *Journal of Economic Perspectives* 23, 27-50.

[33] Mayer, C., E. Morrison, T. Piskorski, and A. Gupta. 2014. Mortgage Modification and Strategic Behavior: Evidence from a Legal Settlement with Countrywide, *American Economic Review* 104, 2830-57.

[34] Mian, A., and A. Sufi, 2009. The Consequences of Mortgage Credit Expansion: Evidence from the US Mortgage Default Crisis, *Quarterly Journal of Economics* 124, 1449–1496.

[35] Mian, A. R. and A. Sufi, 2011. House Prices, Home Equity-Based Borrowing, and the U.S. Household Leverage Crisis, *American Economic Review* 101, 2132–2156

[36] Mian, A., A. Sufi, and F. Trebbi, 2014. Foreclosures, House Prices, and the Real Economy, *Journal of Finance*, forthcoming.

[37] Mian, A., A. Sufi, 2015. Fraudulent Income Overstatement on Mortgage Applications During the Credit Expansion of 2002 to 2005. NBER WP 20947.

[38] Pence, Karen M., 2006. Foreclosing on Opportunity: State Laws and Mortgage Credit, *Review of Economics and Statistics* 88, 177–82.

[39] Piskorski, T., A. Seru, and V. Vig, 2010. Securitization and Distressed Loan renegotiation. Evidence from the Subprime Mortgage Crisis, *Journal of Financial Economics* 97, 369–397.

[40] Purnanandam, A., 2011. Originate-to-distribute Model and the Subprime Mortgage Crisis, *Review of Financial Studies* 24, 1881–1915.

[41] Rao, J., and G. Walsh, 2009. Foreclosing a Dream: State Laws Deprive Homeowners of Basic Protections, Technical Report. Boston, MA: National Consumer Law Center.

[42] Rosen, R. J., 2010. The Impact of the Originate-to-Distribute Model on Banks Before and During the Financial Crisis, FRB of Chicago Working Paper No. 2010-20

[43] Saiz, A., 2010. The Geographic Determinants of Housing Supply, *Quarterly Journal of Economics* 125, 1253–1296.

[44] Scharfstein, D. S., and A. Sunderam, 2015. Market Power in Mortgage Lending and the Transmission of Monetary Polity. mimeo HBS.

[45] Shleifer, A. and R. Vishny, 1992. Liquidation Values and Debt Capacity: a Market Equilibrium Approach, *Journal of Finance* 47, 1343–1366.

[46] Shleifer, A., and R. Vishny, 2011. Fire Sales in Finance and Macroeconomics, *Journal of Economic Perspectives* 25, 29-48.

Figure 1. Nationwide coverage of U.S. urban counties in the estimation sample

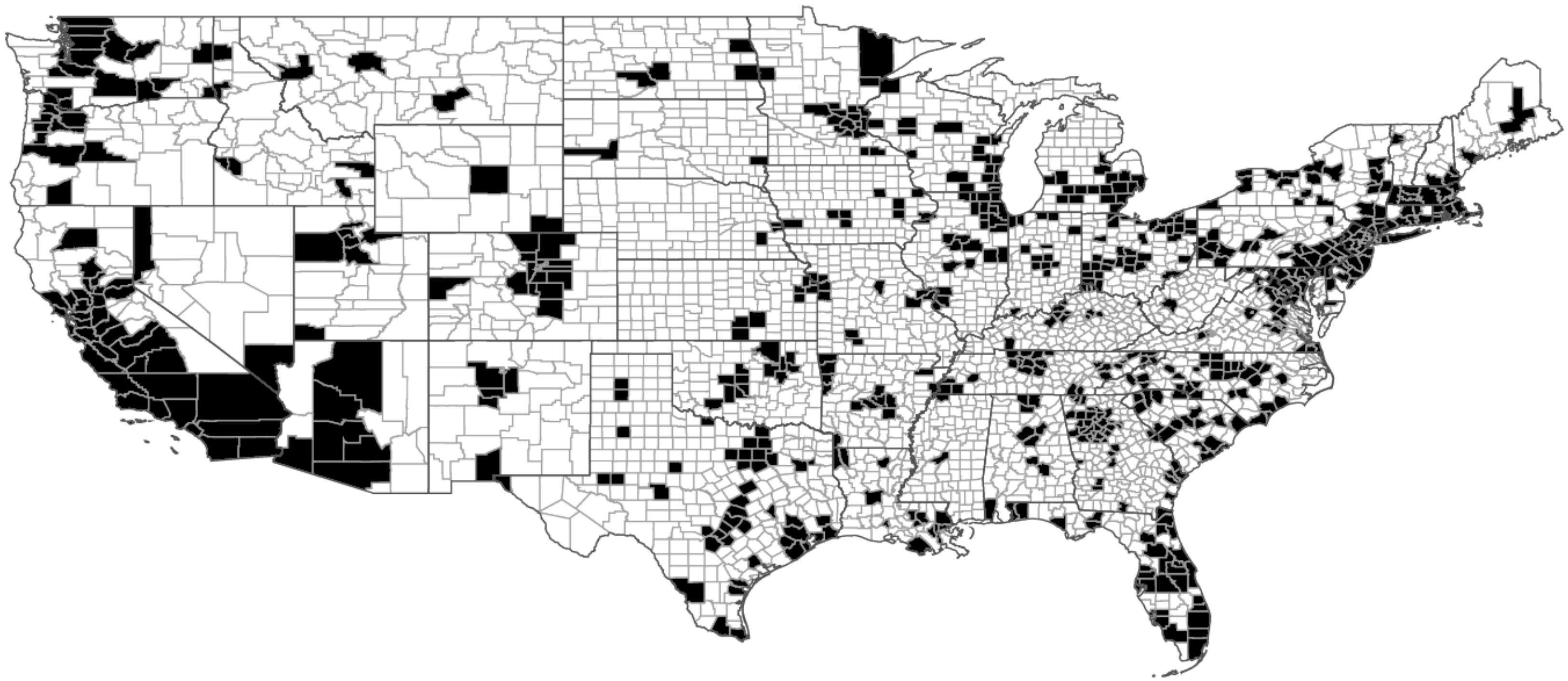

Figure 2. 2004-2006 Nationwide Distribution of Top 4 Index

distribution computed for 6,570 zip codes

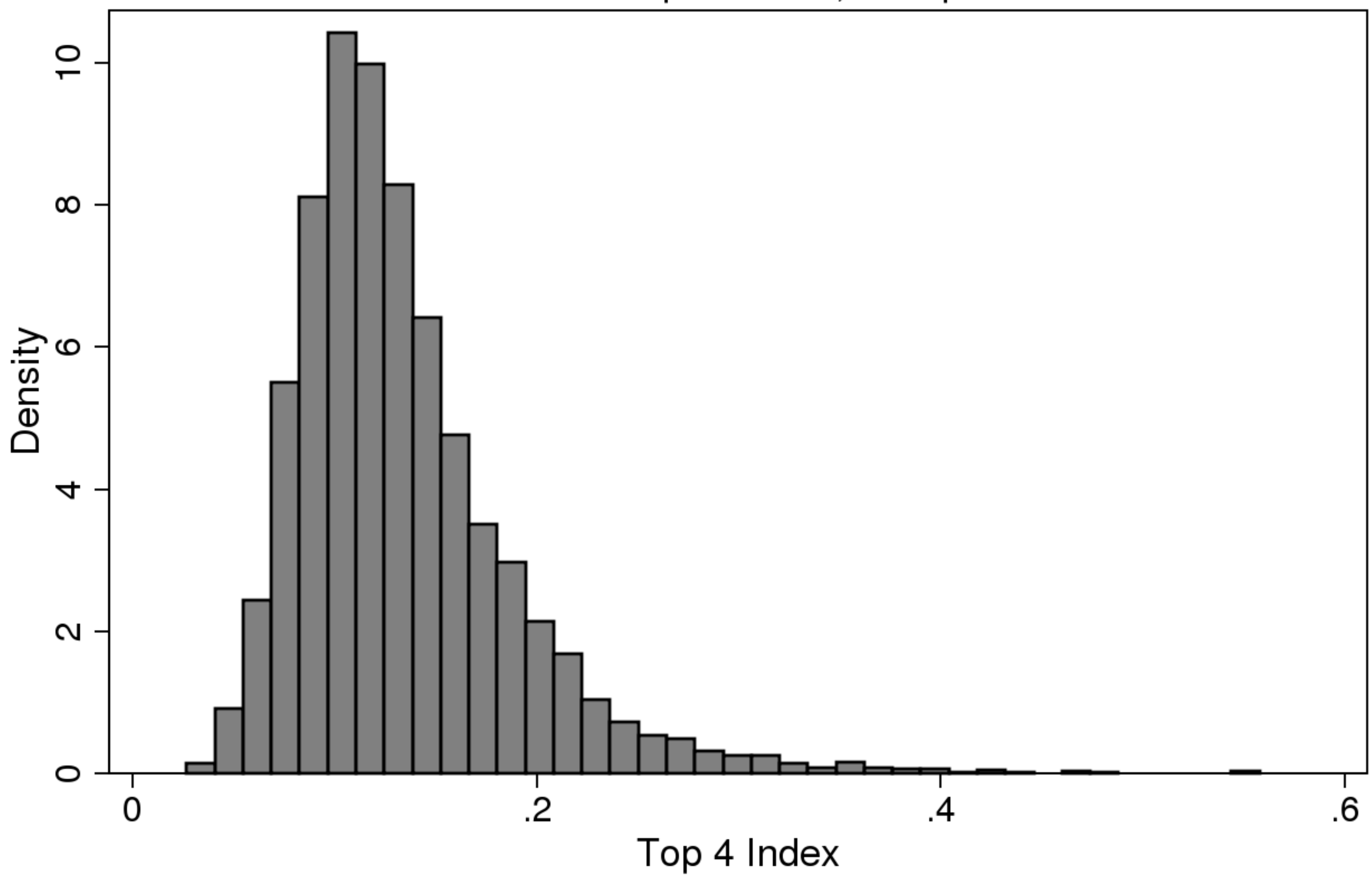

Figure 3. 2004-2006 Distribution of Top 4 Index for Selected U.S. counties

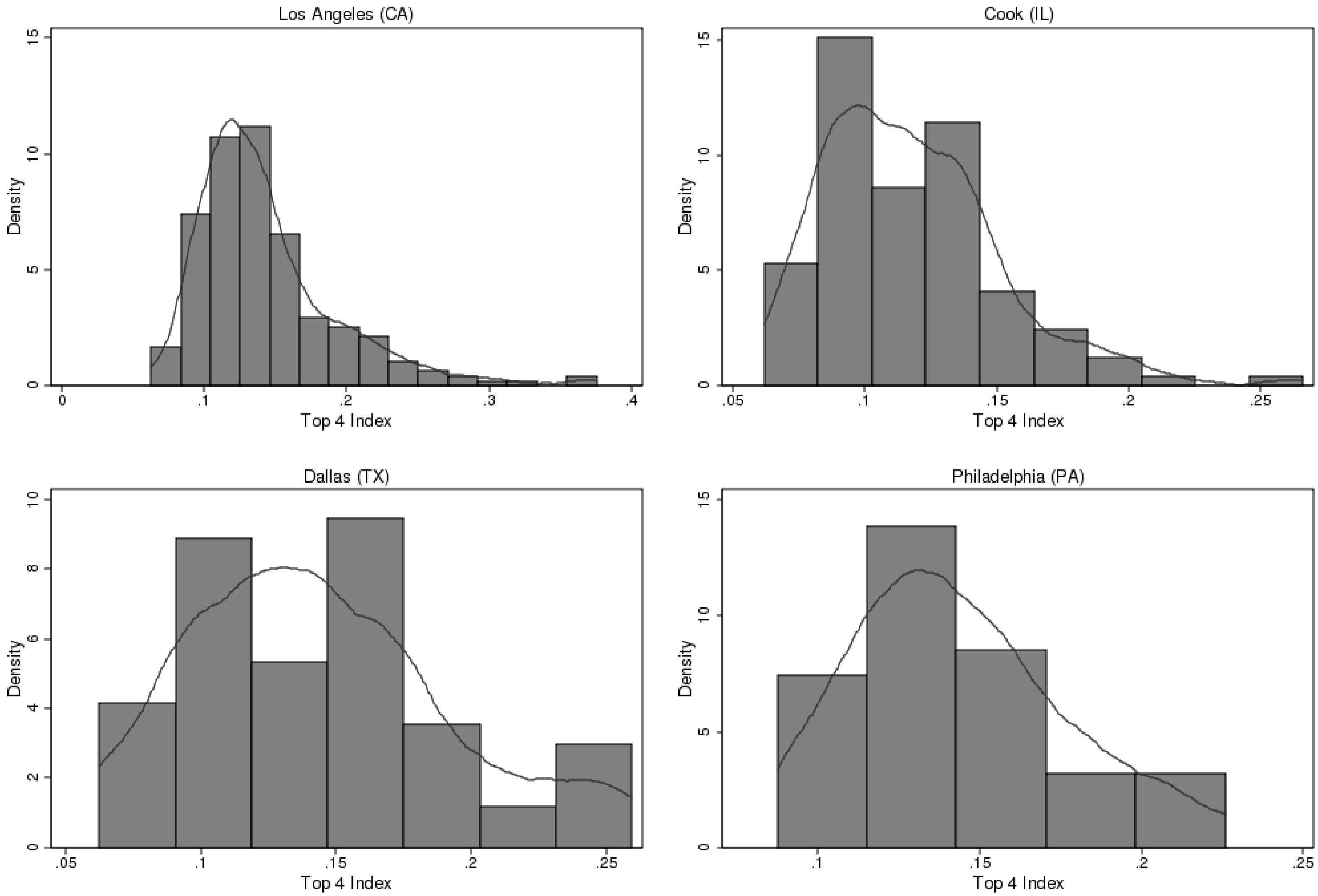

Figure 4. 2004-2006 Nationwide Distribution of Top 4 retained Index
distribution computed for 6,570 zip codes

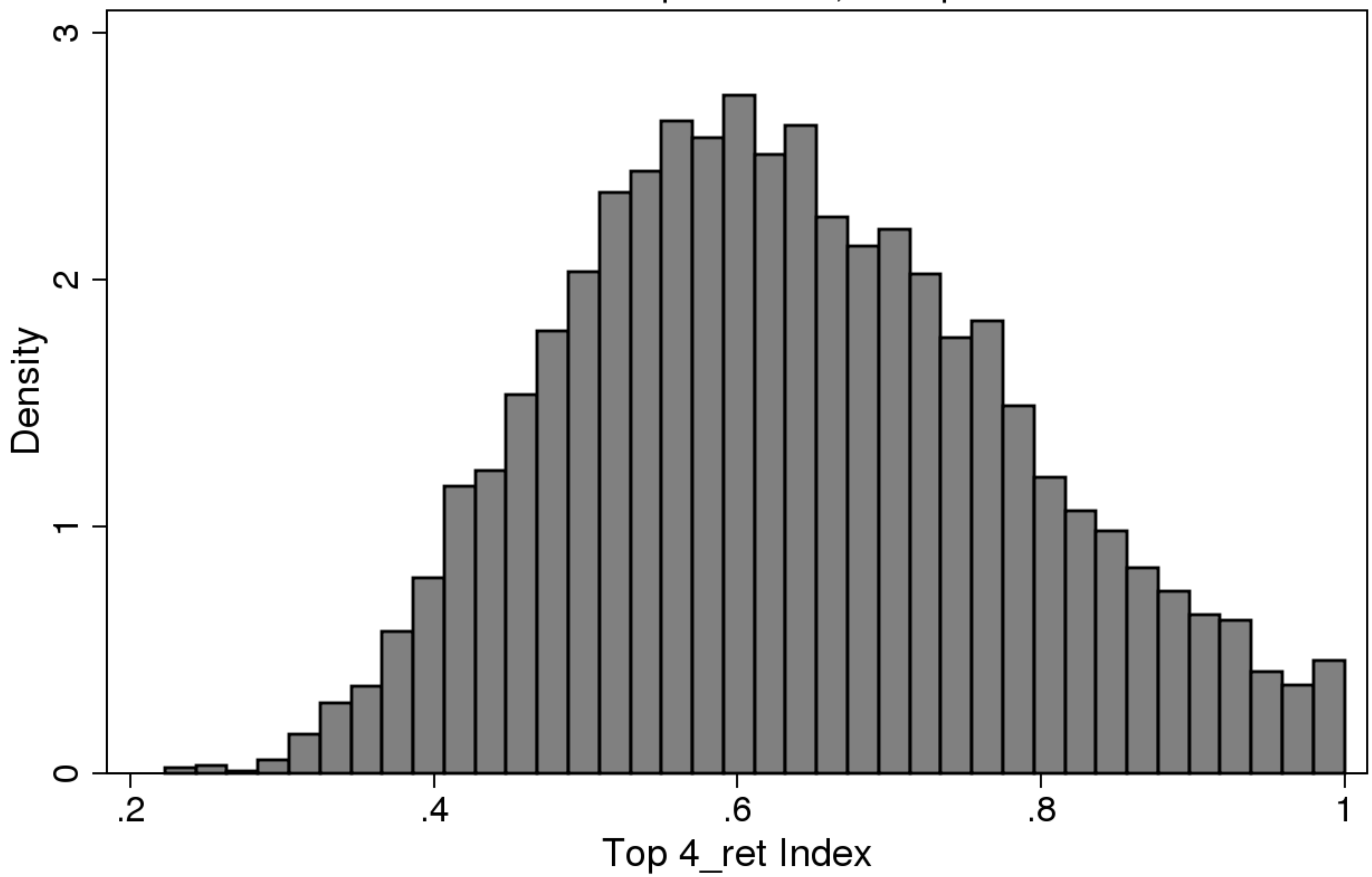

Table 1
Variable Definitions and Summary Statistics

This table provides definitions and descriptive statistics for the main variables used in the empirical analysis.

Variable Name	Variable Description	Source	Mean	p10	p90	sd	Obs.
90+ days delinquency ratio	Zip code number of outstanding mortgages that are 90 days or more delinquent divided by number of outstanding mortgages. Average from 2007 to 2010.	Equifax	0.0258	0.0078	0.0543	0.0197	6534
Census income, 2000	Logarithmic median income in the zip code in 2000.	U.S. Census Bureau	10.9509	10.5412	11.3671	0.3279	6570
Debt per capita	The value of first mortgages, home equity lines, auto loans, credit card outstanding in the zip code divided by the number of households (divided by 100,000). Average between 2004 and 2006.	Equifax	0.8134	0.3700	1.3855	0.4469	6506
Foreclosure rate	The zip code's number of foreclosures, averaged from 2007 to 2010, divided by the number of single family, owner occupied housing units in the zip code in 2000. Foreclosures are defined as the sum of notices of trustee sale and notices of sales.	RealtyTrac.com and U.S Census Bureau	0.0198	0.0014	0.0489	0.0273	6159
Fraction of Subprime	Fraction of households in a zip code with FICO score below 620. Average between 2004 and 2006.	Equifax	0.1103	0.0310	0.2159	0.0810	6489
Fraud	The zip code growth in income reported on home-purchase mortgage applications minus the average IRS-reported income growth from 2004 to 2006	HMDA and IRS SOI Tax Stats	0.0132	-0.0898	0.1257	0.1066	6454
GSE Securitization	The zip code's fraction of loans originated for purchase of single family owner occupied houses sold within the year of origination to government-sponsored housing enterprises. Average between 2004 and 2006.	HMDA	0.3555	0.2626	0.4504	0.0765	6570
High cost mortgages	Fraction of mortgages originated with mortgage rates 3 percentage points above the rate of a comparable maturity Treasury security. Average between 2004 and 2006.	HMDA	0.2163	0.0687	0.4125	0.1360	6570

Variable Name	Variable Description	Source	Mean	p10	p90	sd	Obs.
House price change, 2007-2010	Logarithmic change between 2007 and 2010 of the zip code house price index for single-family owner-occupied houses.	CoreLogic	-0.2411	-0.5320	-0.0341	0.1878	6420
House price change, 2004-2006	Logarithmic change between 2004 and 2006 of the zip code house price index for single-family owner-occupied houses.	CoreLogic	0.1795	0.0266	0.3568	0.1266	5547
Judicial Foreclosure	Dummy variable that takes value equal to one for zip codes in states with a judicial requirement for foreclosures.	Rao and Walch (2009)	0.5339	0	1	0.4989	6570
Loans originated (fraction of housing stock)	Average of the number of loans originated between 2004 and 2006, divided by the number of single family, owner occupied housing units in the zip code in 2000.	HMDA and U.S. Census Bureau	0.0645	0.0049	0.0471	2.3209	6185
Non-GSE securitization	The zip code's fraction of loans originated for purchase of single family, owner occupied housing units sold to non-government-sponsored housing enterprises within the year of origination. Average between 2004 and 2006.	HMDA	0.3920	0.2892	0.4951	0.0822	6570
Population	The logarithm of the zip code's population in 2000.	U.S. Census Bureau	8.4719	8.0953	8.8649	0.3325	6570
Renegotiations-Portfolio Loans	Number of modifications for mortgages held on lenders' balance sheets that are 60 days delinquent, divided by the total number of defaulting loans. Zip code average between 2007 and 2010. Modifications refer to a change in mortgages' interest rates, principal balances and loan terms, using the algorithm developed by Adelino, Gerardi and Willen (2013).	Lender Processing Services (LPS)	0.0242	0.0000	0.0546	0.0270	6548
Renegotiations-Securitized Loans	Defined as Renegotiations-Portfolio Loans, but considering renegotiations of securitized loans at the numerator.	Lender Processing Services (LPS)	0.1412	0.0611	0.2314	0.0814	6548

Variable Name	Variable Description	Source	Mean	p10	p90	sd	Obs.
Securitization	The zip code’s fraction of loans originated for purchase of single family, owner occupied housing units sold within the year of origination to other non-affiliated financial institutions or government-sponsored housing enterprises. Average between 2004 and 2006.	HMDA	0.7476	0.6493	0.8260	0.0767	6570
Top 4	Ratio of the number of mortgages retained by the top 4 holders in a zip code, divided by the total number of mortgages originated in that zip code. Loans originated and retained are measured from 2004 to 2006. Loans are conventional mortgages for purchase of single-family owner-occupied houses. Lenders include commercial banks, thrifts, credit unions and mortgage companies.	HMDA	0.1331	0.0790	0.1997	0.0535	6570
Top 4_ret	Defined as Top 4, but considering only the number of loans originated and retained in a zip code to compute the denominator.	HMDA	0.6356	0.4502	0.8391	0.1463	6570
Zip Code % Black or Hispanic in 2000	The percentage of Black and Hispanic in the zip code in 2000.	U.S. Census Bureau	0.2231	0.0280	0.5840	0.2276	6185

Table 2
Foreclosures

Cross-sectional zip code level regressions of the foreclosure rate between 2007 and 2010 on the index of outstanding mortgage concentration, Top4, between 2004 and 2006. All variable definitions and sources are reported in Table 1. Regressions include MSA fixed effects in columns 1 and 3 and county fixed effects in columns 2 and 4. Standard errors in parenthesis are clustered at the county level and corrected for heteroskedasticity. Estimates followed by ***, **, and * are statistically different from zero with 0.01, 0.05 and 0.10 significance levels, respectively.

	(1)	(2)	(3)	(4)
Top4, 2004-2006	-0.084***	-0.099***	-0.032**	-0.047***
	(0.018)	(0.016)	(0.013)	(0.011)
Securitization, 2004-2006	0.038***	0.047***	0.033***	0.042***
	(0.011)	(0.008)	(0.010)	(0.007)
90+ days delinquency ratio, 2007-2010			0.761***	0.765***
			(0.140)	(0.096)
Loans originated (fraction of housing stock), 2004-2006	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Fraction of subprime, 2004-2006	0.121***	0.115***	0.023	0.019
	(0.026)	(0.019)	(0.027)	(0.019)
Debt per capita, 2004-2006	0.029***	0.032***	0.026***	0.030***
	(0.005)	(0.005)	(0.005)	(0.004)
House price change, 2004-2006	-0.020*	-0.013	-0.025***	-0.023***
	(0.010)	(0.010)	(0.009)	(0.009)
Census income, 2000	-0.026***	-0.031***	-0.022***	-0.027***
	(0.005)	(0.005)	(0.005)	(0.004)
Zip Code % Black or Hispanic in 2000	0.010	0.010*	-0.008*	-0.010**
	(0.006)	(0.005)	(0.005)	(0.005)
Constant	0.257***	0.301***	0.200***	0.254***
	(0.050)	(0.047)	(0.042)	(0.042)
Fixed Effects	MSA	County	MSA	County
Obs	5212	5212	5212	5212
R2	.648	.691	.695	.732

Table 3
The Concentration of Outstanding Mortgages and Zip Code Characteristics

Cross-sectional zip code level regressions of Top 4 on selected zip code characteristics on. All variable definitions and sources are reported in Table 1. Regressions include county fixed effects. Standard errors in parenthesis are clustered at the county level and corrected for heteroskedasticity. Estimates followed by ***, **, and * are statistically different from zero with 0.01, 0.05 and 0.10 significance levels, respectively.

	(1) Census income, 2000	(2) Zip Code % Black or Hispanic, 2000	(3) Fraction of subprime, 2004-2006	(4) Debt per capita, 2004-2006	(5) Loans originated (fraction of housing stock), 2004-2006	(6) House price change, 2004-2006
Top4, 2004-2006	-0.010	0.013	0.014	-0.000	-0.000	-0.010
	(0.007)	(0.011)	(0.016)	(0.006)	(0.000)	(0.021)
Constant	0.240***	0.131***	0.132***	0.133***	0.134***	0.133***
	(0.078)	(0.003)	(0.002)	(0.005)	(0.000)	(0.004)
Obs	6570	6185	6489	6506	6185	5547
R2	.561	.559	.559	.557	.557	.571

Table 4
Propensity Scores and Instrumental Variable Estimates

Cross-sectional zip code level regressions of the foreclosure rate between 2007 and 2010 on the index of outstanding mortgage concentration, Top4, between 2004 and 2006. All variable definitions and sources are reported in Table 1. In column 1, estimates are obtained using a propensity score methodology and the regression includes county dummies. In columns 2 to 4, we present first stage estimates of Top4 on the instruments, which are the numbers of mergers between large financial institutions and the percentage of population above 65. Columns 5 and 6 present the second stage estimates. In the 2SLS estimates, we include MSA dummies. Estimates followed by ***, **, and * are statistically different from zero with 0.01, 0.05 and 0.10 significance levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
Estimation method	Propensity Scores		OLS		2SLS	
Dependent variable	Foreclosure rate		Top 4		Foreclosure rate	
Top4, 2004-2006	-0.074***				-0.334***	-0.290***
	(0.025)				(0.048)	(0.062)
Securitization, 2004-2006	0.004					-0.067***
	(0.021)					(0.024)
90+ days delinquency ratio, 2007-2010	0.510***				0.547***	0.493***
	(0.151)				(0.028)	(0.073)
Loans originated (fraction of housing stock), 2004-2006	0.001***					0.001***
	(0.000)					(0.000)
Fraction of subprime, 2004-2006	0.018					0.027***
	(0.026)					(0.009)
Debt per capita, 2004-2006	0.020***					0.024***
	(0.004)					(0.002)
House price change, 2004-2006	-0.028**					-0.027***
	(0.012)					(0.005)
Census income, 2000	-0.017***					-0.027***
	(0.005)					(0.003)
Zip Code % Black or Hispanic in 2000	0.004					0.001
	(0.007)					(0.003)
Number of mergers, 2004		0.002***		0.002***		
		(0.000)		(0.000)		
% Population above 65, 2004			0.004***	0.004***		
			(0.001)	(0.001)		
Constant	0.187***	0.127***	0.087***	0.082***		
	(0.054)	(0.001)	(0.011)	(0.011)		
Fixed Effects	County	County	MSA	MSA	MSA	MSA
F Test Excluded Instruments					48.61	38.49
Obs	5109	6570	6570	6570	6127	5168
R2	.748	.563	.442	.447		

Table 5
Securitization and the Concentration of Outstanding Mortgages

Cross-sectional zip code level regressions of the foreclosure rate between 2007 and 2010 on the index of outstanding mortgage concentration, Top4, between 2004 and 2006. All variable definitions and sources are reported in Table 1. All regressions include county dummies. Standard errors in parenthesis are clustered at the county level and corrected for heteroskedasticity. Estimates followed by ***, **, and * are statistically different from zero with 0.01, 0.05 and 0.10 significance levels, respectively.

	(1)	(2)	(3)
Top4, 2004-2006	-0.044***	-0.047***	
	(0.011)	(0.011)	
Top4_ret, 2004-2006			-0.068***
			(0.014)
Top 4 _ret ×Securitization, 2004-2006			0.053***
			(0.016)
Securitization, 2004-2006		0.040***	
		(0.007)	
Non GSE securitization, 2004-2006	0.045***		
	(0.008)		
GSE securitization, 2004-2006	0.037***		
	(0.008)		
90+ days delinquency ratio, 2007-2010	0.726***	0.776***	0.722***
	(0.093)	(0.105)	(0.092)
Loans originated (fraction of housing stock), 2004-2006	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)
Fraction of subprime, 2004-2006	0.022	0.044**	0.017
	(0.018)	(0.022)	(0.018)
Debt per capita, 2004-2006	0.029***	0.029***	0.027***
	(0.004)	(0.004)	(0.004)
House price change, 2004-2006	-0.024***	-0.023***	-0.024***
	(0.008)	(0.008)	(0.009)
Census income, 2000	-0.026***	-0.028***	-0.028***
	(0.004)	(0.004)	(0.004)
Zip Code % Black or Hispanic in 2000	-0.010**	-0.008*	-0.008*
	(0.005)	(0.004)	(0.005)
High Cost Mortgages, 2004-2006		-0.025**	
		(0.012)	
Constant	0.244***	0.261***	0.305***
	(0.040)	(0.045)	(0.042)
Obs	5212	5212	5212
R2	.726	.727	.724

Table 6
Loan Modifications

Cross-sectional zip code level regressions for Renegotiations of Portfolio Loans (columns 1 and 3), and Renegotiations of Securitized Loans (columns 2 and 4) between 2007 and 2010 on the index of outstanding mortgage concentration, Top4, between 2004 and 2006. All variable definitions and sources are reported in Table 1. Regressions in columns 1 and 2 are obtained by ordinary least squares and include county fixed effects. Regressions in columns 3 and 4 are obtained by instrumental variables and include MSA fixed effects; the first stage is presented in column 4 of Table 4. Standard errors in parenthesis are clustered at the county level and corrected for heteroskedasticity. Estimates followed by ***, **, and * are statistically different from zero with 0.01, 0.05 and 0.10 significance levels, respectively.

	(1)	(2)	(3)	(4)
	OLS		2SLS	
	Portfolio Loans	Securitized Loans	Portfolio Loans	Securitized Loans
Top4, 2004-2006	0.081***	-0.086*	0.165***	-0.402***
	(0.019)	(0.048)	(0.047)	(0.124)
Loans originated (fraction of housing stock), 2004-2006	-0.000***	-0.001***	-0.000***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Fraction of subprime, 2004-2006	0.076***	0.075**	0.079***	0.047*
	(0.009)	(0.035)	(0.009)	(0.024)
Debt per capita, 2004-2006	0.010***	-0.013**	0.008***	-0.015***
	(0.002)	(0.007)	(0.002)	(0.005)
House price change, 2004-2006	-0.008	0.014	0.014**	0.024
	(0.008)	(0.024)	(0.007)	(0.018)
Census income, 2000	0.010***	0.061***	0.015***	0.059***
	(0.003)	(0.009)	(0.003)	(0.009)
Zip Code % Black or Hispanic in 2000	-0.008***	-0.025*	-0.007***	-0.029***
	(0.003)	(0.014)	(0.003)	(0.007)
Constant	-0.115***	-0.510***		
	(0.032)	(0.103)		
Fixed Effects	County	County	MSA	MSA
F Test Excluded Instruments			67.95	67.95
Obs	5232	5232	5189	5189
R2	.325	.388		

Table 7
Cross-Sectional Variation in the Effect of the Outstanding Mortgage Concentration

Cross-sectional zip code level regressions of the foreclosure rate between 2007 and 2010 on the index of outstanding mortgage concentration, Top4, between 2004 and 2006. T2 Top 4 and T3 Top 4 are dummy variables that take value equal to 1 for zip codes that are respectively in the second and third terciles of the Top 4 index's distribution. All variable definitions and sources are reported in Table 1. All regressions include county dummies. In column 1 to 5 we present ordinary least squares regressions. In columns 6 and 7, we present weighted least squares, with observations weighted by total population and the loans originated, respectively. Standard errors in parenthesis are clustered at the county level and corrected for heteroskedasticity. Estimates followed by ***, **, and * are statistically different from zero with 0.01, 0.05 and 0.10 significance levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			OLS			WLS	
Top4, 2004-2006	-0.036***	-0.038***		-0.085***	-0.044***	-0.040***	-0.040***
	(0.011)	(0.014)		(0.016)	(0.011)	(0.010)	(0.010)
Top 4, 2004-2006×High delinquency, 2007-2010	-0.017**						
	(0.007)						
Top 4, 2004-2006×High Elasticity		-0.016					
		(0.027)					
T2 Top 4, 2004-2006			-0.002***				
			(0.001)				
T3 Top 4, 2004-2006			-0.003***				
			(0.001)				
Top 4, 2004-2006×Judicial foreclosures				0.066***			
				(0.026)			
Fraud, 2002-2005					-0.012*		
					(0.007)		
Securitization, 2004-2006	0.041***	0.041***	0.049***	0.041***	0.041***	0.040***	0.039***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
90+ days delinquency ratio, 2007-2010	0.745***	0.732***	0.744***	0.721***	0.744***	0.734***	0.748***
	(0.095)	(0.093)	(0.094)	(0.092)	(0.091)	(0.095)	(0.095)
Loans originated (fraction of housing stock), 2004-2006	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	
Fraction of subprime, 2004-2006	0.026	0.022	0.022	0.025	0.025	0.025	0.023
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
Debt per capita, 2004-2006	0.029***	0.029***	0.029***	0.030***	0.029***	0.029***	0.028***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
House price change, 2004-2006	-0.025***	-0.023***	-0.024***	-0.023***	-0.022***	-0.023***	-0.025***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Census income, 2000	-0.027***	-0.026***	-0.026***	-0.026***	-0.027***	-0.025***	-0.025***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Zip Code % Black or Hispanic in 2000	-0.011**	-0.010**	-0.010**	-0.010**	-0.010**	-0.009**	-0.010**
	(0.005)	(0.005)	(0.005)	(0.004)	(0.005)	(0.004)	(0.005)
Constant	0.255***	0.245***	0.229***	0.243***	0.247***	0.232***	0.227***
	(0.042)	(0.040)	(0.040)	(0.040)	(0.040)	(0.038)	(0.039)
Obs	5212	5212	5212	5212	5210	5212	5212
R2	.727	.726	.725	.727	.726	.724	.711

Table 8
Lender's Propensity to Foreclose and Fraction of Outstanding Mortgages

Panel A. Summary Statistics

This table provides definitions and descriptive statistics for the main variables used in the loan level regressions estimated in Panel B.

Variable Name	Variable Description	Source	Mean	p10	p90	sd	Obs.
Black-Hispanic borrower	Dummy variable equal to 1 if the borrower is Black or Hispanic, and zero otherwise.	HMDA	0.403	0	1	0.490	143,647
Borrower debt to income	Borrower's debt to income ratio at origination.	LPS	36.26	21	49	11.316	143,647
Borrower fico score	Borrower's Fico score at origination.	LPS	689.87	620	763	56.024	143,647
Foreclosure portfolio loans	Foreclosure probability of portfolio loans conditional on being 90 plus days delinquent.	LPS	0.612	0	1	0.487	27,011
Foreclosure securitized loans	Foreclosure probability of securitized loans conditional on being 90 plus days delinquent.	LPS	0.738	0	1	0.439	116,636
High cost loan	Dummy variable equal to one if the loan is originated with an interest rates 3 percentage points above the rate of a comparable maturity Treasury security, and zero otherwise.	HMDA	0.234	0	1	0.423	143,647
Interest-only loan	Dummy variable equal to one if the loan is interest only, and zero otherwise.	LPS	0.301	0	1	0.459	143,647
Loan LTV ratio	Borrower's loan to value ratio at origination	LPS	81.75	74.67	96.71	9.161	143,647
$Ret_{04\text{-}06}$ portfolio loans	Number of mortgages retained by lender *l* in zip code *z* as a fraction of the total number of mortgages originated in the same zip code between 2004 and 2006. Sample of portfolio loans only.	HMDA	0.027	0.004	0.052	0.029	27,011
$Ret_{04\text{-}06}$ securitized loans	Number of mortgages retained by lender *l* in zip code *z* as a fraction of the total number of mortgages originated in the same zip code between 2004 and 2006. Sample of securitized loans only.	HMDA	0.007	0	0.022	0.013	116,636

Panel B. Results

Loan level regressions for the probability that a delinquent loan is foreclosed. The dependent variable is a dummy variable that takes a value equal to 1 if a delinquent loan has been foreclosed during 2007-2010 and equal to zero otherwise; only loans that have been delinquent 90 days or more during this period are considered. All variables are defined in Panel A. In columns 1 to 3 we consider only portfolio loans; in column 4 we consider only securitized loans. T2 Ret_{04-06} and T2 Ret_{04-06} are dummy variables that take value equal to 1 if Ret_{04-06} is respectively in the second and third tercile of the distribution of Ret_{04-06} among portfolio loans. Standard errors in parenthesis are clustered at the zip code level and corrected for heteroskedasticity. Estimates followed by ***, **, and * are statistically different from zero with 0.01, 0.05 and 0.10 significance levels, respectively.

	(1)	(2) Portfolio Loans	(3)	(4) Securitized Loans
Ret_{04-06}	-0.6066***		-0.361***	0.0418
	(0.1947)		(0.130)	(0.1654)
T2 Ret_{04-06}		-0.0177		
		(0.0196)		
T3 Ret_{04-06}		-0.0580***		
		(0.0206)		
Loan level controls				
Borrower fico score	-0.0001***	-0.0001**	-0.0001*	-0.0001***
	(0.0000)	(0.0001)	(0.0000)	(0.0000)
Loan LTV ratio	-0.0005	-0.0006*	-0.0008**	0.0011***
	(0.0003)	(0.0004)	(0.0003)	(0.0002)
Borrower debt to income	-0.0004	-0.0004	-0.0003	-0.0009***
	(0.0003)	(0.0003)	(0.0003)	(0.0001)
Interest-only loan	0.0135	0.0131	0.0215**	0.0324***
	(0.0083)	(0.0083)	(0.0087)	(0.0033)
High cost loan	0.1126***	0.1120***	0.1003***	0.0724***
	(0.0129)	(0.0129)	(0.0137)	(0.0037)
Black-hispanic borrower	-0.0094	-0.0116*	0.0642**	0.0345***
	(0.0066)	(0.0066)	(0.0260)	(0.0113)
Zip code level controls				
90+ days delinquency ratio, 2007-2010			1.113***	1.587***
			(0.243)	(0.1347)
Loans originated (fraction of housing stock), 2004-2006			-0.0028	0.0058
			(0.0073)	(0.0036)
Fraction of subprime, 2004-2006			-0.0175	-0.276***
			(0.0738)	(0.0372)
Debt per capita, 2004-2006			-0.0001***	-0.0001***
			(0.0000)	(0.0000)
House price change, 2004-2006			0.1430***	0.0376**
			(0.0310)	(0.0158)
Census income, 2000			0.00127	-0.0264**
			(0.0208)	(0.0120)
Zip Code % Black or Hispanic in 2000			0.0237***	0.0085***
			(0.0059)	(0.0026)
Constant	0.7832***	0.8008***	0.6107**	0.9847***
	(0.0518)	(0.0557)	(0.2393)	(0.1365)
Obs	27011	27011	24010	102460
R2	.0840	.0841	.0929	.0573

Table 9
Change in House Prices and Outstanding Mortgage Concentration

Cross-sectional zip code level regressions of the logarithmic change in house prices between 2007 and 2010 on the index of outstanding mortgage concentration, Top4, between 2004 and 2006. All variable definitions and sources are reported in Table 1.. All regressions include county dummies. In all columns, but columns 3 and 4 we report ordinary least squares regressions. In column 3, estimates are obtained using a propensity score methodology. In column 4, estimates are obtained by two stage least squares; the first stage is reported in column 4 of Table 4. Standard errors in parenthesis are clustered at the county level and corrected for heteroskedasticity. Estimates followed by ***, **, and * are statistically different from zero with 0.01, 0.05 and 0.10 significance levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS		Propensity Scores	2SLS	OLS		
Top4, 2004-2006	0.205***	0.128***	0.129***	0.553**			0.281***
	(0.046)	(0.036)	(0.046)	(0.272)			(0.056)
Top4_ret, 2004-2006					0.131***		
					(0.043)		
Top 4 _ret ×Securitization, 2004-2006					-0.112**		
					(0.052)		
T2 Top 4, 2004-2006						0.008***	
						(0.003)	
T3 Top 4, 2004-2006						0.013***	
						(0.004)	
Top 4, 2004-2006×Judicial foreclosures							-0.241***
							(0.070)
Securitization, 2004-2006		0.013	-0.001	0.154	0.015	-0.002	0.014
		(0.023)	(0.048)	(0.105)	(0.024)	(0.025)	(0.022)
90+ days delinquency ratio, 2007-2010		-1.301***	-0.834*	-1.460***	-1.267***	-1.319***	-1.267***
		(0.182)	(0.454)	(0.284)	(0.179)	(0.178)	(0.179)
Loans originated (fraction of housing stock), 2004-2006		-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Fraction of subprime, 2004-2006		0.087**	-0.015	0.048	0.084**	0.090**	0.078**
		(0.036)	(0.082)	(0.031)	(0.036)	(0.036)	(0.035)
Debt per capita, 2004-2006		0.013**	0.011	0.014***	0.014**	0.013**	0.012**
		(0.006)	(0.007)	(0.005)	(0.006)	(0.006)	(0.006)
House price change, 2004-2006		0.055**	0.031	0.028	0.055**	0.057**	0.052**
		(0.024)	(0.047)	(0.021)	(0.024)	(0.023)	(0.023)
Census income, 2000		-0.005	-0.011	0.005	-0.003	-0.005	-0.005
		(0.009)	(0.013)	(0.010)	(0.009)	(0.009)	(0.009)
Zip Code % Black or Hispanic in 2000		0.008	0.012	0.019	0.005	0.007	0.009
		(0.011)	(0.020)	(0.015)	(0.011)	(0.011)	(0.011)
Constant	-0.227***	-0.175*	-0.055		-0.206**	-0.147	-0.175*
	(0.006)	(0.101)	(0.158)		(0.101)	(0.099)	(0.100)
Fixed Effects	County	County	County	MSA	County	County	County
F Test Excluded Instruments				36.91			
Obs	5547	5233	5138	5190	5233	5233	5233
R2	.882	.887	.872		.887	.887	.888

Table A.1
The Exclusion Restriction

This table provides evidence supporting the validity of the fraction of senior as an instrument for Top 4. We run cross-sectional regressions of the variables listed in each column, measured between 2004 and 2006, on the % Population above 65. All regressions include MSA fixed effects. Standard errors in parenthesis are clustered at the MSA level and corrected for heteroskedasticity. Estimates followed by ***, **, and * are statistically different from zero with 0.01, 0.05 and 0.10 significance levels, respectively.

	(1) Loans Originated	(2) Median Income	(3) Income Growth	(4) Loans Retained
% Population above 65, 2004	0.000	-0.006	-0.001	0.003**
	(0.002)	(0.007)	(0.001)	(0.001)
Constant	0.730***	11.026***	0.067***	0.219***
	(0.018)	(0.078)	(0.009)	(0.016)
Obs	6570	6570	6454	6570
R2	.31	.223	.0723	.414