

# Credit Supply and Housing Speculation

Atif Mian

Princeton University and NBER

Amir Sufi

University of Chicago Booth School of Business and NBER

March 2019

## **Abstract**

Speculation is a critical channel through which credit supply expansion affects the housing cycle. The surge in private label mortgage securitization in 2003 fueled a large expansion in mortgage credit supply by lenders financed with non-core deposits. Areas more exposed to these lenders experienced a large relative rise in transaction volume driven by a small group of speculators, and these areas simultaneously witnessed an amplified housing boom and bust. Consistent with the importance of belief heterogeneity, house price growth expectations of marginal buyers rose during the boom, while housing market pessimism among the general population increased.

---

\*This research was supported by funding from the Julis Rabinowitz Center For Public Policy and Finance at Princeton, and the Fama-Miller Center and Initiative on Global Markets at Chicago Booth. We are especially grateful to Neil Bhutta, Nicola Cetorelli, Neale Mahoney, Stefan Nagel, Stavros Peristiani, Tomasz Piskorski, Philipp Schnabl, Amit Seru, Anne Thompson, and Eric Zwick for help with data access. We received helpful comments from John Griffin, Adam Guren, Christopher Palmer, Andrei Shleifer, Paul Willen, Eric Zwick, and seminar participants at Chicago Booth, DeNederlandscheBank, the European Central Bank, Harvard Business School, John Hopkins University, Northwestern University, the NBER Corporate Finance, Real Estate, and Monetary Economics meetings, and Princeton University. Nicolas Andreoulis, Tu Cao, Sebastian Hanson, Michael Varley, and Hanbin Yang provided excellent research assistance. Mian: (609) 258 6718, atif@princeton.edu; Sufi: (773) 702 6148, amir.sufi@chicagobooth.edu

The interaction between credit and speculation plays a leading role in the narrative of financial cycles advanced by scholars such as Charles Kindleberger (1978) and Hyman Minsky (1986). The Global Financial Crisis has led researchers to focus more on this narrative. This study exploits a natural experiment centered on the rise of the private label mortgage securitization market (the “PLS market”) in order to investigate the link between credit supply expansion and housing speculation.

In the language of Kindleberger and Minsky, a boom and bust cycle in asset prices often begins with a “displacement” such as financial innovation or financial liberalization that expands credit to speculators. The speculators use leverage to bid for an asset, and such bidding boosts the asset’s price. The increase in price brings in more speculators creating a positive feedback between credit and speculation. This feedback effect generates “euphoria” or “mania” in the market as prices and trading volume rise rapidly.

These dynamics imply that speculators are overly optimistic during the euphoric state with many holding the asset purely on the belief that prices will continue to grow. Adam Smith (1776) referred to this as “overtrading” where market-implied expectations are in “error.” Consequently, the boom in credit and asset prices leads to a predictable bust. Evidence from recent business cycle history is consistent with these ideas where housing assets in particular are critical. For example, growth in mortgage credit is strongly correlated with growth in house prices, and housing-related credit growth predicts financial crises and a decline in GDP growth (e.g., Jordà et al. (2016) and Mian et al. (2017)).

However, formally testing the details of the ideas in Kindleberger and Minsky remains difficult for a number of reasons. First, identifying a credit supply shock is hard. It is difficult to isolate a financial shock that might trigger a speculative cycle. Second, tracing the cycle of speculation requires granular data on transaction volume, prices, and the characteristics of individual traders who come into the market due to credit expansion. Third, an important element of the Kindleberger-Minsky narrative is the expectations of speculators, and how these expectations differ from average or rational expectations. This necessitates data on

expectations both for speculators who trade the underlying asset and the broader population who stays out of the market.

This study addresses these challenges by exploiting the global surge in shadow banking in 2003 that enabled lenders relying on non-core deposit financing to increase mortgage credit supply more aggressively. Existing research shows how events in the late summer of 2003 led to a sudden and large surge in the PLS market (e.g., [Justiniano et al. \(2017\)](#)). The sudden acceleration of the PLS market created a natural experiment in which housing markets in the United States that were more exposed to non-core deposit financed lenders (which we refer to as non-core liability lenders, or “high NCL lenders”) experienced an immediate and large relative rise in mortgage originations.

Section 1 discusses the background of the natural experiment and explains why exposure to high NCL lenders provides a plausibly exogenous source of variation in mortgage credit supply expansion across housing markets. There was no differential pre-trend in any outcome variable in more exposed markets, and the relative rise in mortgage originations in more exposed markets occurred exactly in the same months as the aggregate surge in the PLS market. Furthermore, as in [Khwaja and Mian \(2008\)](#), confounding shocks can be absorbed at a granular census-tract level to show that the relative rise in mortgage originations in locations more exposed to high NCL lenders was not driven by spurious credit demand shocks.

The empirical strategy uses administrative data from TransUnion to construct a new individual-level data set of mortgage originations by both speculators and non-speculators. Individuals buying a house are classified as speculators if they (a) buy multiple houses in a short time period, (b) buy and sell a given house within one year, or (c) buy a house when already having 2 or more first-lien mortgages on their balance sheet. The TransUnion data also allows for an exploration of the age and credit score of the individuals buying homes with a mortgage. In addition, data on average (population-wide) housing market expectations at the MSA level from the Michigan Survey of Consumers are combined with

data on housing market expectations of actual home buyers from [Case et al. \(2012\)](#). This allows for a contrast between the marginal expectations of home buyers with average housing market expectations in the broader population.

The results from the natural experiment show that housing markets more exposed to high NCL lenders witnessed a sharp increase in home-purchase mortgage originations and housing transaction volume from 2002 to 2006. The magnitude of the effect was large. Housing markets in the top quartile of NCL exposure experienced home purchase mortgage origination growth that was 23.0 percentage points higher than mortgage origination in the bottom quartile at the peak of the cycle. This effect translates into a 19.1 percentage point larger increase in transaction volume.

Who bought during the boom? The relative increase in transaction volume in high NCL exposed areas was driven mainly by speculators. Depending on the precise definition of a speculator, the share of the total relative increase in volume driven by speculators was between 40% and 70%. If a speculator is defined in the broadest terms as fitting any of the three definitions, then almost 100% of the relative increase in transaction volume in high NCL lender exposed areas was driven by speculators. An alternative measure of speculation based on how rapidly a given property is traded (following [DeFusco et al. \(2018\)](#)) yields similar results. The speculators that were brought into the market tended to be younger and riskier from an ex ante perspective. We interpret these findings as showing how an expansion of mortgage credit supply helped instigate a speculative trading frenzy in areas most exposed to high NCL lenders.

Both house prices and construction rose more substantially in more exposed areas from 2003 to 2006. Moving from the bottom to top quartile of the NCL-lender exposure distribution led to a 12.1 percentage point increase in house prices and a 19.0 percentage point increase in construction of new housing units.

However, the boom was short-lived; beginning in 2007, more exposed housing markets suffered a larger collapse in mortgage origination, transaction volume, house prices, and

construction activity. In fact, house prices and construction activity in more exposed areas over-corrected relative to their pre-boom levels, consistent with the prediction of models of speculation such as [Glaeser et al. \(2008\)](#). The origins of the mortgage default crisis were closely linked to this cycle. Mortgage default rates began to rise as early as late 2006 in high NCL lender exposed areas, and rose even further through 2009. In 2007, the share of total mortgage delinquencies coming from zip codes in the top quartile of the NCL exposure distribution increased by 5 percentage points.

Speculators had a large effect on housing markets more exposed to high NCL lenders even though they made up a small part of the overall population. Even under the broadest definition, speculators made up less than 1.5% of the overall population. Furthermore, the evidence suggests that high credit score traditional home-buyers experienced a relative decline in home purchase activity in areas exposed to high NCL lenders. Taken together, these facts motivate us to explore the idea that heterogeneity in beliefs about house price growth may have been important in explaining how credit affected the housing market (e.g., [Geanakoplos \(2010\)](#), [Simsek \(2013\)](#)).

The results show that disagreement about future house price growth was a salient feature of the housing boom from 2002 to 2006. Recent home-buyers had high expectations of house price growth ([Case et al. \(2012\)](#)), while the broader population actually became more pessimistic about the prospect of buying a house ([Piazzesi and Schneider \(2009\)](#)). In the cross-section, the average individual became more pessimistic about the direction of future house prices in cities most exposed to high NCL lenders. Consistent with models of heterogeneous beliefs, credit expansion fueled purchases by more optimistic speculators while pessimism increased among the general population.

Overall, the results on expectations suggest that greater availability of credit allowed optimistic speculators to increase demand and hence asset prices as in [Geanakoplos \(2010\)](#). The increase in asset prices may have convinced more people to become speculators, as in [Burnside et al. \(2016\)](#), which further boosted asset prices. However, the rest of the population

agreed to disagree, becoming more pessimistic about the market as prices rose. In general, the results are supportive of the idea that belief heterogeneity is critical to understanding the relationship between credit, speculation, and asset prices.<sup>1</sup>

There is a large body of empirical work aimed at understanding the causes and consequences of the credit boom during the 2000s. Much of this research is discussed in the next section in the context of the acceleration of the PLS market. This study is related to research on the role of investors in explaining the housing cycle of 2000 to 2010. [Bhutta \(2015\)](#), [Gao et al. \(2017\)](#), [DeFusco et al. \(2018\)](#), [Haughwout et al. \(2014\)](#), and [Albanesi et al. \(2017\)](#) focus on investors, showing that areas with larger boom-bust cycle witnessed greater increase in investor share and speculative behavior, and short-term investors amplified volume and price movements. [Purnanandam \(2011\)](#) shows that banks relying more on the originate-to-distribute model increased originations of riskier credit.

[Mian and Sufi \(2009\)](#) use a within-county empirical strategy to show that zip codes with a higher initial share of subprime borrowers witnessed larger relative growth in mortgage originations for home purchase and house prices from 2002 to 2005. The measure of exposure to PLS used in this study, the 2002 NCL share of a zip code, is strongly correlated with the share of subprime borrowers in 2000 within MSAs used in [Mian and Sufi \(2009\)](#).<sup>2</sup> This implies that many of the patterns shown in [Mian and Sufi \(2009\)](#) are related to the speculation channel emphasized here.

There is also prior work on credit expansion and house price growth, e.g. [Adelino et al. \(2014\)](#), [Favara and Imbs \(2015\)](#), and [Di Maggio and Kermani \(2017\)](#). These research studies use different instruments for credit growth to show a causal effect of credit growth on house

---

<sup>1</sup>There is a large theoretical literature using asymmetric information and belief heterogeneity to generate bubbles, e.g. [Miller \(1977\)](#), [Harrison and Kreps \(1978\)](#), [Scheinkman and Xiong \(2003\)](#), [Abreu and Brunnermeier \(2003\)](#), [Simsek \(2013\)](#), [Caballero and Simsek \(2018\)](#), [Piazzesi and Schneider \(2009\)](#) and [DeFusco et al. \(2018\)](#). Empirical support for the importance of leverage and belief heterogeneity from the 18th century can be found in [Koudijs and Voth \(2016\)](#). [Allen and Gorton \(1993\)](#) and [Allen and Gale \(2000\)](#) argue that easy credit further encourages speculators to pay more than the fundamental value of an asset because they can shift downside risk to lenders.

<sup>2</sup>Figure [A1](#) in the appendix plots the share of subprime borrowers in 2000 against the 2002 NCL share in a zip code after demeaning these variables at the MSA level. Each zip code is weighted by its population size. The within-MSA pairwise correlation coefficient is 0.57 and highly significant.

prices.

To the best of our knowledge, this study is the first to test how a specific source of variation in mortgage credit supply expansion fuels speculative trading activity, and as a result affects house prices, construction, and delinquency.<sup>3</sup> It is also the first to link credit supply expansion to the evolution of differences in beliefs about the housing market between homebuyers and the rest of the population. This study also contributes to the recent literature on the role of financial and behavioral factors in driving economic cycles. This literature builds on the seminal insights introduced by scholars such as Irving Fisher, Charles Kindleberger, and Hyman Minsky by providing theoretical reasoning and empirical evidence on the mechanisms behind financial boom-bust cycles (see e.g., [Bordalo et al. \(2017\)](#), [Greenwood and Hanson \(2013\)](#), [Krishnamurthy and Muir \(2017\)](#), and [López-Salido et al. \(2017\)](#)).

## 1 Research Design and First Stage

### 1.1 The surge in private label mortgage securitization

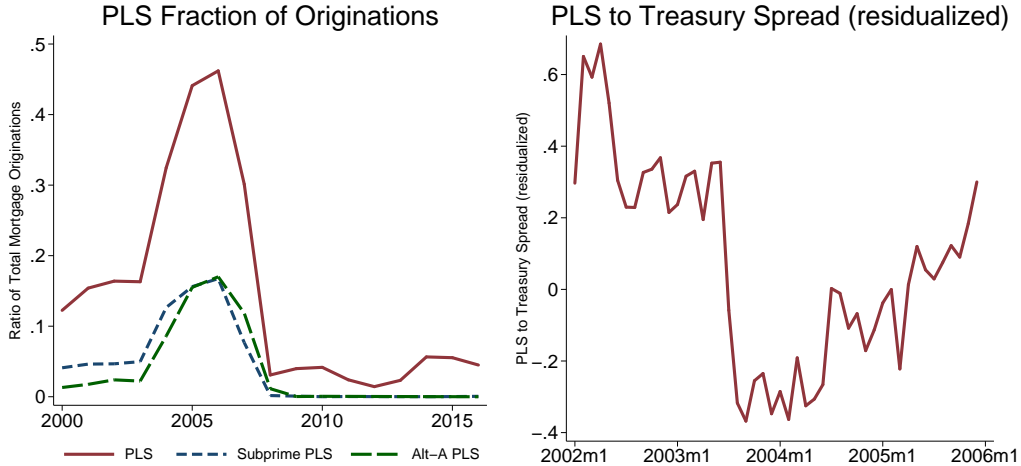
The starting point for the analysis is the large expansion in private label mortgage securitization documented by [Justiniano et al. \(2017\)](#). The left panel of Figure 1 shows the surge in the PLS market starting in late 2003. The share of mortgage originations that were securitized in the PLS market went from 16% in 2002 to 46% in 2006. The right panel shows that the rise in private label securitization corresponded with a sharp fall in the PLS mortgage interest spread relative to Treasuries. The spread is calculated by [Justiniano et al. \(2017\)](#) in the spirit of [Gilchrist and Zakrajšek \(2012\)](#) by controlling for risk and other contractual features.

The simultaneous rise in quantity and decline in spread implies an expansion in the sup-

---

<sup>3</sup>The empirical strategy used here is closest to [Nadauld and Sherlund \(2013\)](#) who measure a zip code's exposure to the growth in securitization of mortgages by the five largest broker/dealer investment banks during the 2003 to 2005 period. They find that securitization affected mortgage originations and default rates, but they do not focus on house prices, construction, housing market optimism, or speculation.

Figure 1: Acceleration in Private Label Securitization



The left panel plots the share of total mortgage originations that were sold into private label securitization (PLS), subprime PLS, and Alt-A PLS. The data on originations come from SIFMA, divided by total originations from NY Fed Equifax data. The right panel shows the average spread between mortgage interest rates in the private label securitization market and U.S. Treasuries from Justiniano et al. (2017), where characteristics of the mortgage are absorbed and the residual is plotted.

ply of credit. The expansion in total supply of mortgage credit to the economy can be seen in Figure 2. It separates mortgage originations for home purchases into three types of originations: GSE-backed originations, originations held on bank balance sheets, and originations securitized in the private label market.<sup>4</sup> The flow of mortgage origination amounts increased by \$600 million in 2006 relative to 2002, and \$500 million of the increase was mortgages originated for the PLS market. The aggregate increase in mortgage originations from 2003 to 2006 was driven by the surge in the PLS market.<sup>5</sup>

What lies behind the rise in the PLS market starting in 2003? Justiniano et al. (2017)

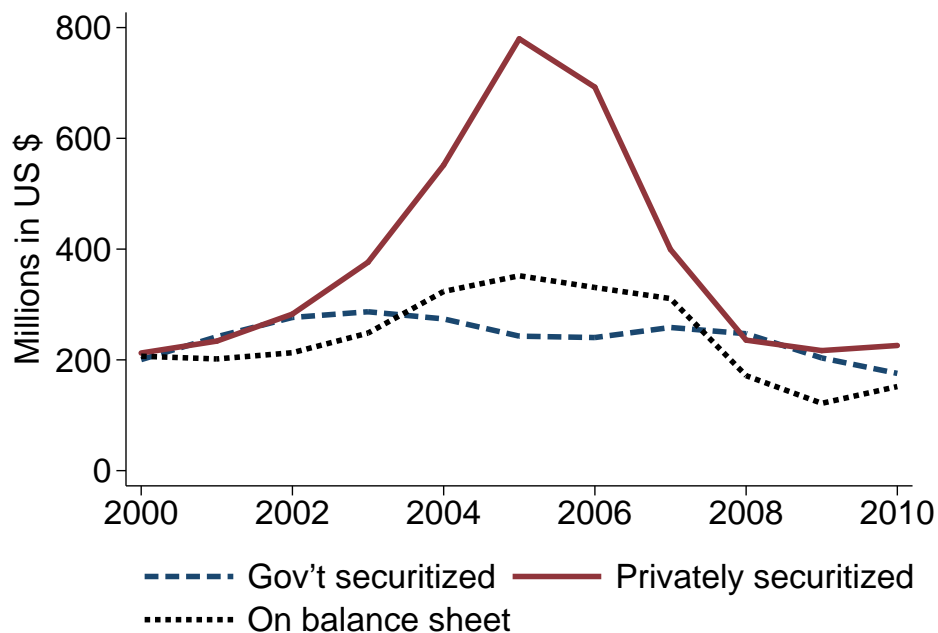
<sup>4</sup>HMDA requires lenders to report to whom an originated loan is sold if it is sold within one year of origination. We group together five categories as a measure of mortgages sold into the PLS market.: mortgages sold (1) into private securitization, (2) to a commercial bank, savings bank, or savings affiliation affiliate, (3) to a life insurance company, credit union, mortgage bank, or finance company, (4) to an affiliate institution, or (5) to other type of purchaser. Section A.1 in the appendix provides more details on HMDA data construction.

<sup>5</sup>A recent body of research explores the rise of the *level* of household debt across the income and credit score distribution during the housing boom (Mian and Sufi (2017b); Adelino et al. (2017); Foote et al. (2016); Albanesi et al. (2017)). The results in this study focus instead on the extensive margin of house purchases, whereas the rise in the level of household debt from 2000 to 2007 was due primarily to existing homeowners borrowing against home equity (e.g., Mian and Sufi (2011), Mian and Sufi (2015), Bhutta and Keys (2016), Mian and Sufi (2017b)).



attribute the rise of PLS to the response of mortgage lenders to the end of the refinancing wave. The authors argue that mortgage originators shifted their attention “to previously underserved segments of the mortgage market by keeping mortgage rates low especially for those borrowers.”<sup>6</sup> [Gurun et al. \(2016\)](#) show evidence of a greater push by the mortgage industry to “sell” mortgages. For example, advertisement by the mortgage industry increased six-fold from 2003 onward and induced people to take on mortgages that they otherwise would not have. [Mian et al. \(2013\)](#) show political lobbying for sub-prime mortgage expansion increased and resulted in favorable legislation, such as the safe harbor provision that facilitated shadow financing (e.g., [Perotti \(2013\)](#)).

Figure 2: Mortgage Originations for Home Purchase by Type



Source: HMDA. See text for precise definition of a mortgage sold into the private securitization.

Mortgage originators found investors in the private label market who “neglected risk” and bought mortgages that were riskier than their implied prices ([Gennaioli et al. \(2012\)](#),

<sup>6</sup>Other related work showing declining spreads and rising origination volume includes [Levitin and Wachter \(2013\)](#) and [Demyanyk and Van Hemert \(2011\)](#). [Drechsler et al. \(2017\)](#) and [Xiao \(2018\)](#) provide theory and evidence that suggests that growth in the shadow banking sector was related to the end of the monetary policy easing cycle in the summer of 2003. See also [Landier et al. \(2015\)](#) and [Nagel \(2016\)](#).

Chernenko et al. (2018) and Coval et al. (2009)). A BIS study shows that European banks bought a large fraction of the private label non-agency mortgage debt (McCauley (2018)). In highly influential work, Keys et al. (2010) and Keys et al. (2012) uncover a shift towards lax screening in the PLS market around 2003. This shift encouraged lenders to originate new non-conforming mortgages that the private label market was increasingly willing to absorb. Rajan et al. (2015), Dell’Ariccia et al. (2012) and Ashcraft et al. (2010) show evidence that holders of PLS mortgages took on mortgages with riskier soft-information attributes. Piskorski et al. (2015), Griffin and Maturana (2016) and Mian and Sufi (2017a) show that lenders and borrowers took advantage of neglected risk in the PLS market through fraudulent behavior such as misreporting owner-occupied status and borrower income.

The broader rise in shadow financing went beyond the PLS market to impact other sectors as well (see Gorton and Metrick (2013) for a review). For example, Figure A2 in the appendix shows that originated amounts in the collateralized loan obligation market, which focuses exclusively on corporate debt with no direct link to residential mortgages, increased from less than \$20 billion to almost \$90 billion from 2002 to 2006. The empirical strategy does not need to take a stand on the precise source of the expansion in shadow financing availability in 2003, as long as it reflects an expansion in credit supply.

## 1.2 Data on mortgage lenders and mortgage originations

The empirical strategy outlined below seeks to estimate the effect of the aggregate expansion of the PLS market on lenders and geographic regions in the United States. It therefore requires microeconomic data on mortgage lending and mortgage lenders. Mortgage origination information comes from HMDA. The HMDA data set records the universe of mortgage originations for mortgage originators that have an office within metropolitan statistical areas (MSAs).<sup>7</sup> We identify each mortgage originator in the HMDA data, and we classify them

---

<sup>7</sup>See guidelines for HMDA issued by the Federal Reserve in 2005: “a lender does not have to report HMDA data unless it has an office in a metropolitan statistical area (MSA). As a result, reporting of home loans in some rural areas may be relatively low.”

as either a “bank” or a “non-bank” based on whether they are regulated by the Federal Reserve as a deposit-taking institution. Furthermore, we link these financial institutions to Call Report data using a key provided to us by the Federal Reserve Board.

Given the reporting restriction for originators in the HMDA data, we isolate our sample to zip codes that are located within metropolitan statistical areas. For these zip codes, we aggregate all HMDA originations by year, which gives us a zip-year level data set on mortgage originations. There are two additions to the standard HMDA data. First, we use an MSA by month level version of the HMDA data set below in some specifications. Second, for home purchase mortgages, the HMDA data split first- and second-lien mortgages beginning in 2004. For the years prior to 2004, we use data from [Bhutta and Keys \(2018\)](#) that split first- and second-liens based on a methodology explained in their study.<sup>8</sup> Summary statistics are shown in Table [A1](#) in the appendix.

### **1.3 First stage: High NCL lenders and the PLS expansion**

The rise in the PLS market allowed lenders to originate more mortgages as they were able to more easily originate and sell mortgages into securitization. However, lenders differed in their funding models, which generated variation across lenders in the increase in the ability to finance mortgages through securitization. For example, certain commercial banks relied mainly on core deposits raised through their branch network for funding their balance sheets. However, other banks relied less on core deposits (say because of a weaker depositor base) and more on wholesale shadow financing. Non-bank mortgage lenders relied entirely on non-core deposits to finance mortgage originations.

The strategy employed in this study constructs a measure of dependence on shadow financing at the lender level and tests if the rise in PLS boosted mortgage originations more for shadow-financed lenders. The lender level measure is the ratio of non-core liabilities to total liabilities (NCL for short) as of 2002. NCL is defined as one minus the ratio of core

---

<sup>8</sup>We are extremely grateful to Neil Bhutta who provided us access to the key linking the HMDA Report ID’s to the TFR bank ID’s and to the two additions to the standard HMDA data.

deposits to total liabilities, where core deposits are defined as FDIC insured deposits.<sup>9</sup>

The rise in the PLS market would lead to stronger origination growth by high NCL lenders under the assumption that it was difficult for lenders to change their funding structure relatively quickly. This assumption is supported by the fact that there was a large degree of persistence in the ranking of lenders by the NCL ratio during the boom. For example, a regression of NCL ranking by lenders as of 2002 on NCL ranking from 2001 through 2007 yields a coefficient between 0.9 and 1 for every year.

The dependence of lenders with a high NCL ratio on the PLS market for origination can be seen in Figure 3. The figure splits lenders in the HMDA data set as “high NCL” if their NCL ratio in 2002 was above the median and “low NCL” otherwise. We then plot the share of total mortgage originations in 2002 by each lender type across the three categories shown earlier in Figure 2. Lenders with a high NCL ratio sold 43.4% of their originated mortgages into the PLS market as of 2002, while lenders with low NCL only sold 23.5% into the PLS market. On the other hand low NCL lenders kept 39.5% of their originated mortgages on balance sheet compared to only 19.2% for high NCL. The two types of banks did not differ in the fraction of mortgage originations sold into GSE backed pools. Even before the expansion of the PLS market, high NCL lenders heavily utilized the PLS market for funding mortgage origination volume.

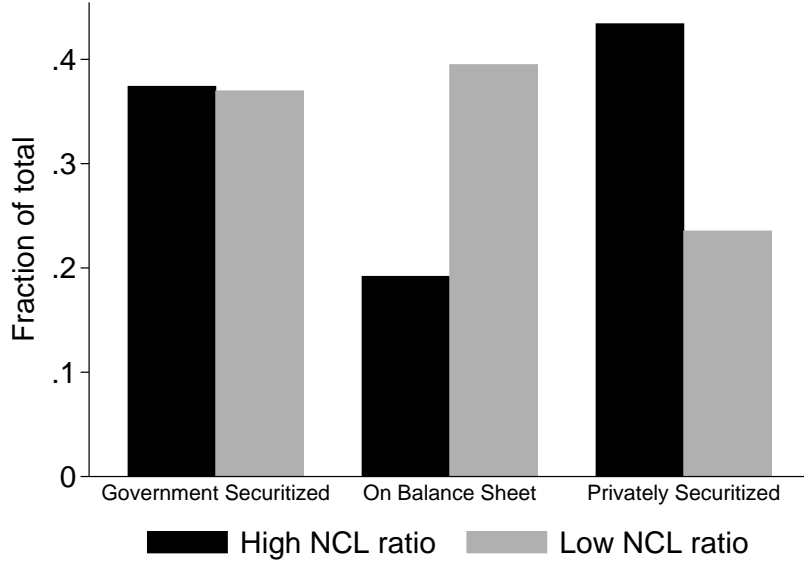
The expansion in the PLS market led to a significant relative expansion of mortgage lending by high NCL lenders. The upper panel in Figure 4 plots total mortgage originations and home purchase mortgage originations separately for high and low NCL lenders. There was almost no difference in mortgage origination between the two types of lenders prior to 2003. However, starting in 2003, high NCL lenders expanded mortgage supply by more.

The bottom left panel uses the full distribution of exposure to NCL in 2002 and presents

---

<sup>9</sup>Notable high NCL banks as of 2002 were Countrywide Bank NA and IndyMac BK FSB. Notable non-bank mortgage lenders, with an NCL equal to one by definition, as of 2002 were Ameriquest Mortgage Company, New Century Mortgage Corp, and American Home Mortgage Company.

Figure 3: Mortgage Originations as of 2002, by Lender Type



The figure shows the fraction of total mortgage originations that are government securitized, privately securitized, and on balance sheet for lenders that have above and below the median non-core liability ratio as of 2002. Source: HMDA. See text for precise definition of a mortgage sold into the private securitization.

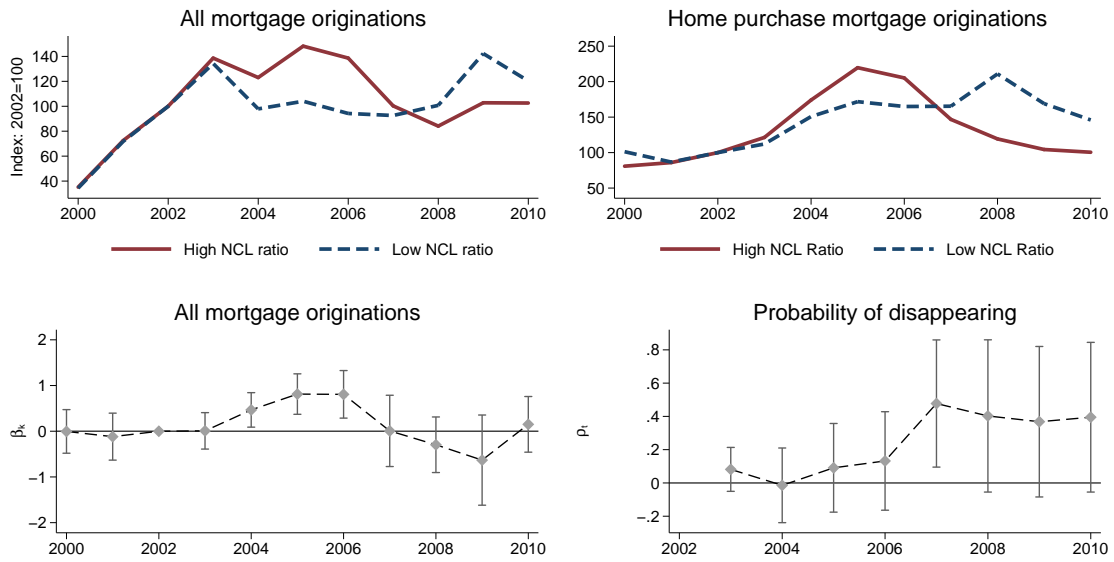
estimates  $\{\beta_k\}$  from the regression specification:

$$\ln(y_{b,t}) = \alpha_b + \gamma_t + \sum_{k \neq 2002} \mathbb{1}_{t=k} \beta_k NCL_{b,2002} + \varepsilon_{b,t} \quad (1)$$

where  $\ln(y_{b,t})$  is the natural logarithm of total amount originated by lender  $b$  in year  $t$ . The coefficients  $\{\beta_k\}$  give the relative growth in mortgage amount originated by high NCL lenders since 2002 (which is the omitted year). The estimated coefficients show that there was no pre-trend and a sharp relative rise for high NCL lenders starting in 2003 and accelerating during 2004 and 2005. These results confirm the presence of a strong first stage at the lender level. Lenders with less access to core deposits relied more on the PLS market for mortgage origination and consequently originated more mortgages with the rise of the shadow banking sector.<sup>10</sup>

<sup>10</sup>In appendix figure A3, we present results separately for refinancing originations. The results are similar: there is no significant pre-trend, and high NCL ratio lenders see stronger relative growth in refinancing originations starting in 2003.

Figure 4: Mortgage Origination by Dependence on NCL



The top panel plots total mortgage originations (total and home purchase) for lenders above and below the median NCL ratio as of 2002. The bottom left panel plots coefficients  $\{\beta_k\}$  from equation 1. The specification that generates the bottom right panel uses as a dependent variable  $Exit_{b,t}$ , an indicator variable for whether a lender that is in sample in 2002 is still in sample in year  $t$  for years 2003-2009. Regressions use HMDA data and are weighted by the mortgage amount originated in 2002 by lender  $b$ . Robust standard errors reported.

There was a relative decline in mortgage lending by high NCL lenders when the PLS market collapsed in 2007. However, this decline is underestimated in the regression specification because high NCL lenders were more likely to disappear from the sample after 2006. If a lender disappeared, then it is not included in the sample for that year in the bottom left panel. The bottom right panel presents regression coefficients for a linear probability model that is similar to equation 1 except the left hand side variable is the probability of the lender being absent from the HMDA data in that year. Results show that high NCL lenders have a higher probability of exiting the sample in 2007.

Column 1 of Table 1 summarizes the first stage effect of rise in PLS on shadow-financed lenders by regressing:

$$\Delta y_{b,02-05} = \alpha + \beta NCL_{b,2002} + \varepsilon_b \quad (2)$$

where  $\Delta y_{b,02-05}$  is the percentage change in origination growth from 2002 to 2005 for lender  $b$ . A lender NCL ratio going from 0 to 1 led to a 95.7 percentage point increase in mortgage originations in 2005 relative to 2002.

Columns 2 through 4 decompose the total effect into the three ways a lender can finance a loan origination. Column 2 changes the dependent variable to the change in originated mortgages that were sold into the PLS market divided by all 2002 originations. Columns 3 and 4 do likewise for originations sold to GSEs and originations kept on a lender's balance sheet. This decomposition ensures that the estimates in columns 2 through 4 add up to the total effect in column 1. Decomposing the total effect shows that almost the entire effect is driven by an increase in originations by high NCL banks that were sold into the PLS market. This further supports the view that the rise in the PLS market enabled high NCL lenders to expand credit supply by originating and selling loans in the PLS market.

Table 1: High NCL Ratio Predicts Growth in Mortgage Originations, 2002-2005

	Total Growth	Contribution from, 02 to 05			Total Growth, 02 to 05	
	02 to 05 (1)	PLS (2)	GSE (3)	On Balance Sheet (4)	MSA (5)	Census Tract (6)
2002 NCL Ratio	0.957* (0.377)	0.910** (0.311)	0.052 (0.100)	-0.005 (0.127)	1.107* (0.547)	1.010* (0.491)
N	3,957	3,957	3,957	3,957	39,378	330,244
R-sq	0.004	0.005	0.001	0.000	0.069	0.332
Level	Bank	Bank	Bank	Bank	Bank-MSA	Bank-tract
FE	None	None	None	None	MSA	Tract

Standard errors clustered at bank level in columns 1 to 4.

Standard errors clustered at bank level and geographic level in columns 5 and 6.

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

## 1.4 Exclusion restriction tests

The first stage impact of non-core deposit dependence on mortgage origination growth is strong, and the timing lines up with the aggregate expansion in the shadow banking sector. The identifying exclusion restriction assumption is that lenders with a higher NCL ratio did not receive a spurious credit demand shock contemporaneously with the expansion of the PLS market. For example, perhaps high NCL lenders focused on areas that for some reason experienced a stronger housing demand shock in 2003. If that were the case, then the exclusion restriction assumption would no longer hold and the first stage could not be interpreted as identifying a credit supply effect.

This section tests for concerns about spurious credit demand shocks for high NCL lenders. One technique to address these concerns is saturating the main regression specification with granular fixed effects to absorb credit demand shocks in the spirit of [Khwaja and Mian \(2008\)](#). In particular, the following specification utilizes disaggregated mortgage origination data at the lender-MSA and lender-census-tract level, and it includes MSA and census-tract fixed effects:

$$\Delta y_{b,g,02-t} = \alpha_{g,t} + \beta_t NCL_{b,2002} + \varepsilon_{b,g,t} \quad (3)$$

where the outcome variable is the growth in originated mortgage amount by lender  $b$  in



geography  $g$  from 2002 to  $t$ . The geographical unit may be an MSA or a census tract. Census tracts are small contiguous subdivisions of a county, with typical population size between 1,200 and 8,000 with an average of about 4,000. The definition of census tracts can potentially change every decennial census, and 2003 was the first year that HMDA used 2000 census definitions. The estimation therefore restricts itself to census tracts that did not change materially between the 1990 and 2000 census definitions to compute mortgage origination growth at the census tract level between 2002 and  $t$ .<sup>11</sup>

The estimates of  $\alpha_{g,t}$  absorb any credit demand shocks at geographic level  $g$  between 2002 and year  $t$ . The coefficient  $\beta_t$  is estimated from variation between two lenders that differ in their exposure to the PLS market but service the same narrowly defined housing market. Columns 5 and 6 in Table 1 estimate equation 3 for origination growth between 2002 and 2005. Column 5 estimates the equation at the bank-MSA level with MSA fixed effects, and column 6 at the bank-census-tract level with census-tract fixed effects. The estimated  $\beta$  is similar to the lender-level estimate in column 1, even though the specification is now saturating possible credit demand shocks at a highly granular level. For example, the R-squared in column 6 rises to 0.332 from 0.004 in column 1.<sup>12</sup>

Another important point is that there was considerable overlap in the borrowers financed by high versus low NCL lenders prior to the acceleration of the PLS market in 2003. Figure 5 compares the credit score distribution for originated mortgages in 2002 for high and low NCL lenders. The credit score distribution is constructed using information from HMDA on mortgage origination by a lender in a zip code and matching it with the average credit score of borrowers in that zip code from the TransUnion credit bureau data. While high NCL lenders tended to lend to lower credit score zip codes on average as of 2002, there was

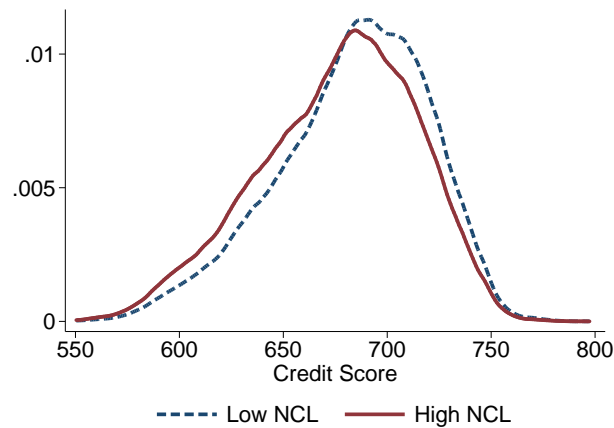
---

<sup>11</sup>We impose a restriction that a bank-census tract must have at least 10 mortgage originations in 2002 to be included in the sample. We also weigh each bank-census-tract observation within a census tract by its origination share in 2002. This way the correct dollar-weighted census-tract-level credit demand shock is differenced out.

<sup>12</sup>It is important to recognize that the specifications in columns 5 and 6 are “over controlling” in the sense that any spillovers in credit expansion due to competitive effects from one lender to another within a narrowly defined housing market would tend to reduce the estimated coefficient.

considerable overlap in the credit score distribution of high and low NCL lenders.

Figure 5: Credit Score Density in 2002 by Lender Type



The figure plots the credit score density in 2002 of borrowers from lenders that have above and below the median non-core liability ratio as of 2002. Lending data are from HMDA and credit score data from TransUnion.

## 1.5 Geographic exposure to high NCL lenders

The rise of the PLS market led high NCL lenders to increase mortgage originations significantly more than low NCL lenders starting in 2003. In order to analyze its impact on local housing markets, the rest of this study uses zip code or MSA level variation in exposure to high NCL lenders. For each geographical area  $g$ , the exposure to the expansion of the PLS market is calculated as the average of the 2002 NCL ratios of mortgage lenders in the area, where the average is weighted by a lender's amount of mortgage originations in 2002. More specifically:

$$NCLShare_{g,2002} = \sum_b \omega_{g,b,2002} * NCL_{b,2002}$$

where:

$$\omega_{g,b,2002} = \frac{Originations_{g,b,2002}}{\sum_b Originations_{g,b,2002}}$$

Table A.3 in the appendix shows the correlation matrix between the 2002 NCL share at the zip code level to various zip code level attributes. The strongest correlation is with the deposit to mortgage origination ratio. Zip codes with a stronger presence of high NCL lenders are “deposit-poor,” in the sense that the supply of local retail deposits is low compared to the demand for mortgage originations. Consequently these areas are more likely to rely on lenders with non-core deposit sources of financing. Areas with high NCL lenders also tend to have less elastic housing supply.

Areas with high NCL exposure tend to have lower income, lower homeownership, lower credit scores, and a younger population. The fact that high NCL lenders have large market share in zip codes with a lower fraction of individuals over the age of 65 is consistent with Becker (2007), who shows that seniors tend to save via deposits in local banks. Older zip codes are therefore “deposit-rich,” and are less reliant on outside sources of funding. All of these correlations make economic sense: deposit-poor areas are more likely to rely on mortgages originated by lenders that rely on external funding. In this sense, we do not want to control for these factors; they are the underlying source of variation in exposure to high NCL lenders.

The variation across geographical areas in the 2002 NCL share is used to test whether the lender level first stage shown above passes through to the geographic level. The first stage at the geographic level may not be strong if high NCL lenders simply captured existing market share from low NCL lenders due to their easier access to the PLS market. Similarly, a first stage at the geographic level would not be observed if lenders could costlessly start lending in new areas and there was no segmentation in lending markets. However, there is evidence of such segmentation. For example, the dependence of a geographical area on high NCL lenders is highly persistent. As shown in the first row of Table A.3 in the appendix, the correlation between the 1998 NCL share of the zip code and the 2002 NCL share of the zip code is 0.84 with a small standard error.

In order to test for the first stage at the geographical level, we first estimate specifications

at the MSA by month level. We focus on the MSA level initially because of a new data set available from the Federal Reserve which tracks HMDA mortgage originations at the county by month level, which can easily be aggregated to the MSA by month level. It is not possible at this time to obtain zip code by month level data. The use of monthly HMDA data allows for the precise identification of the timing when high NCL share MSAs experienced stronger mortgage origination growth. The exact specification is:

$$\ln(y_{m,t}) = \alpha_m + \gamma_t + \sum_{k \neq 2002m1} \mathbb{1}_{t=k} \beta_k NCLShare_{m,2002} + \varepsilon_{m,t} \quad (4)$$

The dependent variable is log of mortgage originations in MSA  $m$  in month  $t$ . The coefficients  $\beta_k$  trace the relative growth of originated amounts in MSAs with a high NCL share as of 2002. One issue with monthly data is seasonality in the home purchase market that varies from region to region (e.g., depending on weather). We address seasonality by including a month of year fixed effect from the mortgage origination data for each MSA.<sup>13</sup>

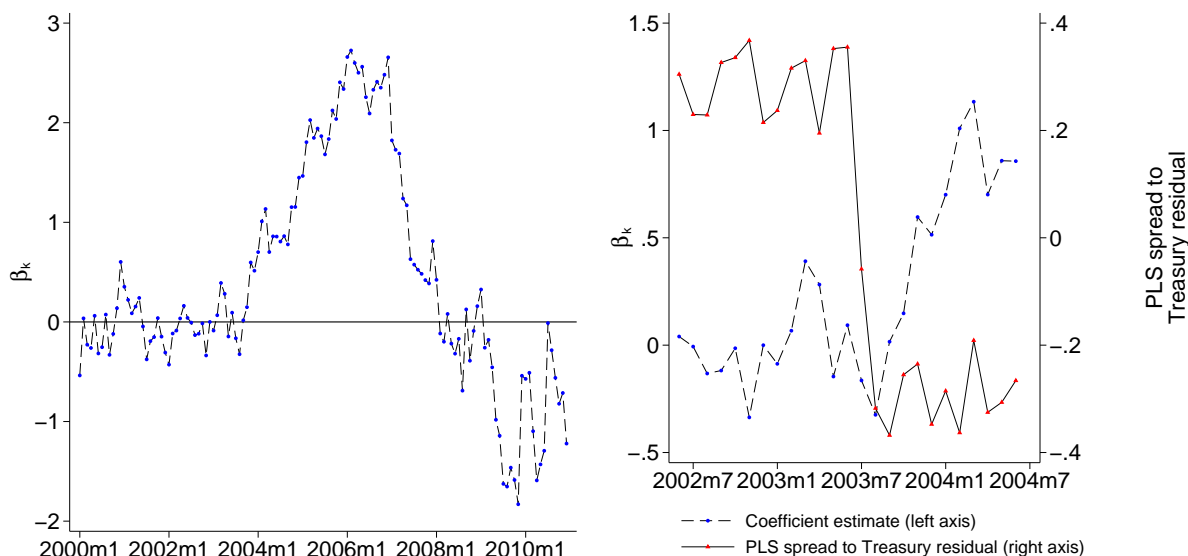
Figure 6a shows the estimated  $\beta_k$  for home purchase mortgage originations, and Figure 6b does the same for refinancing mortgage originations. Both figures show a strong first stage, with the estimated coefficient rising sharply in September and October of 2003. Moreover, there is no pre-trend prior to this period in the estimated coefficient. The magnitude of the effect is large. Housing markets in the top quartile of the 2002 NCL share distribution experienced home purchase mortgage origination that was 23.0 percentage points higher than mortgage origination in the bottom quartile at the peak of the cycle.

The right panel of both figures zooms in on 2003 and also includes the PLS mortgage spread to Treasury rate residual from Justiniano et al. (2017) shown earlier in Figure 1. The relative rise in amount originated in high NCL share MSAs starts at almost the exact same time as the PLS spread drops. Our interpretation of this pattern is that the acceleration

---

<sup>13</sup>The monthly series have been seasonally adjusted by using the residuals  $\ln(y_{m,t})$  of the following regression:  $\ln(y_{m,t}) = \gamma_t + \sum_{k=January}^{December} \mathbb{1}_{t=k} NCL_{m,2002}$ . The specifications with no seasonal adjustment are reported in Figure A4 in the appendix.

Figure 6a: PLS Driven Credit Supply Effect on MSA level: Monthly Frequency, Home Purchases



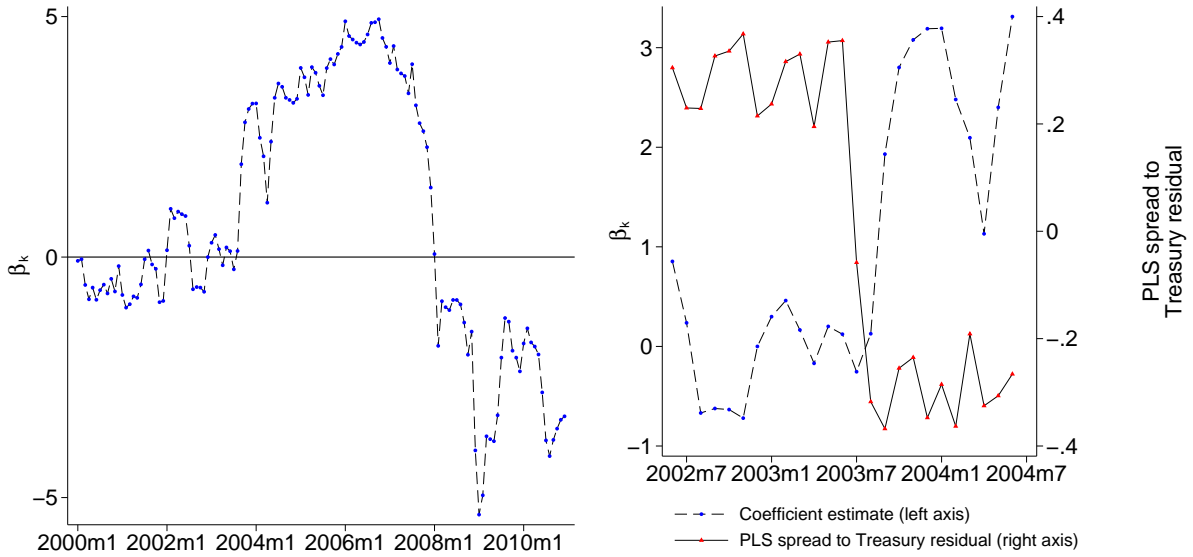
The left panels plots  $\{\beta_k\}$  from equation 4 for home purchase mortgage originations. The monthly series have been seasonally adjusted. Regressions are weighted by the share of total occupied housing units in MSA  $m$  in 2000. Standard errors are clustered by MSA.

of the PLS market lowered mortgage interest spreads and led to a sudden relative rise in originations in high NCL share MSAs. The high frequency analysis supports the view that high NCL share MSAs experienced a sudden rise in originations because of the acceleration of the PLS market; it is unlikely that income prospects or housing market optimism increased by more in high NCL share MSAs suddenly in August, September, and October of 2003.

Given that the analysis in the rest of the study focuses primarily on variation at the zip code level as opposed to MSA level, Figure A5 in the appendix shows estimates from the zip code by year level version of the specification in equation 4. As mentioned above, zip code level data cannot be measured in HMDA at a higher frequency than yearly. Nonetheless, the results using the zip code by year level data set are similar. There was no differential pre-trend in mortgage originations for high NCL share zip codes, and then a sudden rise in 2003 followed by a collapse starting in 2007.

Table 2 further tests for pre-trends by regressing the change in various additional variables

Figure 6b: PLS Credit Supply Effect on MSA level: Monthly Frequency, Refinancing



The left panels plots  $\{\beta_k\}$  from equation 4 for refinancing originations. The monthly series have been seasonally adjusted. Regressions are weighted by the share of total occupied housing units in MSA  $m$  in 2000. Standard errors are clustered by MSA

of interest on a local housing market's exposure to high NCL lenders in 2002.<sup>14</sup> Columns 1 and 2 show that there was no differential change in housing market sentiment between 2000 and 2002. The dependent variable in column 1 is the overall change in optimism about the housing market, where optimism is measured as the percentage of respondents in the Michigan Survey of Consumer who think it is a favorable time to buy a house.<sup>15</sup> Column 2 uses the change in the percentage of respondents who say that it is a good time to buy a house because of price considerations. Both columns show that there is no evidence of a change in housing market optimism prior to the sudden expansion of mortgage lending in 2003. The estimated coefficients also have small standard errors. Columns 3 through 6 show that there is no differential growth in house prices or construction activity in high NCL areas prior to 2003.

The first stage analysis presented here focuses on the amount of mortgages originated.

<sup>14</sup>The sources of the data sets used in Table 2 are described in more detail in subsequent sections.

<sup>15</sup>The construction of the dependent variable in columns 1 and 2 are described in more detail below in Section 4.

Table 2: NCL Share and Pre-Trends

	Good time to buy		Ln(House Prices)		Ln(Construction)	
	$\Delta_{00,02}$ (1)	$\Delta_{00,02}$ due to Prices (2)	$\Delta_{00,02}$ (3)	$\Delta_{00,02}$ (4)	$\Delta_{00,02}$ (5)	$\Delta_{00,02}$ (6)
2002 NCL Share	0.083 (0.313)	-0.143 (0.230)	0.221 (0.140)	0.079 (0.117)	-0.323 (0.413)	0.022 (0.409)
Elasticity				-0.050*** (0.006)		0.045** (0.014)
Level of Observation	MSA	MSA	Zip	Zip	MSA	MSA
N	338	338	5,540	5,540	259	259
R-sq	0.000	0.001	0.010	0.268	0.003	0.049

Standard errors are clustered at the MSA level for zip-level specifications.

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

However, there is also evidence that mortgage contract terms were affected by the rise of the PLS market. Motivated by the empirical analysis in this study, [Dokko et al. \(2019\)](#) show that high NCL share counties witnessed a sudden relative increase in the use of non-traditional mortgage features such as variable interest rates or interest only payments exactly in the late summer and early fall of 2003.<sup>16</sup>

In summary, there are a number of findings that support the exclusion restriction assumption that the rise of the PLS market led to a relative credit supply expansion toward high NCL share areas that was independent of spurious shifts in household demand. First, the lender-level regressions can be saturated by census-tract fixed effects to control for demand side factors in the spirit of [Khwaja and Mian \(2008\)](#), and high NCL ratio lenders expand mortgage lending significantly more even with these detailed controls for demand. Second, there was no evidence of differential pre-trends in mortgage originations, housing market optimism, house price growth, or construction prior to 2002 in high NCL share areas. Third, the relative increase in mortgage originations in high NCL share MSAs is perfectly timed with the aggregate rise in the PLS market in August, September, and October of 2003.

<sup>16</sup>The fact that high NCL share counties experienced a sharp rise in the use of mortgages with non-traditional features is further support for the idea that mortgages originated in the PLS market facilitated speculation in housing markets, e.g., [Barlevy and Fisher \(2018\)](#).

With this evidence in hand, the next three sections explore how this mortgage credit supply expansion affected speculative trading, house prices, construction, and housing market optimism in areas most exposed to high NCL ratio lenders.

## 2 Speculative Trading Activity

### 2.1 Data and measurement of speculation

An advantage of the data set used in this study is that it allows for the measurement of the marginal buyers brought into the market by the PLS-driven expansion in mortgage credit supply. Such measures are calculated based on credit data provided by TransUnion, a global information solutions company, through a relationship with the Kilts Center for Marketing at The University of Chicago Booth School of Business

The underlying TransUnion data set is at the individual-account level, and is available from June 2000 through the end of the sample period in 2010. The data set records the outstanding balance and delinquency status for each credit account that an individual has every month. The available TransUnion data set is a 10% random sample of the universe, where individuals are in the universe if they appear in the data at any point between June 2000 and 2016. This study uses a 5% random sample of the 10% sample for computational reasons, giving us a 0.5% sample of the universe.

We use the individual-account data set to construct a mortgage origination data set similar to HMDA. However, the advantage of the TransUnion data is that it includes important attributes of the borrower taking on the mortgage, such as her credit score, age, credit market behavioral patterns, and delinquency status. The TransUnion data do not contain any explicit flag for whether a mortgage is a refinancing, a home purchase loan, or a first-lien. In the appendix, we describe how we classify mortgages into these categories, and we provide evidence that our methodology produces aggregate statistics in line with HMDA.<sup>17</sup>

---

<sup>17</sup>See Section A.2 and Figure A6 in the appendix for details. We are happy to share this code with any



The appendix also describes how zip codes of the house being purchased are assigned to an individual in the TransUnion data. As shown below, specifications that can be estimated with either the HMDA or TransUnion data produce similar results.

Summary statistics for the TransUnion mortgage origination level data are shown in Table A1 in the appendix. The average first-lien purchase mortgage in the TransUnion data set is \$188 thousand. The average VantageScore of an individual taking out a first-lien mortgage between 2001 and 2010 is 692, which is in the prime category. The average age at origination is 42 years, and 24% of individuals that take out a first-lien mortgage default on a mortgage at some point in 2006 or after.

The TransUnion data set allows us to construct measures of speculation. We measure speculators in the TransUnion data in three ways. First, a mortgage origination is classified as being taken out by a speculator if the individual taking out the mortgage in question also takes out another distinct first-lien purchase mortgage in a two year period around the origination in question. We refer to this as the “multiple houses” categorization of a speculator.

Second, a given first-lien purchase origination is classified as being taken out by a speculator if the first-lien purchase mortgage is subsequently closed within a year, and there is no associated refinancing for the individual in the six months after the purchase mortgage is closed. We refer to such an individual as a “short-term” trader, where we are making the assumption that the closed mortgage reflects a sale. Third, a given first-lien purchase origination is classified as being taken out as a speculator if the individual taking out the mortgage already has at least two existing first-lien mortgages on his balance sheet at the time of the new origination. We refer to such an individual as a “2+ mortgage” speculator.

How common are speculators in the data? Just over 4% of individuals in the TransUnion data as of 2005 or 2006 obtain a new first lien purchase mortgage. For the “multiple houses,” “short-term,” and “2+ mortgage” categories, 0.5%, 0.7%, and 0.6% of all individuals in the

---

other researchers that have permission to use the TransUnion data set at the Kilts Center at Chicago Booth.

TransUnion data in 2005 or 2006 are classified as speculators, respectively. If a speculator is defined as an individual having any of these attributes, speculators make up 1.3% of individuals in the TransUnion data in 2005 or 2006. By any measure, speculators represent a small fraction of the overall population, a point to which we return to below in Section 4.

These three definitions of speculation are not mutually exclusive, and indeed they are highly correlated. If we condition the sample to individuals who take out a first lien mortgage in 2005 or 2006, there is a 0.37 correlation across individuals for the 2+ mortgage and multiple houses definitions, a 0.29 correlation for the short-term and multiple houses definitions, and a 0.15 correlation between the short-term and 2+ mortgage definitions. All of these correlations are significant at the 1% confidence level.

This section also uses a measure of total housing transaction volume, which comes from CoreLogic, a private vendor that collects and standardizes publicly available tax assessments and deed records from municipalities across the United States. Using the original CoreLogic files, we replicate the DeFusco et al. (2018) version of this data set, which is filtered to get the most accurate measure of volume at the zip code level.<sup>18</sup> In robustness tests, we use the measure of speculation from DeFusco et al. (2018), which is based on how frequently a given property is sold.

## 2.2 The effect on transaction volume

Cross-sectional variation across zip codes in the 2002 NCL share is used to estimate the effect of PLS-driven credit expansion on trading volume. Figure 7 plots  $\beta_k$  from the regression,

$$\ln(y_{z,t}) = \alpha_z + \gamma_t + \sum_{k \neq 2002} \mathbb{1}_{t=k} \beta_k NCLShare_{z,2002} + \varepsilon_{z,t} \quad (5)$$

---

<sup>18</sup>We are extremely grateful to the authors for sharing the zip-year level version of their total volume variable, which we were able to successfully replicate using the original CoreLogic files. Please see the appendix of DeFusco et al. (2018) for more information on the data construction.

where  $y_{z,t}$  is the number of first lien mortgage originations for home purchase in zip code  $z$  in year  $t$  in the left panel, and the number of housing transactions in the right panel.<sup>19</sup>

Both panels in Figure 7 show a large relative increase in first-lien mortgage growth and the volume of housing transactions from 2003 to 2006 in high NCL share zip codes. The relative increase began exactly when the PLS market accelerated in 2003. The magnitude of the impact on volume is quite large. Going from a zip code in the bottom quartile of 2002 NCL share distribution to a zip code at the top quartile leads to an increase in transaction volume of 19.1%.

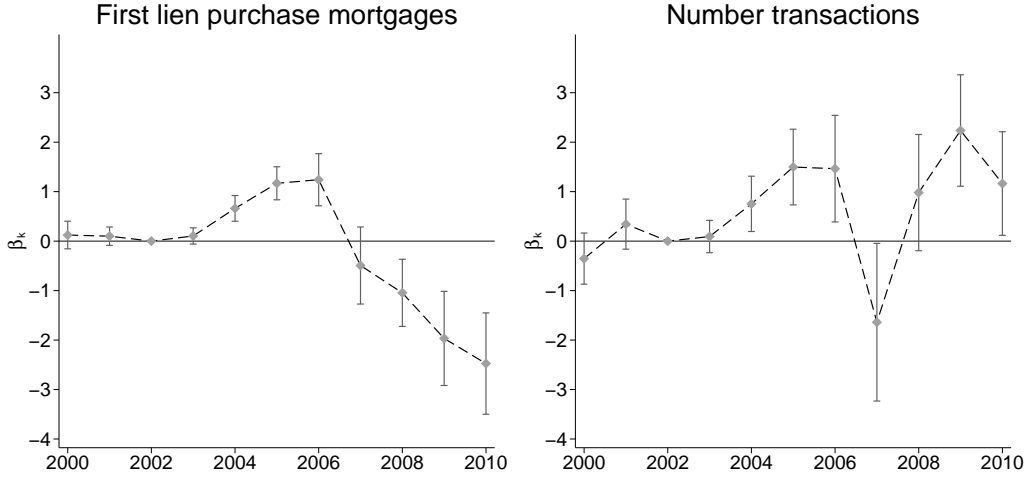
Moreover, the increase in the number of first-lien purchase mortgages was almost identical to the increase in housing transaction volume in terms of magnitude. This implies that the relative rise in transaction volume in high NCL share zip codes was driven entirely by a rise in transactions using a first-lien purchase mortgage, highlighting the importance of the credit supply expansion in driving the rise in volume.

Both the number of first-lien purchase mortgage originations and volume collapsed in 2007 in high NCL share zip codes. In 2009 and 2010, volume again rose in high NCL share zip codes, but the number of first-lien mortgages continued to collapse. This suggests that cash buyers moved into the market after the crash. The results are in-line with the “leverage cycles” of [Geanakoplos \(2010\)](#). Growth in housing transactions in high NCL areas was driven by mortgage financing during the boom but a larger share of transactions in these areas were financed with cash during the bust.

---

<sup>19</sup>First lien mortgages can be measured in both the HMDA and the TransUnion data. HMDA splits home purchase mortgages into first- and second-lien mortgages beginning in 2004. For the years prior to 2004, we use data from [Bhutta and Keys \(2018\)](#) that split first- and second-liens based on a methodology explained in their study. We are extremely grateful to Neil Bhutta for sharing with us his data. In TransUnion, first-lien home purchase mortgages capture a transaction financed with a mortgage, and does not include second-lien “piggy-back” mortgages. See appendix for more details on data construction and comparison of the HMDA and Transunion measure of first-lien mortgages.

Figure 7: The Effect of PLS Credit Supply Expansion on Transaction Volume



The figure plots the coefficients  $\{\beta_k\}$  from equation 5. The dependent variable in the left panel is the natural logarithm of first-lien home purchase mortgages and the dependent variable in the right panel is housing transaction volume. The regressions are weighted by the share of total occupied housing units in zip code  $z$  in 2000. 95% confidence intervals are from clustered standard errors at the MSA level.

### 2.3 Speculators as marginal buyers

The mortgage transaction level TransUnion data set allows us to understand the marginal buyers brought into the housing market by the acceleration of the PLS market. Formally, let  $y_{z,t}$  represent the number of first-lien mortgage originations for home purchase in zip code  $z$  at time  $t$ . As shown above,  $y_{z,t}$  closely tracks the growth in overall housing transactions during the housing boom period. For any partition of the population, e.g. partitioning the population into high and low credit score individuals, the total change in  $y_{z,t}$  between years  $t - h$  and  $t$  can be decomposed as,

$$\frac{\Delta y_{z,t}}{y_{z,t-h}} = \sum_i \frac{\Delta y_{z,t}^i}{y_{z,t-h}}$$

where  $\Delta y_{z,t} = y_{z,t} - y_{z,t-h}$ . Each term on the right hand side of the above expression represents the contribution to growth coming from each sub-group  $i$  of the partition. This decomposition can be used to estimate the contribution of each sub-group  $i$  to the total

increase in origination volume driven by the expansion in the PLS market.

Table 3 presents estimates from the following specification for first-lien purchase mortgages,

$$\frac{y_{z,0506}^i - y_{z,0102}^i}{y_{z,0102}} = \alpha + \beta^i NCLShare_{z,2002} + \varepsilon_z \quad (6)$$

where 0102 and 0506 represent the sum of  $y$  from 2001 to 2002 and 2005 to 2006, respectively.<sup>20</sup> Columns 1 and 2 in panel A report specifications where  $i$  reflects the universe of first lien mortgages in the HMDA data and the TransUnion data, respectively. For the specifications reported in these columns, the left hand side variable is just the percentage change in first-lien mortgages in zip code  $z$  during the boom. A zero to one change in the 2002 NCL share is associated with a 176 percentage point increase in first-lien mortgage originations in the HMDA data set, and 168 percentage point increase in the TransUnion data set. The estimates are similar across the two types of data sets, which gives us confidence that the measurement of first-lien purchase mortgages in the TransUnion data set is similar to the HMDA data.<sup>21</sup> The TransUnion data estimate implies that housing markets in the top quartile of NCL exposure experience an increase in mortgage-financed home purchases that is 23.0 percentage points higher than mortgage-financed purchases in the bottom quartile at the peak of the cycle.

Columns 3 through 5 of panel A examine how much of this total effect is coming from speculators. A substantial fraction of the overall rise in volume is driven by speculators, with the largest fraction of the overall effect coming from speculators according to the multiple houses definition. The point estimates suggest that between 30 and 70% of the relative rise in volume in high NCL share zip codes was due to speculators. The estimation in column 6 classifies a first-lien mortgage as being taken out by a speculator if the individual taking out

---

<sup>20</sup>The TransUnion data set begins in June 2000, which means 2001 is the first full year available to calculate  $y_{z,t}$ .

<sup>21</sup>Appendix Figure A7 shows the entire time series of the NCL effect for TransUnion and HMDA data sets; the results are similar across the two data sets.

Table 3: Who Are the Marginal Buyers?

*Panel A: Total and by speculation measures*

	$\Delta$ first-lien, Total		$\Delta$ first-lien, Speculators			
	HMDA (1)	TransUnion (2)	Multiple Houses (3)	Short Term (4)	2+ mortgages (5)	Union (6)
2002 NCL Share	1.763*** (0.268)	1.680*** (0.374)	1.026*** (0.219)	0.817*** (0.118)	0.473*** (0.122)	1.643*** (0.244)
N	9,020	9,023	9,023	9,023	9,023	9,023
R-sq	0.033	0.004	0.007	0.010	0.003	0.011

*Panel B: By measures of risk*

	$\Delta$ first-lien, By ex ante risk			
	No Score (1)	Subprime (2)	Near Prime (3)	Prime (4)
2002 NCL Share	0.604*** (0.109)	0.872*** (0.097)	0.263* (0.118)	-0.059 (0.196)
R-sq	0.005	0.010	0.001	0.000

*Panel C: By measures of risk and speculation*

	$\Delta$ first-lien, Speculators		$\Delta$ first-lien, Non-speculators	
	Non-prime (1)	Prime (2)	Non-prime (3)	Prime (4)
2002 NCL Share	1.239*** (0.168)	0.405** (0.130)	0.500*** (0.151)	-0.464*** (0.127)
R-sq	0.015	0.001	0.002	0.002

*Panel D: By borrower age at origination*

	$\Delta$ first-lien, Borrower Age at Origination				
	Missing (1)	Under 30 (2)	30-40 (3)	41-50 (4)	Over 50 (5)
2002 NCL Share	0.063 (0.052)	0.187 (0.136)	0.948*** (0.154)	0.407** (0.138)	0.075 (0.128)
R-sq	0.000	0.000	0.006	0.001	0.000

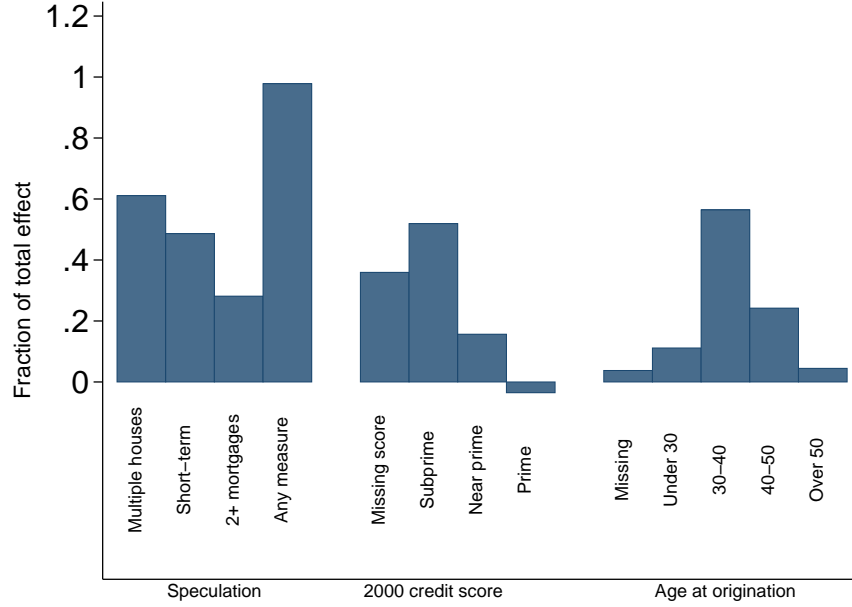
Standard errors clustered at MSA level.

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

the mortgage fits any of the three definitions. As the estimate shows, under the broadest definition of a speculator, the entire rise in transaction volume in high NCL share zip codes is driven by speculative trading activity.

Figure 8 presents the share of the total relative growth in first-lien mortgages in high NCL share zip codes during the boom for different sub-groups. The bars are the estimated

Figure 8: Decomposing Growth in First-lien Purchase Mortgages in High NCL Zip Codes



This figure plots the share of the relative growth in first-lien purchase mortgage originations in high NCL share zip codes by certain groups. It is constructed by first estimating  $\beta^i$  from the following specification:  $y_{z,m,BOOM}^i = \alpha_m + \beta^i NCL_{z,m,2002} + \varepsilon_{z,m}$  from zip code  $z$  and MSA  $m$ . The variable  $y_{z,m,BOOM}^i = \frac{firstlienmortgages_{z,m,0506}^i - firstlienmortgages_{z,m,0102}^i}{firstlienmortgages_{z,m,0102}^i}$  where  $firstlienmortgages^i$  is a subset of first-lien mortgages such as first lien mortgages taken out by speculators. For each group  $i$ , the coefficient  $\beta^i$  is divided by the total relative effect  $\beta$  estimated from:  $\frac{firstlienmortgages_{z,m,0506} - firstlienmortgages_{z,m,0102}}{firstlienmortgages_{z,m,0102}} = \alpha_m + \beta NCL_{z,m,2002} + \varepsilon_{z,m}$ . Plotted above is  $\frac{\beta^i}{\beta}$ , which is the share of the relative growth coming from group  $i$ .

$\beta^i$  for each group  $i$  from equation 6 scaled by the total effect of 1.680 reported in column 2 of Panel A of Table 3. As it shows, under the broadest definition, speculators accounted for nearly 100% of the relative growth in volume in high NCL share zip codes.

Table 3 also shows that the marginal house buyers brought in by the PLS market were riskier from an ex ante perspective. As Panel B shows, most of the effect comes from individuals with a subprime credit score as of 2000 and individuals with no credit score as of 2000.<sup>22</sup> Individuals with a prime credit score as of 2000 in high NCL share zip codes did not witness stronger growth in first-lien mortgage originations relative to individuals with a prime credit score as of 2000 in low NCL share zip codes.

<sup>22</sup>When individuals with no credit score as of 2000 eventually enter the TransUnion data set, the average and median score are both in the near prime category.

In order to learn more about these speculators, Panel C of Table 3 does a double sort of the data based on the broad measure of speculation and the credit score of the individual as of 2000. Non-prime in Panel C includes those with a missing score as of 2000 and those with a near-prime or subprime credit score as of 2000. As Panel C shows, most of the rise in speculative trading activity came from individuals with a non-prime credit score as of 2000. Interestingly, as column 3 shows, the PLS market appears to have brought in non-prime individuals who did not engage in speculative trading activity. Furthermore, column 4 of Panel C indicates that there was a relative decline in house purchases in high NCL share zip codes by individuals with a prime credit score as of 2000 who were not engaging in speculative trading activity. In other words, individuals in high 2002 NCL share zip codes with high credit scores that were not engaging in speculative trading activity experienced a relative decline in house buying relative to their counter-parts in low NCL share zip codes. High credit score traditional home-buyers in zip codes seeing a trading frenzy appear to have been sitting out, a point to which we will return in Section 4.

As Panel D shows, there was a slight tilt toward young individuals in accounting for the relative increase in first-lien purchase mortgage origination growth in high NCL share zip codes. The median age at origination of a first-lien mortgage from 2001 to 2010 was 40, and the results in Panel C indicate that 65% of the relative growth in first-lien purchase originations in high NCL share zip codes was driven by individuals below 40 at origination.

Table A3 in the appendix replicates Table 3 with the inclusion of MSA fixed effects. Such specifications exploit only within MSA variation in the NCL share of a zip code as of 2002 to see the effect of credit supply expansion on the marginal buyers brought into the housing market. As the table in the appendix shows, the coefficient estimate on the 2002 NCL share is similar and even somewhat larger in the specifications with MSA fixed effects. Omitted variables at the MSA level do not appear to be responsible for the coefficient estimates in Table 3.



## 2.4 A property-based measure of speculation

An alternative approach to measuring speculative trading activity is presented in [DeFusco et al. \(2018\)](#). More specifically, using the CoreLogic deed records transactions, one can measure how many times a given property is bought and sold in some short time period. This approach is entirely independent of the TransUnion data, and as a result provides a useful check on the results above. Following [DeFusco et al. \(2018\)](#), we classify each house transaction by whether the house is subsequently sold within one, two, or three years of the purchase. This allows for a decomposition of total transaction volume by whether the house will be quickly “flipped” again.

Table 4: Property-based Measures of Speculation

	$\Delta$ volume, 01-05, sold again within x years			
	Total (1)	Within 1 year (2)	Within 2 years (3)	Within 3 years (4)
2002 NCL Share	2.265** (0.779)	0.475*** (0.132)	0.906*** (0.198)	1.182*** (0.240)
N	3,673	3,673	3,673	3,673
R-sq	0.009	0.041	0.063	0.068

Standard errors clustered at MSA level.

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

This table presents zip-code level regressions of the percentage change in volume from 2001 to 2005 on the 2002 NCL Share. Following [DeFusco et al. \(2018\)](#), we split transactions by whether the property is subsequently sold within a certain number of years. More specifically, the coefficients reported in columns 2 through 4 are  $\beta^i$  from the following specification:  $y_{z,01-05}^i = \alpha + \beta^i NCLShare_{z,2002} + \varepsilon_z$  from zip code  $z$ .

The variable  $y_{z,01-05}^i = \frac{volume_{z,05}^i - volume_{z,01}^i}{volume_{z,01}^i}$  where  $volume^i$  is a subset of all volume based on whether and when the property is subsequently sold.

Table 4 focuses on the increase in volume from 2001 to 2005. Given the manner in which the [DeFusco et al. \(2018\)](#) measure is constructed, the left hand side variable in the regressions represents the increase in the number of transactions of properties that subsequently transacted again within a certain number of years. As in Table 3 above, the specifications follow equation 6 in decomposing the total rise in volume into the share of the rise coming from properties that were transacted again within 1, 2, and 3 years.

As Table 4 shows, 40% of the differential increase in volume for high NCL share zip codes comes from the rise in transactions of houses that will be subsequently sold within two years, and over half of the total effect comes from those that will be subsequently sold within three years. Figure A8 in the appendix shows estimates of equation 5 for total volume and volume of houses subsequently sold again in one, two, or three years. As the figure shows, the coefficients on the 2002 NCL share of the zip code are two to three times larger for transaction volume of houses that are sold again in one or two years relative to all transactions. Property-based measures of speculation confirm that exposure to high NCL lenders led to a credit-driven rise in speculative trading activity in the housing market.

### 3 House prices, Construction, and Defaults

How did the rise in credit supply-driven speculative trading activity affect housing markets more broadly? The left panel of Figure 9 presents estimates of  $\beta_k$  from the estimation of equation 5 with the natural logarithm of house prices as the left hand side variable.<sup>23</sup> This estimation is conducted using the zip by year level data set. As the estimates show, high NCL share zip codes experience positive relative growth in house prices in 2003, which then accelerates rapidly in 2004, 2005, and 2006. Going from a zip code in the bottom quartile of NCL exposure to a zip code in the top quartile of NCL exposure leads to a 12.1 percentage point increase in house prices.

The PLS market collapsed in 2007, which corresponds to a collapse in house prices in high NCL share zip codes. In fact, the collapse was severe enough that the log house price level ended up lower in 2009 and 2010 than its 2002 level relative to low NCL share zip codes.

The right panel of Figure 9 presents estimates of  $\beta_k$  using the natural logarithm of new housing units as the left hand side variable. This specification is estimated at the MSA by year level, given that new housing permit data is only available from the Census at the

---

<sup>23</sup>House price data used in Figure 9 comes from CoreLogic and is recorded at the zip code level. Housing units data comes from the Census Building Permits Survey and is recorded at the county level, which we aggregate up to the MSA level. Housing supply elasticity is measured as in Saiz (2010).

Figure 9: The Effect of PLS Credit Supply Expansion on House Prices and Construction



The figure plots the coefficients  $\{\beta_k\}$  from equation 5. The left panel uses the natural logarithm of house prices in a zip code and the right panel uses the natural logarithm of new housing permits in an MSA as the dependent variable. The regressions are weighted by the share of total occupied housing units in zip code  $z$  or MSA  $m$  in 2000. Standard errors are clustered at the MSA level.

county level. It shows a similar pattern: high 2002 NCL share MSAs witnessed a substantial boom and bust in construction. As with house prices, construction during the bust fell even below the 2002 level relative to low 2002 NCL share MSAs.

The evidence is supportive of the argument in Glaeser et al. (2008). They argue that in the presence of speculators, the boom in house prices will lead to excess speculative investment in construction of new homes during the boom. Consequently when prices collapse, they will “over shoot” by going below their pre-boom levels. This is needed to absorb the excess inventory built up during the boom.

Table 5 presents regression estimates corresponding to Figure 9. The boom and bust in house prices and construction activity was significantly pronounced in high NCL share areas. The third and sixth column of Table 5 test the long run “over-correction” hypothesis put forth in Glaeser et al. (2008). While the statistical power is weaker, there does appear to have been lower growth in both house prices and construction activity from 2002 to 2010 in high NCL share areas. Not only did house prices and construction activity collapse in high

NCL share areas from 2006 to 2010, but the collapse was large enough to bring the relative level below where they began in 2002.<sup>24</sup>

Table 5: NCL Share, House Prices, and Construction

	Ln(House Prices)			Ln(New Construction)		
	02 to 06 (1)	06 to 10 (2)	02 to 10 (3)	02 to 06 (4)	06 to 10 (5)	02 to 10 (6)
2002 NCL Share	0.946** (0.326)	-1.620*** (0.315)	-0.674* (0.341)	1.245** (0.462)	-2.595** (0.968)	-1.350 (1.103)
Elasticity	-0.122*** (0.017)	0.081*** (0.016)	-0.041* (0.019)	-0.001 (0.019)	0.119** (0.037)	0.119** (0.043)
Level	Zip	Zip	Zip	MSA	MSA	MSA
N	5,540	5,540	5,540	259	259	259
R-sq	0.345	0.231	0.051	0.036	0.145	0.089

Standard errors clustered at MSA level for zip-level specifications.

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

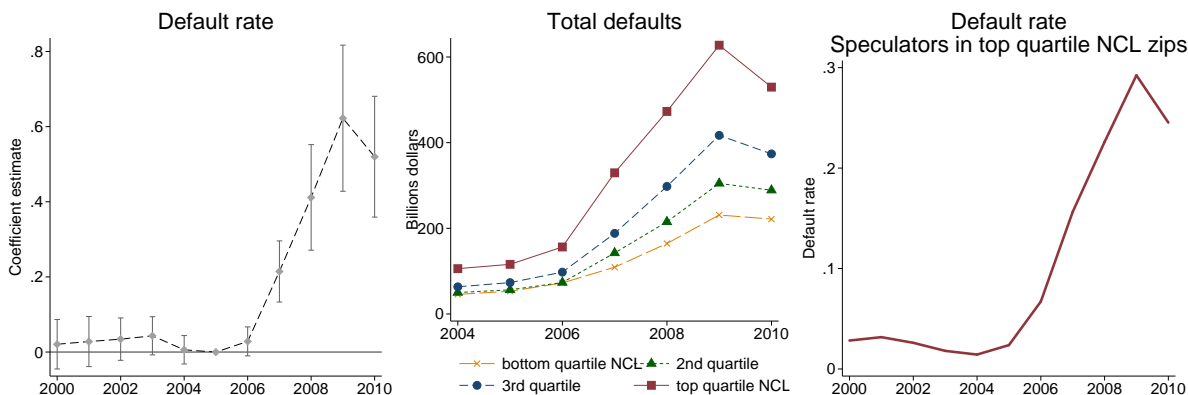
As a side note, it is worth mentioning that credit expansion fueled by the PLS market boosted house prices even in cities such as Las Vegas and Phoenix that experienced substantial new construction during the boom. The phenomenon of high house price growth in such cities has puzzled researchers given that in standard models the ability to cheaply construct more housing units should put a lid on house price growth (e.g. [Davidoff \(2013\)](#), [Nathanson and Zwick \(2017\)](#)). There is a related body of research focusing on anomalous elastic housing supply MSAs with both a boom in construction and house prices.<sup>25</sup> [Nathanson and Zwick \(2017\)](#) point to the importance of supply-side speculation and [Chinco and Mayer \(2015\)](#) point to the importance of out-of-town investors. The results presented here suggest that the expansion of the PLS market also played an important role.

The collapse of the housing market from 2006 to 2010 led to a dramatic rise in mortgage default rates in high NCL share zip codes. The left panel of [Figure 10](#) plots  $\beta_k$  from the estimation of [equation 5](#) with the mortgage default rate of the zip code as the outcome

<sup>24</sup>An important caveat is that the negative effect of defaults and foreclosures on house prices during the bust may bring prices even below long-run fundamental value. This force is not contained in the [Glaeser et al. \(2008\)](#) model.

<sup>25</sup>See, e.g., [Davidoff \(2013\)](#), and [Gao et al. \(2017\)](#).

Figure 10: Mortgage Defaults by NCL Share



The left panel plots the coefficients  $\{\beta_k\}$  from equation 5 with the mortgage default rate in a zip code as the left hand side variable. The regression is weighted by the share of total occupied housing units in zip code  $z$ , and standard errors are clustered at the MSA level. The middle panel plots total delinquent mortgage debt in zip codes by  $NCLShare_{z,2002}$  quartiles. The quartiles are weighted by total mortgage debt outstanding as of 2006. The right panel plots the mortgage default rate for individuals that speculated in 2005 and 2006 in the top quartile NCL zip codes.

variable. As early as year end 2006, there was an uptick in defaults in high NCL share zip codes that is marginally statistically distinct from zero. By the end of 2007, there was a large relative increase in mortgage default rates in high NCL share zip codes that continued to rise until 2010.

The middle panel shows total delinquent mortgage debt by zip code quartiles of the 2002 NCL share ratio. From 2006 to 2007, total mortgage delinquent debt rose substantially more in high NCL share zip codes relative to the rest of the country. As a share of total delinquencies, the top quartile zip codes by the 2002 NCL share increased by almost 5 percentage points from 2006 to 2007. By 2008 and 2009, the mortgage default crisis spread to the rest of the country (e.g., [Ferreira and Gyourko \(2015\)](#)), but Figure 10 shows that the earliest stage of the mortgage default crisis was driven by defaults in zip codes most exposed to the PLS market. This is consistent with widespread evidence that the mortgage default crisis was initiated with defaults in the PLS market.<sup>26</sup>

The right panel of Figure 10 examines the default rate of individuals classified as specu-

<sup>26</sup>See for example Section 5 of [Mian and Sufi \(2017b\)](#).

lators in 2005 or 2006 according to any of the three definitions above. The sample is limited to the top quartile zip codes by the 2002 NCL share. The default rate of speculators in high NCL share zip codes increased substantially from 2005 to 2006, and reached nearly 20% in 2007. This is another piece of evidence supporting the importance of speculation driven by the PLS market in instigating the mortgage default crisis.

## 4 Heterogeneity in Beliefs

It is difficult to discuss speculative trading activity without discussing house price growth expectations. Indeed, models in which credit availability affects asset prices often assume heterogeneity in beliefs about future asset values (e.g., [Geanakoplos \(2010\)](#) and [Simsek \(2013\)](#)). There is also some evidence of this type of heterogeneity in [Table 3](#); the results in that table suggest that individuals with prime credit scores not engaging in speculative trading actually reduced purchases in high NCL share zip codes.

We are also motivated to examine evidence on belief heterogeneity given the fact that only a small part of the overall population engaged in speculative trading activity during the housing boom of 2005 and 2006. As mentioned above, even under the most broad definition of a speculator, only 1.3% of all individuals in TransUnion in 2005 and 2006 engaged in speculative trading activity in the housing market. A small group of individuals had a large effect on the overall housing market, which suggests that an overall rise in optimism about the housing market may not have been necessary to generate a boom in prices and volume.

### 4.1 Data on beliefs

Unfortunately, there is no information on house price expectations in the data sets used above. Furthermore, there is limited individual level data on house price beliefs during the housing boom from 2002 to 2006 that can be linked to measures of speculation at the

individual level.<sup>27</sup> As a result, the analysis in this section is conducted at the aggregate and MSA level.

Two data sets are used to measure beliefs. First, following [Piazzesi and Schneider \(2009\)](#), the Michigan Survey of Consumers is used to measure beliefs about the housing market. We use county identifiers in the individual level Michigan survey data to construct housing market belief data at the MSA-year level. We measure average housing market beliefs using questions such as the fraction of respondents saying whether it is good time to buy a house, and for what reason. See Section [A.3](#) in the appendix for more details.

Second, data on house price expectations of recent home-buyers come from [Case et al. \(2012\)](#), who survey individuals who have bought a house in the preceding year.<sup>28</sup> The main advantage of this data set is that it provides house price expectations of the marginal trader. That is, it measures the expectations of the individuals that have recently bought a house. It therefore allows a comparison of the expectations of the marginal buyer with average expectations taken from the Michigan survey. One important limitation of the [Case et al. \(2012\)](#) data is that it only covers four metropolitan areas that the authors considered representative of the broader U.S. market. Summary statistics for all of these data sets are shown in Table [A1](#) in the appendix.

## 4.2 Evolution of beliefs during the housing boom

The aggregate evidence on housing market optimism from the Michigan survey is consistent with the importance of belief heterogeneity during the housing boom. As the left panel of Figure [11](#) shows, during the heart of the PLS acceleration period of 2003 to 2006, the fraction of individuals saying that now is a bad time to buy a house rose from 15% to 35%. This finding replicates the analysis in [Piazzesi and Schneider \(2009\)](#), which is difficult to reconcile with the view that widespread optimism about housing among the general population was

---

<sup>27</sup>There has been a substantial increase in data availability on house price expectations since 2009 (e.g., [Bailey et al. \(2016\)](#), [Fuster et al. \(2018\)](#), and [Adelino et al. \(2018\)](#)) but almost all of these new data sets cover information only after the 2002 to 2006 housing boom.

<sup>28</sup>We are grateful to Anne Thompson for sharing these data with us.

responsible for the rise in house prices from 2003 to 2006.<sup>29</sup>

The right panel uses information from the Michigan survey covering why an individual believes it is either a good or bad time to buy a house. In particular, individuals can specifically think it is a good time to buy a house because of positive house price growth expectations, or they can also specifically think it is a bad time to buy a house because of negative house price growth expectations.<sup>30</sup> Even though the fraction of the population saying that it is a good time to buy a house is falling (which is one minus the fraction saying it is a bad time in the left panel), the fraction of the population saying it is a good time to buy a house because of positive price expectations is actually rising (right panel). The right panel shows that there is a rise in both the fraction of the population saying it is a good time to buy because of positive price expectations and the fraction saying it is a bad time to buy because of negative price expectations, which is also consistent with the findings of [Piazzesi and Schneider \(2009\)](#).

The same heterogeneity in beliefs can be seen when using the [Case et al. \(2012\)](#) data set, as is done in [Figure 12](#). As the left panel shows, from 2003 to 2006, the house price expectations of individuals who recently bought a home increased. However, the fraction of the population saying it is a bad time to buy a home because of negative price expectations also rose. The marginal home-buyers in the market had high house price growth expectations, while the average individuals in the economy were becoming more pessimistic about the evolution of house prices. This suggests that heterogeneity in beliefs was an important part of the housing boom.

The right panel of [Figure 12](#) uses MSA-by-year level variation in both the Michigan survey and the [Case et al. \(2012\)](#) data set.<sup>31</sup> It plots the house price growth expectations of

---

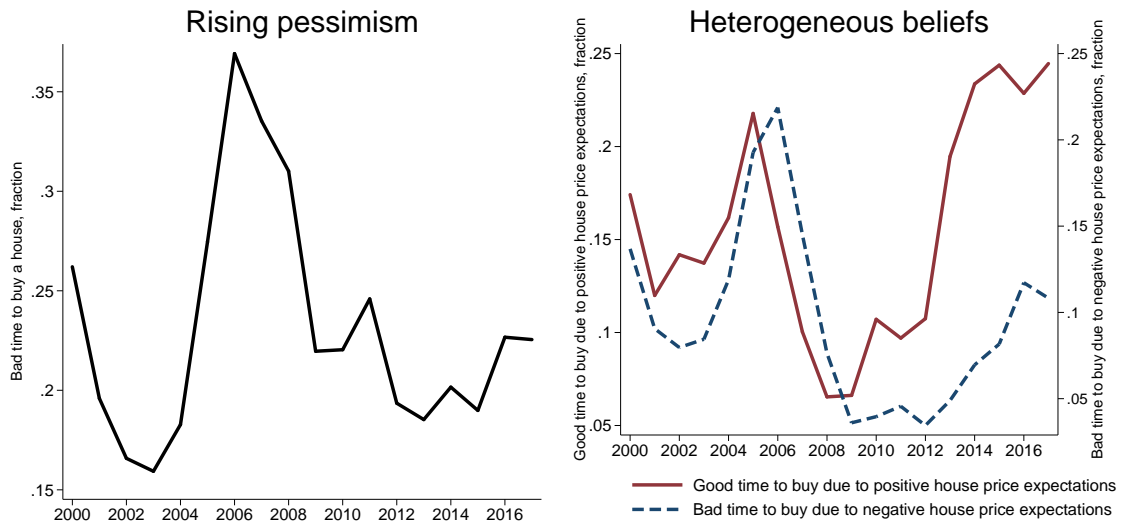
<sup>29</sup>This does not dispute the observation in the literature that a change in expectations by *lenders* was an important part of the lending boom, see, e.g., [Gerardi et al. \(2008\)](#), [Gennaioli et al. \(2012\)](#), [Landvoigt \(2016\)](#), [Kaplan et al. \(2017\)](#). For example, an overall neglect of downside risks as in [Gennaioli et al. \(2012\)](#) could explain the rise in asset-backed securitization across many markets from 2003 to 2006.

<sup>30</sup>Please see [Section A.3](#) in the appendix for more details on how answers to the Michigan survey questions are classified.

<sup>31</sup>Recall that there are only four MSAs in the [Case et al. \(2012\)](#) data set.

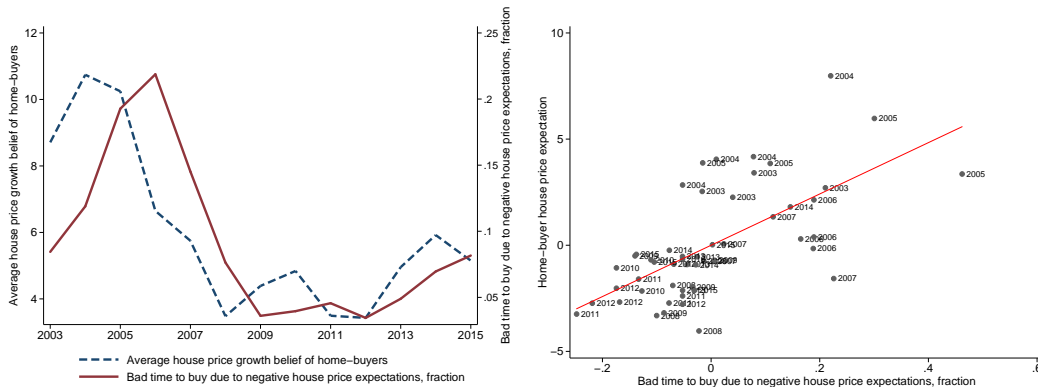


Figure 11: Measures of Optimism on Housing Market from the Michigan Survey



The left panel plots the fraction of people reporting that it is a bad time to buy a house. The right panel reports the fraction saying it is a bad time to buy a house due to negative house price expectations and the fraction of people reporting it is a good time to buy buying a house due to positive house price expectations.

Figure 12: Measures of Optimism on Housing Market from the Michigan Survey



The left panel plots the average of long and short term price expectations of home buyers, together with the share of individuals who report that it is a bad time to buy a house due to negative house price expectations. The right panel plots the correlation of home buyers price expectations and the share of home buyers who believe that it is a bad time to buy due to prices.

recent home-buyers against the fraction of individuals in the general population saying now is a bad time to buy because of negative house price growth expectations. There is a robust positive correlation, providing further support the importance of belief heterogeneity.

How was housing market optimism related to the acceleration of the PLS market? The Michigan data allow for the use of cross-sectional variation in housing market optimism across MSAs by the 2002 NCL share to answer this question. Table 6 reports the following specification on the evolution of optimism on the housing market in high versus low NCL share MSAs during the housing boom:

$$\Delta Optimism_{m,BOOM} = \alpha_m + \beta * HPGrowth_{m,0206} + \varepsilon_{z,m}$$

where  $\Delta Optimism_{m,BOOM}$  is the MSA-level average of the survey responses to a given Michigan question in MSA  $m$  in years 2004 through 2006 minus the average of the survey responses to the same question in MSA  $m$  in years 2000 to 2002. Columns 1 and 3 present the OLS estimates, and columns 2 and 4 present instrumental variable estimates where the instrument for house price growth is the 2002 NCL share of the MSA.

The OLS and IV estimates convey a consistent message: the average household in high house price growth areas became more pessimistic about the housing market in 2004 through 2006 relative to 2000 to 2002. In terms of magnitudes, a one standard deviation increase in house price growth leads to a 6 to 8 percentage point increase in the share of individuals expressing pessimism on the housing market. The increasing pessimism was driven by people who became more pessimistic because of house price considerations.

In Table A4 of the appendix, we split the “bad time to buy because of prices considerations” into the two separate subcomponent answers: “bad time to buy because prices are too high,” and “bad time to buy because prices will fall.” For both components, we find that there was a relative increase in the fraction of individuals expressing pessimism in high house price growth MSAs during the boom.

Table 6: NCL Share and Housing Market Optimism: CBSA-Level

	$\Delta_{boom}$ Bad time to buy	$\Delta_{boom}$ Bad time to buy	$\Delta_{boom}$ Bad time to buy bc of prices	$\Delta_{boom}$ Bad time to buy bc of prices
	(1)	(2)	(3)	(4)
HP growth, 02 to 06	0.247*** (0.050)	0.245 (0.140)	0.314*** (0.045)	0.241* (0.107)
Type	OLS	IV	OLS	IV
N	727	727	727	727
R-sq	0.062	0.062	0.181	0.171

Standard errors in parentheses.

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

Regression results for the specification  $y_m = \alpha + \beta HPgrowth_{m,0206} + \varepsilon_m$  from MSA  $m$ . For columns 1 and 2,  $y_m$  is the 2004-2006 average share of respondents saying it is a bad time to purchase a home minus the 2000-2002 average share. For columns 3 and 4,  $y_m$  is the 2004-2006 average share of respondents saying it is a bad time to purchase a home because of price considerations minus the 2000-2002 average share.  $HPgrowth_{m,0206}$  is house price growth in MSA  $m$  from 2002 to 2006. The regressions are weighted by the number of survey participants in an MSA  $m$ . In columns 2 and 4,  $HPgrowth_{m,0206}$  is instrumented using the NCL share of the MSA as of 2002. The NCL at the MSA level is defined as the weighted average of NCL at the lender-level where the weights are the share of loans originated in 2002 by a lender  $b$  in msa  $m$ .

The results are consistent with the view that higher house price growth, fueled by the acceleration of the PLS market, made the average individual in these MSAs more pessimistic about house prices. This provides further evidence that the PLS market affected the housing market not through a general rise in house price expectations, but instead through boosting the buying power of a smaller group of individuals.<sup>32</sup>

## 5 Conclusion

The narrative approach to economic history by Charles Kindleberger and Hyman Minsky shows a close connection between credit and speculation. Both scholars point to several

<sup>32</sup>We also examine whether there is evidence of a small cluster of optimists in high house price growth MSAs using the Michigan survey. We find that there is a positive correlation between house price growth and the share of individuals saying it is a good time to buy because of positive house price growth expectations, but it is not statistically distinct from zero at a reasonable confidence level. It is not surprising that the Michigan survey at the MSA level cannot detect a rise in optimistic speculators in high NCL share MSAs; recall that these speculators make up a small fraction of the overall population.

examples in history in which a sudden change in credit conditions was the spark that led to a speculative boom and bust in asset prices. This study tests for such a connection using the surge of the PLS market in 2003 and cross-sectional variation across geographical areas in the United States in exposure to this surge. The natural experiment allows for the isolation of a plausibly exogenous increase in mortgage credit supply, and it shows that this expansion in mortgage credit supply affected the housing market primarily through a speculation channel.

A small group of speculators used newly available credit to buy houses, which boosted both house prices and construction activity in the exposed areas. The crash was particularly painful in these areas, and defaults coming from the PLS market initiated the mortgage default crisis. Differences in beliefs about the evolution of house prices appear to be important in explaining these overall patterns. The marginal buyers of homes became more optimistic during the boom, whereas the average individuals in the overall economy became more pessimistic about the housing market. There are a number of new data sets recording house price growth expectations of individuals; the results presented in this study suggest that a promising avenue for future empirical research is to explore the importance of heterogeneous beliefs in more detail using these new data sets.

## References

- Abreu, Dilip and Markus K Brunnermeier, “Bubbles and crashes,” *Econometrica*, 2003, 71 (1), 173–204.
- Adelino, Manuel, Antoinette Schoar, and Felipe Severino, “Credit supply and house prices: Evidence from mortgage market segmentation,” *Available at SSRN 1787252*, 2014.
- , – , and – , “Dynamics of Housing Debt in the Recent Boom and Great Recession,” Technical Report, National Bureau of Economic Research 2017.
- , – , and – , “Perception of house price risk and homeownership,” 2018.
- Albanesi, Stefania, Giacomo De Giorgi, and Jaromir Nosal, “Credit growth and the financial crisis: A new narrative,” Technical Report, National Bureau of Economic Research 2017.
- Allen, Franklin and Douglas Gale, “Bubbles and crises,” *The economic journal*, 2000, 110 (460), 236–255.
- and Gary Gorton, “Churning bubbles,” *The Review of Economic Studies*, 1993, 60 (4), 813–836.
- Ashcraft, Adam B., Paul Goldsmith-Pinkham, and James Vickery, “MBS ratings and the mortgage credit boom,” Staff Reports 449, Federal Reserve Bank of New York 2010.
- Bailey, Michael, Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel, “Social networks and housing markets,” 2016.
- Barlevy, Gadi and Jonas Fisher, “Mortgage Choices During the U.S. Housing Boom,” 2018.
- Becker, Bo, “Geographical segmentation of US capital markets,” *Journal of Financial economics*, 2007, 85 (1), 151–178.
- Bhutta, Neil, “The ins and outs of mortgage debt during the housing boom and bust,” *Journal of Monetary Economics*, 2015, 76, 284–98.
- and Benjamin J Keys, “Interest rates and equity extraction during the housing boom,” *American Economic Review*, 2016, 106 (7), 1742–74.
- and Benjamin Keys, “Eyes Wide Shut? The Moral Hazard of Mortgage Insurers During the Housing Boom,” Technical Report, National Bureau of Economic Research 2018.
- Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, “Diagnostic Expectations and Credit Cycles,” *Journal of Finance*, 2017.
- Burnside, Craig, Martin Eichenbaum, and Sergio Rebelo, “Understanding booms and busts in housing markets,” *Journal of Political Economy*, 2016, 124 (4), 1088–1147.

- Caballero, Ricardo and Alp Simsek**, “A Risk-centric Model of Demand Recessions and Macroprudential Policy,” Working Paper 23614, National Bureau of Economic Research January 2018.
- Case, Karl E, Robert J Shiller, and Anne K Thompson**, “What Have They Been Thinking? Homebuyer Behavior in Hot and Cold Markets [with Comments and Discussion],” *Brookings Papers on Economic Activity*, 2012, pp. 265–315.
- Chernenko, Sergey, Samuel G. Hanson, and Adi Sunderam**, “Who Neglects Risk? Investor Experience and the Credit Boom,” *Journal of Financial Economics*, 2018.
- Chinco, Alex and Christopher Mayer**, “Misinformed speculators and mispricing in the housing market,” *The Review of Financial Studies*, 2015, 29 (2), 486–522.
- Coval, Joshua D., Jakub W. Jurek, and Erik Stafford**, “Economic Catastrophe Bonds,” *American Economic Review*, June 2009, 99 (3), 628–66.
- Davidoff, Thomas**, “Supply Elasticity and the Housing Cycle of the 2000s,” *Real Estate Economics*, 2013, 41 (4), 793–813.
- DeFusco, Anthony A, Charles G Nathanson, and Eric Zwick**, “Speculative dynamics of prices and volume,” Technical Report, National Bureau of Economic Research 2018.
- Dell’Ariccia, Giovanni, Luc Laeven, and Deniz Igan**, “Credit Booms and Lending Standards: Evidence from the Subprime Mortgage Market,” *The Journal of Money, Credit and Banking*, 2012, 44 (2-3), 367–384.
- Demyanyk, Yuliya and Otto Van Hemert**, “Understanding the subprime mortgage crisis,” *Review of financial Studies*, 2011, 24 (6), 1848–1880.
- Dokko, Jane, Benjamin J Keys, and Lindsay Relihan**, “Affordability, Financial Innovation, and the Start of the Housing Boom,” 2019.
- Drechsler, Itamar, Alexi Savov, and Philipp Schnabl**, “The deposits channel of monetary policy,” *The Quarterly Journal of Economics*, 2017, 132 (4), 1819–1876.
- Favara, Giovanni and Jean Imbs**, “Credit supply and the price of housing,” *The American Economic Review*, 2015, 105 (3), 958–992.
- Ferreira, Fernando and Joseph Gyourko**, “A new look at the US foreclosure crisis: Panel data evidence of prime and subprime borrowers from 1997 to 2012,” Technical Report, National Bureau of Economic Research 2015.
- Foote, Christopher L, Lara Loewenstein, and Paul S Willen**, “Cross-sectional patterns of mortgage debt during the housing boom: evidence and implications,” Technical Report, National Bureau of Economic Research 2016.
- Fuster, Andreas, Ricardo Perez-Truglia, and Basit Zafar**, “Expectations with Endogenous Information Acquisition: An Experimental Investigation,” 2018.

- Gao, Zhenyu, Michael Sockin, and Wei Xiong**, “Economic consequences of housing speculation,” Technical Report, Working paper 2017.
- Geanakoplos, John**, “The leverage cycle,” *NBER macroeconomics annual*, 2010, *24* (1), 1–66.
- Gennaioli, Nicola, Andrei Shleifer, and Robert Vishny**, “Neglected risks, financial innovation, and financial fragility,” *Journal of Financial Economics*, 2012, *104* (3), 452–468.
- Gerardi, Kristopher, Andreas Lehnert, Shane M Sherlund, and Paul Willen**, “Making sense of the subprime crisis,” *Brookings Papers on Economic Activity*, 2008, *2008* (2), 69–159.
- Gilchrist, Simon and Egon Zakrajšek**, “Credit Spreads and Business Cycle Fluctuations,” *American Economic Review*, June 2012, *102* (4), 1692–1720.
- Glaeser, Edward L, Joseph Gyourko, and Albert Saiz**, “Housing supply and housing bubbles,” *Journal of urban Economics*, 2008, *64* (2), 198–217.
- Gorton, Gary and Andrew Metrick**, “Securitization,” in “Handbook of the Economics of Finance,” Vol. 2, Elsevier, 2013, pp. 1–70.
- Greenwood, Robin and Samuel G Hanson**, “Issuer quality and corporate bond returns,” *The Review of Financial Studies*, 2013, *26* (6), 1483–1525.
- Griffin, John M. and Gonzalo Maturana**, “Who Facilitated Misreporting in Securitized Loans?,” *The Review of Financial Studies*, 2016, *29*, 384–419.
- Gurun, Umit G., GREGOR MATVOS, and AMIT SERU**, “Advertising Expensive Mortgages,” *The Journal of Finance*, 2016, *71* (5), 2371–2416.
- Harrison, J Michael and David M Kreps**, “Speculative investor behavior in a stock market with heterogeneous expectations,” *The Quarterly Journal of Economics*, 1978, *92* (2), 323–336.
- Haughwout, Andrew, Donghoon Lee, Joseph S Tracy, and Wilbert Van der Klaauw**, “Real estate investors, the leverage cycle, and the housing market crisis,” 2014.
- Jordà, Òscar, Moritz Schularick, and Alan M Taylor**, “The great mortgaging: housing finance, crises and business cycles,” *Economic Policy*, 2016, *31* (85), 107–152.
- Justiniano, Alejandro, Giorgio E. Primiceri, and Andrea Tambalotti**, “The Mortgage Rate Conundrum,” 2017. Working Paper.
- Kaplan, Greg, Kurt Mitman, and Giovanni L Violante**, “The housing boom and bust: Model meets evidence,” Technical Report, National Bureau of Economic Research 2017.

- Keys, Benjamin J, Amit Seru, and Vikrant Vig**, “Lender Screening and the Role of Securitization: Evidence from Prime and Subprime Mortgage Markets,” *The Review of Financial Studies*, 2012, 25 (7), 2071–2108.
- , **Tanmoy Mukherjee, Amit Seru, and Vikrant Vig**, “Did securitization lead to lax screening? Evidence from subprime loans,” *The Quarterly journal of economics*, 2010, 125 (1), 307–362.
- Khwaja, Asim Ijaz and Atif Mian**, “Tracing the impact of bank liquidity shocks: Evidence from an emerging market,” *The American Economic Review*, 2008, 98 (4), 1413–1442.
- Kindleberger, Charles**, *Manias, panics and crashes: A history of financial crises*, New York : Basic Books, 1978.
- Koudijs, Peter and Hans-Joachim Voth**, “Leverage and beliefs: personal experience and risk-taking in margin lending,” *American Economic Review*, 2016, 106 (11), 3367–3400.
- Krishnamurthy, Arvind and Tyler Muir**, “How credit cycles across a financial crisis,” Technical Report, National Bureau of Economic Research 2017.
- Landier, Augustin, David Alexandre Sraer, and David Thesmar**, “The risk-shifting hypothesis: Evidence from subprime originations,” in “AFA 2012 Chicago Meetings Paper” 2015.
- Landvoigt, Tim**, “Financial intermediation, credit risk, and credit supply during the housing boom,” *Working paper*, 2016.
- Levitin, Adam and Susan Wachter**, “Explaining the Housing Bubble,” *Georgetown Law Journal*, 2013, 100 (4), 1177–1258.
- López-Salido, David, Jeremy C Stein, and Egon Zakrajšek**, “Credit-market sentiment and the business cycle,” *The Quarterly Journal of Economics*, 2017, p. qjx014.
- Maggio, Marco Di and Amir Kermani**, “Credit-induced boom and bust,” *The Review of Financial Studies*, November 2017, 30 (11), 3711–58.
- McCauley, Robert**, “The 2008 crisis: transpacific or transatlantic?,” *Quarterly Review*, Bank for International Settlements 2018.
- Mian, Atif, Amir Sufi, and Emil Verner**, “Household debt and business cycles worldwide,” *The Quarterly Journal of Economics*, 2017, 132 (4), 1755–1817.
- , – , and **Francesco Trebbi**, “The Political Economy of the Subprime Mortgage Credit Expansion,” *Quarterly Journal of Political Science*, 2013, 8 (4), 373–408.
- and – , “The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis,” *The Quarterly Journal of Economics*, 2009, 124 (4), pp. 1449–1496.



- **and** – , “House prices, home equity–based borrowing, and the us household leverage crisis,” *The American Economic Review*, 2011, *101* (5), 2132–2156.
  - **and** – , *House of debt: How they (and you) caused the Great Recession, and how we can prevent it from happening again*, University of Chicago Press, 2015.
  - **and** – , “Fraudulent income overstatement on mortgage applications during the credit expansion of 2002 to 2005,” *The Review of Financial Studies*, 2017, *30* (6), 1832–1864.
- Mian, Atif R and Amir Sufi**, “Household Debt and Defaults from 2000 to 2010: The Credit Supply View,” *Evidence and Innovation in Housing Law and Policy*, 2017.
- Miller, Edward M**, “Risk, uncertainty, and divergence of opinion,” *The Journal of finance*, 1977, *32* (4), 1151–1168.
- Minsky, Hyman**, *Stabilizing an Unstable Economy*, New Haven, Yale University Press, 1986.
- Nadauld, Taylor D and Shane M Sherlund**, “The impact of securitization on the expansion of subprime credit,” *Journal of Financial Economics*, 2013, *107* (2), 454–476.
- Nagel, Stefan**, “The liquidity premium of near-money assets,” *The Quarterly Journal of Economics*, 2016, *131* (4), 1927–1971.
- Nathanson, Charles G and Eric Zwick**, “Arrested development: Theory and evidence of supply-side speculation in the housing market,” Technical Report, National Bureau of Economic Research 2017.
- Perotti, Enrico**, “The roots of shadow banking,” *CEPR policy insight*, December 2013, (69).
- Piazzesi, Monika and Martin Schneider**, “Momentum traders in the housing market: survey evidence and a search model,” *American Economic Review*, 2009, *99* (2), 406–11.
- Piskorski, Tomasz, Amit Seru, and James Witkin**, “Asset quality misrepresentation by financial intermediaries: evidence from the RMBS market,” *The Journal of Finance*, 2015, *70* (6), 2635–2678.
- Purnanandam, Amiyatosh**, “Originate-to-Distribute Model and the Subprime Mortgage Crisis,” *The Review of Financial Studies*, 2011, *24* (6), 1881–1915.
- Rajan, Uday, Amit Seru, and Vikrant Vig**, “The failure of models that predict failure: Distance, incentives, and defaults,” *Journal of Financial Economics*, 2015, *115* (2), 237 – 260.
- Saiz, Albert**, “The geographic determinants of housing supply,” *The Quarterly Journal of Economics*, 2010, *125* (3), 1253–1296.
- Scheinkman, Jose A and Wei Xiong**, “Overconfidence and speculative bubbles,” *Journal of political Economy*, 2003, *111* (6), 1183–1220.

**Simsek, Alp**, “Belief disagreements and collateral constraints,” *Econometrica*, 2013, 81 (1), 1–53.

**Smith, Adam**, *An Inquiry Into The Nature And Causes Of The Wealth Of Nations*, W. Strahan and T. Cadell, London, 1776.

**Xiao, Kairong**, “Monetary Transmission through Shadow Banks,” 2018.

# Online Appendix

## A Further details on the data

### A.1 HMDA match to lenders

The HMDA data set records the universe of mortgage originations for mortgage originators that have an office within metropolitan statistical areas (MSAs).<sup>33</sup> We identify each mortgage originator in the HMDA data, and link these financial institutions to Call Report data using a key provided to us by the Federal Reserve Board.

In the HMDA data, financial institutions report any mortgage loans to their regulatory agency, and they are given a unique ID number and agency code in the loan-level data. The financial institutions we focus on are regulated by the Office of the Comptroller of the Currency (OCC, agency code 1), the Federal Reserve System (FRS, agency code 2), and the Federal Deposit Insurance Corporation (FDIC, agency code 3); thrifts regulated by the Office of Thrift Supervision (OTS, agency code 4); and independent mortgage companies regulated by the Department of Housing and Urban Development (HUD, agency code 7). This only leaves out credit unions, who make up a small portion of lending, and institutions regulated by the Consumer Financial Protection Bureau, an agency that was only created at the end of the decade as a response to the financial crisis. What we call banks are those institutions regulated by agencies corresponding to agency codes 1-4 since these are associated with a depository institution. Non-banks are non-depository independent mortgage lending companies and correspond to agency code 7.

We link lenders in the HMDA data to regulatory data filed by banks in the Report of Condition and Income (the Call Report) and by thrifts in the OTS Thrift Financial Report (TFR). Financial institutions that submit one of these forms are given a unique ID. If a

---

<sup>33</sup>See guidelines for HMDA issued by the Federal Reserve in 2005: “a lender does not have to report HMDA data unless it has an office in a metropolitan statistical area (MSA). As a result, reporting of home loans in some rural areas may be relatively low.”

bank (thrift) is part of a multi-bank (multi-thrift) holding company then each form provides a holding company ID which corresponds to the regulatory high holding company of the institution. For our analysis, we use the bank-holding or thrift-holding company ID when an institution is part of a holding company and its unique ID otherwise. The HMDA ID is used for independent mortgage companies. Using a key of the HMDA Report ID's and the Call Report and TFR bank ID's provided to us by the Federal Reserve Board, we match the loan-level HMDA data to the bank level report data.

## A.2 TransUnion details

There are three main challenges in using the TransUnion data relative to HMDA. First, the HMDA data set records the address of the property being purchased, whereas in the TransUnion data we can only see the address of the individual taking out the mortgage. We impute the address of the house being purchased in TransUnion as the address of the individual six months after the home purchase mortgage origination. An obvious concern is that many individuals are buying second houses and therefore the address in the credit bureau data does not reflect the address of the house being purchased. While this is a valid concern, we find remarkably similar zip-code level results using the HMDA and TransUnion data, which suggests that this issue does not materially affect our results. For example, Figure A7 below shows the evolution of first-lien mortgages by the NCL share of a zip code as of 2002, and the coefficients are similar across the two data sets.

The second and third challenges are related: the TransUnion data do not contain information on whether the originated mortgage is a refinancing or a home purchase, nor does it contain information on whether the mortgage is a first-lien or second-lien. Home equity loans are recorded separately, so the latter concern is only with “piggy-back” second liens used at purchase.

We begin by classifying a mortgage origination as a refinancing if any of the following four conditions are met:

- for the individual taking out the mortgage, the TU data set records the same census tract for the individual one month before to six months after the origination and the number of mortgages outstanding for the individual is the same from one month before to six months after the origination.
- for the individual taking out the mortgage, the TU data set records the same census tract for the individual one month before to three months after the origination and the number of mortgages outstanding for the individual is the same from one month before to three months after the origination.
- for the individual taking out the mortgage, the TU data set records the same census tract for the individual one month before to six months after the origination and the number of mortgages outstanding for the individual is reduced from one month before to six months after the origination. The assumption is that such a refinancing was used to pay off multiple mortgages outstanding prior to the refinancing.
- for the individual taking out the mortgage, the TU data set records the same census tract for the individual one month before to three months after the origination and the number of mortgages outstanding for the individual is zero one month before the origination and one six months after the origination. This last step ensures we capture cash-out refinancings by individuals who had no mortgage prior to the refinancing.

The complement of this set is defined to be a purchase mortgage. To separate a “piggy-back” second lien from a first lien purchase mortgage, we classify a second-lien as a mortgage that is less than or equal to 30% of the total mortgage amount taken out by the individual in the same month.

The top left panel of Appendix Figure [A6](#) shows total mortgages originated in the HMDA and TransUnion data sets, which match closely. The rest of the panels show the share of refinancing, home-purchase, and first-lien home-purchase, to total originations in both data

sets. Our methodology seems to over-estimate the refinancing share during the heart of the housing boom, but otherwise it is a relatively close match.

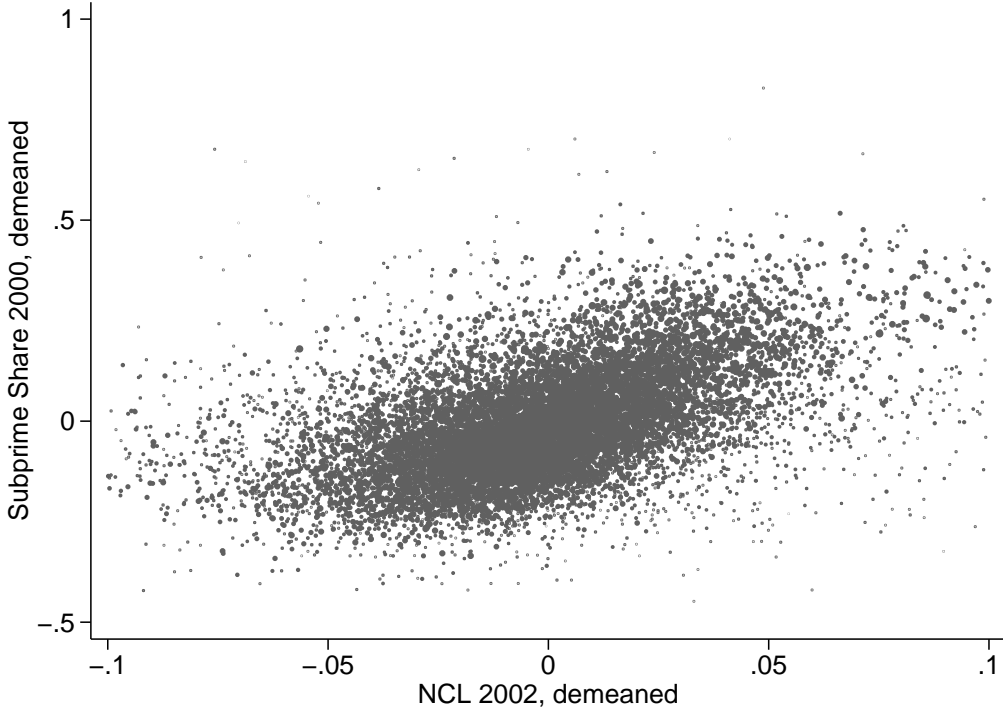
### **A.3 Measuring housing market optimism in Michigan survey**

The Michigan survey asks the following question to respondents: “Generally speaking, do you think now is a good time or bad time to buy a house?” Almost 99% of survey respondents answer this question by saying it is either a “good” or “bad” time. In addition to this question, respondents are asked a follow up question: “Why do you say so?” In response to this question, respondents give up to two answers. The survey questioners record a number of different responses, which we classify into sub-groups as follows.

If an individual responds that now is a good time to buy a house because “prices are going up,” “prices are low/stable/not too high,” “prices won’t get any lower,” or “good buys available,” we classify the response as being favorable because of house price considerations. If an individual answers that now is a bad time to buy a house because “prices will fall later, will come down” or “prices are too high, houses cost too much,” we classify the response as being unfavorable because of house price considerations. If an individual answers that now is a good time to buy a house because “credit easy to get, easy money,” “credit will be tighter later,” or “lower down payment,” we classify the response as being favorable because credit is loose. Note that there are many other responses to the follow up question; the most common response given is that now is a good time to buy a home because “interest rates are low.” As a result, the fractions of respondents saying now is a good time versus a bad time to buy because of house prices considerations will not add up to one. Further, the total fraction mentioning prices as a consideration will vary over time. From 2000 onward, it is possible to match individual respondents to the county in which they live. We use this match to construct MSA-by-year level measures of housing market optimism.

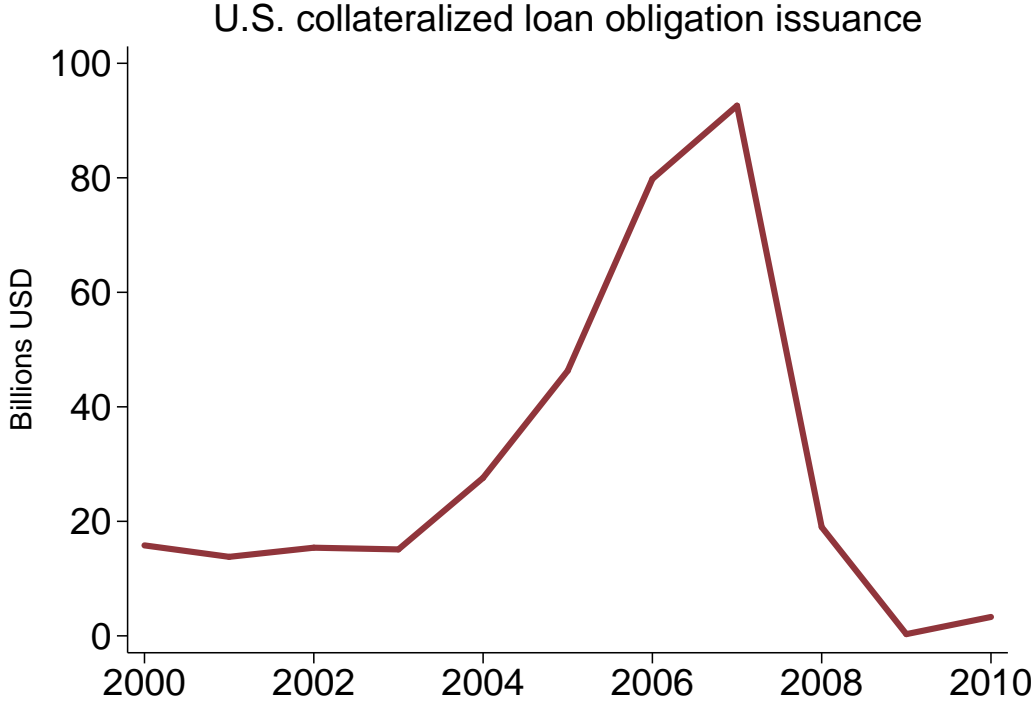
# Appendix Tables and Figures

Figure A1: Within-MSA Fraction of population under 660 in 2000 versus NCL share in 2002



The figure plots fraction of population in a zip code with a credit score below 660 in 2000 against the NCL share in the zip code in 2002. The variables are de-meaned at the MSA level before plotting, and each observation is weighted by its population size.

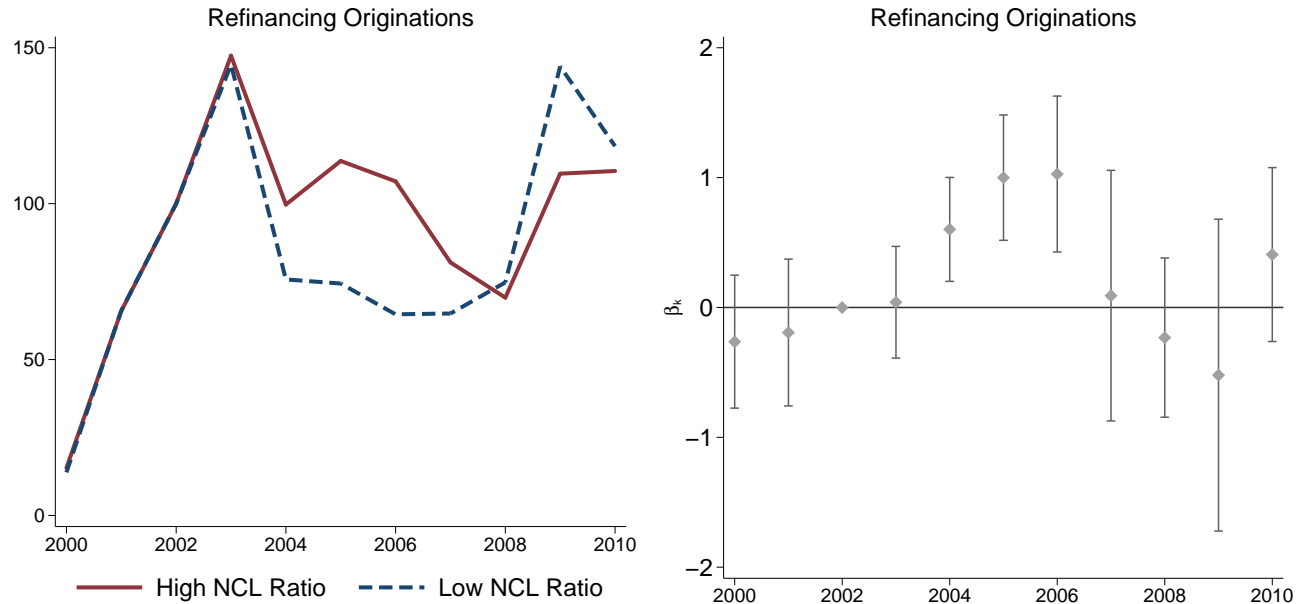
Figure A2: Rise in Collateralized Loan Obligations



The figure plots the amount of U.S. collateralized loan obligations issued in billions of US dollars.

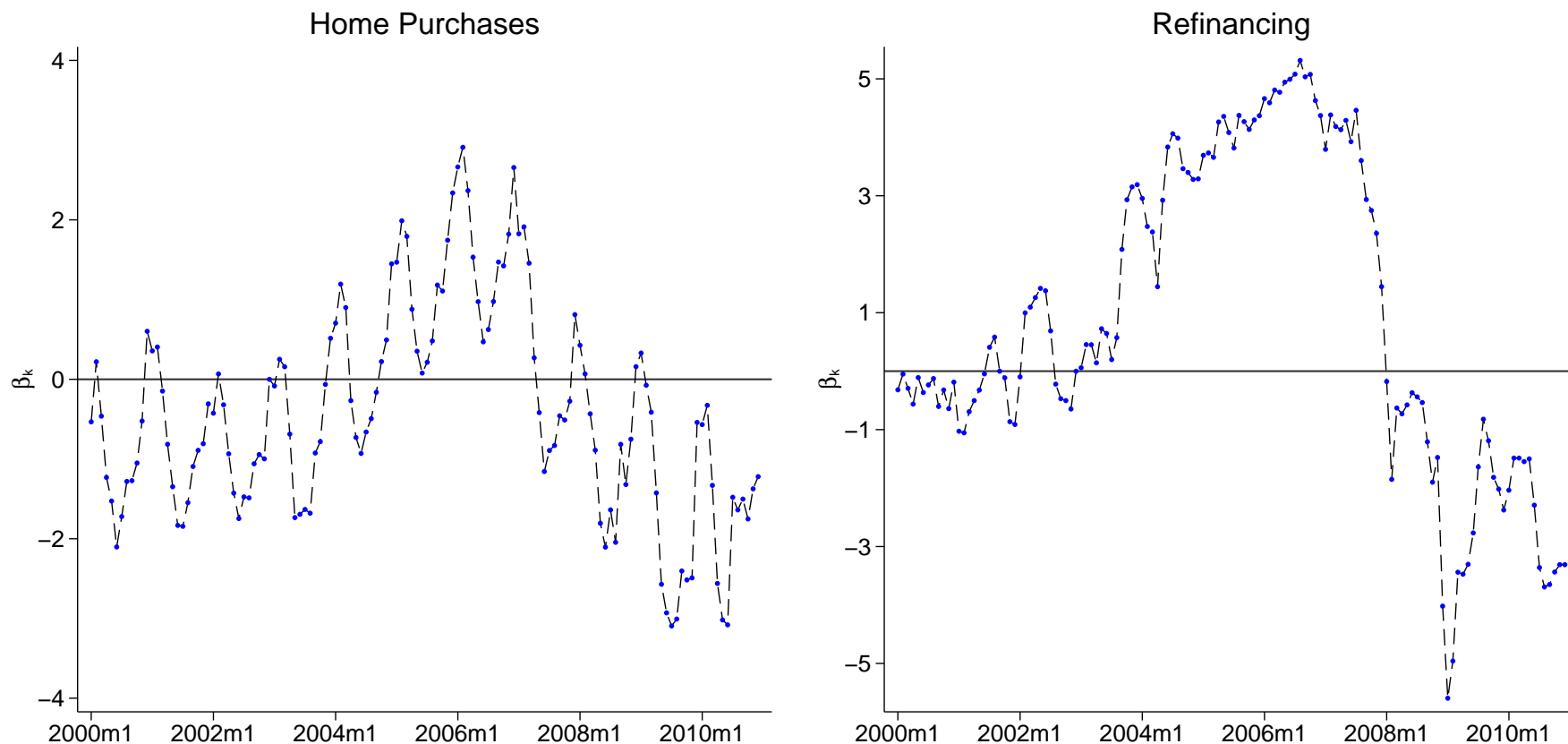


Figure A3: Lender Level Results for Refinancing Originations



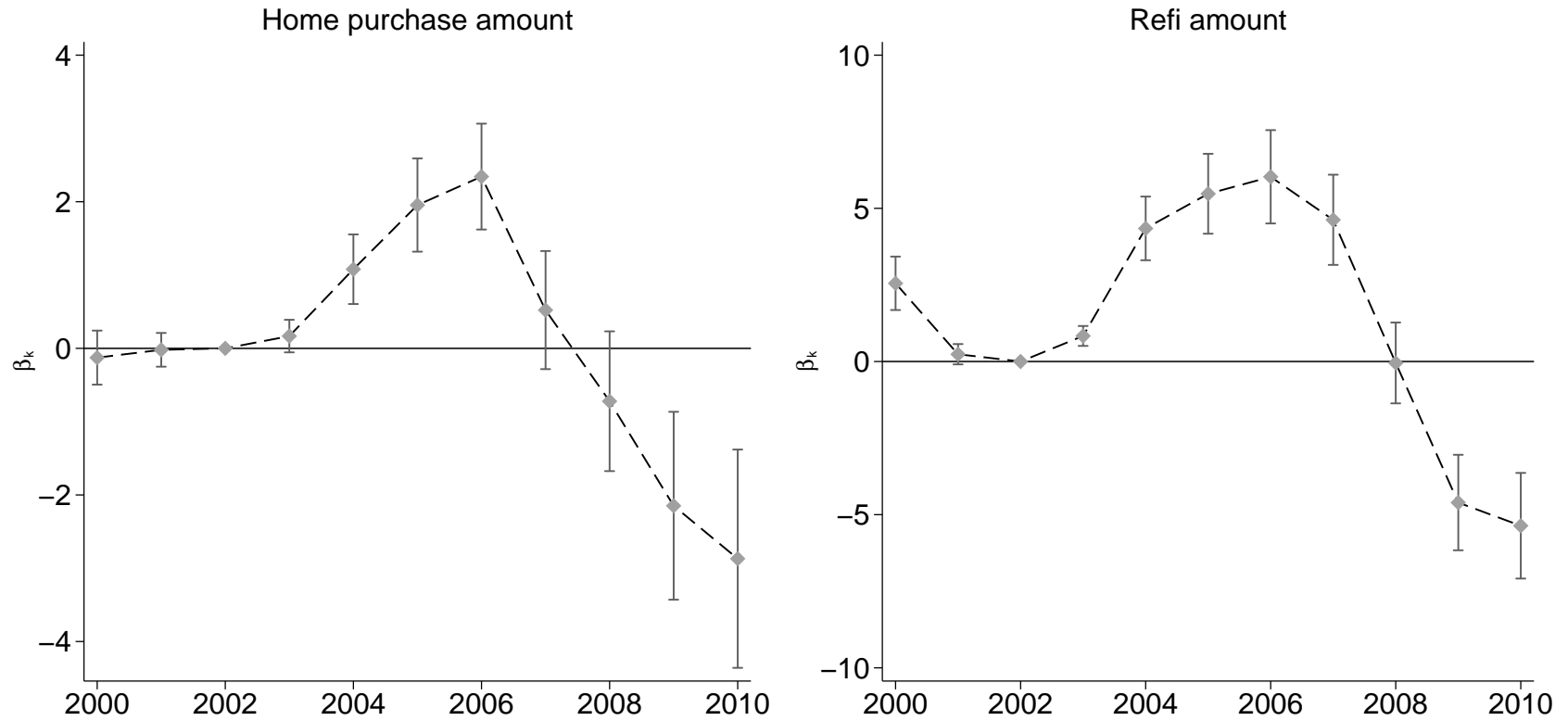
The left panel plots the growth rates of mortgage originations for refinancing of lenders above and below the median non-core liabilities share (NCL). The right panel plots the coefficients  $\{\beta_k\}$  of the specification  $y_{b,t} = \alpha_b + \gamma_t + \sum_{k \neq 2002} \mathbb{1}_{t=k} \beta_k NCL_{b,2002} + \varepsilon_{b,t}$  for lender  $b$  at time  $t$ .  $y_{b,t}$  is equal to total mortgage originations for refinancing by a lender  $b$  in year  $t$ . NCL is defined here as one minus the proportion of liabilities that are federally insured deposits for institutions that are in the FFIEC Call Reports and one for institutions regulated by the Department of Housing and Urban Development (HUD). The regressions are weighted by the share of refinancing loans originated in 2002 by lender  $b$ . 95% confidence intervals from robust standard errors are also plotted. Lender fixed effects included in panel regression.

Figure A4: PLS Credit Supply Effect on MSA level: Monthly Frequency without Seasonal Adjustment



The left panels plot the coefficients  $\{\beta_k\}$  of the specification  $\ln(y_{m,t}) = \alpha_m + \gamma_t + \sum_{k \neq 2002} \mathbb{1}_{t=k} \beta_k NCL_{m,2002} + \varepsilon_{m,t}$  for MSA  $m$  at time  $t$ .  $y_{m,t}$  refers to the total volume of home purchase mortgages originated in MSA  $m$  in month  $t$ . The right panel uses the same specification but  $y_{m,t}$  refers to the total volume of refinancing mortgages. NCL at the MSA-level is defined as the weighted average of NCL at the lender-level where the weights are the share of loans originated in 2002 by a lender  $b$  in MSA  $m$ . The regressions are weighted by the share of total occupied housing units in MSA  $m$  in 2000. MSA and time fixed effects are included in the panel regressions.

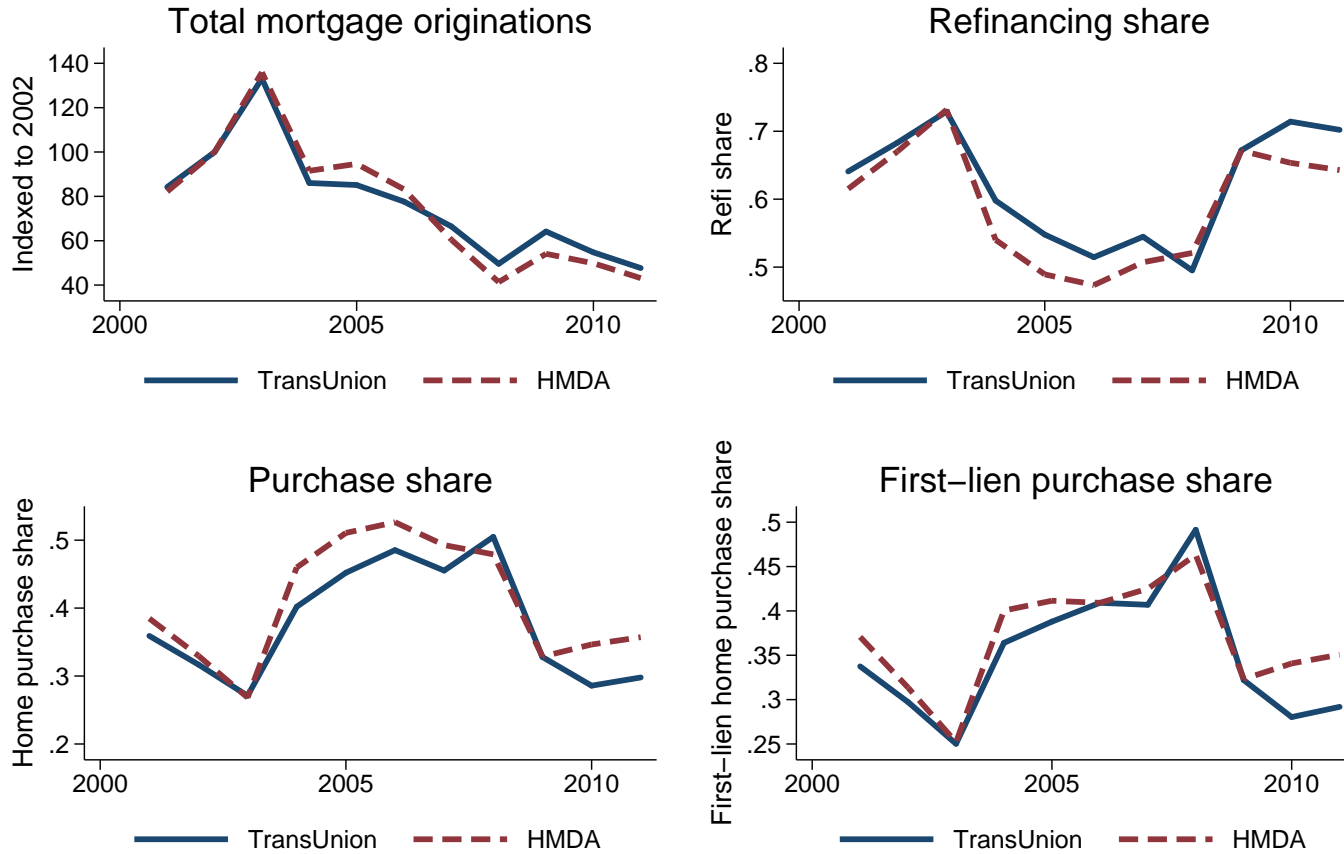
Figure A5: PLS Driven Credit Supply Effect at Zip Code level



58

The panels plot the coefficients  $\{\beta_k\}$  of the specification  $\ln(y_{z,t}) = \alpha_z + \gamma_t + \sum_{k \neq 2002} \mathbb{1}_{t=k} \beta_k NCL_{z,2002} + \varepsilon_{z,t}$  for zip code  $z$  in year  $t$ .  $y_{z,t}$  is total mortgage amount originated in zip code  $z$  in year  $t$ . NCL at the zip code-level is defined as the weighted average of NCL at the lender-level where the weights are the share of loans originated in 2002 by a lender  $b$  in zip code  $z$ . The regressions are weighted by the share of total occupied housing units in zip code  $z$  in 2000. 95% confidence intervals from MSA clustered standard errors are also plotted. Zip code and time fixed effects are included.

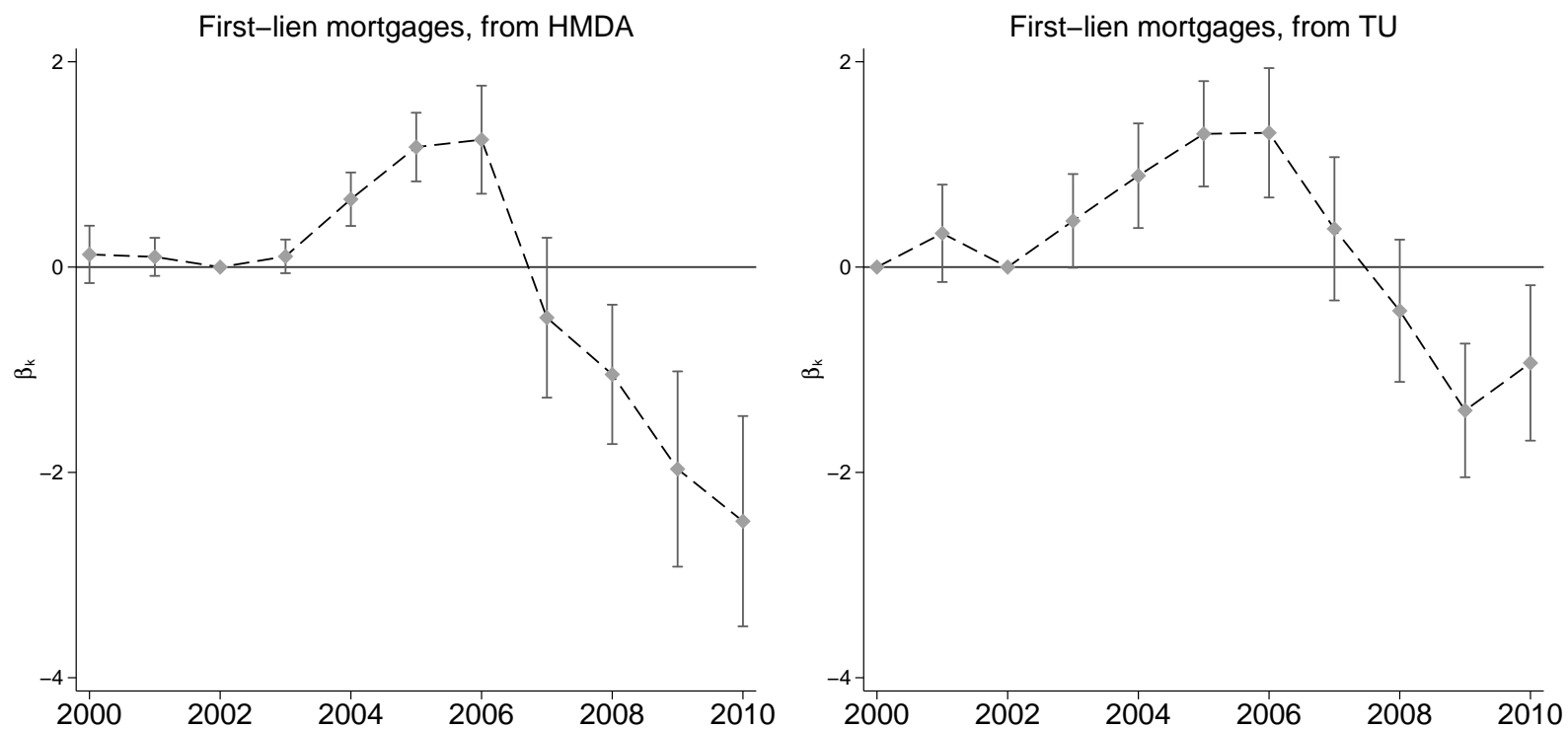
Figure A6: Comparison of TransUnion and HMDA Aggregates



69

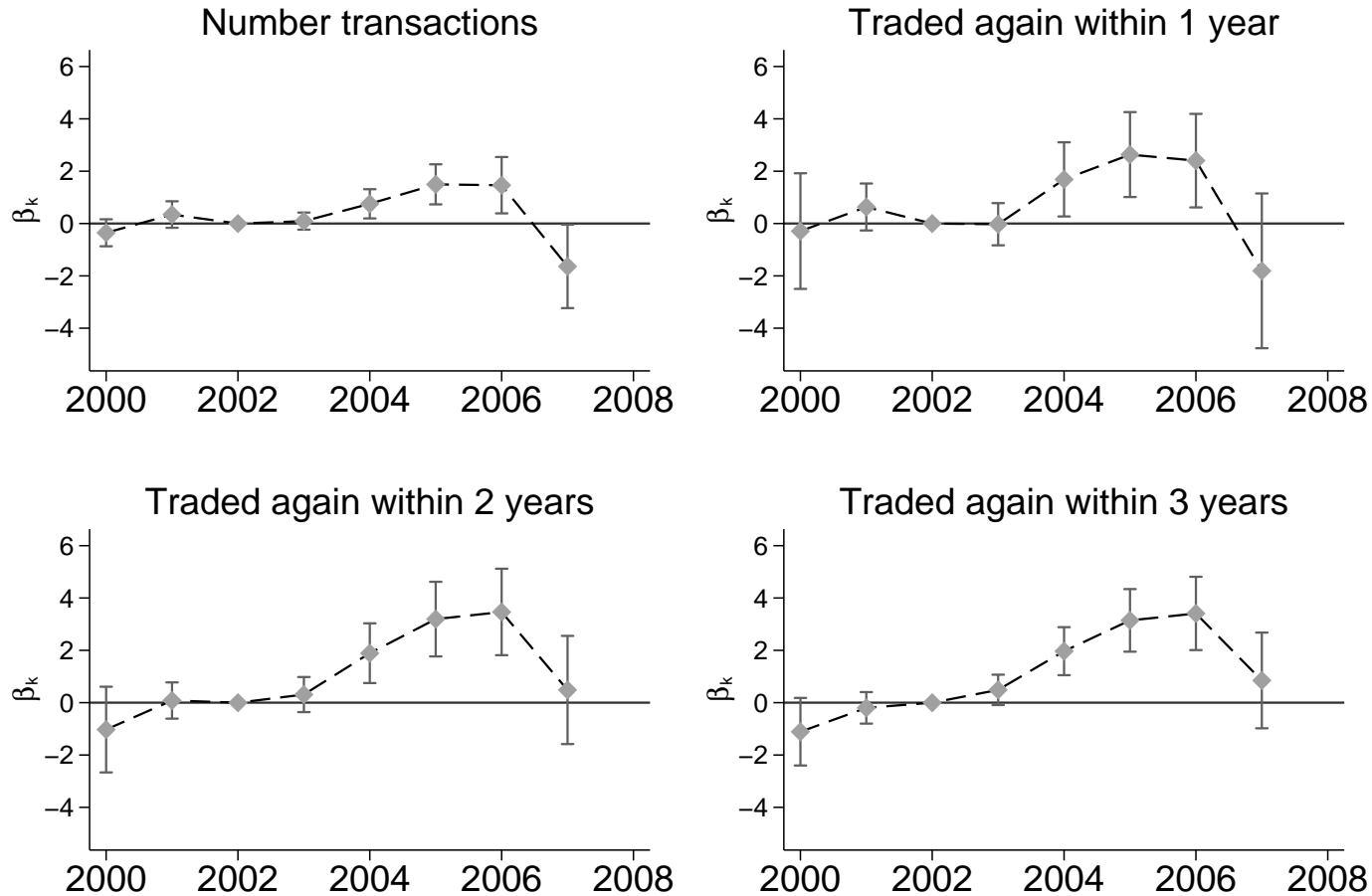
The top left panel plots total mortgage originations from the HMDA data set and TransUnion data set, where both series are indexed to 2002. The other three panels plot the share of total mortgage originations that are refinancing, purchase, and first-lien purchase for the HMDA and TransUnion data sets. The classification of the mortgage as refi, purchase, and first-lien purchase for the TransUnion data is described in the text of the appendix.

Figure A7: Zip-code First-Lien Mortgages by NCL Share: Panel Regressions for HMDA and TransUnion data sets



The panels plot the coefficients  $\{\beta_k\}$  of the specification  $\ln(y_{z,t}) = \alpha_z + \gamma_t + \sum_{k \neq 2002} \mathbb{1}_{t=k} \beta_k NCL_{z,2002} + \varepsilon_{z,t}$  for zip code  $z$  in year  $t$ .  $y_{z,t}$  in the left panel is the number of first lien mortgage originations for home purchase according to HMDA.  $y_{z,t}$  in the right panel is the number of first lien mortgage originations for home purchase according to TransUnion. NCL at the zip code-level is defined as the weighted average of NCL at the lender-level where the weights are the share of loans originated in 2002 by a lender  $b$  in zip code  $z$ . The regressions are weighted by the share of total occupied housing units in zip code  $z$  in 2000. 95% confidence intervals from MSA clustered standard errors are also plotted. Zip code level fixed effects are included.

Figure A8: Property Level Measure of Speculation



61

The panels plot the coefficients  $\{\beta_k\}$  of the specification  $\ln(y_{z,t}) = \alpha_z + \gamma_t + \sum_{k \neq 2002} \mathbb{1}_{t=k} \beta_k NCL_{z,2002} + \varepsilon_{z,t}$  for zip code  $z$  in year  $t$ . The outcome variables in all specification are the logarithm of transaction volume from CoreLogic. Following DeFusco et al. (2018), we separate transactions based on whether the same property is transacted again with a certain number of years. NCL at the zip code-level is defined as the weighted average of NCL at the lender-level where the weights are the share of loans originated in 2002 by a lender  $b$  in zip code  $z$ . The regressions are weighted by the share of total occupied housing units in zip code  $z$  in 2000. 95% confidence intervals from MSA clustered standard errors are also plotted. Zip code level fixed effects are included.

Table A1: Summary Statistics

	Obs	Mean	SD	Median	P10	P90
<b>Lender level</b>						
2002 NCL ratio	5,026	0.74	0.20	0.68	0.49	1.00
$\Delta_{02,05}$ ln (Amount originated)	3,950	-0.02	0.73	-0.09	-0.46	0.62
<b>Zip level</b>						
2002 NCL Share	12,427	0.77	0.05	0.77	0.71	0.82
$\Delta_{BOOM}$ (Home purchase amount originated)	12,419	0.57	0.36	0.54	0.18	1.01
$\Delta_{BOOM}$ (Refinancing amount originated)	12,400	0.32	0.53	0.23	-0.25	1.05
$\Delta_{BOOM}$ (First-lien mortgages, HMDA)	12,418	0.14	0.28	0.12	-0.15	0.47
$\Delta_{BOOM}$ (Volume of housing transactions)	3,703	0.17	0.31	0.11	-0.14	0.51
$\Delta_{02,06}$ (House Prices)	6,619	0.37	0.22	0.36	0.10	0.67
<b>Zip level: TransUnion data</b>						
$\Delta_{BOOM}$ (First-lien mortgages, TransUnion)	9,023	0.09	0.67	0.05	-0.69	0.92
$\Delta_{BOOM}$ (First-lien mortgages, HMDA)	9,019	0.12	0.29	0.09	-0.21	0.47
<b>Mortgage level: First-lien purchase</b>						
Origination amount (thousands USD)	347,905	187.86	176.31	149.20	40.55	360.00
Vantage score in 2000	295,547	695.01	84.93	707.00	573.00	800.00
Age in year of origination	338,304	42.18	12.68	40.00	27.00	59.00
Default in 2006 or after	347,905	0.24	0.43	0.00	0.00	1.00
<b>MSA Level</b>						
2002 NCL Share	259	0.76	0.04	0.77	0.71	0.81
Housing Supply Elasticity	259	1.96	1.18	1.65	0.76	3.47
$\Delta_{BOOM}$ Constructed units	259	0.16	0.26	0.11	-0.13	0.48
$\Delta_{BOOM}$ Bad time to buy	259	0.07	0.13	0.06	-0.07	0.23
$\Delta_{BOOM}$ Bad time bc of prices	259	0.08	0.11	0.07	-0.04	0.21
$\Delta_{BOOM}$ Good time bc of prices	259	0.01	0.10	0.00	-0.08	0.11

This table presents summary statistics at the lender, zip, mortgage, and MSA level.  $\Delta_{BOOM}$  is defined as the log change in outcome  $y$  from 2000-2002 to 2004-2006. For the Zip level: TransUnion data panel,  $\Delta_{BOOM}$  is defined as the log change in outcome  $y$  from 2001-2002 to 2005-2006. This change is made because the TransUnion data are not available for 2000. Summary statistics at the lender level are weighted by total mortgage origination amount as of 2002. Summary statistics for the zip level and MSA level are weighted by total number of households as of 2000.

Table A2: 2002 NCL Share Correlations with Observable Variables

Covariates	Across Zipcode
1998 NCL share	0.83
Saiz elasticity	-0.12
2000 Deposits/Purchase amount originated	-0.16
2000 Log median household income	-0.25
2000 Log median home value	-0.22
2000 Fraction hispanic or black	0.50
2000 Subprime share	0.48
2000 Fraction renters	0.18
2000 Fraction age 65+	-0.14

Correlation coefficients of the non-core liabilities share (NCL) in 2002 to observable variables at the zip-code level. The NCL ratio at the geographical-level is defined as the weighted average of NCL at the lender-level where the weights are the share of loans originated in 2002 by lender  $b$  in geography  $g$ . The Saiz elasticity measure and the deposit to purch amount originated ratio are measured at the MSA level given data availability.



Table A3: Who Are the Marginal Buyers? With MSA FE

*Panel A: Total and by speculation measures*

	$\Delta$ first-lien, Total		$\Delta$ first-lien, Speculators			
	HMDA (1)	TransUnion (2)	Multiple Houses (3)	Short Term (4)	2+ mortgages (5)	Union (6)
2002 NCL Share	3.438*** (0.351)	3.077*** (0.626)	1.455*** (0.318)	1.473*** (0.235)	0.607** (0.205)	2.458*** (0.402)
N	9,020	9,023	9,023	9,023	9,023	9,023
R-sq	0.280	0.076	0.076	0.075	0.064	0.076

*Panel B: By measures of risk*

	$\Delta$ first-lien, By ex ante risk			
	No Score (1)	Subprime (2)	Near Prime (3)	Prime (4)
2002 NCL Share	1.164*** (0.199)	1.858*** (0.198)	0.209 (0.197)	-0.153 (0.367)
R-sq	0.065	0.073	0.059	0.065

*Panel C: By measures of risk and speculation*

	$\Delta$ first-lien, Speculators		$\Delta$ first-lien, Non-speculators	
	Non-prime (1)	Prime (2)	Non-prime (3)	Prime (4)
2002 NCL Share	2.120*** (0.277)	0.338 (0.226)	1.111*** (0.242)	-0.492* (0.218)
R-sq	0.084	0.059	0.070	0.068

*Panel D: By borrower age at origination*

	$\Delta$ first-lien, Borrower Age at Origination				
	Missing (1)	Under 30 (2)	30-40 (3)	41-50 (4)	Over 50 (5)
2002 NCL Share	0.054 (0.104)	0.512** (0.191)	1.513*** (0.289)	0.659** (0.225)	0.339 (0.206)
R-sq	0.049	0.066	0.059	0.067	0.060

Standard errors clustered at MSA level.

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

Table A4: NCL Share and Housing Market Optimism: MSA-Level

	$\Delta_{boom}$ Bad time to buy bc prices too high (1)	$\Delta_{boom}$ Bad time to buy bc prices too high (2)	$\Delta_{boom}$ Good time to buy bc of prices will fall (3)	$\Delta_{boom}$ Good time to buy bc of prices will fall (4)
HP growth, 02 to 06	0.120*** (0.035)	0.096 (0.087)	0.089*** (0.019)	0.097* (0.039)
Type	OLS	IV	OLS	IV
N	727	727	727	727
R-sq	0.043	0.042	0.090	0.089

Standard errors in parentheses.

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