How Do Banks Compete With Non-Banks in the Conforming Residential Mortgage Market? The Role of Balance Sheet

By

Lan Shi, Yan Zhang, and Xinlei Zhao

First version: August 2022

This version: January 2023

JEL classification: G21; G23; G29; G51

Key words: Non-bank; residential mortgages; originate to distribute; GSE guarantee; balance sheet

Lan Shi (lan.shi@occ.treas.gov) is from the Retail Credit Risk Analysis Division, Office of the Comptroller of the Currency, and both Yan Zhang (yan.zhang@occ.treas.gov) and Xinlei Zhao (email: Xinlei.Zhao@occ.treas.gov) are from the Commercial Credit Risk Analysis Division, Office of the Comptroller of the Currency, Washington, D.C.

We thank the helpful comments from Daniel Grodzicki, Feng Li, Stephen Karolyi, Natalie Tiernan, Ken Ueda, and participants at the research seminar at the Office of the Comptroller of the Currency. The views expressed in this paper are those of the authors alone and do not necessarily reflect those of the Office of the Comptroller of the Currency or the U.S. Department of the Treasury and do not establish supervisory policy, requirements, or expectations. The authors take responsibility for any errors.

How Do Banks Compete With Non-Banks in the Conforming Residential Mortgage Market? The Role of Balance Sheet

Abstract

We find evidence that banks respond to increasing non-bank competition in the post-2012 conforming residential mortgage market by making use of their balance sheet financing capability and the cross-subsidization in the GSE pricing schemes. Loan retention decision post 2012 is in stark contrast with that before 2012 when non-bank competition was only modest. Banks retain on their balance sheets higher proportions of low-risk conforming mortgages for which the guarantee-fees might be high relative to these loans' credit risk, and banks have been keeping increasingly higher fractions of such mortgages over time. Finally, banks charge lower interest rates on mortgages retained on their balance sheets relative to those sold to the GSEs.

Section 1: Introduction

One of the most prominent developments of U.S. residential mortgage market was the sharply rising role of non-banks in origination and servicing. Figure 1 demonstrates that non-bank share of the residential mortgage origination has been rather stable before 2011 and shows substantial growth starting from 2012.[1] Non-banks have more than doubled their share of the residential mortgage origination market from 2011 to 2019. This phenomenon has attracted a great deal of interest from academics and policy makers. Researchers have proposed various explanations for non-banks' growth in the residential mortgage market, such as non-banks' expansion into the risky

[1] Figure 1 is generated based on the Home Mortgage Disclosure Act (HMDA) data using all conventional loans (including both conforming loans and jumbo loans) and it is by origination amount. Results using origination counts show similar pattern. We check whether each lender is a bank or subsidiary of a bank. If a lender does not have a RSSD and a bank charter, nor does it have a bank parent, it is considered a non-bank.

market segment (Buchak et al., 2018a; Kim et al., 2022), fintech firms' technology innovation (Fuster et al., 2018 and Buchak et al., 2018a), or pricing (Guo et al., 2021).

All the above-mentioned studies rely on publicly available agency data provided by government sponsored enterprises (GSEs). However, a significant proportion (ranging from 10% to 20%) of all U.S. residential mortgages are held on lenders' balance sheets (according to Urban Institute).[2] The overwhelming majority of the portfolio loans resides on banks' balance sheets because of non-banks' lack of balance sheet capacity.[3] Therefore, academic investigations of securitized mortgages only can leave a significant hole in our understanding of the competition between banks and non-banks in the residential mortgage market. This study aims to fill this gap and move the literature forward.

It is widely accepted that banks have capacities for balance-sheet intensive activities, such as originations of jumbo mortgages that are not eligible for GSE purchase (Buchak et al., 2018b; Haughwout et al., 2022). Consistent with such a proposition, Panel A of Figure 2 (which is constructed from the HMDA data) shows that over the period from 2013 to 2019, banks originate the majority of all jumbo mortgages. Unreported results show that banks keep nearly all those jumbo loans on their balance sheets.[4]

Although banks can make up for the lost business on residential mortgage origination by lending to non-banks (see for example, Jiang (2022)), another possible way to boost their income is to retain loans on their balance sheet instead of selling the loans to GSEs if it is more profitable

[2] https://www.urban.org/sites/default/files/publication/103539/housing-finance-at-a-glance-a-monthly-chartbook-january-2021_1.pdf

[3] Although credit unions also have balance sheet capacity, their share of the residential mortgage market is very small, at around 4.6 percent as of 2022Q1 (Urban Institute Housing Finance Monthly Chartbook, July 2022.)

[4] Credit unions are excluded in Figure 2. Non-banks have increased their jumbo loan origination over time, and they largely sell these loans to commercial banks and credit unions.

to retain such loans. We can indeed see from Panel B of Figure 2 that,[5] besides jumbo mortgages, banks retain a considerable proportion of conforming mortgages on their balance sheets as well. Further, the fraction of conforming mortgages kept on banks' balance sheet has been rising from 2013 to 2019,[6] and by 2019, banks have retained about half of conforming mortgages they originate. This empirical regularity has not been documented in the literature before. Having the option to sell the loans to GSE or keep them on their balance sheets, what are the features of conforming loans that banks kept and why did they retain a rising proportion of these conforming loans during the sample period of 2013-2019?

Loans sold to GSEs are subject to the guarantee fees (G-fee), which provides guarantee to investors in GSE mortgage backed securities (MBS) that principal and interest will be paid in the event of borrower non-performance.[7] GSEs have been implementing a system of cross-subsidies in the G-fee grid to support low-income mortgage borrowers,[8] and Federal Housing Finance Agency (FHFA) reports indeed show that the G-fees are high relative to the credit risks for adjustable-rate mortgages (ARMs),[9] as well as fixed-rate mortgages (FRMs) with high credit score (CS) and low CLTV.[10] Because of the size of GSEs in the conforming residential mortgage

[5] Panel B is constructed using the Y14 data as we can clearly see whether a loan is sold or retained in the Y14 data. Y14 data start from 2013, and we will provide more details of the Y14 data in Section 3.
[6] Private label mortgage securitization (PLS) market during our sample was small compared to pre-2008 levels. For example, only 2 percent of the 2013 conforming loan vintage ended up in PLS; the majority were either held on banks' balance sheet or sold to GSE.
[7] There are two types of guarantee fees: ongoing and upfront. Ongoing fees are factored into each loan's interest rate and collected each month over the life of a loan. Upfront fees are one-time payments made by sellers upon loan delivery to an Enterprise that are similarly factored into the interest rate paid by the borrower and thus recouped by the seller. Upfront fees are converted to an annual g-fee equivalent by dividing the upfront fee by the expected number of years a loan will remain outstanding. Fannie Mae refers to upfront fees as "loan level price adjustments," while Freddie Mac refers to them as "credit fees in price" (FHFA, 2020).
[8] See Stegman and Cooperstein (2019).
[9] We will show later that the ARMs post 2012 are offered to high credit score borrowers, instead of the high-risk ARMs before 2007.
[10] Goodman, et al. (2022) examines the working of cross-subsidization in GSE pricing and offers suggestions to refine it to enhance house ownership in U.S.

market, GSE pricing can serve as the benchmark for lenders even if the latter plans to retain the mortgages on their balance sheets, and thus mortgages with higher G-fees relative to their credit risks can be more profitable for the lenders if kept on their balance sheets. We thus hypothesize that banks hold more ARMs and FRMs with higher CS/lower CLTV on their balance sheets.

Our analyses confirm that the mortgages retained on banks' balance sheets are mostly ARMs and medium-to-high CS and low CLTV FRMs during 2013-2019. The fractions of both ARMs and FRMs kept on banks' balance sheets have markedly been on the rise over time, and the interest rates on those retained loans are also lower than those sold to GSE.

We argue that banks' response to competition from non-banks, coupled with the existence of cross-subsidy in the G-fee, is a primary reason behind banks' decision to hold these low-risk loans, and we provide evidence from two aspects. First, during 2008-2012 when non-bank competition was still not very fierce, banks held only 4.3% of ARMs and 3.3% of 30-year FRMs, and banks held high-risk loans in this period. As non-bank competition increased, banks' holding pattern started to change and they began to hold more low-risk loans.

Second, we include among the explanatory variables non-bank share of origination at the national level and interact this variable with the CS and CLTV buckets. We find that these interactive terms in the 2013-2019 regressions on banks' retention behavior can completely absorb 1) the pattern of rising proportion of loans kept on banks' balance sheets over time, and 2) the positive coefficients of the high credit score and low CLTV buckets. These results suggest the rising non-bank competition during 2013-2019 is a primary reason behind our finding of banks' increasing holding of less risky conforming mortgages on their balance sheets.

The rest of the paper is organized as follows. We provide a brief literature review in Section 2 and describe the data in Section 3. We investigate banks' decision to keep a loan on their balance sheet versus selling the loan to GSEs in Section 4 and conclude in Section 5.

Section 2: Literature review and our hypotheses development

The existing literature mainly focuses on the reasons behind non-banks' expansion in the residential mortgage market and proposes two main explanations. The first factor is non-banks' expansion into the riskier residential mortgage market (for example, Lux and Greene (2015) and Buchak et al. (2018a)). The other factor is improvement in technological innovation (for instance, Buchak et al. (2018a) and Fuster et al. (2018)).[11]

The literature has also investigated banks' decision to finance mortgages using their balance sheet capacity. For example, Buchak et al. (2018b) finds that the conforming mortgage market fits the business model of non-banks, i.e., origination-to-distribution, and banks have an advantage in the activities that are balance sheet intensive, for instance, the jumbo loans. Banks' decision to keep or sell their mortgages has been empirically investigated before, for example, by Agarwal, Chang, and Yavas (2012), but their sample period is from 2004 to 2007, a period in which the landscape in the residential mortgage market is drastically different from the latest decade.

[11] The empirical evidence for the role played by mortgage pricing is more ambiguous. Using the 2010-2015 public GSE data, Buchak et al. (2020a) finds that, controlling for mortgage characteristics at origination, non-banks on average charge slightly higher interest rates than banks do, while Guo et al. (2022) find that non-banks charge lower interest rate than banks do for mortgages with similar ex post credit risks.

This paper aims to move the literature forward by focusing on conforming mortgages, where the prior literature has not identified obvious balance sheet capacity advantage for banks relative to non-banks. Further, we examine conforming loans originated post the latest housing market crisis, which is not covered in Agarwal, Chang, and Yavas (2012).[12]

Our argument starts from the cross-subsidization in GSE pricing schedule, which was put in place to expand homeownership by charging low-credit-risk borrowers a higher G-fee than justified by their mortgages' credit risks (see Stegman and Cooperstein (2019)). We reproduce in Panel A of Figure 3 of this paper Chart 4 of the FHFA (2020) report, which suggests that the profitability among ARMs is persistently positive and higher than that of FRMs. Further, Panels B and C of Figure 3, which are also re-produced from the FHFA (2020) report, suggest the G-fees are high among high CS and low CLTV loans relative to the costs of guaranteeing such loans.[13]

We can see from Fig.1 of Bartlett et al. (2021) that there is little differentiation in the GSE grid among mortgages with CS>=720. Further, the GSE grid stops at CS=740, and borrowers with CS>740 do not have lower G-fee even though their credit risks are supposedly lower than those with CS=740. In addition, the cross-subsidization also occurs across the CLTV groups, as Fig.1 of Bartlett et al. (2021) shows little variation in the first two rows from low CLTV to high CLTV. Not surprisingly, Panels B and C of Figure 3 clearly show positive returns to GSEs for loans with CS>720 and, especially, CLTV<70%.

[12] In addition, we are not aware of any study investigating the interest rates between bank-originated mortgages kept on their balance sheet versus those sold to GSEs.

[13] This pricing pattern is also documented in Figure 3 of FHFA (2014), which includes a comprehensive discussion of the framework of how G-fee is set considering the expected credit losses, unexpected losses and the resulting capital costs and operational costs.

Table 1 shows the sample breakdown by CS and CLTV buckets from the HMDA data among conforming mortgages originated (by both banks and non-banks) in 2018-2019. We can see that 68% of conforming mortgage borrowers have CS>=740, and 52% of the borrowers have CS>=740 and CLTV<80%. Therefore, the size of the market where banks might have competitive opportunities because of the GSE cross-subsidy is reasonably large.

Academic studies have discussed cross-subsidization between high and low risk borrowers (see, for example, Gerardi (2017), Stegman and Cooperstein (2019), and Tsai (2019)), but these papers do not assess the impact on lenders' behavior from such cross-subsidization. This G-fee cross-subsidization might present opportunities for banks, since the prevailing primary market mortgage rates (which should be heavily affected by the G-fee due to the dominating role of GSEs in U.S. mortgage market) might be high relative to credit risks for some product types or loan types, and it thus might be profitable for banks to originate and keep these loans on their balance sheets. They could set the same levels of interest rates for these loans as non-banks do and pocket the profits, lower the interest rates on these loans to boost their competitive position, or do both.

Although there is a suspicion among practitioners that the cross-subsidization might be motivating banks to retain their high-quality loans,[14] there is lack of empirical evidence on banks' strategies in response to the cross-subsidization. By contrast, Agarwal, Chang, and Yavas (2012) finds that banks generally retain higher risk mortgages in their portfolios; such a finding was based on mortgages originated between 2004-2007, when the competition from non-banks was at most modest.

[14] For example, https://www.urban.org/sites/default/files/publication/22841/413202-Guarantee-Fees-An-Art-Not-a-Science.PDF.

Following the discussion above, we have the following hypothesis for banks' mortgage retention strategy facing the competition from non-banks:

H1: Facing competition from non-banks in the conforming mortgage market, banks hold more loans on their balance sheets, particularly those for which the G-fees might be high relative to their credit risks, such as ARMs and FRMs with medium-to-high CS and/or low CLTV. Since non-bank competition largely took off after 2012, the hypothesis should hold after 2012.

H2: Inclusion of non-bank shares in the conforming mortgage market, which captures the competition from non-banks, should be able to largely explain banks' loan retention decision post 2012.

Section 3: Data description

Our analyses primarily utilize the FRB Y-14 First Lien data, which include all first-lien residential mortgages that are serviced by the largest 19 national banks. Covering both the bank-held and serviced loans that are updated monthly, the Y-14 data contain a rich set of borrower- and loan-level variables both at origination and for loan performance including delinquency status, loss mitigations, and liquidations, etc.

Since our focus is the conforming mortgage market, we first exclude non-conventional loans. We then exclude from our analysis all jumbo loans, that is, the mortgages with origination amount above the conforming loan limits. We obtain the conforming loan limit from FHFA and

merge these limits with the Y14 mortgage data by origination year, state/county, and number of units, as the conforming loan limits vary by these factors.[15]

The remaining conforming mortgages in the Y14 data come from different origination sources, and we exclude correspondent loans, loans for which servicing rights were purchased, and loans via bulk purchases or other unknown sources, because these loans are overwhelmingly sold to GSE and rarely kept on banks' balance sheet, as shown in Table A1 in the Appendix.

Further, we exclude from our analysis on banks' loan retention decisions mortgages that are not eligible for sale to GSE, as banks have to hold such loans on their balance sheets. These loans include interest only loans, negative amortization loans, balloon loans, low- or no-doc loans, and mortgages with prepayment penalties. We also exclude loans with servicing rights transferred within the 24 months since loan origination, and those with CLTV>97% or CS<620.[16]

Panel A of Table 2 shows the breakdown of the first-lien conforming residential mortgage originations during 2013-2019 by product type and their average kept rate during our sample period after the filters discussed above. Over two-thirds of the conforming loans in the Y14 data are 30-year FRMs (FRM30), and one quarter of the conforming loans are 15-year FRMs (FRM15). ARMs constitute roughly 4% of the conforming loans, and the remaining 1.5% are of miscellaneous types. Banks keep roughly 46% of the ARMs and half of the miscellaneous types by count on their balance sheet. The retention rates are much lower for the FRMs at 8.4% for FRM30 and 11.7% for FRM15.

[15] The historical conforming loan limits can be found at https://www.fhfa.gov/DataTools/Downloads/Pages/Conforming-Loan-Limits.aspx

[16] The largest observation loss is due to the exclusion of correspondent loans; the next largest contributor is exclusion of low- and no-doc loans which were prevalent in refinance loans in earlier part of our sample.

The numbers in Panel A of Table 2 only report the averages of the first-lien conforming residential mortgage originations during 2013-2019, and Panels A and B of Figure 4 depict the portion of ARMs and FRMs kept on banks' balance sheet over time by count and by balance, respectively. We can see that the retention rates of ARMs are multiple times higher than those for FRM15 and FRM30, and the differences in retention rates between FRM15 and FRM30 are rather minimal. Further, all three lines show upward trends, and the rise in the fraction of ARMs kept on banks' balance sheet has been especially noticeable, exceeding 70% in 2019. The proportions retained on banks' balance sheet have also doubled for both FRM15 and FRM30 from 2013 to 2019 by both count and balance.

We will focus on ARMs and FRM30s for the rest of the analysis. Following the filters afore-mentioned, we have 201,371 ARMs and 2,453,281 FRM30s originated from January 2013 to December 2019 for subsequent analyses. Panel B of Table 2 reports the distribution of ARMs and FRM30s by the CS and CLTV groups. We can see that about 78% of ARMs and roughly 65% of FRM30s in our sample are offered to borrowers with CS greater than 740. Approximately 8% FRM30 and less than 2.5% ARMs are offered to borrowers with CS lower than 680. Therefore, the ARMs originated during our sample period tend to have higher quality than FRM30s. Further, for both ARM and FRM30, higher proportions of refinance (both rate refinance and cash-out refinance) loans have low CLTVs than purchase loans.

Panel A of Table 3 shows the loan characteristics by ARM and FRM30 on a broader spectrum than just CS and CLTV,[17] [18] which again suggests that the ARMs in our sample are generally of higher credit quality than FRM30s. Compared with FRM30s, bank ARMs have

[17] Because of space limitations, we focus on FRM30 when analyzing fixed-rate mortgages.
[18] Note that a few other categories – home improvements, others, and unknowns – are not reported in the table.

higher CS, lower CLTV, and lower DTIs. Over half of the ARMs are refinance loans, and most of the refinance loans are rate refinance. Roughly 7% of the ARMs are from the wealth management channel, while only 2% of FRM30s are from wealth management. FRM30s are less likely held on balance sheets than ARMs with a kept rate of 14% vs. 49%. The interest rates are roughly 0.8 percentage points lower, and the origination amount is almost $100K higher among ARMs than among FRM30s. Further, we find that the delinquency rates among ARMs are substantially lower than FRM30s: the ARM 60 days past due or worse (DPD60) rates within 24 months after origination are only about one fourth of those of FRM30s. The same conclusion is drawn when DPD90+ is used to measure credit risk. On the other side, ARMs are more likely to be prepaid than FRM30s[19]

Section 4: Empirical results

Section 4.1: Banks' retention decision, 2013-2019

Panel B of Table 3 reports the summary statistics by whether the loan is retained or sold, for ARMs and FRM30s, respectively. We can see that, regardless of ARMs or FRM30s, the conforming loans kept on banks' balance sheet have higher CSs, lower CLTVs and much larger origination amounts than loans sold. The DTIs of the FRM30s kept on banks' balance sheets are lower than those sold, while the DTIs of the ARMs sold and kept are comparable at around 32.6%. Banks are also inclined to keep loans originated from wealth management and larger loans on their balance sheet. So, in general, loans retained on banks' balance sheets tend to be less risky in terms of origination variables, and the last rows of the table show that loans held on

[19] The number of observations for DPD60 and other performance variables is smaller than that for the loan or borrower characteristics because of the 24-month window – originations in 2018 and 2019 have to be dropped.

banks' balance sheets have lower delinquency rates, consistent with their loan and borrower characteristics.

The finding of banks' tendency to keep high CS and/or low CLTV loans is further illustrated in Figure 5, and we can see that the differences in the percentages kept on bank's balance sheet between different CS/CLTV buckets are substantial. Panel A of Figure 5 shows that, along the CS dimension, the retention rate for FRM30s with CS<680 is very low at 5%, and this number largely stays at 5% throughout our sample period. For the medium and high CS FRM30s, the retention rate was low in single digits in early 2013, but the retention rates for both groups grow continuously from 2013 to 2019. The retention rate for the medium CS group increases to over 15% by the end of 2019, and that of the high CS group exceed 20% in 2019. Panel B of Figure 5 shows the retention rate by CLTV groups among FRM30s with CS >=680, and this graph suggests that banks are not very interested in keeping loans with CLTV>80% even if the borrowers' CSs are reasonably high.

Panel C of Figure 5 depicts the retention rates among ARMs. We can see that the retention rates are close for medium and high CS ARMs, reaching 80% in mid-2019 before declining slightly. The retention rates for ARMs with CS<680 are volatile, which might be driven by the low number of loans in this group. The CS<680 line also deviates from those for high and medium CS loans starting from 2017, suggesting that banks' willingness to keep low CS ARMs on their balance sheets started to decline since 2017. Panel D of Figure 5 suggests that for medium and high CS borrowers, banks showed clear inclination in keeping low CLTV ARMs from late 2013 to mid-2016 relative to high CLTV ARMs, and such preference towards low CLTV ARMs declined slightly afterwards.

We next resort to regression method to investigate banks' decision to retain loans on their balance sheets. The econometric specification for the "Retained" regression is as follows:

$$Retained = \beta_G group(CS, CLTV) + \beta_X X + \mu_{orig\ year} + \mu_{state} + \mu_{lender} + \varepsilon \quad (1)$$

where "Retained" is an indicator variable that takes the value of 1 if a loan is kept on banks' balance sheet over the course of the loan's life, and 0 if the loan is sold to GSE.[20] Group(CS, CLTV) is nine groups by interacting borrower origination CS bands (<=679, 680-739, >=740), and CLTV bands (<=70%, 70.1-80%, >80%). The reference CS/CLTV group is CS>=740 and CLTV>80%, shown as csH_cltvH. CS 680-739 are represented as csM, CS <=679 as csL, CLTV 70.1-80% as cltvM, and CLTV <=70% as cltvL. *X* contains other borrower and loan characteristics, including DTI ratio bands, loan purpose (home purchase, rate refinance, cash-out refinance), occupancy (owner occupied, second home, or investment properties), property type (single family, townhouse, condo), loan origination source (retail, broker, wealth management), and loan origination amount. For the regression within ARMs, *X* also include ARMs product type dummies, including ARM2, ARM3, ARM5, ARM7, ARM10, and other ARMs.[21] Note that ARM2 refers to an ARM loan where the borrower pays an initial fixed rate in the first two years and then faces floating rates, and so on. Debt-to-income ratio is missing for a sizable portion of borrowers; we thus create indicator variables for missing DTI and create DTI groups using the 25th, 50th, and 75th percentiles as cut-off points. The loan amount variable has both large and small values; to help reduce the impact of extreme values, we create group indicator variables by utilizing the 5th, 10th, 25th, 50th, 75th, 90th, and 95th percentiles as cut offs.

[20] Analyses have shown that close to 95% of loans that are sold to GSE occurred during the first 6 months of a loan's life.

[21] Other ARMs include adjustable-rate loan where the loan term does not fall under any of the ARM categories (ARM 2, ARM 3, ARM 5, ARM 7, ARM 10). An example would be an Option ARM or ARM 15. Our data has few ARM2s and all of them are kept on banks' balance sheet.

$\mu_{orig\ year}$, μ_{state}, and μ_{lender} are fixed effects for origination year, state, and lender, respectively.

This regression is at the loan level, and the equation is estimated using Linear Probability Model (LPM).[22] The standard errors are corrected for clustering at lender*origination year-month level. The regression results are reported in Tables 4.

The first two columns report the regression results for bank retention of ARM loans. We can see from the first eight rows of the ARM model in Table 4 that the coefficients of the four medium-to- high CS and medium-to-low CLTV groups are all positive, indicating that banks are more likely to retain ARMs with CS>=680 and CLTV<=80%, compared to the reference group (i.e., CS>=740 and CLTV>80%). The coefficients are the most positive in the first two rows of the ARM model, suggesting that banks are particularly inclined to keep high CS and medium to low CLTV ARMs. The economic magnitude of the coefficient estimates is large. If a high CS ARM has its CLTV change from above 80% to 70.1-80%, its chance of being retained by banks increases by six percentage points on average. This magnitude is economically significant as Panel A of Table 3 shows that banks keep roughly 49% of their ARMs. If its CLTV falls to the low CLTV group, its probability of being retained by banks will rise by another one percentage point. The coefficients of the three low CS groups, as well as the medium CS and high CLTV group, may not be reliable or stable because of the low number of observations in these ARM groups as Table 2 Panel B suggests.

Not shown for brevity reason, we find that ARMs with missing DTI or higher DTIs are less likely to be kept on banks' balance sheets, and the coefficients of the >50% DTI groups are

[22] As robustness checks, we conducted estimation using logistic regression method; results are available upon request.

especially negative. Banks are less likely to retain rate refinance ARMs, loans backed by second homes or investment properties, or multi-unit properties, and more likely to keep loans originated from brokers or from wealth management, and loans with larger origination amounts. ARM loans with a fixed-rate period of 3 years are more likely to be kept than those with a longer fix-rate period; this can be rationalized since the bank faces lower interest rate risk in the former.

The last few rows of Table 4 report the year dummy variables, and loans originated in 2013 constitute the reference group. We can see that, for ARMs, the year dummy coefficients rise steadily from 0.07 in 2014 to 0.13 in 2019, and the coefficients are all statistically significant. In summary, the multivariate regression results show that even after controlling for loan characteristics, i) banks tend to retain generally less risky ARM loans during the sample period from 2013-2019, ii) banks' propensity to retain conforming ARMs on their balance sheet has been rising over time.

Columns 3 and 4 of Table 4 show the retention regression results for FRM30s. Like the ARM results, the coefficients of the medium- to-high CS and medium-to-low CLTV dummies are significantly positive, again suggesting that banks have a higher tendency to retain medium-to-high CS and medium-to-low CLTV FRM30s. Given the average retention rate of only 14% among FRM30, as reported in Table 3, the economic magnitude of the coefficients reported here is much larger than that of ARM. If a high CS FRM30 sees its CLTV fall from above 80% to below 70%, its chance of being retained on banks' balance sheets increases by seven percentage points, which is nearly 0.07/0.14=50%.

The coefficients of the low-to-medium CS and high CLTV loans are both significantly positive. This result might be driven by the fact that we only exclude loans with CLTV>97%. For some loan types, such as cash-out refinance and multi-unit property types, the CLTV limits

can be much lower than 97%, so some of these over-limit loans might be included in our data. However, the CLTV limits vary across different loan types and change over time, and it is difficult to completely exclude all these non-eligible loans. The coefficients of the low CS and medium-to-low CLTV dummies are both significantly negative, implying that banks are more likely to sell such FRM30s.

In unreported results, we find banks are less likely to retain FRM30s with high or missing DTIs or those backed by second homes or investment properties or multi-unit properties, and they are more likely to keep loans originated through the wealth management business and larger FRM30 loans in our sample.[23]

From the last few rows, we can see that the estimated coefficients of the year dummies for the FRM30 model turn from close to zero in 2014 and 2015 to significantly positive in 2016, and the coefficients increase further to 0.04 – 0.06 in 2017 – 2019. So, the over-time upward trend in bank's propensity to retain more FRM30 is clear.

In summary, we have uncovered evidence that banks have a higher propensity to keep low risk conforming loans, and they have become increasingly inclined in keeping such loans during the period 2013-2019. Such findings provide support to H1. We next conduct more analyses to understand the reasons behind such retention decisions.

Section 4.2 Banks' retention decisions, 2008-2013

[23] The effect of the wealth management indicator and large loan size indicators are especially large for both ARM and FRM30 in Table 4. Such a finding suggests that banks may have considered cross-selling or have used private information in their decision to retain loans on their balance sheet.

Our main dataset, FRB Y-14, started full year reporting in 2013. Examining banks' holding behavior in earlier period could help shed light on changes in factors that could have shaped banks' holding behavior 2013-2019. Non-bank competition did not take off until after 2012, as can be seen from Figure 1.

For this exercise, we use OCC Mortgage Metrics (MM) dataset, which is a loan-level dataset maintained by the largest bank servicers reporting loan performance starting in January 2008 and ending in 2016. A pre-cursor of the FRB Y-14 data, OCC MM data contain rich borrower- and loan-level characteristics and performance information. It differs from FRB Y-14 in that some servicers were dropped in Y14 and some other servicers appeared in Y14 but not in MM, but most of the servicers are the same in the two databases. Variable definitions also contain further refinements from MM to Y14 data, and OCC MM does not have a variable on loan purpose. We use the OCC MM data from 2008 to 2012.[24]

The summary statistics of the MM data are presented in Table A2, and we depict the retention rates in Figure 6. We can see that the retention rates of both ARM and FRM30s are low from 2010 to 2012. Further, the low CS lines were at the highest levels for both ARM and FRM30, followed by the medium CS lines, while the high CS lines are at the bottom of both graphs in Figure 6W largely throughout 2008-2012.[25] Figure 6 thus contrasts sharply with Figure

[24] Using the overlapping period of 2013-2016, we find noticeable differences in the Y14 and MM data from the same banks. Since Y14 data are available starting from 2013 following regulatory requirement, it is likely the incentive to maintain the MM data has declined and MM data quality might have decayed post 2013. However, even using MM 2013-2016 data, we see a higher propensity for banks to hold less risky loans over time. Such results are available but not reported for both brevity and the concern that the MM data might not be as reliable as the Y14 data post 2013.

[25] The slight discrepancy of the kept levels at the end of 2012 depicted in Figure 6 (using MM data) and the beginning of 2013 depicted in Figure 5 (using Y14 data) can be traced to i) the difference in reporting banks and in variable definitions across the MM data and the Y14 data, and ii) the low number of monthly ARM originations, and results can be sensitive to small changes in the sample.

5. These findings provide the first piece of evidence that the bank mortgage retention decisions might be significantly different before 2012 and after 2013.

Retention rate regression results from 2008-2012 are reported in Table 5 following model specification (1). To further help capture the variations over time, we separate the sample into two periods: 2008-2010 originations (or vintages) vs 2011-12 originations (with the last column for 2012 originations only). We aim to examine how banks' loan retention decision evolved during this period of 2008-2012.

Results for ARM are reported in Panel A of Table 5. We can see that during the period 2008-2010, the coefficients of the csH groups are negative. This result suggests banks held less of the high CS and medium-to-low CLTV ARMs relative to the omitted group (i.e., high CS but high CLTV loans). On the contrary, the coefficient estimates of the csL groups are significantly positive for the high and medium CLTV groups, suggesting that banks had a higher propensity to retain the riskiest ARMs from 2008 to 2010.

However, starting in 2011, the coefficient estimates of the csH groups and of the csM low-to-medium CLTV groups become significantly positive. These results suggest that banks' propensity to hold ARMs with medium-to-high CS and low-to-medium CLTVs increased starting from 2011 and continued into 2012. However, different from Table 4, the coefficients of the low CS groups continue to be significantly positive during 2011-2012, which explains why the three CS lines intertwine in Figure 6. Nevertheless, even after 2011, Panel A of Table 5 does not show a higher propensity of banks to retain high CS and low CLTV ARMs.

Results for FRM30s are reported in Panel B of Table 5. In column 1 of Table 5 Panel B, we find that, during the period 2008-2010, the coefficient estimate is 0.06 for the medium CS and

high CLTV group and 0.26 for the low CS and high CLTV group, suggesting that banks were more likely to hold onto medium or low CS FRM30s among all high CLTV loans. In addition, across the CS bands, the coefficient estimates decline with CLTV, which suggests that banks were more likely to hold high CLTV loans, controlling for the same CS. However, this pattern started to change for the 2011-12 vintages. For example, we see that the coefficient for the high CLTV and low CS group is 0.26 during 2008-2010 and 0.20 during 2012, indicating that, among the high CLTV loans and relative to high CS loans, the propensity to hold a low CS loan is 26 percentage points higher during 2008-2010 but only 20 percentage points higher during 2012.

In summary, banks' loan retention pattern from the 2008-2012 period reported in Table 5 is distinctively different from that during 2013-2019, as shown in Table 4, and banks tend to hold riskier loans before 2012, a finding consistent with that in Agarwal, Chang, and Yavas (2012) using data from 2004-2007. However, this tendency to hold risky loans on banks' balance sheet diminishes, and banks hold less risky loans after 2013. One of the key differences between the two periods is the increasingly higher level of non-bank competition as is illustrated in Figure 1. So, results from Table 5 provide the first piece of evidence suggesting that the different levels of non-bank competition might be a factor driving the difference in banks' loan retention behavior before and after 2012.

Section 4.3 Direct impact from non-bank share in loan retention regressions

In this exercise, we utilize the time-series variation in the share of non-banks in conforming loan originations (namely, PctNbk) as the measure of non-bank competition, and estimate the following augmented equation:

$$Retained = \beta_G gr(CS, CLTV) + \beta_n gr(CS, CLTV) * PctNbk_{t-1} + \beta_X X + \mu_{orig\ yr} + \mu_{st} + \mu_{ldr} + \varepsilon \quad (2)$$

The coefficient β_n examines whether banks' retention behavior regarding CS/CLTV increases with heightened non-bank competition, measured by PctNbk. The PctNbk, i.e., non-bank share, is computed using the conforming loans sold to GSE in the period of 2011-2019 using HMDA data. In Table 6, we report results using the one-year lag of the PctNbk variable among the right-hand side (RHS) variables.[26]

We can see from both models in Table 6 that the PctNbk interactive terms carry significantly positive coefficients for the high and medium CS and low-to-medium CLTV groups, and the coefficient is significantly negative for the low CS and high CLTV group in both the ARM and FRM30 models. The magnitude of these coefficient estimates is also large. If the non-bank share increases from 25th percentile of 27% to 75th percentile of 54%,[27] the likelihood of a high CS and low CLTV FRM30 being kept on banks' balance sheet increase by 6.75 percentage points. The coefficients of the PctNbk interactive terms are larger for the high CS groups among ARMs, suggesting that the impact of non-bank market share on banks' retention decision is even larger among ARMs than among FRM30s. Further, the coefficient estimates of the low CS- PctNbk interactive terms are large and negative for the high CLTV groups for both ARM and FRM30, implying that banks kept less high CLTV and low CS loans because of non-bank competition.

More importantly, the coefficient estimates of the stand-alone medium and high CS and low-to-medium CLTV groups become significantly negative or non-significant for both FRM30 and

[26] We provide results on using different lagging term and alternative granular level (county level) in robustness checks.
[27] = 0.25*(54% - 27%).

ARMs, suggesting that, without the non-bank competition interaction terms, banks would not have held more high or medium CS conforming residential mortgage loans on their balance sheets. Likewise, the coefficient estimates for the stand-alone terms of the low CS groups also increase substantially from Table 4 to Table 6, indicating that without non-bank competitions, banks would have retained more low CS loans than they have done. These results are consistent with the evidence presented in Table 5 from the period 2008-2012, when the non-bank competition was modest.

Finally, the coefficient estimates of the year dummies become largely non-significant in Table 6 for both ARMs and FRM30s, suggesting that, without the rising non-bank competition, banks would not have increased their holding of conforming residential mortgages over time during the 2013-2019 period.

Therefore, the non-bank market share variable almost entirely absorbs 1) banks' decision to retain less risky loans and 2) banks' rising propensity to hold more loans on their balance sheets over time during 2013-2019. Therefore, the results from Tables 5 and 6 provide strong evidence suggesting that non-bank competition in the conforming residential mortgage market might be a primary factor behind banks' loan retention decision during 2013-2019. Such findings render support to H2.

Section 4.4. Interest rates among loans sold vs. retained

We examine here the interest rate differentials between mortgages sold and retained by banks. The interest rates for loans retained on banks' balance sheets might be lower because the G-fees rose by 20 basis points during 2012, reflecting a 10- basis point increase mandated by the Congress to fund the 2012 payroll tax reduction and another 10-basis point increase mandated by

FHFA. Loans sold to GSE are subject to such surcharges, but loans kept on banks' balance sheet are not. Further, the overwhelming majority of loans retained on banks' balance sheets are high CS and low CLV loans, and the G-fees for these loans are high relative to the loans' credit risks. As a result, banks can afford to charge lower interest rates to the loans retained on their balance sheet than they would have charged if the loans were sold to GSE, and such lower interest rates can also facilitate banks' competitive position in the conforming mortgage market.

Table 7 present regression results on interest rates. The regression specification for the interest regression is:

$$Interest\ Rate = Retained + \beta_G group(CS, CLTV) + \beta_X X + \mu_{orig\ year-month} + \mu_{state} + \mu_{lender} + \varepsilon \quad (3)$$

The dependent variable of this equation is the origination interest rate, which is the annual percentage rate as specified on the mortgage note at the time of origination. The interest rate equation is cross-sectional at the loan level. Since the dependent variable is not binary, we estimate Equation (3) using ordinary least squared (OLS) regression with error terms corrected for clustering at lender and origination year-month level. The borrower and loan characteristics included on the RHS of the regression are the same as those in Equation (1). The reference CS/CLTV group is again CS >740 and CLTV>80%, and the regression results are reported in Table 7. We do not report coefficient estimates of the borrower and loan characteristics other than the CS and CLTV groups in this table because of space limitations and these results are available upon request.

The coefficients reported in the first column of Table 7 suggests that banks charge higher interest rates to ARMs with lower CSs and higher CLTVs. Controlling for the nine CS and

CLTV groups and other borrower and loan characteristics, the coefficient of the stand-alone retained indicator is -0.12, suggesting that the interest rates on ARMs kept by banks on their balance sheets have lower interest rates by roughly 12 basis points.

Regression results (unreported for brevity reason) also show that ARM10s have higher interest rates than other ARMs, which is most likely due to the positive term structure of mortgage interest rates. Further, ARMs with lower DTIs, ARMs backed by owner-occupied properties and those originated from wealth management have lower interest rates. After controlling for other origination variables, larger loan amounts are associated with lower interest rates.

The FRM30 interest rate regressions follow the same specifications as Equation (3), except that there are no product dummies on the RHS, as we only include FRM30s in this section. We can see from Table 7 that FRM30s with higher CSs and lower CLTVs have lower interest rates. Results on estimated coefficients on other explanatory variables (unreported for brevity reason) suggest that loans with lower DTIs have lower interest rates and purchase FRM30s and FRM30s for owner-occupied properties have lower interest rates. Different from ARMs, FRM30s from the wealth management business have higher interest rates. Interest rates declines with loan sizes, and FRM30s on townhouses or condos have higher interest rates than those on single family detached homes. Controlling for borrower and loan characteristics, the stand-alone retained indicator is -0.10, suggesting that the interest rates on FRM30s kept on banks' balance sheet are roughly 10 basis points lower than those sold to GSE overall.

As a result, some low-risk borrowers benefit from banks' loan retention decisions and obtain lower pricing on their mortgages.

Section 5: Additional analyses

5.1 By loan purpose

First, we run all the regressions by purchase, rate refinance, and cash-out refinance, separately. Results from such sub-sample analyses are largely qualitatively similar to those reported in Tables 4, 6, and 7.

5.2 By year during the 2013-2019 period

We have conducted all analyses discussed so far by each year for the 2013-2019 period. The results reported in the paper hold for most years. Specifically, banks are always more interested in retaining medium-to-high CS and low-to-medium CLTV loans, and the interest rates on these loans kept by banks are typically lower than those sold to GSEs.

5.3 Subsample analysis

Third, we group the 19 banks into two categories (the largest four banks by consolidated assets and the remainder) and conduct the analyses. The results show that both groups keep on their balance sheet medium-to-high CSs and low-to-medium CLTV loans and they charge lower interest rates to loans they hold.

We have also investigated similar analysis among FRM15s. Results from all these additional analyses are qualitatively similar to those reported from Tables 4, 6, and 7. [28]

5.4 Credit risk among loans sold vs. retained

[28] We conducted a series of additional robustness checks. First, First, we used DPD60 within 12 months as an alternative measure of default and also used DPD90 as an alternative default measure. Results show similar patterns. Second, in Kept regression including PctNbk, we utilized the variable lagged by two years and obtained similar findings.

We have investigated the credit risk differences (as measured by DPD60+ 2 year, DPD60+ 1 year and DPD90+ 1 year) between conforming residential mortgages sold and retained on banks' balance sheet. We find that, controlling for borrower and loan characteristics, the ARMs kept on banks' balance sheets have marginally lower credit risks than those sold to GSEs. For FRM30, there is little difference in the credit risks between FRM30s sold and kept on banks' balance sheets beyond what is captured by the borrower and loan characteristics.[29]

Section 6: Conclusion

Banks' share in the conforming residential mortgage origination market has been declining substantially since 2012, while non-banks have been taking up a progressively larger share over time. How do banks respond to non-bank competition in this market?

Given the cross-subsidy in the G-fees, it might be profitable for banks to retain on their balance sheets loan types for which the G-fees might be high relative to their credit risks, such as adjustable-rate mortgages and medium to high credit score and low-to-medium CLTV fixed-rate mortgages. We find evidence in support of this hypothesis, i.e., banks are increasingly inclined to retain such low-risk loans on their balance sheets. Further analyses indicate that the growing non-bank competition in the conforming residential mortgage market is a primary factor behind banks' rising loan retention rate during 2013-2019.

Therefore, banks respond to non-bank competition in the post-2012 conforming mortgage market by making use of both their balance sheet financing capability and the cross-subsidization

[29] FRM30 loans retained by banks have higher prepayment rates in high CLTV buckets, while the difference in prepayment rate in other buckets of FRM30s and ARMs are not obvious.

in the pricing schedules of the GSEs. Findings from the data suggest that mortgage borrowers also benefit from banks loan retention behavior and enjoy lower pricing on such loans.

References:

Agarwal, Sumit, Yan Chang, and Abdullah Yavas, 2012, Adverse selection in mortgage securitization, *Journal of Financial Economics* 105, 640-660.

Bartlett, Robert, Adair Morse, Richard Stanton, and Nancy Wallace, 2022, Consumer-lending discrimination in the FinTech Era, *Journal of Financial Economics*, 143 (1), 30-56.

Buchak, Greg, Gregor Matvos, Tomasz Piskorski, and Amit Seru, 2018a, Fintech, regulatory arbitrage, and the rise of shadow banks, *Journal of Financial Economics*, 130 (3), 453-483

Buchak, Greg, Gregor Matvos, Tomasz Piskorski, and Amit Seru, 2018b, Beyond the balance sheet model of banking: Implications for bank regulation and monetary policy, NBER Working paper 25149.

Federal Housing Finance Agency (FHFA), 2020, Fannie Mae and Freddie Mac single-family guarantee fees in 2019, https://www.fhfa.gov/AboutUs/Reports/ReportDocuments/GFee-Report-2019.pdf

FHFA, 2014, Fannie Mae and Freddie Mac guarantee fees: request for input, https://www.fhfa.gov/PolicyProgramsResearch/Policy/Documents/GfeeRFI060514F.pdf

Fuster, Andreas, Matthew Plosser, Schnabl Philipp, and James Vickery, 2019, The role of technology in mortgage lending, *The Review of Financial Studies* 32 (5), 1854-1899

Fuster, Andreas , Laurie Goodman, David Lucca, Laurel Madar, Linsey Molloy, and Paul Willen, 2013, The Rising Gap between Primary and Secondary Mortgage Rates, FRBNY Economics Policy Review, 19 (2): 17-39.

Gerardi, Kris, 2017, GSE mortgage insurance pricing, Working paper, Federal Reserve Bank of Atlanta.

Goodman, Laurie, Jim Parrott, Bob Ryan, and Mark Zandi, 2022, How Fannie and Freddie can use pricing to expand affordable homeownership, moodysanalytics.com

Guo, Zhengfeng, Hongyan Liang, Zilong Liu, and Xinlei Zhao, 2022, How do non-banks compete by pricing in the Government-Sponsored Enterprise (GSE) mortgage market? Working paper, Office of the Comptroller of the Currency.

Haughwout, Andrew, Donald Morgan, Michael Neubauer, Maxim Pinkovskiy, and Wilbert van der Klaauw, 2022, Nonconforming Preferences: Jumbo Mortgage Lending and Large Bank Stress Tests, Federal Reserve Bank of New York Staff Reports, no. 1029.

Jiang, Erica Xuewei (2022), Financing Competitors: Shadow Banks' Funding and Mortgage Market Competition, USC Marshall School of Business Research Paper.

Kim, You Suk, Karen Pence, Richard Stanton, Johan Walden, and Nancy Wallace, 2022, Nonbanks and mortgage securitization, *Annual Review of Financial Economics*, Vol 14.

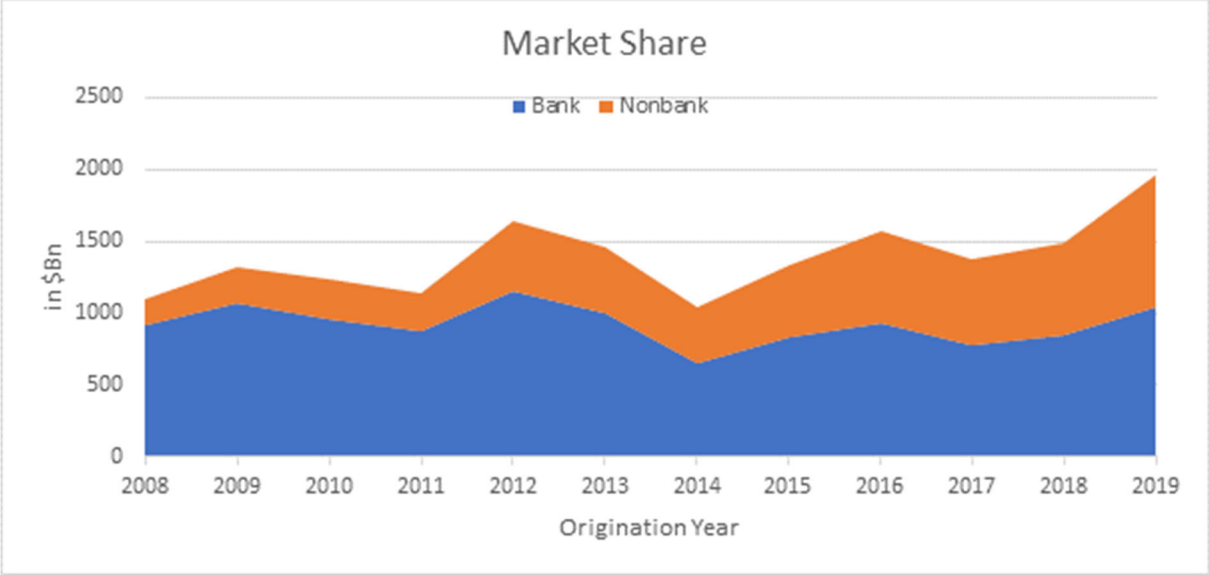
Lux, Marshall, and Robert Greene, 2015, What's behind the non-bank mortgage boom? M-RCBG Associate Working Paper Series No. 42.

Stegman, Michael, and Richard Cooperstein, 2019, A missing piece of the administration reform puzzle: How the GSEs generate cross-subsidies, working paper, Harvard University.

Tsai, Hsin-Tien, 2019, Insurance pricing and market structure: A study of GSE-securitized mortgage loans, Working paper, University of Berkeley.

Urban Institute Housing Finance Policy Center, Monthly Chartbook, urban.org/research/publication/housing-finance-glance-monthly-chartbook-january-2022

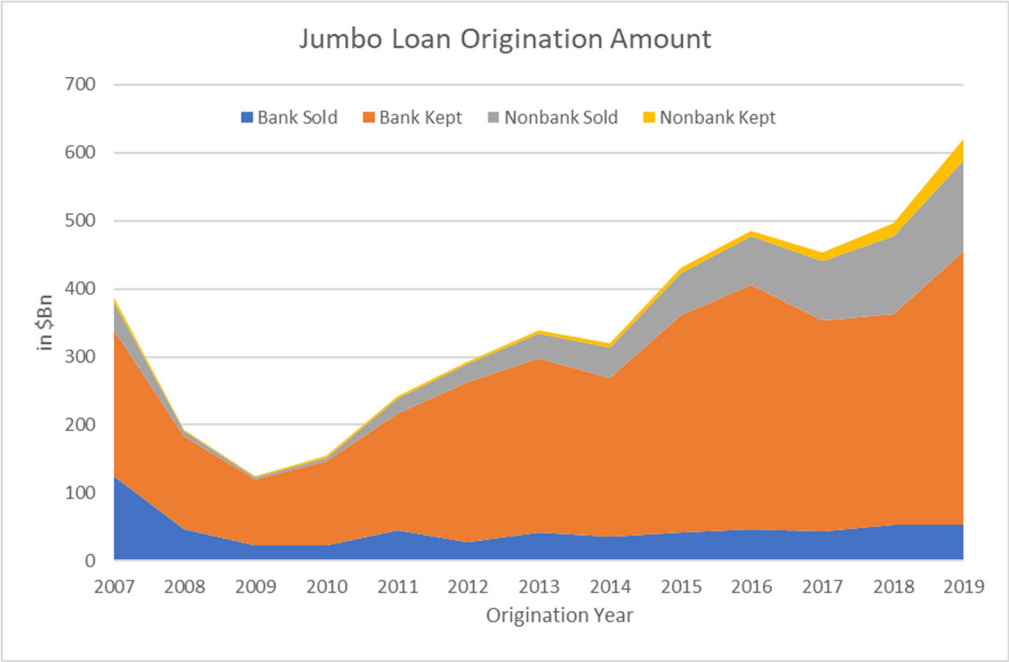
Figure 1: Residential Mortgage Origination Volume by Banks vs Non-Banks



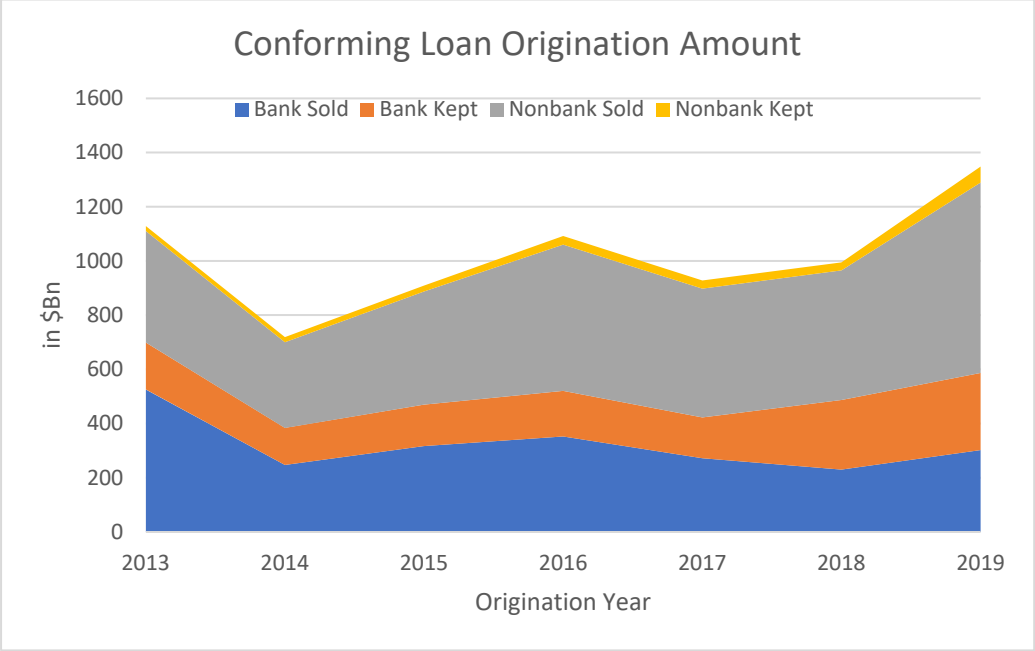
Notes: based on HMDA conventional residential mortgage loans originated in 2008-2019, excluding loans originated by credit unions.

Figure 2: HMDA Loan Distribution by Lender Type/Investor Type

Panel A: Jumbo Loans by Lender Type



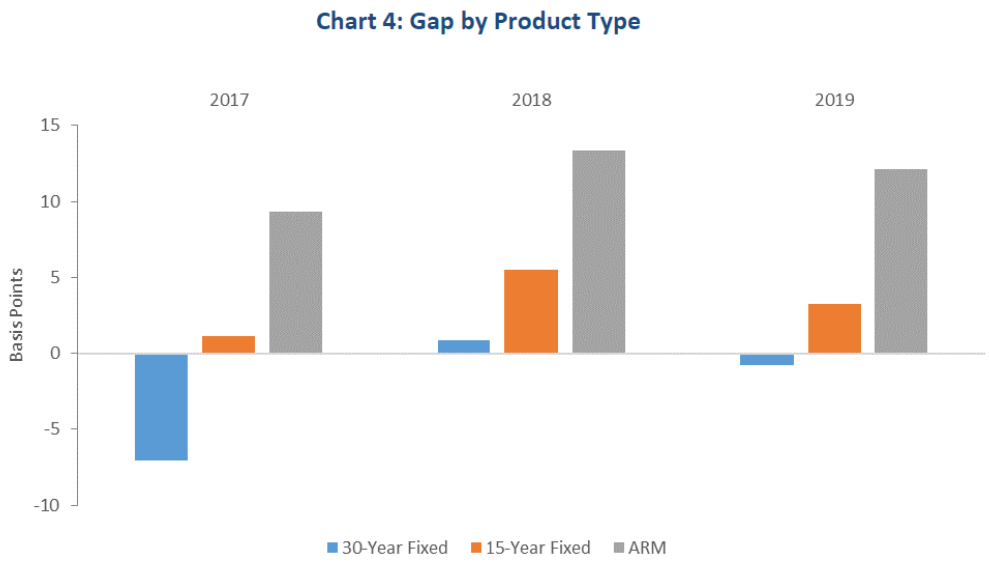
Panel B: Conforming Loans by Lender/Investor Type



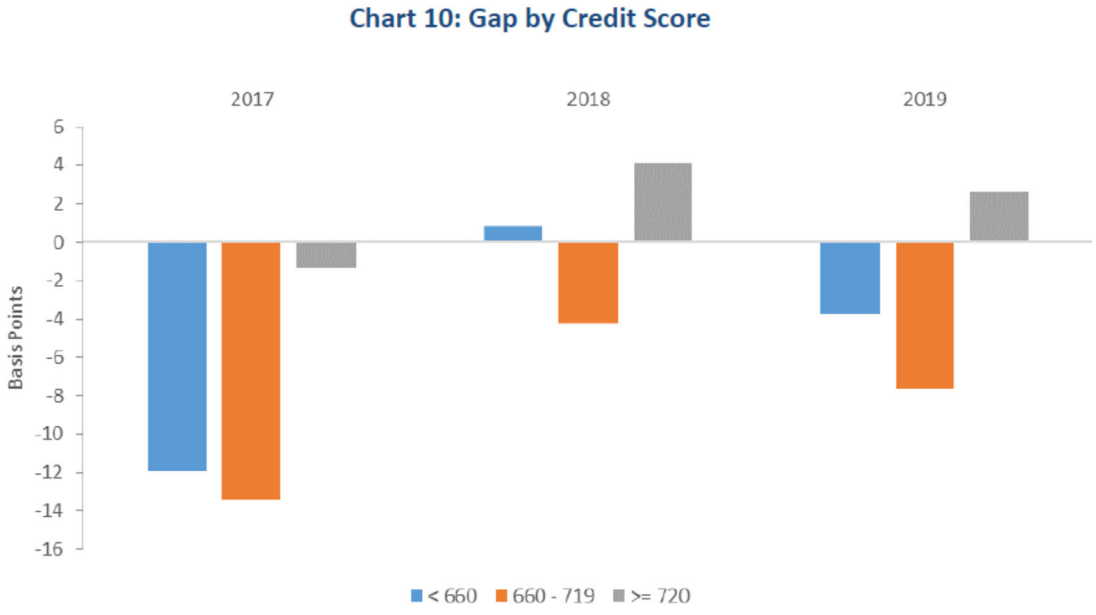
Notes: based on HMDA conventional loans originated in 2013-2019, excluding loans originated by credit unions. Conforming loans are conventional residential mortgages whose amount is not above the amount threshold for it to be eligible for sale to GSEs; jumbo loans are conventional residential mortgages with amount exceeding the limit. Bank Sold are mortgages originated and sold by banks in the HMDA report year. Bank Kept are mortgages originated and kept by banks in the HMDA report year.

Figure 3: GSE Guarantee Fees

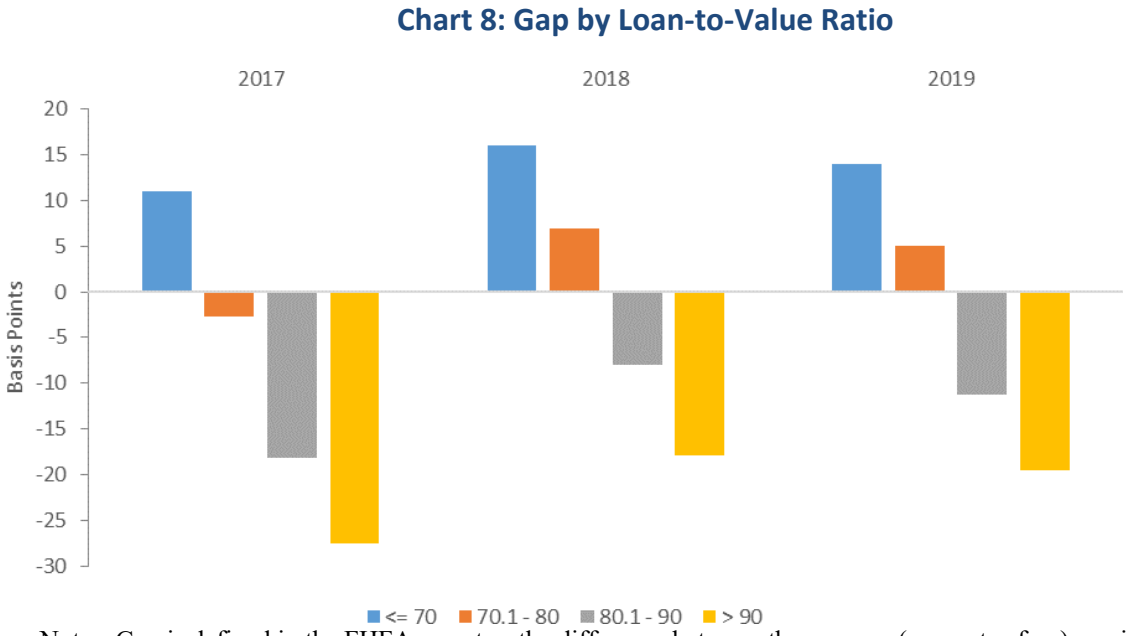
Panel A:



Panel B:



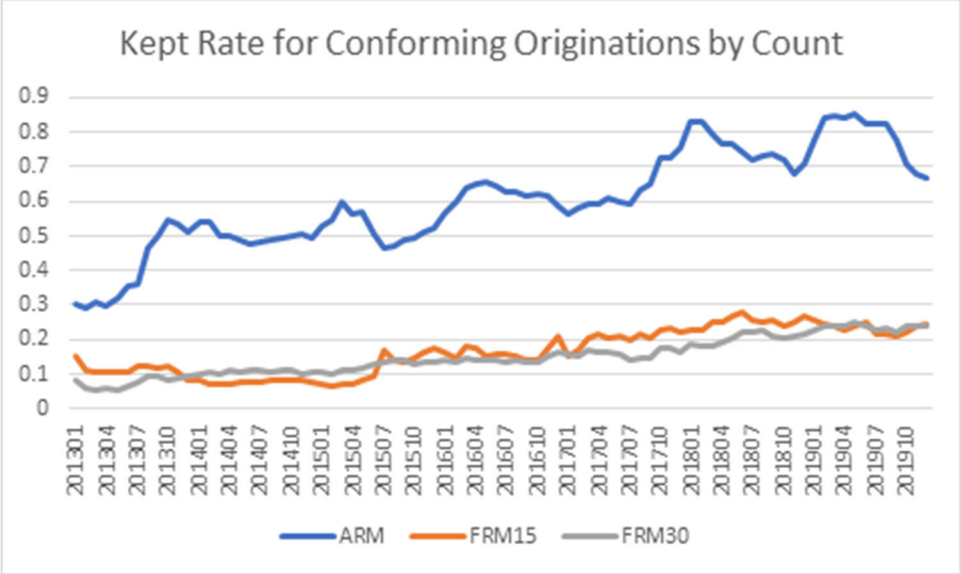
Panel C:



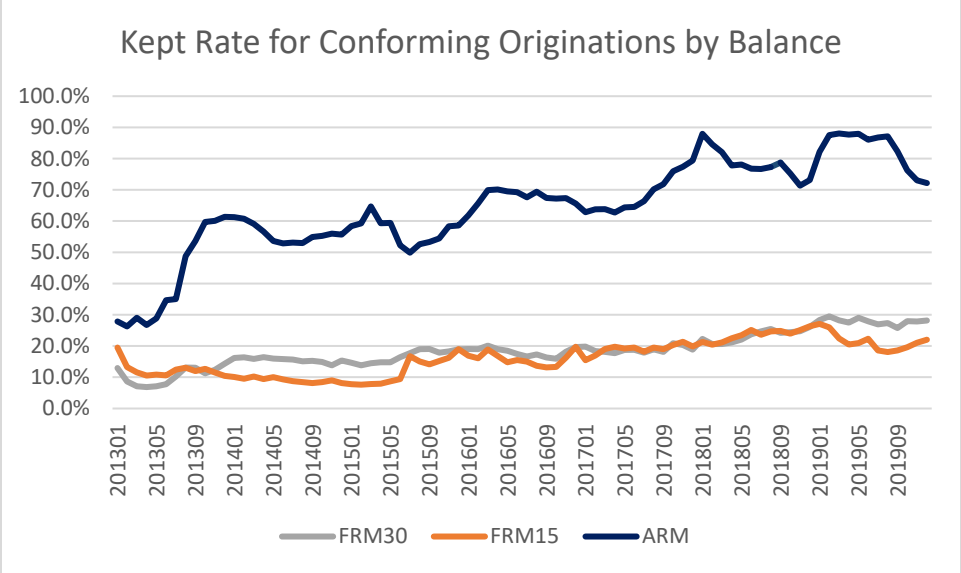
Notes. Gap is defined in the FHFA report as the difference between the revenue (guarantee fees) received and the estimated cost. The gap thus serves as the measure of estimated profitability of the loan acquisition based on the cost of its risk and the revenue charged.
Source: https://www.fhfa.gov/PolicyProgramsResearch/Policy/Documents/GfeeRFI060514F.pdf

Figure 4: Kept Rate for Conforming Loans by Product Type

Panel A: by Loan Count



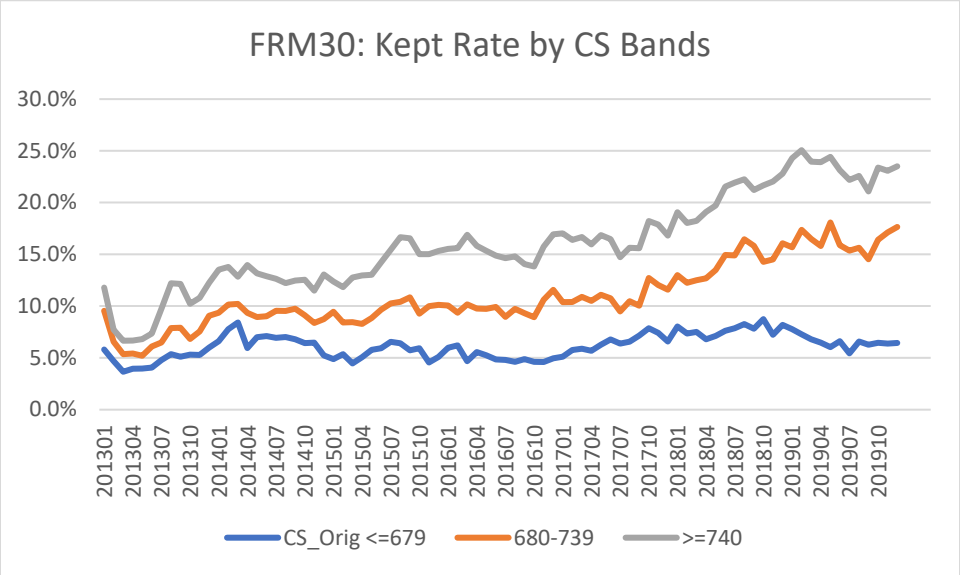
Panel B: by Loan Balance



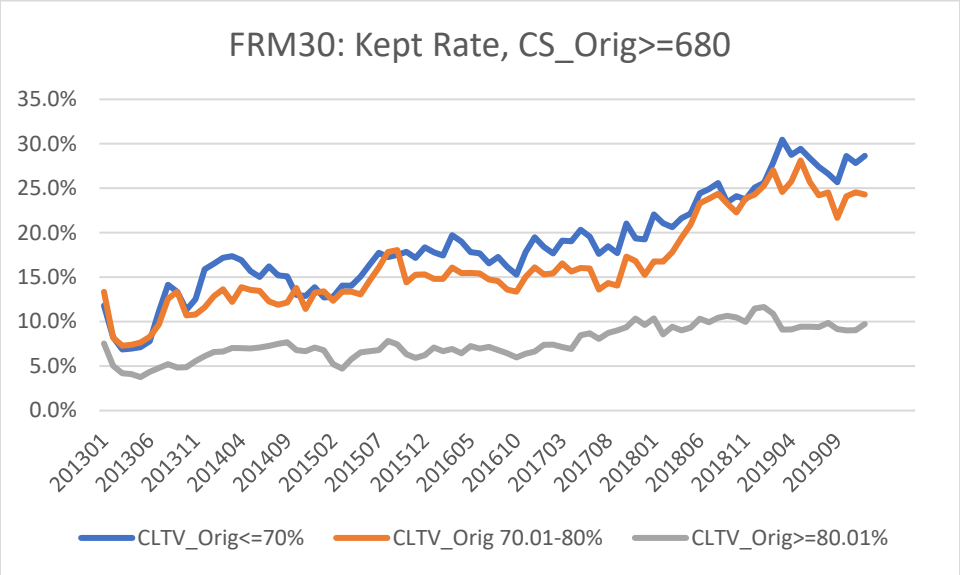
Notes: FRB Y-14 First Lien data. Conventional conforming originations 2013-2019. Kept equals 1 if the originated conforming loan is kept on the lender's balance sheet, and 0 if sold to GSE.

Figure 5: Kept Rate by CS and CLTV Bands

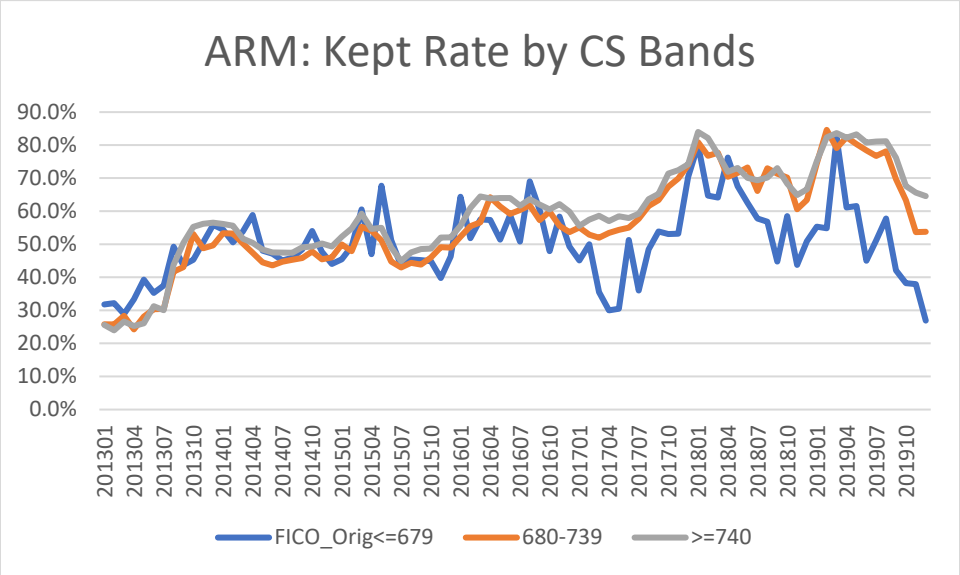
Panel A: FRM30 Kept Rate by CS Bands



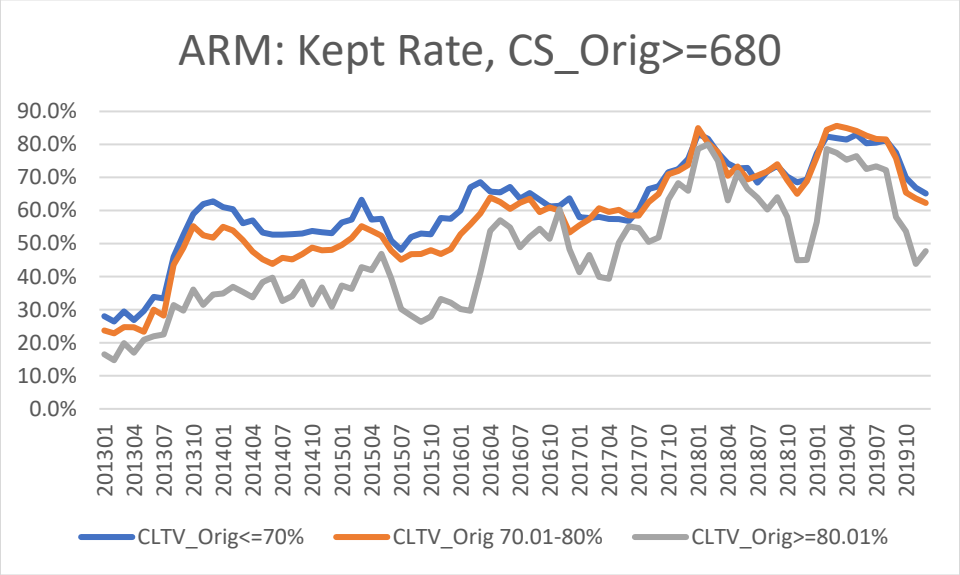
Panel B: FRM 30 Kept Rate by CLTV Bands, Credit Score (CS_Orig)>=680



Panel C: ARM Kept Rate by CS Bands



Panel D: ARM Kept Rate by CLTV Bands, CS>=680

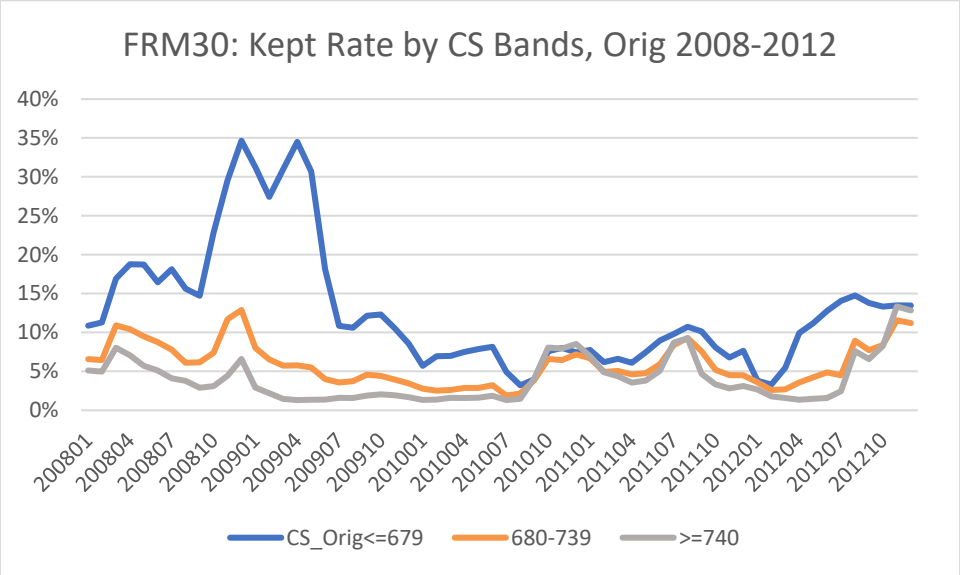


Data source: FRB Y-14 first-lien.

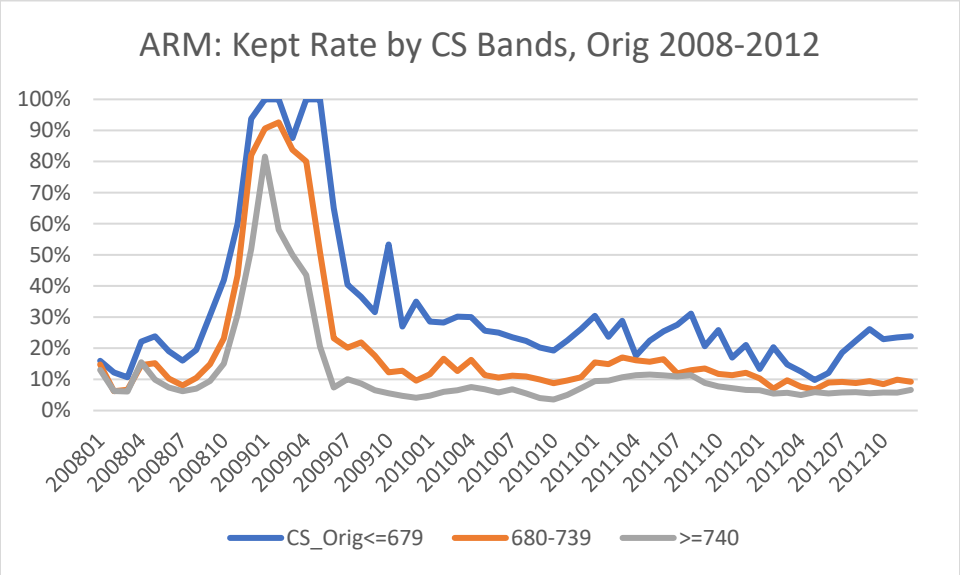
Notes: Removed loans with source channels being correspondent, servicing right purchased, or bulk purchases, borrowers of credit score lower than 620 or CLTV greater than 97%, interest only, balloon, negative amortization, low- or no-doc loans, or loans with prepayment penalty. Performance is measured on loans with 24 months of performance, those with last performance being servicing right transferred are dropped. Reported are percent of conforming loan originations that are kept on balance sheet by count.

Figure 6: Kept Rate by CS and CLTV Bands, Originations 2008-2012

Panel A: FRM30 Kept Rate by CS Bands



Panel B: ARM Kept Rate by CS Bands



Data source: OCC Mortgage Metrics Origination years 2008-2012. Removed loans with source channels being correspondent, servicing right purchased, or bulk purchases, borrowers of credit score lower than 620 or CLTV greater than 97%, interest only, balloon, negative amortization, or loans with prepayment penalty. Reported are percent of conforming loan originations that are kept on balance sheet by count.

Table 1: Breakdown of all conforming loans originated during 2018-2020

Credit score	CLTV			
	0-60	60-75	75-80	80-97
620-679	2%	2%	1%	2%
680-719	3%	4%	3%	5%
720-739	2%	3%	2%	4%
740-950	19%	18%	15%	16%

Notes: HMDA 2018-2020 conventional conforming 1st lien single family purchase or refinance loans, sold to GSE or kept on bank balance sheets, excluding HELOC, reverse mortgage loans, interest-only, balloon, negative amortization, credit union originations, or loans with invalid credit score, CLTV, and DTI.

Table 2: Loan characteristics by loan product and purpose

Panel A				
Loan product types	Count	kept		
ARM	468,275	45.90%		
FRM15	2,765,932	11.70%		
FRM30	7,597,689	8.40%		
Other	183,768	49.40%		
All	11,015,664	11.50%		
Panel B	All Purposes	Purchase	Rate Refi	CashOut Refi
	Frequency (%)			
ARM				
CS 740+; CLTV>80% (csH_cltvH)	8.7	13.9	5.4	2.3
CS 740+; CLTV 70.1%-80% (csH_cltvM)	30.8	41.8	22.1	20.5
CS 740+; CLTV<=70% (csH_cltvL)	38.3	23.3	50.3	52.4
CS 680-739; CLTV>80% (csM_cltvH)	2.9	4.6	2.1	0.6
CS 680-739; CLTV 70.1%-80% (csM_cltvM)	8.2	9.3	7.3	7.2
CS 680-739; CLTV<=70% (csM_cltvL)	8.8	5.0	10.5	14.0
CS<=679; CLTV>80% (csL_CltvH)	0.2	0.3	0.2	0.1
CS<=679; CLTV 70.1%-80% (csL_CltvM)	0.7	0.9	0.5	0.6
CS<=679; CLTV<=70% (csL_CltvL)	1.4	0.9	1.5	2.3
Observations	201.371	91,693	57,836	47,051
FRM30				
CS 740+; CLTV>80% (csH_cltvH)	17.0	24.9	11.0	3.4
CS 740+; CLTV 70.1%-80% (csH_cltvM)	27.1	31.1	21.8	21.5
CS 740+; CLTV<=70% (csH_cltvL)	21.6	12.0	34.5	33.9
CS 680-739; CLTV>80% (csM_cltvH)	9.0	13.7	5.6	1.0
CS 680-739; CLTV 70.1%-80% (csM_cltvM)	10.2	8.8	9.3	14.3
CS 680-739; CLTV<=70% (csM_cltvL)	7.4	2.8	10.8	15.4
CS 679-; CLTV>80% (csL_cltvH)	2.2	3.4	1.4	0.2
CS<=679; CLTV 70.1%-80% (csL_cltvM)	2.7	2.3	2.2	4.0
CS<=679; CLTV<=70% (csL_cltvL)	2.7	0.9	3.5	6.4
Observations	2,453,281	1,388,413	443,570	575,012

Source: FRB Y-14 first lien 2013-2019 originations. Panel A includes all observations; Panel B contains data for regression analyses only.

Table 3: Summary statistics for regression sample
Panel A: by ARM vs. FRM30

	ARM		FRM30	
	Mean	Std. Dev	Mean	Std. Dev
Proportion kept	0.49	0.50	0.14	0.35
Interest rate (in %)	3.35	0.54	4.19	0.58
Origination loan amount (in $1,000)	347	162	249	140
Credit score at origination	766	37	753	45
CLTV at origination	0.66	0.18	0.74	0.17
With a 2nd lien	0.09	0.29	0.07	0.26
DTI at origination (in %)	32.49	10.07	34.17	9.44
Purpose				
Purchase	0.46	0.50	0.57	0.50
Rate refi	0.29	0.45	0.18	0.38
Cash-out refi	0.23	0.42	0.23	0.42
ARM3	0.01	0.09		
ARM5	0.25	0.43		
ARM7	0.47	0.50		
ARM10	0.24	0.43		
ARM Other	0.03	0.16		
Occupied				
By owner	0.84	0.36	0.87	0.34
2nd home	0.09	0.28	0.05	0.22
Investment property	0.07	0.25	0.08	0.27
Loan source:				
Retail	0.89	0.32	0.92	0.27
Broker	0.04	0.20	0.05	0.22
Wealth management	0.07	0.25	0.02	0.14
Property:				
Single family	0.72	0.45	0.81	0.39
Townhouse	0.04	0.20	0.04	0.20
Condo	0.22	0.41	0.12	0.32
Observations	201,371		2,453,281	
DPD60+ (24 month)	0.06%	2.48%	0.26%	5.10%
DPD90+ (24 month)	0.04%	2.02%	0.18%	4.21%
Prepaid (24 month)	19.96%	39.97%	12.19%	32.72%
Unemployment Rate (%)	4.27	1.34	4.35	1.48
HPI percentage change (%)	1.11	0.08	1.11	0.07
Observations	146, 754		1,598,171	

Panel B: Mean values by kept or sold (to GSE) of FRM30 and ARM

	ARM		FRM30	
	Sold	Kept	Sold	Kept
Interest rate (in %)	3.32	3.38	4.22	4.04
Origination loan amount (in $1,000)	285	413	233	344
Credit score at origination	766	767	752	763
CLTV at origination	0.66	0.65	0.75	0.70
DTI at origination (in %)	32.39	32.60	34.33	33.22
Purpose:				
Purchase	0.46	0.45	0.57	0.54
Rate or term refi	0.25	0.32	0.18	0.21
Cash-out refi	0.26	0.20	0.24	0.22
ARM3	0.00	0.01		
ARM5	0.30	0.20		
ARM7	0.48	0.46		
ARM10	0.20	0.29		
ARM Other	0.02	0.04		
Occupied:				
By owner	0.85	0.84	0.87	0.86
2nd home	0.09	0.09	0.05	0.06
Investment property	0.06	0.08	0.08	0.08
Loan source				
Retail	0.98	0.79	0.94	0.82
Broker	0.02	0.07	0.05	0.02
Wealth management	0.00	0.14	0.01	0.11
Property:				
Single family	0.78	0.66	0.82	0.76
Townhouse	0.03	0.06	0.05	0.02
Condo	0.18	0.26	0.11	0.18
Observations	103,275	98,096	2,104,680	348,601
DPD60+ (24m)	0.07%	0.05%	0.28%	0.16%
DPD90+ (24m)	0.05%	0.03%	0.19%	0.11%
Prepaid (24m)	20.57%	19.22%	12.17%	12.82%
Observations	84,500	61,930	1,386,489	203,796

Source: FRB Y-14 first-lien. Notes: Removed loan with source channels being correspondent, servicing right purchased, or bulk purchases and borrowers of credit score lower than 620 or CLTV greater than 97%, interest only, balloon, negative amortization, low- or no-doc loans, or loans with prepayment penalty. Performance is measured on loans with 24 months of performance, those with last performance being servicing right transferred are dropped.

Table 4: Loan retention regressions for 2013-2019 originations

	ARM		FRM30	
	Coef.	Std. Err.	Coef.	Std. Err.
csH_cltvM	0.06***	(0.00)	0.05***	(0.00)
csH_cltvL	0.07***	(0.01)	0.07***	(0.00)
csM_cltvH	-0.02***	(0.00)	0.01***	(0.00)
csM_cltvM	0.03***	(0.01)	0.03***	(0.00)
csM_cltvL	0.02***	(0.01)	0.03***	(0.00)
csL_cltvH	0.02	(0.02)	0.05***	(0.00)
csL_cltvM	0.04***	(0.01)	-0.00**	(0.00)
csL_cltvL	-0.04***	(0.01)	-0.02***	(0.00)
Origination year: 2014	0.07***	(0.02)	0.00	(0.02)
Origination year: 2015	0.05***	(0.02)	0.00	(0.01)
Origination year: 2016	0.10***	(0.01)	0.02*	(0.01)
Origination year: 2017	0.12***	(0.02)	0.04***	(0.01)
Origination year: 2018	0.14***	(0.02)	0.06***	(0.01)
Origination year: 2019	0.13***	(0.02)	0.05***	(0.01)
Intercept	0.49***	(0.04)	0.21***	(0.02)
Observations	198,025		2,368,858	
R-squared	0.577		0.355	

Source: FRB Y-14 first-lien originations 2013-2019. All columns include borrower/loan characteristics, state fixed effects (FE), lender FE, origination year FE. Standard errors are clustered lender*origination year-month level. ***, **, * represent statistical significance at the 1 , 5 , and 10 percent levels, respectively.

Table 5: Loan retention regression for 2008-2012 originations

Panel A: ARM

	2008-10		2011-12		2012	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
csH_cltvM	-0.03**	(0.01)	0.02***	(0.00)	0.02***	(0.00)
csH_cltvL	-0.03*	(0.01)	0.03***	(0.00)	0.02***	(0.00)
csM_cltvH	-0.01	(0.01)	0.00	(0.00)	0.00	(0.00)
csM_cltvM	-0.03**	(0.01)	0.02***	(0.01)	0.01**	(0.01)
csM_cltvL	-0.02	(0.01)	0.03***	(0.01)	0.02***	(0.01)
csL_cltvH	0.04*	(0.02)	0.11**	(0.05)	0.09*	(0.05)
csL_cltvM	0.03*	(0.02)	0.03**	(0.02)	0.04*	(0.02)
csL_cltvL	-0.01	(0.02)	0.02	(0.01)	0.01	(0.01)
Intercept	1.05***	(0.05)	0.01	(0.02)	0.05	(0.04)
Observations	49,824		83,595		40,824	
R-squared	0.093		0.12		0.092	

Panel B: FRM30

	2008-10		2011-12		2012	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
csH_cltvM	-0.04***	(0.00)	-0.04***	(0.01)	-0.03***	(0.01)
csH_cltvL	-0.01	(0.02)	0.00	(0.02)	-0.03**	(0.01)
csM_cltvH	0.06***	(0.01)	0.06***	(0.01)	0.04***	(0.01)
csM_cltvM	-0.03***	(0.00)	-0.04***	(0.01)	-0.03***	(0.01)
csM_cltvL	-0.01	(0.01)	0.00	(0.02)	-0.03***	(0.01)
csL_cltvH	0.26***	(0.02)	0.33***	(0.03)	0.20***	(0.02)
csL_cltvM	0.02***	(0.00)	-0.02***	(0.00)	-0.01**	(0.01)
csL_cltvL	0.00	(0.01)	0.01	(0.02)	-0.02**	(0.01)
Intercept	1.37***	(0.06)	-0.31***	(0.04)	-0.03	(0.03)
Observations	1,016,694		611,206		318,256	
R-squared	0.178		0.15		0.215	

Data source: OCC Mortgage Metrics, originations 2008-2012. All columns include borrower/loan characteristics, state fixed effects (FE), lender FE, origination year FE. Standard errors are clustered at lender*origination year-month level. ***, **, * represent statistical significance at the 1, 5, and 10 percent levels, respectively.

Table 6: Loan retention regressions for 2013-2019 originations, controlling for nonbank share

	ARM		FRM30	
	coefficient	Std. Err.	coefficient	Std. Err.
csH_cltvM	-0.00	(0.02)	-0.05***	(0.01)
csH_cltvL	-0.03	(0.02)	-0.06***	(0.02)
csM_cltvH	0.05**	(0.02)	0.05***	(0.01)
csM_cltvM	-0.06***	(0.02)	-0.08***	(0.01)
csM_cltvL	-0.05***	(0.02)	-0.09***	(0.02)
csL_cltvH	0.27***	(0.06)	0.18***	(0.02)
csL_cltvM	0.15***	(0.04)	0	(0.01)
csL_cltvL	0.04	(0.03)	-0.03**	(0.01)
PctNbk*csH_cltvH	0.20***	(0.07)	-0.04	(0.03)
PctNbk*csH_cltvM	0.34***	(0.07)	0.20***	(0.04)
PctNbk*csH_cltvL	0.44***	(0.08)	0.25***	(0.04)
PctNbk*csM_cltvH	0.05	(0.07)	-0.12***	(0.04)
PctNbk*csM_cltvM	0.41***	(0.08)	0.20***	(0.03)
PctNbk*csM_cltvL	0.37***	(0.07)	0.23***	(0.03)
PctNbk*csL_cltvH	-0.39***	(0.14)	-0.30***	(0.06)
PctNbk*csL_cltvM	-0.07	(0.09)	-0.05**	(0.02)
PctNbk*csL_cltvL	0.00	(0.00)	0.00	(0.00)
Origination year: 2014	0.04**	(0.02)	-0.01	(0.02)
Origination year: 2015	-0.02	(0.02)	-0.03	(0.02)
Origination year: 2016	0.02	(0.02)	-0.01	(0.01)
Origination year: 2017	0.03	(0.02)	0.01	(0.02)
Origination year: 2018	0.03	(0.02)	0.02	(0.02)
Origination year: 2019	0.01	(0.03)	0.01	(0.02)
Intercept	0.47***	(0.04)	0.26***	(0.02)
Observations	198,025		2,368,858	
R-squared	0.578		0.357	

Source: FRB Y-14 first-lien originations 2013-2019. All columns include borrower/loan characteristics, state fixed effects (FE), lender FE, origination year FE. Standard errors are clustered at lender*origination year-month. ***, **, * represent statistical significance at the 1, 5, and 10 percent levels, respectively.

Table 7: Interest rate regressions for 2013-2019 originations

Dep Var: rate (in bps)	ARM		FRM30	
	coefficient	Std. Err.	coefficient	Std. Err.
Retained	-0.12***	(0.01)	-0.10***	(0.01)
csH_cltvM	-0.08***	(0.00)	-0.07***	(0.00)
csH_cltvL	-0.18***	(0.01)	-0.17***	(0.00)
csM_cltvH	0.10***	(0.01)	0.13***	(0.00)
csM_cltvM	0.07***	(0.01)	0.10***	(0.00)
csM_cltvL	-0.08***	(0.01)	-0.09***	(0.00)
csL_cltvH	0.44***	(0.04)	0.37***	(0.02)
csL_cltvM	0.32***	(0.01)	0.46***	(0.01)
csL_cltvL	0.08***	(0.01)	0.08***	(0.01)
Intercept	3.20***	(0.09)	3.69***	(0.16)
Observations	200,191		2,440,355	
R-squared	0.613		0.569	

Source: FRB Y-14 first-lien originations 2013-2019. All columns include borrower/loan characteristics, state fixed effects (FE), lender FE, origination year-month FE. Standard errors are clustered at lender*origination year-month level. ***, ** and * represent statistical significance at the 1, 5, and 10 percent levels, respectively.

Appendix

Table A1: Break-down of Y14 conforming loans by origination sources

Loan source type	Count	Kept rate
Retail	5,920,926	15.4%
Broker	308,870	12.4%
Correspondent	3,408,962	3.6%
Servicing right purchased	840,373	1.1%
Bulk purchase	204,174	9.6%
Wealth Management	170,096	81.4%
Unknown	160,985	16.3%
All	11 million	11.5%

Notes: Conventional conforming originations 2013-2019. Keep equals 1 if the originated (conforming) loan is kept on the lender's balance sheet, 0 if sold to GSE. Source: FRB Y-14 first lien data.

Table A2: Summary statistics for regression sample using OCC MM data

	ARM		FRM30	
Mean Values of	Sold	Kept	Sold	Kept
Interest rate (in %)	3.38	3.82	4.76	4.57
Origination loan amount (in $1,000)	265	271	215	187
Credit score at origination	769	763	758	735
CLTV at origination	0.64	0.61	0.70	0.74
DTI at origination (in %)	30.69	32.35	33.87	33.74
ARM3	0.01	0.02		
ARM5	0.41	0.35		
ARM7	0.28	0.19		
ARM10	0.15	0.11		
Owner Occupied	0.85	0.76	0.84	0.90
Loan Source: Retail	0.90	0.92	0.85	0.60
Broker	0.10	0.08	0.15	0.40
1-unit	0.99	0.98	0.98	0.98
2-units	0.01	0.01	0.01	0.01
>2 units	0.00	0.01	0.01	0.00
Observations	151,742	6,861	1,513,970	124,527

Source: OCC Mortgage Metrics loan level data for first-lien mortgage originations, 2008-2013. Notes: Removed loan with source channels being correspondent, servicing right purchased, or bulk purchases and borrowers of credit score lower than 620 or CLTV greater than 97%, interest only, balloon, negative amortization, or low- or no-doc loans.