The Edge of Banks is Still Sharp: Evidence from Market Segmentation in the Conforming Loan Market *

Peter Han †

July 27, 2023

Abstract

Contrary to the prevalent perception of the conforming loan market as intensely competitive and homogeneous, my research reveals substantial market segmentation. I identify that for mortgages with comparable ex-ante risk attributes, banks that finance these mortgages using their balance sheets impose a premium varying between 9.1 and 12.8 basis points. I construct a model of lender competition that elucidates two key conditions necessary for yielding a positive premium: imperfect competition and heterogeneity in the cross-elasticity of demand among lenders. My empirical findings align with the first condition, as I illustrate that the premium elevates by 5.5 to 6.5 basis points with every standard deviation increase in local market concentration. Concurrently, in keeping with the second condition, I show that the premium ascends by 12.0 to 12.4 basis points for every ten-percentage point rise in local market demand for mortgages.

Keywords: Mortgage pricing, portfolio lending, market concentration, cross-elasticity of demand, product differentiation

*I am indebted to George Pennacchi, Charles Kahn, Jialan Wang, Julia Fonseca, and Yufeng Wu for their continuous support and guidance. I am also grateful to Joshua Pollet, Tatyana Deryugina, Qiping Xu, Heitor Almeida, Jaewon Choi, Haoyang Liu, Filipe Correia, Elizaveta Sizova, Justin Mohr, Don Fullerton, and participants at the Gies College of Business Brownbag and CES North American Conference for their valuable suggestions.

†University of Illinois at Urbana-Champaign, 1206 South 6th Street, Champaign, IL 61820, weitong2@illinois.edu

I Introduction

”Banking is necessary, banks are not.”

– Bill Gates, 1994

The statement made decades ago could not be more relevant today given the burgeoning digital transformation of the finance industry. With the advancement of technology, more aspects of financial service have moved from brick-and-mortar physical locations to online, where the new generation of nonbank intermediaries has a first-mover advantage. One market that has faced nonbank erosion is the conforming loan market. As of 2017, the market share of the bank lenders in this market has dropped from around twenty percent in 2007 to about sixty percent in 2017[1].

In this paper, I investigate the banks' competitiveness in the face of non-bank competition in the conforming loan market. This sector, widely considered highly competitive (Fuster, Goodman, Lucca, Madar, Molloy and Willen (2013), Hurst, Keys, Seru and Vavra (2016), Nguyen (2019)), offers highly standardized loan products adhering to the GSEs securitization guidelines (Ambrose and Buttimer Jr (2005)). Under such conditions, one would anticipate banks to have limited pricing power in the conforming loan market.

However, my findings reveal that portfolio lenders (banks that fund lending on their balance sheets) can charge a premium of approximately ten basis points over lenders who primarily fund their lending via securitization (OTD lenders)[2]. Henceforth, I refer to this premium as the portfolio lender premium (PL premium). Furthermore, I show that the portfolio lenders are able to charge this premium because these lenders can exert market power in the local market and the bank lenders can achieve product differentiation by appeal to a group of borrowers who are less price sensitive.

The portfolio lenders' market power seemingly originates from their branch networks. As

[1]Both Buchak, Matvos, Piskorski and Seru (2018b) and the Gies Consumer Credit Panel provide similar numbers.

[2]While bank lenders can adopt a mixed strategy encompassing both securitization and balance sheet financing, nonbank lenders predominantly engage in originate-to-distribute (OTD) lending, in which they securitize the majority of the loans. Notably, even though most nonbank lenders are OTD lenders, there is still heterogeneity within the bank lenders and many large bank lenders also securitize a considerable amount of mortgages (Purnanandam (2011)).

banks are almost exclusively portfolio lenders, often deposit-rich with significant retail deposits relative to their mortgage demand (Han, Park and Pennacchi (2015),Pennacchi (2019)), they typically have extensive branch networks. These networks can boost a bank lender's bargaining power in multiple ways[3]. For instance, branch locations attract borrowers preferring face-to-face interactions, which not only provide additional utility but also increase the cost of shopping for lower prices for such borrowers. Furthermore, physical branches can gather more detailed soft information from borrowers, strengthening lenders' negotiation power during loan term determination[4].

My study underscores that certain bank lenders retain a competitive edge, even amidst rapid digitization. While my analysis primarily focuses on the conforming loan market, it's reasonable to infer that similar market segmentation could occur within subprime and jumbo loan markets. The presence of portfolio lenders' market power emphasizes the importance of understanding market concentration effects on the lending market. For instance, regulatory bodies often emphasize deposit market concentration during bank mergers, but the potential impacts of lending market concentration receive less attention.

To drive my empirical analysis, I build an industrial organization model of lender competition. In this model, two types of lenders with different funding costs engage in Cournot competition in a partially-segmented market. An equilibrium with a positive PL premium can exist if two necessary conditions are met: (1) lenders exert market power in the local market, and (2) portfolio lenders have a lower cross-elasticity of demand than the OTD lenders. The first condition determines the level of competition within the sector, and the second condition determines the relative level of

[3]It should be noted that in the context of increasing a bank's bargaining power, a strong branch network does not necessarily mean a larger number of branch locations. For example, a bank with many branch locations in highly competitive regions may have contributed little to increase its bargaining power with the borrowers. Similarly, a bank with a few locations in a highly concentrated market may derive strong bargaining power from these branches.

[4]Research in the relationship lending literature, many studies have shown that prior relationships affect lending decisions in corporate lending settings. For example, Petersen and Rajan (1994) finds that businesses are more likely to obtain credits from lenders whom they have close ties with. and Petersen and Rajan (1995) provides a framework showing that creditors are more likely to provide credits when the market is concentrated.

market power between the two sectors [5].

For my empirical analysis, I utilize loan-level data from the Gies Consumer Credit Panel (GCCP). This dataset, consisting of approximately one million 30-year conforming mortgages from 2004 to 2017, provides more accurate information on mortgage securitization, consumer credit scores, and demographic information compared to HMDA, the most commonly used loan-level mortgage market data source. This research represents the first use of credit bureau data to analyze loan-level mortgage market outcomes. Moreover, the GCCP includes data on mortgage inquiries, enabling me to create a time-varying mortgage demand proxy at the county level.

I begin my empirical analysis by documenting polarization in the financing strategy of the lenders. Across the sample years, more lenders choose to either securitize most mortgages or finance a large portion of loans on balance sheets than those who choose a more balanced financing strategy. Motivated by this finding, I create a time-varying classification for portfolio lenders based on the percentage of mortgages securitized in a given time window. Next, I show that the portfolio lenders charge a PL premium of 9.1 to 12.8 bps on conforming loans over the other lender after controlling for all observable risk characteristics. Importantly, a PL premium of a similar magnitude can be found between the bank PL lenders and the bank OTD lenders, suggesting that the premium is not merely caused by the differences between banks and the nonbanks alone.

I show pieces of evidence that are consistent with the two necessary conditions for a positive PL premium. First, I estimate that the portfolio premium is 5.5 to 6.5 bps higher per one standard deviation increase in market concentration. This result shows that market concentration has a first-order impact on the relative pricing power between portfolio lenders and OTD lenders, even though concentration does not necessarily impact mortgage pricing at the average level. Second, using a Bartik-type instrument for mortgage demand, I estimate that PL premium increases by 12.0 to 12.4 bps per a ten percentage points increase in the demand for mortgages in the local markets.

[5]The separation of the two types of market power is critical because the pricing power of the portfolio lenders cannot be explained by higher concentration within the portfolio lenders alone. There is lower concentration within the OTD lenders than within the portfolio lenders, with the former type of lenders having one-tenth the number of the latter but contributing to two-thirds of total loan origination.

On the other hand, the relative difference in origination volume and market share of the two types of lenders do not change by demand shocks. The combination of the results on the interest rate and origination volume suggests that the portfolio lenders have a steeper supply curve relative to the OTD lenders in the local markets. Together, the two results jointly support the mechanism proposed in my theoretical framework.

This paper offers substantial contributions to several branches of literature on credit market structure. Primarily, it brings fresh insights to the dialogue on competition between banks and nonbanks, a topic that has gained considerable attention due to the rise of nonbank intermediaries over the last decade. For example, Demyanyk and Loutskina (2016), Buchak, Matvos, Piskorski and Seru (2018a), Buchak et al. (2018b), and Pennacchi (2019) examined nonbank competition in the context of regulatory burden and funding cost. Gertler, Kiyotaki and Prestipino (2016), Fuster, Plosser, Schnabl and Vickery (2019), and Bartlett, Morse, Stanton and Wallace (2022) studied the flexibility of the nonbank lenders in the adoption of new technologies. A related paper, Shi, Zhang and Zhao (2023), finds that competition forces banks to increase balance sheet financing due to frictions in guarantee fees. In contrast, my paper focuses on the competition between sectors of lenders based on financing methods and highlights the heterogeneity within the bank lenders. Although prior works have investigated the merits of branch networks, my paper focuses on their influence on mortgage demand, as opposed to merely their role in deposit market dominance.

My results on the relationship between market concentration and PL premium contribute to the literature on the effects of market concentration in the mortgage market. Until recently, the current consensus in the literature is that market concentration does not impact pricing at the local level. Fuster et al. (2013) argues that in a market with a large number of fringe firms, concentration should not lead to higher pricing. Hurst et al. (2016) and Amel, Anenberg and Jorgensen (2018) find market concentration does not impact the interest rates of GSE-insured mortgages and their sensitivity to monetary policy. On the other hand, Scharfstein and Sunderam (2016) shows that the interest rate in more concentrated markets responds less sensitively to monetary policy. A more recent paper, Buchak and Jørring (2021) finds that non-interest costs, such as rebates and loan rejections, have a strong relationship with market concentration. To my knowledge, this paper is the first evidence of the differential effect of market concentration between heterogeneous sets of lenders. My results provide a new angle from which market concentration affects mortgage pricing, which provides empirical support for policy-making for the mortgage market.

This paper also contributes to discussions on market segmentation within the mortgage market, most notably between conforming, jumbo, and subprime loan markets based on preset conforming loan guidelines. For example, some works, such as Courchane, Surette and Zorn (2004), Chomsisengphet and Pennington-Cross (2006), and Adelino, Schoar and Severino (2012), study the frictions related to jumbo-conforming segmentation, while other works, such as DeFusco and Paciorek (2017), explore the mechanical nature of the conforming loan limits as an identification strategy. In my paper, I explore the understudied segmentation within the conforming loan market, which is often viewed as homogeneous (Ambrose and Buttimer Jr (2005)).

Finally, this paper connects to the literature examining changes in incentives induced by loan securitizations. Benveniste and Berger (1987) argues that securitization with recourse improves the allocation of risk sharing among a bank's liability holders. Pennacchi (1988) builds a model where banks use loan sales to reduce regulatory cost, Gorton and Pennacchi (1995) studies the incentive-compatible contract that facilitates loan sales in the face of moral hazard problem, Gorton and Souleles (2007) discusses the use of SPV as a measure to reduce bankruptcy cost. Keys, Mukherjee, Seru and Vig (2010) finds that the subprime mortgages that are easier to be securitized are 10-25% more likely to default. While the previous literature mainly discussed how securitization impacts lender incentives, my paper takes a unique direction by examining the link between financing methods and changes in borrower preferences[6].

The rest of the paper is organized as follows: Section II lays out the institutional background and introduces the theoretical framework. Section III describes the dataset. Section IV presents the evidence of the PL premium. Section V and VI show empirical evidence that is consistent with the

[6] A related paper, An, Deng and Gabriel (2011) finds that securitized commercial loans are priced at a higher interest rate compared to portfolio loans due to adverse selection, while my paper shows an opposite relationship using data from the residential mortgage market and is motivated by a different market mechanism.

two necessary conditions that generate a PL premium, as predicted by the theoretical framework. Section VII concludes.

II Background and Theoretical Framework

II.1 Institutional Details

This paper focuses on the conforming mortgage market. A conforming mortgage is a mortgage that meets the dollar limits set by the Federal Housing Finance Agency and the funding criteria set by Government-Sponsored Enterprises (GSEs), allowing them to be securitized into mortgage-backed securities (MBS) by the latter two institutions.

There are three major types of players in the mortgage market: lenders, GSEs, and investors. The lenders, or the originators in the context of this paper, can be a variety of financial intermediaries. The traditional mortgage originators are the banks[7]. In the recent twenty years, nonbank lenders have gained an increasingly larger share of the market in terms of origination volume. There are two GSEs in the conforming mortgage market: Fannie Mae and Freddie Mac[8]. The GSEs are quasi-governmental entities that are established to provide liquidity to the housing market. They do so by pooling the mortgage loans originated by lenders into MBS and selling the MBS in the secondary market. Typically, the investors in the mortgage market are institutional investors, such as commercial banks, mutual funds, life insurance companies, and the US government.

There are two main approaches used to finance mortgages. One is the portfolio lending model. In this model, the lender owns the mortgages and funds them with its own debt (deposits) and eq-

[7]In this paper, I define the banks as the financial intermediaries that are deposit-taking intuitions, such as commercial banks, thrift banks, and credit unions. I define the nonbanks as non-deposit-taking institutions.

[8]After the financial crisis of 2008, the volume of private-labeled MBS has dwindled considerably and the two GSEs, Fannie Mae and Freddie Mac, become the only two dominant players in the conforming loan market. In principle, it is still possible for a conforming mortgage to be securitized by non-GSE intermediaries into private-label MBS. According to HMDA data, there are only about 0.5% of conforming mortgages that are securitized into private-labeled MBS according to HMDA data between 2008 and 2017. Even before the financial crisis, this number is only 3.1% in 2007.

uity, and receives the interest payment until the mortgages are paid off. Thus, the lenders are directly exposed to the credit risk of the mortgages. The second type is the originate-to-distribute (OTD) model. In the OTD model, a lender sells the mortgages it originates to the GSEs or other financial intermediaries. When the mortgages are sold to a GSE, the GSE charges the originator a small guarantee fee (the g-fee). The guarantee fee varies between 20 to 50 bps and is based on a number of factors, including mortgage characteristics (mortgage terms, mortgage purpose, etc), basic borrower characteristics (credit score, LTV ratio, etc), as well as lender characteristics. The GSEs guarantee to buy back a mortgage at par value in the case of default. When the mortgages are sold to another financial intermediary, the purchaser will face the same decision as the mortgage originator: either finance the mortgages on its own balance sheet or sell them to another financial intermediary. Typically, the mortgages will not change hands more than twice before it is securitized.

The majority of the lenders that finance mortgage lending via the portfolio lending model are deposit-taking institutions. The reason is that these lenders have access to retail deposits, which gives them the balance sheet capacity to hold the loans to maturity. However, many deposit-taking institutions also finance mortgage lending via the OTD method due to a number of different reasons, such as high loan demand relative to deposit supply, high corporate tax burden, (Han et al. (2015)), and regulatory burdens (Buchak et al. (2018b), Seru (2020)). On the other hand, nonbank lenders are predominately lending via the OTD model. In practice, many bank lenders use a mixed financing strategy. Due to the lack of bank status, nonbank lenders can only rely on competitively priced wholesale funding or warehouse lines of credit (Stanton, Walden, Wallace et al. (2014), Jiang (2019)), which are typically more expensive than deposits. In summary, bank lenders can fund their lending via either the portfolio lending method or the OTD method, while nonbank lenders fund their lending exclusively via the OTD method.

II.2 Theoretical Framework

In this section, I present a standard industrial organization model to explain how market segmentation drives the PL premium. The model is stylized and is only used as a conceptual framework to guide my empirical analysis in the following sections.

II.2.1 A Model of Lender Competition

A local market (a county) has both portfolio lenders and OTD lenders. Without a loss of generality, I assume that there are no lenders that adopt mixed financing methods. i is the index for lender type. $i = 1$ represents portfolio lenders and $i = 2$ represents OTD lenders. There is a total of N lenders, in which $N_1 = \delta N$ are portfolio lenders and $N_2 = (1-\delta)N$ are OTD lenders. Lenders of each type are identical to each other. r_i is the type-i lender's funding cost. The funding cost of the portfolio lenders, r_1, equals the weighted average of debt and equity of the lenders. The funding cost of the OTD lenders, r_2, would be the interest rate at which they sell the MBS to the secondary market. To capture the empirical fact that the MBS rate is higher than the deposit rate, assume $r_1 < r_2$.

The market is partially segmented in the sense that the mortgages originated by portfolio lenders and OTD lenders are not perfect substitutes, i.e. there is product differentiation in the mortgage market. Portfolio lenders are usually deposit-rich banks with a larger share of retail deposits relative to their mortgage demand. These banks often possess a robust deposit demand due to their expansive branch network. Borrowers, especially senior borrowers, may gain utility from engaging with loan officers in person at branch locations. Furthermore, widespread branch networks may serve to collect more comprehensive, nuanced information from borrowers, enabling lenders to strengthen their negotiation power when determining loan terms. On the other hand, many OTD lenders, particularly the specialized nonbank ones, operate exclusively online without branch networks. Due to a lack of physical locations, there could be less differentiation in terms of customer service and more intensive price competition amongst the OTD lenders[9]. Importantly, the model distinguishes the two dimensions of market power, with the first condition governing the level of competition within the sector and the second condition governing the relative level of market power between the two sectors. This is because the heterogeneity of market power across lenders cannot be explained by higher concentration with the portfolio lenders. The level of market concentration within the OTD lenders is actually higher than that within the portfolio lenders. As shown in Figure 3, the total number of OTD lenders is only less than one-tenth of that of the portfolio lender, while contributing to two-thirds of total loan origination.

To capture the intuition of market segmentation, I incorporate the cross-elasticity of demand in the demand functions of two types of lenders. Let R_i be the local market's equilibrium mortgage interest rates of the bank and nonbank mortgages. Similarly, Q_i denotes the equilibrium quantity of mortgages of type i in the local market. I assume a set of linear demand functions:

$$R_i = a - bQ_i - b_{i,j}Q_j \tag{1}$$

where if $i = 1$, then $j = 2$ and if $i = 2$, then $j = 1$. a, b, and $b_{i,j}$ are positive constants. The key feature of the demand function is the term $b_{i,j}$, which governs the cross-elasticity of demand faced by lenders of type i from lenders of type j. When $b_{ij} = b_{j,i} = b$, the mortgages offered by lenders of type i are perfect substitutes for the mortgages offered by lenders of type j. If $b_{ij} = b_{j,i} = 0$, the market is perfectly segmented. I assume the two types of mortgages to be partial substitutes, i.e. $b > b_{ij} > 0$ and $b > b_{ji} > 0$. Finally, assume that $a > r_i$, such that the equilibrium interest rates would be positive.

The lenders engage in a Cournot competition. For each loan originated, a lender earns the spread between the mortgage interest rate and its funding cost, $R_i - r_i$. Each lender of type i solves

[9] It should be also emphasized OTD lenders are not necessarily nonbank lenders. Some depository institutions operate exclusively in the OTD model. In the context of this model, the bank OTD lenders and nonbank OTD lenders are identical as they both have the same funding cost r_2 and face the same demand function.

a maximization problem by choosing the optimal quantity of mortgage, q_i:

$$\max_{q_i} R_i q_i - r_i q_i \tag{2}$$

Each type-i lender takes the total number of lenders of each type as given. By doing so, a type-i lender also takes the optimal aggregate quantity originated by the type-j lenders, Q_j, as given. Solving a symmetric Nash equilibrium yields the equilibrium mortgage interest rate for type$-i$ lenders:

$$R_i = \frac{a - b_{ij} Q_j + N_i r_i}{N_i + 1} \tag{3}$$

where

$$Q_i = \frac{N_i(a(b + bN_j - b_{ij}N_j) + b_{ij}N_j r_j - b(1+N_j)r_i)}{-b_{ji}b_{ij}N_j N_i + b^2(1+N_j)(1+N_i)} \tag{4}$$

The PL premium is defined as follows:

$$\Delta^R = R_1 - R_2 :\rightarrow \Delta^R(N, \Delta^b, \Delta^r) \tag{5}$$

where $\Delta^b = b_{12} - b_{21}$ and $\Delta^r = r_1 - r_2$. Note that $N = N_1 + N_2 = N_1/\delta = N_2/(1-\delta)$. Here, I assume δ to be a fixed value such that a change in N proportionally change the number of portfolio lenders and OTD lender. In principle, When δ is small enough (when the portfolio lenders have very high market power), the PL premium is unambiguously positive. Without further clarification, I focus on the more interesting parameter space when δ is large enough such that the sign of the PL premium is ambiguous.

I use a numerical example to illustrate the intuitions from the model. Panel A of Table 3 exhibits the parameter values of the example, while Panel B shows the solutions. A PL premium of 21.22 bps is generated from this example, even though the funding cost of the portfolio lenders is lower than that of the OTD lenders.

All the derivations and proofs are provided in Appendix A.5.

II.2.2 Model Predictions

The following propositions can be made regarding the properties of Δ^b:

Proposition 1. *When market concentration decreases, the difference between the interest rates charged by the portfolio and OTD lenders approaches the difference between their funding cost, i.e. when $N \to +\infty$, $\Delta^R \to \Delta^r$.*

Proposition 1 captures the two edge case where market concentration approaches perfect competition, i.e. $N \to +\infty$. Under perfect competition, neither types of lender are able to earn a profit above their funding cost. Regardless of their respective cross-elasticity of demand, the two types of lenders will only supply mortgages to the market to a point where the market equilibrium mortgage interest rates equal their respective funding costs.

Proposition 2. *When the market is concentrated enough and when the cross-elasticity of demand of the portfolio lender is lower enough, the portfolio lenders charge a lower interest rate than the OTD lender, i.e. there exists $\hat{N} > 1$ and $\hat{\Delta^b} > 0$, such that when $N < \hat{N}$ and $\Delta^b < \hat{\Delta^b}$, $\Delta^R > 0$.*

Proposition 2 shows that a positive PL premium can exist under certain conditions even if the funding cost of the portfolio lenders is higher than that of the OTD lenders. Two necessary conditions need to be met in order for the PL premium to be positive:

(1) Lenders have market power in the local market

(2) The portfolio lenders have a smaller cross-elasticity of demand than that of the OTD lenders'

The intuition behind the first necessary condition is that the two types of lenders will need to be able to exert some level of monopolistic power in the local market such that their supply curve is sloped. The intuition behind the second necessary condition is that conditioning the local market is not perfectly competitive, the portfolio lenders must face a less price-sensitive borrower demand compared to that faced by the OTD lenders to be able to charge a higher equilibrium interest rate. If the two types of lenders have the same cross-elasticity of demand, they will be effectively facing an identical demand curve. As a result, the portfolio lenders will be able to charge a lower interest rate regardless of the level of lender market power due to their lower funding cost.

Figure 6 illustrates the intuition behind Proposition 2 by showing how the PL premium changes with b_{12} and N. The parameter values used in the figure are the same as the values in Panel A of Table 3, except for b_{12} and N. The figure shows that when the number of lenders in the market passes the threshold of 490 and the cross-elasticity of demand of the portfolio lenders passes the threshold of 0.0107, the PL premium enters into a region in the top-left side of the parameter space where its value is non-ambiguously positive. In this parameter region, the PL premium is a decreasing function of both b_{12} and N.

While it is challenging to test the two necessary conditions directly, two testable predictions can be derived from Propositions 2 to motivate the empirical analysis in the next Section:

Corollary 1. *When the PL premium is positive, the PL premium decreases as OTD lenders become more concentrated, i.e. When* $N < \hat{N}$ *and* $\Delta^b < \hat{\Delta^b}$, $\partial \Delta^R / \partial N < 0$.

In my model, the different levels of cross-elasticity of demand drive different sensitivities of interest rate to market concentration by the two types of lenders. Therefore, the model predicts that there should be a positive relationship between the level of market concentration and PL premium in the graphical cross-section. Furthermore, market concentration should be able to explain most of the PL premium, since market concentration is a necessary condition of a positive PL premium.

Corollary 2. *When the PL premium is positive, the PL premium increases while the portfolio lenders increase credit supply less than the OTD lenders when there is a positive shock in the demand for mortgages in the local market, i.e. When* $N < \hat{N}$ *and* $\Delta^b < \hat{\Delta^b}$, $\partial \Delta^R / \partial a > 0$ *and* $\partial Q_1 / \partial a - \partial Q_2 / \partial a < 0$.

A key feature of my model is that the two types of lenders face different demand curves. As a result, the model predicts that the two types of lenders should be different responses in credit supply when faced with the same demand shock. If the portfolio lenders have lower cross-elasticity of demand than the OTD lenders, the portfolio lenders will be able to raise interest rates higher while increasing credit supply by a smaller ratio compared to the OTD lenders.

In the next Section, I focus on finding empirical support for the predictions made in Corollaries 1 and 2. If empirical evidence supports the two Corollaries, the two necessary conditions that generate the PL premium found in Section IV.2 likely hold, thereby providing additional support for the mechanism proposed in Section II.2.1.

III Data Description

III.1 Description of the Data Set

The main data source used in this paper is Gies Consumer Credit Panel (GCCP), which is provided by Experian, one of the three largest credit bureaus in the US. The data set is created through random sampling of one percent of all consumers with a credit history at the end of the first quarter of each sample year. The sample period spans a total of 14 full years from 2004 to 2017.

At the consumer level, the dataset contains the credit score, estimated income, and estimated DTI of each consumer. The estimated income and estimated DTI are estimated and validated by a model internally developed by Experian. Starting from 2011, the dataset contains demographic variables, including age, gender, marital status, occupation category, education level, number of adults in a household, number of children in a household, and home ownership status.

At the loan level, the dataset contains all the mortgage loans borrowed by the consumers in the sample. The observable information for each loan includes the mortgage amount, term length, monthly payment, origination date, remaining balance, and delinquency status. A separate categorical variable enables me to identify if a mortgage is guaranteed by Fannie Mae or Freddie Mac. The lender name is anonymized, but each lender is assigned a unique institutional-level lender key and a business classification code. I can also identify the type of mortgage (conventional mortgage, FHA mortgage, VA mortgage, etc) through another categorical variable. Lastly, the borrower of each mortgage can be identified by an anonymized borrower ID, which allows me to link loan-level data to the consumer characteristics of each sample year. To make the mortgages in my analysis comparable, I follow the standard practice in the literature of restricting my sample to 30-year first-lien conforming mortgages[10].

An important feature of the loan-level mortgage data from the GCCP is a variable that identifies whether a mortgage is sold to either Fannie Mae or Freddie Mac at the time of each snapshot. This is a key advantage of GCCP over HMDA, as HMDA only records the securitization information of mortgages sold before the end of the calendar year of origination. Additionally, HMDA only records the first mortgage sale. If a mortgage is sold to another lender prior to securitization, only the buyer type of the first sale is recorded. As a result, HMDA underestimates the percentage of mortgages that are securitized.

Besides mortgage data, GCCP also contains a data set of credit inquiry data. I use the credit inquiry for all mortgage loans[11]. A credit inquiry is recorded when a lender tries to pull out the credit history of a consumer when the consumer tries to apply for a mortgage product. This usually happens on the same or the next day the borrower submits a mortgage application. For each inquiry entry, I am able to observe the ID of the corresponding consumer, the type of credit product the inquiry is associated with (in the context of this paper, mortgage product), as well as the date the inquiry was pulled. I merge mortgage inquiry data with consumer characteristics data using the

[10]It is also common to restrict the sample to fixed-rate mortgages. Unfortunately, I am not able to identify adjustable-rate mortgages from fixed-rate mortgages. As a result, one implicit assumption of my empirical analysis is that there is no systematic variation in adjustable-rate mortgage origination that is correlated with lender concentration. Multi-unit homes are also not able to be identified from the GCCP data. Considering that high balance mortgages (over 1 million USD) only constitute less than 1 percent of the full sample, the impact of multi-unit homes would be unlikely to be large enough to impact my results

[11]Due to the availability of information, I am not able to separate inquiries for conforming mortgages from those for other types of mortgages.

unique consumer ID.

III.2 Summary Statistics

Panel A of Table 1 reports the summary statistics of the mortgage sample. There is a total of 893,253 mortgages in my final sample. The average Vantage Score of the mortgage sample is 728, slightly higher than the national average of 698 in 2021[12]. The average mortgage amount during the final sample is 216,907 dollars. One novel feature of the GCCP data set is that it allows me to identify mortgage interest rates for the majority of the mortgages[13]. The average interest rate of the mortgages in the final sample is 5.50 percent[14] I identify refinance mortgages if one mortgage has been paid off within a one-month window before a new mortgage is originated. There are 37% of the mortgages are classified as refinancing mortgages in the final sample. I classify all conforming mortgages that are not insured by either Fannie Mae or Freddie Mac in the first snapshot date as balance-sheet-financed mortgages[15]. In the final sample, 57% of the mortgages are securitized through the two GSEs. Figure 1 shows that the balance-sheet-financed mortgages have higher interest rates compared to securitized mortgages throughout the sample years.

Panel B of Table 1 reports the summary statistics of the mortgage inquiry sample. The final sample includes 7,237,580 mortgage inquiries. Compared with the demographic of the approved mortgage data, the demographic of the mortgage applicants exhibits a lower financial strength. More specifically, they have a lower credit score, lower income, higher DTI, are less likely to be owning a home, and are less likely to have received college-level education.

Panel C of Table 1 reports the summary statistics at the lender level by collapsing the mortgage

[12]This number is obtained from Equifax (https://www.equifax.com/personal/education/credit/score/average-credit-score-state).

[13]Unfortunately, rebates and fees are not observable in the GCCP data set.

[14]Appendix A.1 provides more details on the interest rate estimation and its validation.

[15]While it is possible that a mortgage is securitized in the later snapshot dates, such occasions are extremely rare in practice. In the data, less than 0.1% of all securitized mortgages are securitized not at the first but at later snapshot dates.

level data by the lender key associated with each loan[16]. There is a total of 4,488 lenders present in the sample, of which about 96% are bank lenders. Bank status is not directly observable in the GCCP data[17]. On average, the lenders have 2,067 credit card accounts and 431 auto loan accounts at the end of a snapshot year, which are reasonable numbers considering that GCCP is a one percent sample of the US population and that most of the lenders are bank lenders.

IV Evidence of the portfolio lender premium

This Section presents empirical evidence for the PL premium. I construct a time-varying binary classification of portfolio lenders based on the observation of polarization in the financing methods of the lenders. I show that the lenders that are classified as portfolio lenders on average charge a higher interest rate than the OTD lenders after controlling for a rich set of factors that may impact interest rates.

IV.1 Classification of Portfolio Lenders

The identification of portfolio lenders is critical for studying market segmentation. Since many lenders finance mortgages via both portfolio lending and securitization at the same time, it is also important to investigate the distribution of financing methods amongst the lenders. Figure 2 shows the distribution of the top 100 lenders by total origination volume with respect to the percentage of balance-sheet-financed mortgages from 2004 to 2017. The Figure shows strong evidence of polarization in the lenders' securitization strategy: during most of the sample years, more lenders choose to either securitize most mortgages or securitized very few mortgages compared to the lenders that choose a mixed strategy. A gradual shift from balance sheet financing to securitization can also be observed from the figure: prior to 2008, a greater number of the top 100 lenders choose

[16] Appendix A.2 discusses the potential measurement errors associated with the lender key variable in the GCCP dataset.

[17] Appendix A.3 provides a detailed description of the methodology to identify bank lenders from the data.

to finance most of the mortgages on their balance sheet. After 2008, the mass of lenders gradually shift towards the other polar, and by 2011, more lenders choose to securitize most mortgages than lenders that choose to balance sheet finance.

To account for both the polarization and the time-varying nature of lenders' financing strategy, I use a time-varying binary classification to identify the heterogeneity of mortgage financing. More specifically, I define a lender to be a portfolio lender at a given quarter if the average percentage of mortgages the lender securitizes over the past 12 quarters is less than 70%. The rationale for using 70% as the cutoff point is that the average time-until-sale for these lenders is typically within one month, meaning that OTD lenders typically hold less than 10% of their newly originated loans at any given time. Thus, the 70% cutoff point serves as a conservative criterion that filters out the majority of the lenders who operate in the OTD model. Figure 3 shows the fraction of lenders that are classified as portfolio lenders as well as the fraction of mortgages that are originated by portfolio lenders. While OTD lenders constitute a relatively small number of lenders, they are responsible for more than half of the total origination volume after 2009.

IV.2 Main Results on PL premium

Figure 4 presents graphical evidence that the portfolio lenders charge a higher interest rate compared to the OTD lenders. Panel A shows that except for a short period of time during the financial crisis, the portfolio lenders charge about 10 to 20 bps higher interest rates than OTD lenders. Panel B uses only mortgages that are originated by the bank lenders, and finds a premium of similar magnitude, implying that the premium cannot be explained by the heterogeneity of bank status alone.

I formally estimate the PL premium using the following loan-level specification:

$$R_i = \sigma_{c(i),q(i)} + \beta \mathbb{1}^{PTF}_{l(i),q(i)} + \eta' \mathbf{X}_i + \gamma' \mathbf{X}_{j(i),q(i)} + \lambda' \mathbf{X}_{l(i),q(i)} + \varepsilon_i \tag{6}$$

For all the specifications throughout this paper, i, j, l, c, q, and y are indices for loan, borrower, lender, county, quarter, and year, respectively. The variable of interest is $\mathbb{1}^{PTF}_{l(i)}$, which is the time-invariant baseline classification of portfolio lender. The variable is a dummy variable that equals one if the average percentage of securitized mortgages by lender l is lower than 70% in the 12 quarters prior to quarter q. $\sigma_{c(i),q(i)}$ is the county-quarter fixed effects, which capture time-varying local economic conditions. $\mathbf{X}_i$ is the vector of loan controls, including the log of loan amount and loan purpose (for refinancing or for a new purchase). $\mathbf{X}_{j(i),q(i)}$ is a vector of borrower controls, which include credit score, income, and debt-to-income ratio. This set of control captures each borrower's risk characteristics, which are important determinants of mortgage rates.

Importantly, $\mathbf{X}_{l(i),q(i)}$ is the vector of lender controls, which include the number of operating states, number of operating counties, number of existing accounts for different credit product types[18], as well as the total amount of outstanding loan balance. This set of controls can capture the differences between the bank and nonbank lenders, as nonbank mortgage lenders typically only operate mortgage lending businesses. It also captures the heterogeneity in business focus within the bank lenders. Additionally, I calculate the total outstanding debt of each lender in each given quarter. This control can be used as a proxy for the size of the lender. Lastly, I also calculate the time-varying numbers of counties and states each lender operates. These two variables capture the differences in lenders who operate at the national level and those who operate at the state level and the heterogeneity in the geographic coverage within the lenders. With the time-varying lender characteristics controls, my estimation can be interpreted as the marginal effect of a mortgage being originated by a portfolio lender, conditioning on the geographical coverage, the economy of scope, and the economy of scale of the lender.

Table 2 reports the results from Equation (6). The estimated coefficient of the portfolio lender dummy captures the magnitude of the PL premium. Panel A reports the results using the full sample. Column (1) reports the estimation that is only conditioned on loan terms. Columns (2) to (4) add additional loan-level credit riskiness controls and time-varying lender controls. Across

[18]More specifically, I calculate the total number of mortgage accounts, credit card accounts, and auto loan accounts.

Columns (1) to (4), the PL premium is consistently positive. The specifications in Columns (5) and (6) add additional lender and lender-county fixed effects, thereby differencing away all variations across lenders. The estimated magnitude of the premium becomes smaller in these two specifications but remains statistically positive. Across all specifications with the baseline cutoff, the PL premium is around 9.12 to 12.77 bps. Panel B reports the results using the sample of mortgages originated by bank lenders[19]. The estimations of the PL premium remain to be significant and the estimated magnitudes, if anything, become even larger than the ones in the pooled sample. Furthermore, Appendix A.4 shows that the PL premium is also robust to alternative cutoffs used for the classification of portfolio lenders.

IV.3 Discussion of the portfolio lender premium

Instead of making a causal interpretation of financing methods on mortgage pricing, I propose that the financing decisions of lenders serve as a proxy of particular lender-specific characteristics. More precisely, I assert that the extent of a lender's branch network significantly influences the premium. Lenders boasting a robust branch network are likely to possess substantial market power within the deposit market. This power facilitates easier access to low-cost retail deposits, providing these lenders with the balance sheet capacity necessary to finance mortgages on their own balance sheets. Simultaneously, the existence of branch networks permits these lenders to differentiate their products through superior service quality. These attributes could be particularly appealing to certain demographics, such as older or more risk-averse borrowers. On the contrary, OTD lenders, predominantly operating online, may appeal to a younger, more tech-savvy, and less risk-averse demographic. In contrast, borrowers preferring in-person service at branch locations are likely to display greater customer loyalty, enabling lenders to extract more profits from them. Therefore, when portfolio lenders offer such advantages to borrowers, those with extensive branch networks can utilize their superior service delivery as leverage to demand higher interest rates, despite their

[19] Lender types are not directly observable in the data. I classify lenders into bank lenders and nonbank lenders using the outstanding balances on non-mortgage consumer credit products in the GCCP data. Appendix A.3 provides more details on my method to identify bank lenders from the data.

access to deposits potentially resulting in reduced funding costs.

Evaluating the capacity of a branch network within a local market can be a complex task. One plausible approach could be assessing the density of a lender's physical branch locations. Nevertheless, limitations within the GCCP dataset prevent me from correlating mortgages to specific branches[20]. One alternative method is to utilize the prevalence of nonbank lenders as a proxy of the local borrowers' dependence on physical branch networks. Most nonbank lenders operate exclusively online. If a substantial portion of mortgages in a county is initiated by nonbank lenders, it suggests a diminished relevance of bank branches in that county due to the competition posed by online lenders. If my measurement of a portfolio lender is representative of branch network strength, an inverse relationship should exist between the PL premium within bank lenders and the degree of nonbank penetration in local markets. That is to say, when nonbank competition intensifies, the premium that bank portfolio lenders charge over bank OTD lenders would likely decrease.

Figure 5 shows the relationship between nonbank penetration and the PL premium. To eliminate potential mechanical relationships attributed to nonbank lenders, I exclusively examine the relationship between nonbank penetration and the PL premium among bank lenders. The vertical axis represents the PL premium among bank lenders residualized by average credit score, average income, average debt-to-income ratio, average loan amount, the percentage of refinance mortgages, county fixed effects, and year fixed effects. The horizontal axis signifies nonbank penetration, measured by the proportion of mortgage loans originated by nonbank lenders. The figure prominently displays a negative correlation between nonbank penetration and the PL premium within bank lenders.

[20] Although the HMDA dataset includes branch network data, it only started including interest rate data from 2017, and information regarding the securitization status of mortgage loans is scant. Moreover, HMDA lacks appended credit score data.

V Empirical Analysis of Corollary 1

Corollary 1 predicts that in the graphical cross-section, the magnitude of the PL premium should also have a positive relationship with lender market power and that market power should be able to explain most of the variation in PL premium. This Section provides empirical evidence that supports this relationship.

V.1 Empirical Strategy for Testing Corollary 1

I calculate the Herfindahl–Hirschman Index (HHI) using the dollar amount of total mortgage origination as the measure of market concentration. To alleviate potential reverse causality, I calculate the baseline HHI measure in each quarter as the average HHI of the previous twelve quarters. More specifically, the baseline HHI is calculated as follows:

$$HHI_{c,q} = \sum_{\tau=1}^{12} \sum_{l} \left(\frac{Vol_{l,c,q-\tau}}{Vol_{c,q-\tau}} \right)^2 / 12 \tag{7}$$

where $Vol_{l,c,q-\tau}$ is the origination volume by lender l in county c in quarter $q-\tau$, and $Vol_{c,q-\tau}$ is the total origination volume by all lenders in county c in quarter $q-\tau$.

Figure 7 presents the geographic distribution of the baseline HHI. Consistent with the distribution of market concentrated in a number of other papers (Stanton et al. (2014), Scharfstein and Sunderam (2016), Yannelis and Zhang (2021), Buchak and Jørring (2021)), the level of market concentration calculated from GCCP also exhibits a positive correlation with population density: the market is more competitive in densely populated coastal areas such as northeastern states and California. Figure 8 shows the time-series trends of average interest rate in counties with top 50% market concentration versus counties with bottom 50% market concentration. The graph shows no visually significant divergence between the average interest rates in the high versus low-concentration counties, suggesting that, at least along the time-series dimension, variation in market concentration is orthogonal to movements in mortgage interest rates.

Figure 9 provides visual evidence of such a relationship. Panels A, B, and C show the relationships between market concentration and average interest rates of all mortgages, mortgages originated by the portfolio lenders, and mortgages originated by the OTD lenders, respectively. The Figure shows that while average mortgage interest rates of all mortgages are not strongly associated with local concentration, there is a strong divergence between the interest rates charged by the two types of lenders as the counties become more concentrated.

I use the following specification to formally estimate the effect of market concentration on the PL premium.

$$R_i = \sigma_{c(i),q(i)} + \beta_1 \mathbb{1}^{PTF}_{l(i),q(i)} + \beta_2 \mathbb{1}^{PTF}_{l(i),i(q)} \times HHI_{c(i),q(i)} + \eta' \mathbf{X}_i + \gamma' \mathbf{X}_{j(i),q(i)} + \lambda' \mathbf{X}_{l(i),q(i)} + \varepsilon_i \quad (8)$$

where $\mathbb{1}^{PTF}_{l(i),q(i)}$, $\sigma_{c(i),q(i)}$, $\mathbf{X}_i$, $\mathbf{X}_{j(i),q(i)}$, and $\mathbf{X}_{l(i),q(i)}$ are defined the same as in Equation (6). In this specification, the variable of interest is the interaction term $\mathbb{1}^{PTF}_{l(i),q(i)} \times HHI_{c(i),q(i)}$, whose coefficient captures the marginal effect on the PL premium when the mortgages are originated in counties of different levels of market concentration.

One identification strategy frequently applied in the literature on the concentration of financial services is the use of bank mergers as an instrument for exogenous variations in local market concentration[21]. One advantage of my empirical strategy is that the identification relies on exploring the variations in the interest rates of mortgages within a county-quarter cell. This is achieved by controlling for county-quarter fixed effects, which absorb any time-varying county-level variations, including shocks to county-level market concentration due to mergers. This makes instrumenting for local-level market concentration unnecessary since the instrument will be absorbed by county-quarter fixed effects. Thus, the identification assumption of Equation (8) is that the interaction between the endogenous components in $\mathbb{1}^{PTF}_{l(i)}$ and the endogenous components in $HHI_{c(i),q(i)}$ is exogenous to mortgage interest rates. This assumption is considerably weaker than assuming the exogeneity of local market concentrations alone.

[21] Some examples are Scharfstein and Sunderam (2016), Yannelis and Zhang (2021), Buchak and Jørring (2021), Avramidis, Mylonopoulos and Pennacchi (2022)

V.2 Main Results on Market Concentration

Table 4 reports the result from Equation (8) using the baseline definition of portfolio lenders. Column (1) reports the result with only quarter-county fixed effects and loan amount and loan purpose controls. Columns (2) to (4) report the results with additional credit risk and time-varying lender controls. The estimations across all specifications show that the portfolio lenders charge a higher interest rate compared to OTD lenders in more concentrated markets. One standard deviation increase in market concentration increases the PL premium by 5.5 to 6.5 bps (15.6 to 18.7 $\times$ 0.35). While adding credit risk and time-varying lender controls does reduce the magnitude of the estimated differential effect, the reduction is relatively small. Importantly, except in Column (1), market concentration explains most of the variations in the PL premium in Columns (2) to (4)[22].

These results are consistent with the predictions in Corollary 1. Therefore, these results provide strong support for the first necessary condition that generates a positive PL premium in Proposition 2, i.e. Lenders have market power in the local conforming mortgage market. And yet, recent literature has found rich evidence suggesting that mortgage lenders exert limited market power in the local conforming mortgage markets (Fuster et al. (2013), Hurst et al. (2016), Nguyen (2019)), at least in terms of mortgage interest rate (Buchak and Jørring (2021)). The explanations provided by earlier findings are that the conforming loans have limited product differentiation and that the GSEs cross-subsidize the borrowers to an extent that many dimensions of credit risks are not priced by the lenders. My findings in this section are not necessarily in conflict with previous findings. In fact, in an auxiliary analysis[23], I also find no relationship between market concentration and mortgage interest rates of both types of lenders as a whole. The main contribution of my findings is the heterogeneous relationships between market competition and lenders with different funding methods in the local markets.

Understanding the market structure at the local level also has important policy implications. At the moment, regulators do not consider the effect of market concentration when evaluating the

[22] Appendix A.6 shows that the result is robust to alternative cutoffs used in the classification of portfolio lenders.
[23] Available upon request.

impact of lender mergers in the local markets, even though market concentration is an important consideration for the evaluation of concentration in the deposit market. However, my findings suggest that though concentration might not have a first-order impact on the average interest rate, it has a strong impact on the relative pricing power between different groups of lenders operating in a local market.

VI Empirical Analysis of Corollary 2

Corollary 2 predicts that when the portfolio lenders face a lower cross-elasticity of demand, the PL premium has a positive relationship with the increase in the demand for mortgages in a local market. This Section provides support for the predictions made in Corollary 2 using a two-stage least squares specification and a Bartik-type instrument for mortgage demand shocks.

VI.1 Empirical Strategy for Testing Corollary 2

To study the difference in the response to credit demand between portfolio lenders and OTD lenders, I need a source of credit demand shocks that the two types of lenders are exposed to at the same time. To identify exogenous variations in credit demand, I construct a Bartik-type instrument for credit demand (Bartik (1991)). I calculate the Bartik instrument as the inner product of the changes in the nationwide number of mortgage inquiries in different borrower groups and the weight of each population group in a given county. To construct population groups, I assign borrowers into twelve credit score bins: below 300, 300-350, 350-400, 400-450, 450-500, 500-550, 550-600, 600-650, 650-700, 700-750, 750-800, and 800 above. Formally, the Bartik instrument for county c in year y is given by:

$$\Delta Bartik_{c,y} = \sum_{b=1}^{11} w_{b,c,3} \Delta ln(Inq)_{b,\underline{c},y} \tag{9}$$

where $\Delta ln(Inq)_{b,\underline{c},y}$ is the change in the national-wide demand in each credit score bin and is calculated as the log change in the total number of mortgage applications in credit score bin b in the whole nation in year y excluding county c. The weight, $w_{b,c,3}$, is calculated as the average of the percentage share of the population who are in each credit score bin b in the first three sample years. Similar to the measurement of market concentration in Section V.1, I use the first three years of the sample to calculate the weight of the Bartik instrument and exclude the observations from the empirical tests to guard against reverse causality concerns. The identification of a Bartik-type instrument lies in the exogenous assignment of either the shocks, the share exposures, or both (Goldsmith-Pinkham, Sorkin and Swift (2020), Borusyak, Hull and Jaravel (2022)).

To investigate how lenders respond to demand shocks, I explore variations of the same lender across different geographic locations that receive different levels of demand shocks. To do so, I collapse the loan-level panel data to the lender-county-year level. I use the following 2SLS specification to estimate the response of mortgage interest rate to demand shock:

$$\Delta \ln(Inq)_{c,y} = \sigma_{l,y} + \eta_c + \beta_1 \Delta Bartik_{c,y} + \mu' \boldsymbol{X}_{l,c,y} + \varepsilon_{l,c,y} \tag{10}$$

$$\Delta Y_{l,c,y} = \sigma_{l,y} + \eta_c + \beta_2 \widehat{\Delta \ln(Inq)}_{c,y} + \mu' \boldsymbol{X}_{l,c,y} + \varepsilon_{l,c,y} \tag{11}$$

where Equations (10) and (11) are the first and second stage specifications, respectively. The outcome variable, $\Delta Y_{l,c,y}$, is the change in the average interest rate or the log change in the total dollar amount of mortgage originations by lender l in county c of year y. The lender-year fixed effects, $\sigma_{l,y}$, capture the time-varying changes in lender-specific characteristics, such as changes in funding lender funding cost. The county fixed effects, η_c capture the changes in local economic conditions. I also include a vector of county-level controls, $\boldsymbol{X}_{c,y}$, which includes the log changes in the following variables: population, average wage, average credit score, and average DTI. I run the 2SLS specification for the portfolio lender and OTD lender samples separately.

To formally test the differential effect of mortgage demand on the two different types of lenders, I explore the differential response to the same mortgage demand shock of two lenders in the same geographic location. Using the lender-county-year level panel data, I estimate the following second-stage specification:

$$\Delta \ln(Inq)_{c,y} = \sigma_{l,y} + \eta_{c,y} + \beta_1 \Delta Bartik_{c,y} \times \mathbb{1}^{\text{PTF}}_{l,y} + \mu' \boldsymbol{X}_{l,c,y} + \varepsilon_{l,c,y} \tag{12}$$

$$\Delta Y_{l,c,y} = \sigma_{l,y} + \eta_{c,y} + \beta_2 \widehat{\Delta \ln(Inq)}_{c,y} \times \mathbb{1}^{\text{PTF}}_{l,y} + \mu' \boldsymbol{X}_{l,c,y} + \varepsilon_{l,c,y} \tag{13}$$

where Equations (12) and (13) are the first and second stage specifications, respectively. Compared to Equations (10) and (11), the key difference in the two specifications above is that the instrumented mortgage demand shock interacts with the lender type dummy, $\mathbb{1}^{\text{PTF}}_{l,y}$. As a result, the coefficient β_2 in (13) estimates the differential effect of being a portfolio lending on the outcome variable following mortgage demand shocks. Since this specification explores the variation between lenders at the county-year level, I am also able to further saturate the dataset by adding county-year fixed effects, $\eta_{c,y}$, which capture the time-varying changes in the local economic conditions.

VI.2 Main Results on Cross-Elasticity of Demand

Table 5 shows the results from the 2SLS specification in Equations (10) and (11). Panel A and B show the results where the outcome variable is the change in the average interest rate at the lender-county level and the change in log change in total dollar amount originated mortgage at the lender-county level, respectively. Columns (1) and (2) show the results using portfolio lenders. The estimated response in rate and volume are both significant. One standard deviation increase in mortgage demand shock (20.9 percentage points) corresponds to a 16.89 bps increase in the mortgage interest rate and a 25.3% increase in origination volume, respectively. Columns (3) and (4) show the results using OTD lenders. While both interest rate and origination volume respond positively to demand shocks, only the response of origination volume is statistically significant. Columns (5) and (6) show the results using bank OTD lenders. The response of the bank OTD lenders acts similarly to the OTD lenders rather than the portfolio lenders, which are predominately banks. The fact that increases in the instrumented mortgage demand increase both the interest rate and origination volume suggests the shocks to the local markets result in the movement of the demand curve along the supply curves of the two types of lenders. Furthermore, the magnitude of the interest rate response of portfolio lenders is considerably higher than that of the OTD lenders, consistent with the prediction in Corollary 2.

Table 6 shows the results from the 2SLS specification in Equations (12) and (13), which formally test the differential effects between the portfolio and OTD lenders following demand shocks. Panel A and B show the results where the outcome variable is the change in the average interest rate at the lender-county level and the change in log change in total dollar amount originated mortgage at the lender-county level, respectively. Columns (1) to (3) show the results using observations of all lenders. Consistent with the prediction in Corollary 2, the portfolio lenders have a strong response in interest rate compared to OTD lenders after receiving demand shocks. Per one standard deviation increase of mortgage demand, the increase in interest rate by portfolio lenders is 12.01 to 12.41 bps higher than that of the OTD lenders. On the other hand, the increase in origination volume of the two types of lenders is not statistically significant. Columns (4) to (6) use the sample of observations that belong to bank lenders, i.e. the sample includes only portfolio lenders and bank OTD lenders, while nonbank lenders are excluded. The results are similar to those using the full sample. Furthermore, after controlling for county-year fixed effects, the origination volume response of the portfolio lenders becomes lower than that of the OTD lenders[24].

Overall, the results shown in this Section are consistent with the predictions in Corollary 2, thereby providing strong support for the second necessary condition that generates a positive PL premium in Proposition 2, i.e. the portfolio lenders face a lower cross-elasticity of demand than the OTD lenders. So far, existing literature on the product differentiation of the conforming loan market in the United States is relatively scarce. One possible reason is that the mortgage contracts in the conforming loan markets are highly standardized. If any product differentiation exists, it

[24] Appendix A.7 presents robustness analysis using panel data collapsed to the county-year level and finds similar results as this section.

is likely coming from differentiated customer service, which is difficult to measure. While this challenge still applies to this paper, the empirical strategy employed in this section can provide support that product differentiation does exist in the conforming loan market.

VI.3 Discussion on the Identification Strategy

One key threat to the identification of Bartik-type instruments is that the correlates of the credit bin shares could predict the changes in the outcome variables. In the 2SLS specification in Equations (12) and (13), the identification assumption is that the correlates of the credit bin shares do not predict the changes in the outcome variables of the portfolio lenders and the OTD lender differentially. One may easily conjecture that other consumer credit characteristics, such as income, debt-to-income ratio, or age, are correlated with credit score. Thus, these variables could also predict the changes in interest rate and origination volume. Even so, it is unlikely that these potential confounding factors will affect the mortgage loans originated by the portfolio lenders and the OTD lenders differentially.

To directly test whether the potential confounding factors predict the changes in the outcome variables for the portfolio lenders and OTD lenders differentially, I collapse the loan-level data to the county-year level and run the following regression:

$$\Delta Y_{c,y}^{\text{PTF}} - \Delta Y_{c,y}^{\text{OTD}} = \sigma_c + \eta_y + \beta_1 \Delta Inc_{c,y-1} + \beta_2 \Delta DTI_{c,y-1} + \beta_3 \Delta age_{c,y-1} + \varepsilon_{c,y} \tag{14}$$

where the outcome variable is the difference in change in the interest rate or log origination volume between the portfolio lenders and the OTD lenders. $\Delta Inc_{c,y}$, $\Delta DTI_{c,y}$, and $\Delta age_{c,y}$. σ_c and η_y are county and year fixed effects, respectively[25]. σ_c and η_y are county and year fixed effects, respectively. If the confounding factor impact the interest rate or origination volume of the two types of lenders differentially, the regression should yield a statistically significant estimation for

[25] I use lagged changes in the independent variables to avoid reverse causality in Equation (14). The results from these two equations still hold when I use contemporary changes in the three independent variables.

β_1, β_2, or β_3.

Table 7 shows the results from Equation (14). Columns (1) to (4) show the results when the outcome variable is the difference in interest rate changes between the portfolio lenders and OTD lenders. Columns (4) to (8) show the results when the outcome variable is the difference in log changes of origination volume between the portfolio lenders and OTD lenders. Across all specifications, the potential confounding factors do not impact the outcome variables of the two types of lenders differentially. Analysis in Appendix A.7.1 shows these confounding factors do impact the two types of lenders separately. This is expected since these factors are presumably correlated with borrowers' credit riskiness, thereby should impact mortgage origination outcomes. However, these factors do not pose threats to the identification assumption as long as they do not impact the two types of lenders differentially.

Since the instrumented mortgage demand shock is a county-year level measurement, one assumption for the interpretation of the results is that the instrumented mortgage demand shock does not impact two different types of lenders differentially. If this assumption is violated, the estimation in the interaction term in Equation (13) could reflect the difference in the change of credit demand between the two groups of lenders. For example, if applicants with college degrees are more skilled in negotiating loan terms with lenders and more applicants with a college degree choose to apply for bank mortgages when credit demand increases, the estimated coefficient might be picking up the effect of change in the proportion college degree borrowers instead of the difference in the credit supply function between banks and nonbanks. To alleviate this concern, I follow the methodology in Armona, Chakrabarti and Lovenheim (2022) and run Equation (12) using the changes in different factors that could potentially impact the demand for the two types of mortgages differentially. Table 8 reports the results. Only three of the estimations statistically significant relationships with Bartik demand shock. In particular, none of the estimated differences between the changes in demographic characteristics between the borrowers from the portfolio lenders and OTD lenders are statistically significant.

VII Conclusion

This paper investigates the competitiveness of bank lenders within the conforming loan market. Contrary to widely accepted notions of this market as highly competitive and homogeneous, I establish that bank lenders financing mortgages on their balance sheets levy an interest rate of 9.12 to 12.77 bps higher than their counterparts. Moreover, I empirically demonstrate the existence of conditions necessitating this market segmentation. Firstly, I show that, in line with lenders possessing market power at the local level, the premium increases by 5.5 to 6.5 bps per standard deviation rise in local market concentration. Secondly, I show that in accordance with portfolio lenders having lower cross-elasticity of demand, the premium ascends by 12.0 to 12.4 bps per ten percentage points increase in mortgage demand. The findings in this paper enhance our understanding of how financing methods affect mortgage interest rates and shed light on the mechanisms of the competition between heterogeneous lenders in local mortgage markets.

References

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

Ambrose, Brent W and Richard J Buttimer Jr, "GSE impact on rural mortgage markets," *Regional Science and Urban Economics*, 2005, *35* (4), 417–443.

Amel, Dean F, Elliot Anenberg, and Rebecca Jorgensen, "On the geographic scope of retail mortgage markets," *Board of Governors of the Federal Reserve System, FEDS Notes, June*, 2018, *15.*

An, Xudong, Yongheng Deng, and Stuart A Gabriel, "Asymmetric information, adverse selection, and the pricing of CMBS," *Journal of Financial Economics*, 2011, *100* (2), 304–325.

Armona, Luis, Rajashri Chakrabarti, and Michael F Lovenheim, "Student debt and default: The role of for-profit colleges," *Journal of Financial Economics*, 2022, *144* (1), 67–92.

Avramidis, Panagiotis, Nikolaos Mylonopoulos, and George G Pennacchi, "The role of marketplace lending in credit markets: Evidence from bank mergers," *Management Science*, 2022, *68* (4), 3090–3111.

Bartik, Timothy J, "Who benefits from state and local economic development policies?," *WE Upjohn Institute for Employment Research Kalamazoo, MI*, 1991.

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

Benveniste, Lawrence M and Allen N Berger, "Securitization with recourse: An instrument that offers uninsured bank depositors sequential claims," *Journal of Banking & Finance*, 1987, *11* (3), 403–424.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel, "Quasi-experimental shift-share research designs," *The Review of Economic Studies*, 2022, *89* (1), 181–213.

Buchak, Greg and Adam Jørring, "Do mortgage lenders compete locally? Implications for credit access," *Implications for Credit Access (January 7, 2021)*, 2021.

_____, Gregor Matvos, Tomasz Piskorski, and Amit Seru, "Beyond the balance sheet model of banking: Implications for bank regulation and monetary policy," *National Bureau of Economic Research*, 2018.

_____, _____, _____, and _____, "Fintech, regulatory arbitrage, and the rise of shadow banks," *Journal of Financial Economics*, 2018, *130* (3), 453–483.

Chomsisengphet, Souphala and Anthony Pennington-Cross, "The evolution of the subprime mortgage market," *Federal Reserve Bank of St. Louis Review*, 2006.

Courchane, Marsha J, Brian J Surette, and Peter M Zorn, "Subprime borrowers: Mortgage transitions and outcomes," *The Journal of Real Estate Finance and Economics*, 2004, *29* (4), 365–392.

DeFusco, Anthony A and Andrew Paciorek, "The interest rate elasticity of mortgage demand: Evidence from bunching at the conforming loan limit," *American Economic Journal: Economic Policy*, 2017, *9* (1), 210–240.

Demyanyk, Yuliya and Elena Loutskina, "Mortgage companies and regulatory arbitrage," *Journal of Financial Economics*, 2016, *122* (2), 328–351.

Federal Reserve, "Report to the Congress on the Effect of Capital Rules on Mortgage Servicing Assets," Technical Report, Board of Governors of the Federal Reserve System 2016.

Fuster, Andreas, Laurie S Goodman, David O Lucca, Laurel Madar, Linsey Molloy, and Paul Willen, "The rising gap between primary and secondary mortgage rates," *Available at SSRN 2378439*, 2013.

——, **Matthew Plosser, Philipp Schnabl, and James Vickery**, "The role of technology in mortgage lending," *The Review of Financial Studies*, 2019, *32* (5), 1854–1899.

Gertler, Mark, Nobuhiro Kiyotaki, and Andrea Prestipino, "Wholesale banking and bank runs in macroeconomic modeling of financial crises," in "Handbook of macroeconomics," Vol. 2, Elsevier, 2016, pp. 1345–1425.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift, "Bartik instruments: What, when, why, and how," *American Economic Review*, 2020, *110* (8), 2586–2624.

Gorton, Gary B and George G Pennacchi, "Banks and loan sales marketing nonmarketable assets," *Journal of monetary Economics*, 1995, *35* (3), 389–411.

—— **and Nicholas S Souleles**, "Special purpose vehicles and securitization," in "The risks of financial institutions," University of Chicago Press, 2007, pp. 549–602.

Han, Joongho, Kwangwoo Park, and George Pennacchi, "Corporate taxes and securitization," *The Journal of Finance*, 2015, *70* (3), 1287–1321.

Hurst, Erik, Benjamin J Keys, Amit Seru, and Joseph Vavra, "Regional redistribution through the US mortgage market," *American Economic Review*, 2016, *106* (10), 2982–3028.

Jiang, Erica Xuewei, "Financing competitors: shadow banks' funding and mortgage market competition," *USC Marshall School of Business Research Paper Sponsored by iORB, No. Forthcoming*, 2019.

Keys, Benjamin J, Tanmoy Mukherjee, Amit Seru, and Vikrant Vig, "Did securitization lead to lax screening? Evidence from subprime loans," *The Quarterly journal of economics*, 2010, *125* (1), 307–362.

Nguyen, Hoai-Luu Q, "Are credit markets still local? Evidence from bank branch closings," *American Economic Journal: Applied Economics*, 2019, *11* (1), 1–32.

Pennacchi, George G, "Loan sales and the cost of bank capital," *The Journal of Finance*, 1988, *43* (2), 375–396.

——, "Banks, taxes, and nonbank competition," *Journal of Financial Services Research*, 2019, *55* (1), 1–30.

Petersen, Mitchell A and Raghuram G Rajan, "The benefits of lending relationships: Evidence from small business data," *The journal of finance*, 1994, *49* (1), 3–37.

—— **and** ——, "The effect of credit market competition on lending relationships," *The Quarterly Journal of Economics*, 1995, *110* (2), 407–443.

Purnanandam, Amiyatosh, "Originate-to-distribute model and the subprime mortgage crisis," *The review of financial studies*, 2011, *24* (6), 1881–1915.

Scharfstein, David and Adi Sunderam, "Market power in mortgage lending and the transmission of monetary policy," *Unpublished working paper. Harvard University*, 2016, *2.*

Seru, Amit, "Regulation, technology and the banking sector," *Macroeconomic Review*, 2020.

Shi, Lan, Yan Zhang, and Xinlei Shelly Zhao, "How Do Banks Respond to Non-Bank Competition in the Conforming Residential Mortgage Market? The Role of Balance Sheet," *The Role of Balance Sheet (January 1, 2023)*, 2023.

Stanton, Richard, Johan Walden, Nancy Wallace et al., "The industrial organization of the US residential mortgage market," *Annual Review of Financial Economics*, 2014, *6* (1), 259–288.

Yannelis, Constantine and Anthony Lee Zhang, "Competition and selection in credit markets," *National Bureau of Economic Research*, 2021.

Table 1: Summary Statistics

Panel A: Mortgage Sample			
	Mean	SD	# of Obs.
Dollar Amount (USD)	216,907.29	111,351.93	893,253
Estimated Interest Rate (bp)	549.77	255.83	686,469
Account Balance in First Snapshot (USD)	182,809.81	126,689.47	893,253
Account Balance in Final Snapshot (USD)	70,210.65	117,420.62	893,253
Refinance	37.2%		893,253
Securitized via GSEs	57.8%		893,253
Bank Originated	83.6%		893,253
Vantage Score	728.02	77.14	893,253
Estimated Income (1k USD)	105.55	66.04	887,553
Estimated Debt-to-Income Ratio (pct)	26.42	18.26	892,219
Female	0.44	0.50	274,902
Marriage Indicator	0.70	0.46	274,902
Homeowner Indicator	0.72	0.45	274,902
College and Above	0.45	0.50	274,902
# of Adults in Household	2.51	1.38	274,902
# of Children in Household	0.48	0.98	274,902

Panel B: Mortgage Inquiry Sample			
	Mean	SD	# of Obs.
Vantage Score	674.93	114.71	7,237,580
Estimated Income (1k USD)	92.07	73.32	7,176,923
Estimated Debt-to-Income Ratio (pct)	24.39	18.28	6,872,367
Female	0.43	0.49	2,474,974
Marriage Indicator	0.64	0.48	2,474,974
Homeowner Indicator	0.60	0.49	2,474,974
College and Above	0.34	0.47	2,474,974
# of Adults in Household	2.46	1.42	2,474,974
# of Children in Household	0.49	0.98	2,474,974

Panel C: Lender Sample			
	Mean	SD	# of Obs.
Avg. # of Credit Card Accounts per Year	2,067.59	56457.75488	4,488
Avg. # of Auto Loan Accounts per Year	431.59	5448.925978	4,488
Avg. Amount of Outstanding Balance ($1m)	37.60	616522338.6	4,488
Avg. Mortgage Interest Rate (bp)	521.18	156.6242523	4,174
% of Balance-Sheet-Financed Mortgages	80.4%	30.4%	4,488
% of Refinance Mortgages	25.2%	24.6%	4,488
Number of Operating Counties	24.13	130.8580306	4,488
Number of Operating States	3.65	7.725649753	4,488
Bank Lender	96%		4,262

Table 2: Loan-Level Evidence of PL premium

	Dependent Variable: Interest Rate (bps)					
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Full Sample					
Portfolio Lender	12.53***	9.49***	9.61***	12.77***	9.95***	9.12***
	(0.62)	(0.61)	(0.62)	(0.85)	(1.28)	(1.35)
Observations	507571	506250	498125	492459	491639	461126
R-Squared	0.35	0.37	0.38	0.39	0.44	0.49
	Panel B: Bank Sample					
Portfolio Lender	15.79***	11.55***	11.93***	24.38***	8.76***	8.14***
	(0.72)	(0.72)	(0.72)	(1.08)	(1.54)	(1.59)
Observations	385020	383950	377650	374697	373921	351043
R-Squared	0.36	0.39	0.39	0.40	0.46	0.50
Loan Controls	Y	Y	Y	Y	Y	Y
Credit Risk Controls		Y	Y	Y	Y	Y
Credit Risk High-Orders			Y	Y	Y	Y
Lender Controls				Y	Y	Y
Quarter-County FEs	Y	Y	Y	Y	Y	Y
Lender FEs					Y	
Lender-County FEs						Y

Note: This table reports the results from Equation (6). The observations are at the loan level. Panels A and B report the estimations using the full sample and the bank sample, respectively. ”Loan Controls” include the loan amount, and loan purposes (refinance or new purchase). ”Credit Risk Controls” include borrower credit score, borrower debt-to-income ratio, and borrower income. ”Credit Risk High-Orders” include the second, the third, and the fourth power of all controls within ”Credit Risk Controls”. ”lender Controls” include the total amount of loan balance, the total number of credit card accounts, and the total number of auto loan accounts at the lender-quarter level. All standard errors are clustered at the county-quarter level and reported in parentheses.

Table 3: Numerical Example for the Conceptual Framework

Panel A: Parameter Values							
a	b	b_{12}	b_{21}	r_1	r_2	N	δ
10	0.03	0.01	0.02	0.02	0.05	400	0.5

Panel B: Model Solution					
	Q_1	Q_2	R_1	R_2	Δ^R
	233.71	73.10	667.42	646.20	21.22

Table 4: PL premium and Market Concentration

	Dependent Variable: Interest Rate (bps)			
	(1)	(2)	(3)	(4)
Balance Sheet Lender	3.88***	2.26*	2.24*	2.00
	(1.25)	(1.24)	(1.25)	(1.30)
Balance Sheet Lender	22.89***	18.68***	19.26***	15.62***
× HHI	(4.51)	(4.46)	(4.50)	(4.52)
Loan Controls	Y	Y	Y	Y
Credit Risk Control		Y	Y	Y
Credit Risk High-Order			Y	Y
Lender Controls				Y
Quarter-County FEs	Y	Y	Y	Y
Observations	465737	464610	457586	452632
R-Squared	0.329	0.355	0.36	0.372

Note: This table reports the results from Equation (8), where the measure of market concentration is defined in Equation (7). The observations are at the loan level. ”Loan Controls” include the loan amount, and loan purposes (refinance or new purchase). ”Credit Risk Controls” include borrower credit score, borrower debt-to-income ratio, and borrower income. ”Credit Risk High-Orders” include the second, the third, and the fourth power of all controls within ”Credit Risk Controls”. ”lender Controls” include the total amount of loan balance, the total number of credit card accounts, and the total number of auto loan accounts at the lender-quarter level. All standard errors are clustered at the county-quarter level and reported in parentheses.

Table 5: Effects of Mortgage Demand Shock

	Portfolio lenders		OTD lenders		Bank OTD lenders	
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Dependent Variable: ΔInterest rate					
$\widehat{\Delta \ln(Inq)}_{c,y}$	83.52**	76.01**	32.23	17.71	53.33**	38.52
	(38.02)	(37.25)	(24.28)	(22.92)	(25.94)	(24.55)
Number of Obs	25023	24975	36637	36580	21540	21510
F-statistics	113.1	113.5	117.8	118.3	149.5	149.9
	Panel B: Dependent Variable: Δln(vol)					
$\widehat{\Delta \ln(Inq)}_{c,y}$	1.21***	1.25***	0.68**	0.68**	0.74***	0.77***
	(0.37)	(0.38)	(0.34)	(0.33)	(0.27)	(0.26)
Number of Obs	31070	30951	40744	40646	23811	23761
F-statistics	112.8	113.1	121.5	121.6	154.3	154.3
Credit Risk Controls		Y		Y		Y
Lender-Year FEs	Y	Y	Y	Y	Y	Y
County FEs	Y	Y	Y	Y	Y	Y

Note: This table reports the results from the 2SLS test in Equations (10) and (11). The observations are at the lender-county-year level. Panel A reports the results when the outcome variable is the change in average interest rate at the lender-county level. Panel B reports the results when the outcome variable is the change in the log of the dollar amount of originated mortgage at the lender-county level. Columns (1) and (2) use the observations that are classified as portfolio lenders by the baseline classification. Columns (1) and (2) use the observations that are classified as OTD lenders by the baseline classification. Columns (1) and (2) use the observations that are classified as both bank and portfolio lenders by the baseline classification. All regressions are weighted by the number of mortgages in each county-year cell. All standard errors are clustered at the county level and reported in parentheses.

Table 6: Differential Effects of Mortgage Demand Shock

	All lenders			Bank lenders		
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Dependet Variable: ΔInterest rate					
$\widehat{\Delta \ln(Inq)}_{c,y} \times \mathbb{1}^{\text{PTF}}_{l,y}$	58.74***	57.46***	59.41***	59.64***	57.71***	59.80***
	(7.21)	(6.93)	(7.27)	(7.35)	(7.33)	(8.03)
Number of Obs	66413	66296	62160	49446	49360	45191
F-statistics	4284.2	4296.2	4434.8	3500.5	3503.1	4171.0
	Panel B: Dependet Variable: Δln(vol)					
$\widehat{\Delta \ln(Inq)}_{c,y} \times \mathbb{1}^{\text{PTF}}_{l,y}$	0.01	0.03	-0.04	-0.07	-0.06	-0.21***
	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)
Number of Obs	77226	76996	72681	57822	57651	53349
F-statistics	4521.9	4539.2	4719.4	3770.6	3777.5	4390.1
Credit Risk Controls			Y			Y
Lender-Year FEs	Y	Y	Y	Y	Y	Y
County FEs		Y			Y	
County-Year FEs			Y			Y

Note: This table reports the results from the 2SLS test in Equations (12) and (13). The observations are at the lender-county-year level. Panel A reports the results when the outcome variable is the change in average interest rate at the lender-county level. Panel B reports the results when the outcome variable is the change in the log of the dollar amount of originated mortgage at the lender-county level. Columns (1) to (3) use the full sample. Columns (4) to (6) use the observations that are classified as bank lenders. Credit risk controls include the change in the average credit score, average income, and average debt-to-income ratio for all consumers who take mortgages from the given lender in the given county from the previous year. All regressions are weighted by the number of mortgages in each county-year cell. All standard errors are clustered at the county level and reported in parentheses.

Table 7: Robustness Test for the Differential Effects of Confounding Factors

	Dep. Var.: ΔInterest rate				Dep. Var.: Δln(Vol)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ Age	-0.51			-0.65	-0.08			-0.09
	(0.45)			(0.48)	(0.12)			(0.13)
Δ Income		-0.02		0.00		0.02		0.03
		(0.13)		(0.13)		(0.03)		(0.03)
Δ DTI			-0.66	-0.75			0.06	0.04
			(0.47)	(0.47)			(0.11)	(0.12)
County FEs	Y	Y	Y	Y	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y	Y	Y	Y	Y
Number of Obs	10458	10471	10471	10458	11305	11330	11329	11304
R-Squared	0.073	0.072	0.073	0.074	0.061	0.062	0.062	0.061

Note: This table reports the results from Equation (14). The observations are at the county-year level. Columns (1) to (4) report the results when the outcome variable the change in average interest rate at lender-county level. Columns (5) to (8) report the results when the outcome variable the change in log of dollar amount of originated mortgage at lender-county level.All regressions are weighted by the number of mortgages in each county-year cell. All standard errors are clustered at the county level and reported in parentheses.

Table 8: Robustness Test for Exclusion Restriction Violation through Demographics

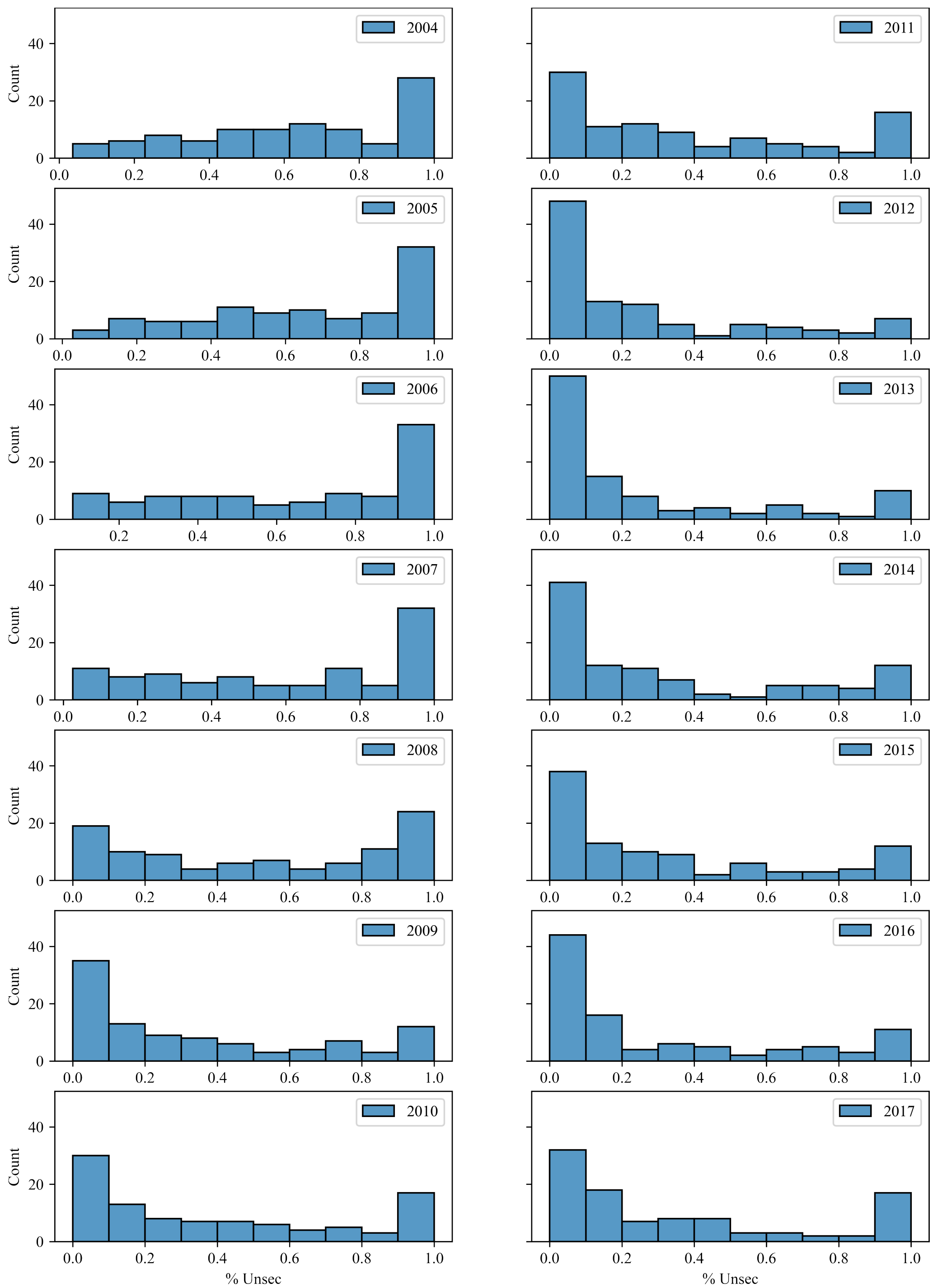
Variable name	Coef.	SE
Δ ln(population)	0.30	(0.32)
Δ dividend income	0.03	(0.02)
Δ interest income	-0.02	(0.07)
Δ wage income	0.01	(0.01)
Δ total income	0.00	(0.00)
Δ debt-to-income ratio	0.00	(0.00)
Δ credit score	0.00	(0.00)
Δ age	-0.02	(0.04)
Δ # of adults in household	0.05	(0.03)
Δ # of children in household	0.04	(0.10)
Δ percentage of homeowners	-0.03	(0.12)
Δ percentage of female	-0.46**	(0.23)
Δ percentage of married	0.12	(0.10)
Δ percentage of college-educated	0.00	(0.11)
Δ age of all approved	0.00	(0.00)
Δ age approved by portfolio lenders	-0.00*	(0.00)
Δ age approved by OTD lenders	0.00	(0.00)
Δ # of adults in household of all approved	0.00	(0.00)
Δ # of adults in household approved by portfolio lenders	0.00	(0.00)
Δ # of adults in household approved by OTD lenders	0.00	(0.00)
Δ # of children in household of all approved	-0.01	(0.01)
Δ # of children in household approved by portfolio lenders	-0.01*	(0.00)
Δ # of children in household approved by OTD lenders	0.00	(0.00)
Δ percentage of homeowners of all approved	-0.02	(0.01)
Δ percentage of homeowners approved by portfolio lenders	0.01	(0.01)
Δ percentage of homeowners approved by OTD lenders	-0.01	(0.01)
Δ percentage of female of all approved	0.01	(0.01)
Δ percentage of female approved by portfolio lenders	0.01	(0.01)
Δ percentage of female approved by OTD lenders	0.00	(0.01)
Δ percentage of married of all approved	0.00	(0.00)
Δ percentage of married approved by portfolio lenders	0.00	(0.00)
Δ percentage of married approved by OTD lenders	0.00**	(0.00)
Δ percentage of college-educated of all approved	-0.02	(0.01)
Δ percentage of college-educated approved by portfolio lenders	-0.01*	(0.01)
Δ percentage of college-educated approved by OTD lenders	0.01	(0.01)

Note: This table reports the results from Equation (12) with different outcome variables. The outcome variables are reported in the "Variable name" Column. All of the regressions reported in this table have the same set of fixed effects (including county fixed effect, year fixed effect) and controls (including change in county-level population, change in the county-level average wage, change in county-level average credit score, and change in county-level average DTI). All of the regressions are weighted by county population. All standard errors are clustered at the county level and reported in parentheses.

Figure 1: Time Series of Avg. Interest Rates

Note: This figure shows the time series of average interest rates of balance-sheet-financed mortgages (blue line) and securitized mortgages (red line) at the national level. Panel 1 shows the time series of the national average interest rates.

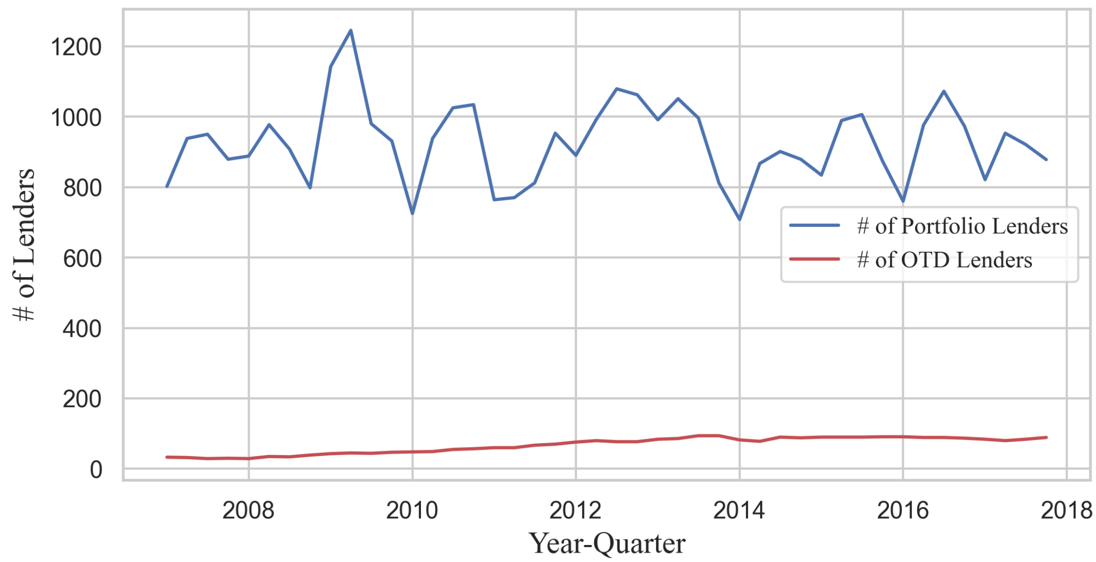
Figure 2: Lender Distribution by % of Balance-Sheet-Financed Mortgages



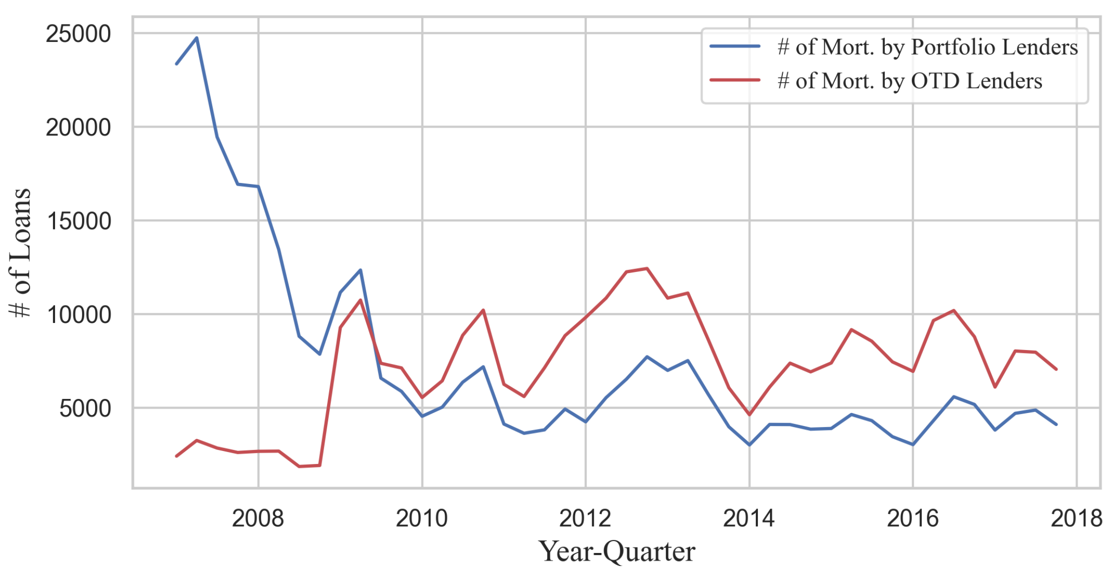
Note: This figure plots the distribution of the top 100 lenders by the percentage of balance-sheet-financed mortgages. The top 100 lenders are selected based on the total number of mortgages the lenders originated from 2004 to 2017. The percentage of balance-sheet-financed mortgages for each of the top 100 lenders for a given year is calculated as the fraction of mortgages that are originated but not securitized via the GSEs amongst all mortgages originated by the lender during the year.

Figure 3: Number of Lenders by financing methods

(a) Number of Lenders by financing methods



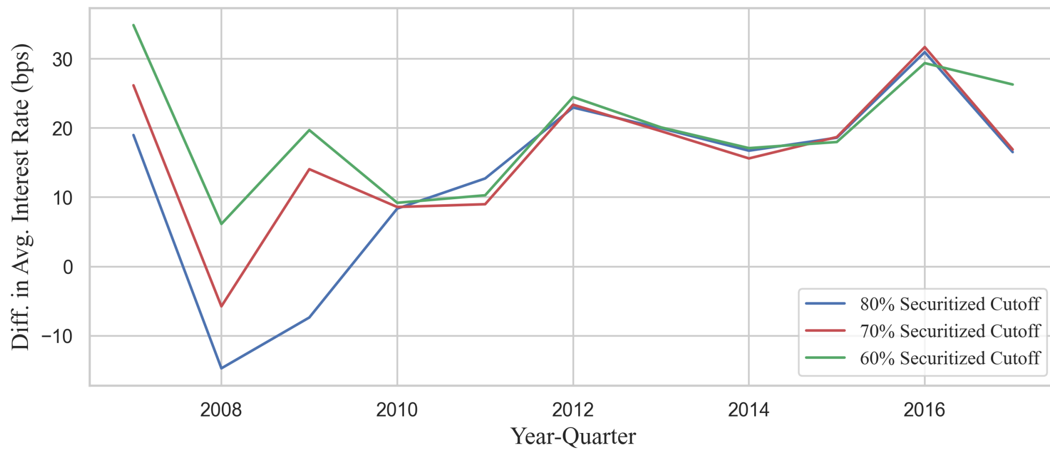
(b) Number of Originated Mortgages by Lender Types



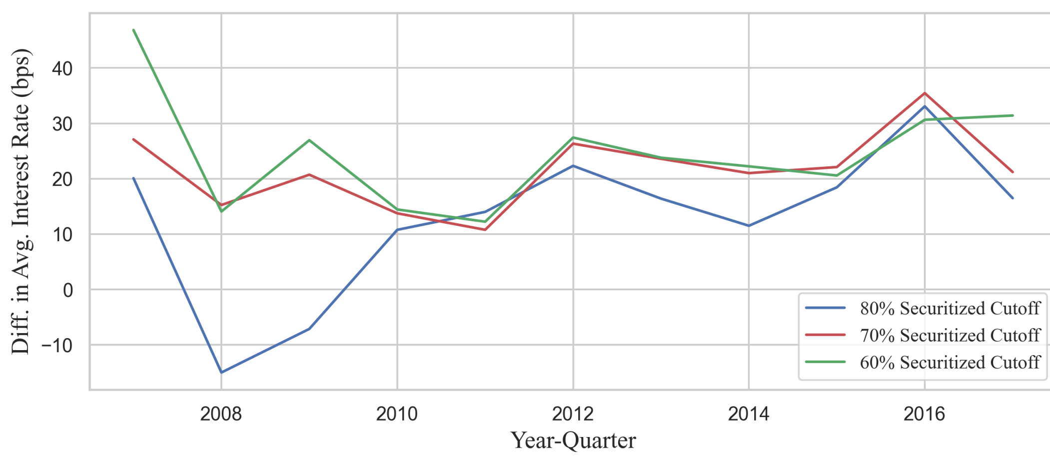
Note: This figure shows the fraction of lenders that are classified as portfolio lenders as well as the fraction of mortgages that are originated by portfolio lenders. Panel 3a shows the number of portfolio lenders and the number of OTD lenders from 2007 to 2017. The classification method is the baseline classification described in Section IV.1. Panel 3b shows the number of mortgages originated by portfolio lenders and the number of mortgages originated by OTD lenders from 2007 to 2017.

Figure 4: Avg. Portfolio Lenders Interest Rate minus Avg. OTD Lenders Interest Rate

(a) All Mortgages

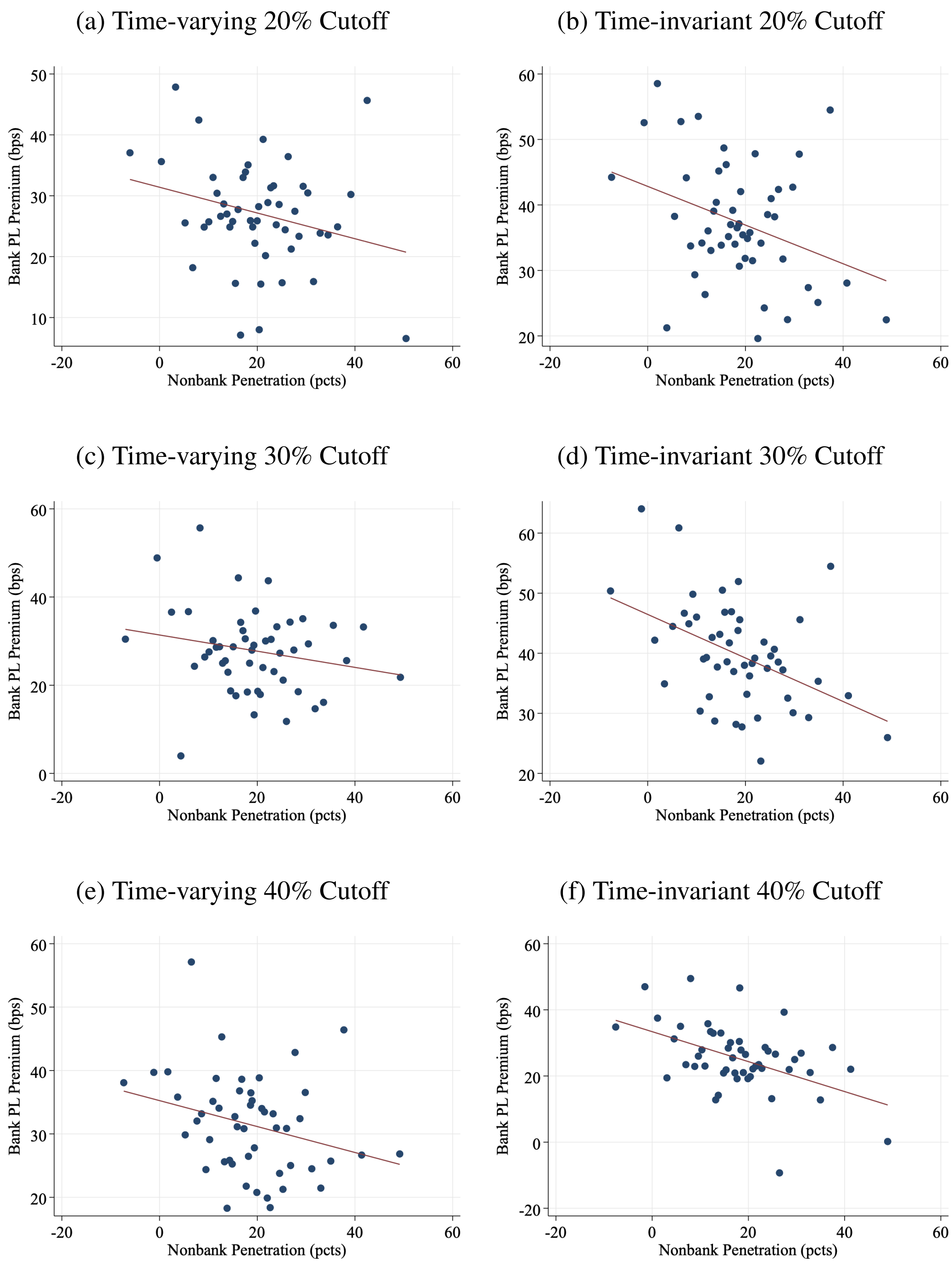


(b) Bank Mortgages



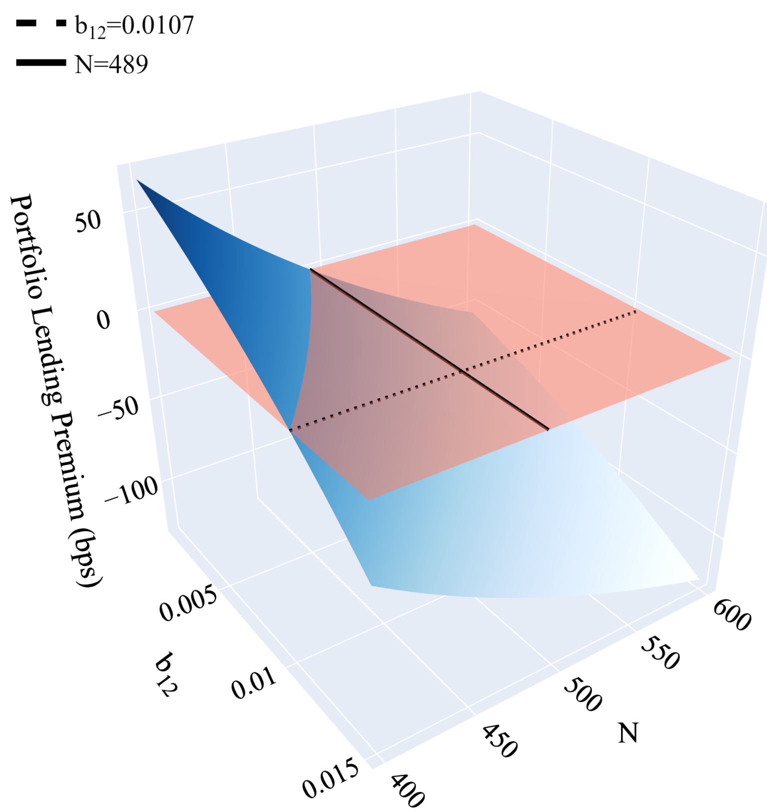
Note: This figure shows the gap between the average interest rate of mortgages originated by portfolio lenders and the average interest rate of mortgages originated by the OTD lenders, i.e. the PL premium. The blue, red, and green lines plot the premium calculated using the 80%, 70%, and 60% cutoffs, respectively. Panel 4a shows the difference in average interest rate in the full sample that includes both bank-originated and nonbank-originated mortgages. Panel 4b shows the difference in average interest rate in the bank-originated subsample. Bank-originated mortgages constituted about 83% of all 30-year conforming mortgages in my sample.

Figure 5: Binscatter plots of the PL premium within banks w.r.t. nonbank penetration



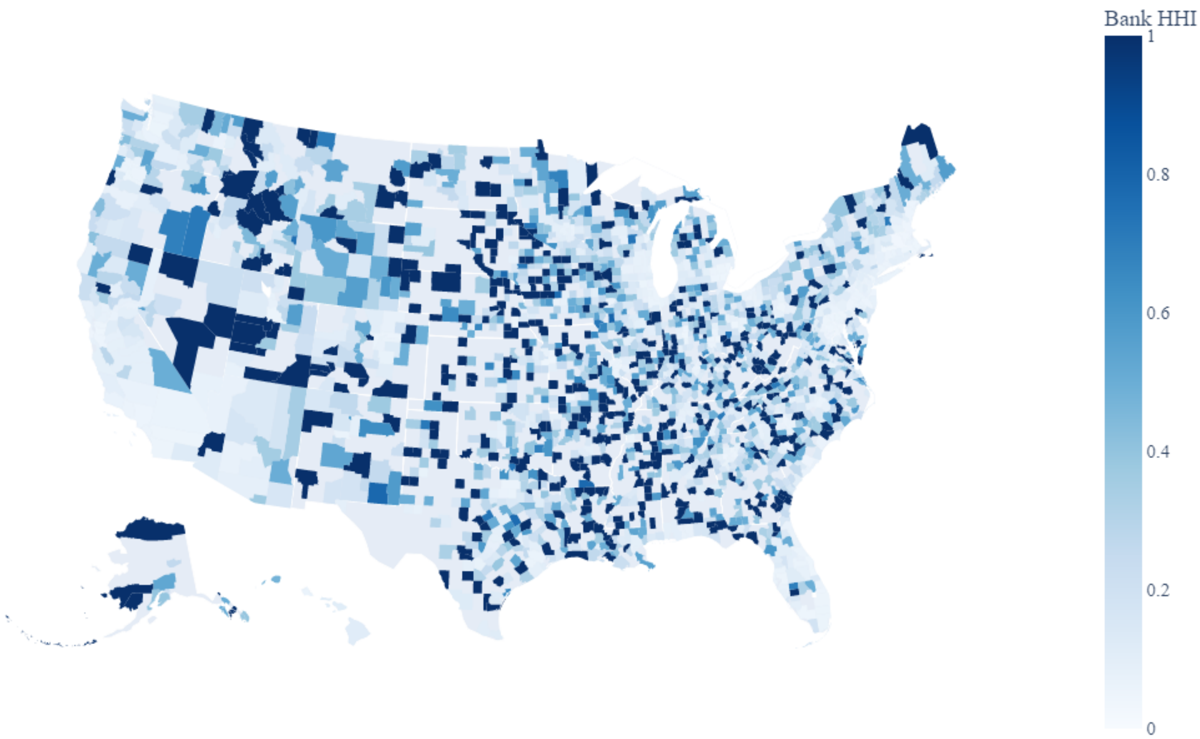
Note: This figure shows the binscatter plots of the PL premium against county-level nonbank penetration. Nonbank penetration is measured by the percentage of mortgage loans that are originated by nonbank lenders. The plots are residualized on average credit score, income, debt-to-income ratio, percentage of refinance mortgages, and the average dollar amount of the originated mortgages of each county-year observation, year fixed effects, as well as county fixed effects. All plots are weighted by the number of loans in each county-year observation. The three plots on the left column use 20%, 30%, and 40% time-varying cutoffs to classify the portfolio lenders, respectively. The three plots on the left column use 20%, 30%, and 40% time-invariate cutoffs to classify the portfolio lenders, respectively.

Figure 6: Numerical Example



Note: This figure illustrates the relationship between the PL premium and the three parameters in the model discussed in Section II.2.1.

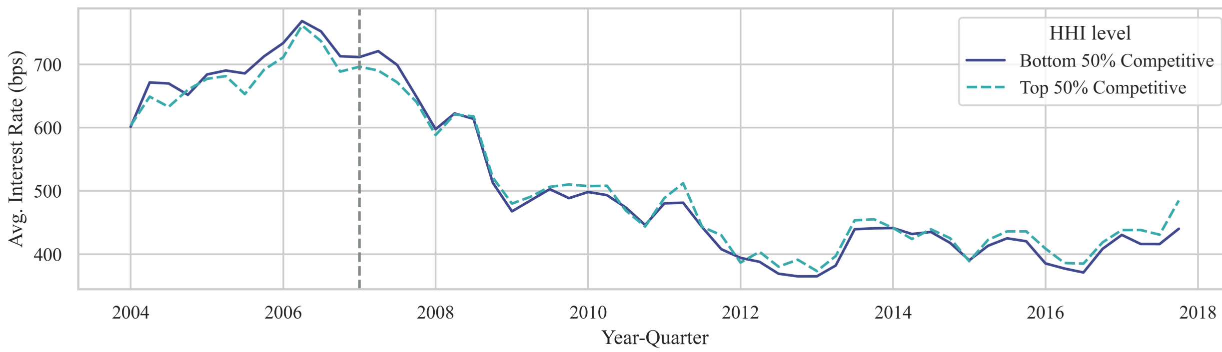
Figure 7: Geographic Distribution of Market Concentration



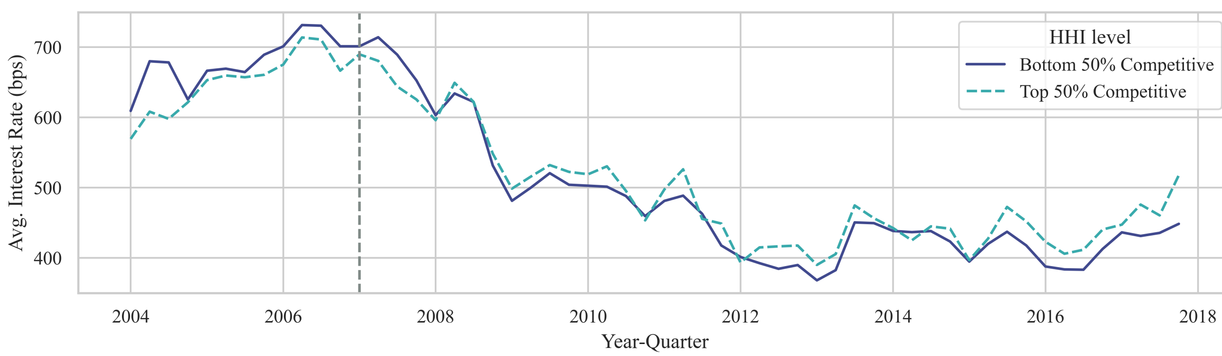
Note: This figure shows the geographic distribution of the county-level average of the baseline HHI as defined in Equation (7) in the year 2015.

Figure 8: Parallel Trends of Avg. Interest Rate by HHI levels

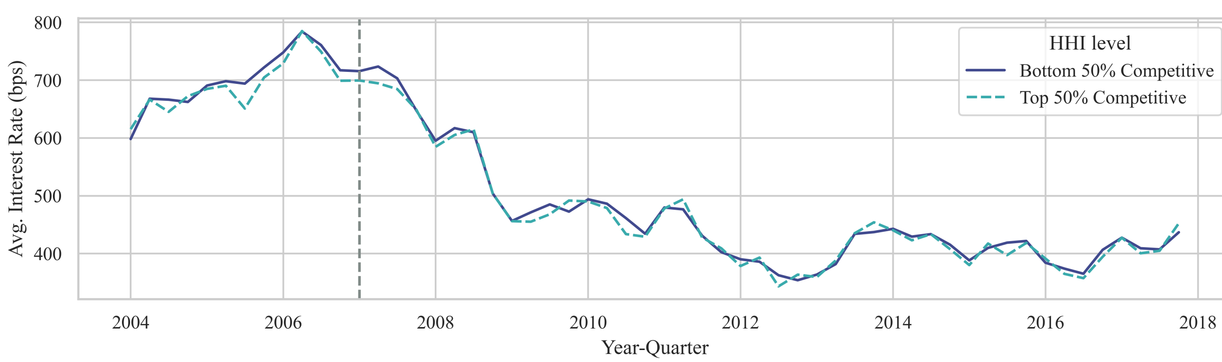
(a) Avg. Interest Rate of All Mortgages by HHI levels



(b) Avg. Interest Rate of Bank Mortgages by HHI levels



(c) Avg. Interest Rate of Nonbank Mortgages by HHI levels

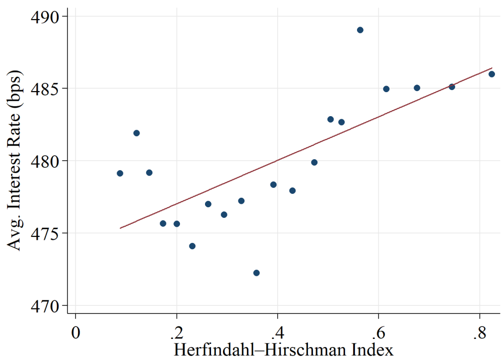


Note: This figure parallels trends of average interest rate in counties of different levels of market concentrations. I define a top-50% concentrated county as a county whose average baseline HHI is in the top 50% amongst all counties. The bottom 50% concentrated counties are defined similarly. The vertical dash line indicates the last quarter of 2006, which is the starting time of the baseline HHI measure calculated using a three-year lagged window as defined in Equation (7).

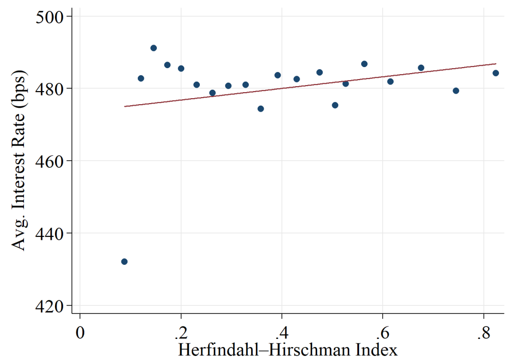
Figure 9: Binscatter plots of average interest rate against market concentration

Panel A: All Mortgages

(a) With year FE

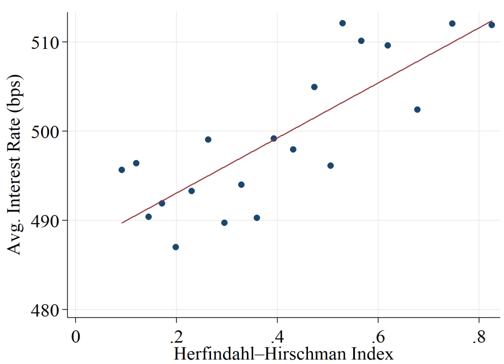


(b) Without year FE

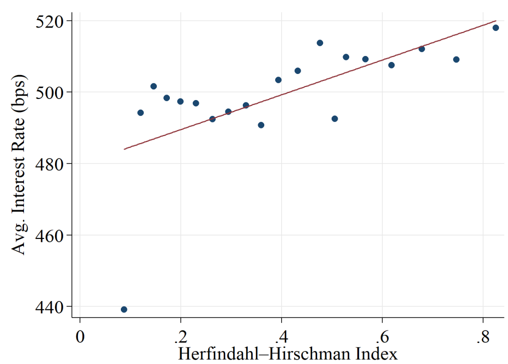


Panel B: Portfolio Lender Mortgages

(c) With year FE

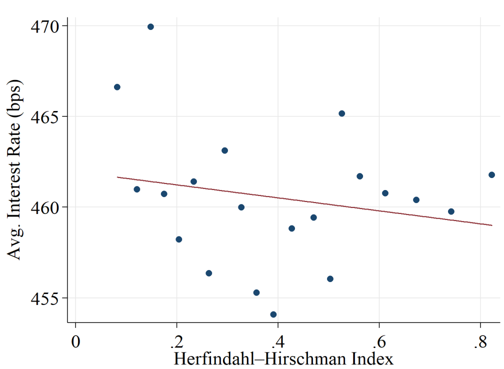


(d) Without year FE

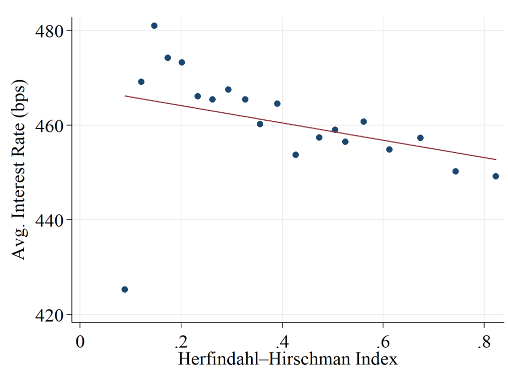


Panel C: OTD Mortgages

(e) With year FE



(f) Without year FE



Note: This figure shows the binscatter plots of county-level average interest rate against county-level Herfindahl–Hirschman Index. Figures 9a, 9c, and 9e are demeaned by yearly average interest rate. Figures 9b, 9d, and 9f are not demeaned.

A Appendix

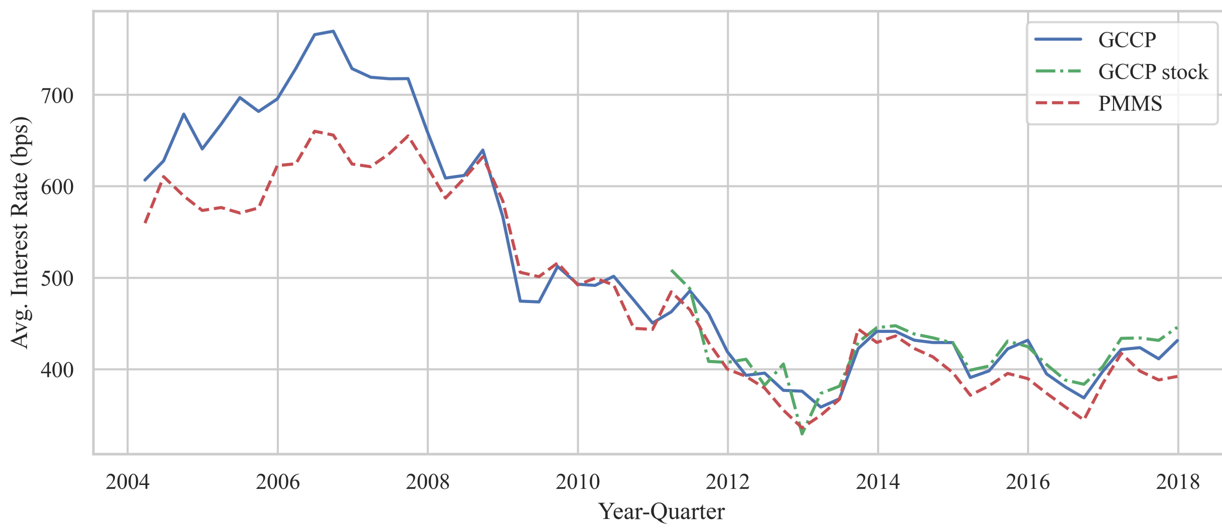
A.1 Validation of Interest Rate

I calculate mortgage interest rates using a root-solving algorithm. Figure A.1 shows the comparison between the quarterly average interest from GCCP and the national average interest rate of the 30-year fixed-rate conforming mortgage obtained from Freddie Mac's Primary Mortgage Market Survey (PMMS). Note that there are mainly three differences between my national average calculation (the GCCP national average) and PMMS: (1) The GCCP national average is calculated using estimated rates from originated loans, while the sample points used in PMMS are obtained from survey responses from the lenders; (2) The GCCP average includes all conventional mortgages that meet the conforming loan amount limit, while PMMS only includes mortgages with LTV equal to or lower than 80%; (3) The GCCP national average is calculated as the simple average of all originated mortgage loans. On the other hand, the PMMS national average is calculated as the weighted average of reported interest rates by all surveyed lenders across the United States, with the weights being the lenders' origination volumes.

In Figure A.1, the time-series trends of the GCCP national average and the PMMS national average move very close to one another. During the earlier years of the comparison period, the average interest rate in GCCP is between 50 to 120 bps higher than the PMMS average. This is likely due to the higher number of originated mortgages with low down payment during the 2008 subprime bubble. GCCP also contains Experian-estimated interest rates (GCCP stock rate) for a small subset of the mortgages. The GCCP stock rate becomes available starting in the year 2011 but only becomes populated in the year 2015. Despite that my estimated interest rate is still around 10 to 20 bps higher than the PMMS national average rate in the later years of the sample, the figure shows that my estimated interest rate moves much closer to the stock interest rate estimated by Experian, suggesting that the difference between my estimation and PMMS is likely due to differences in sampling and/or weighting methods rather than measurement errors.

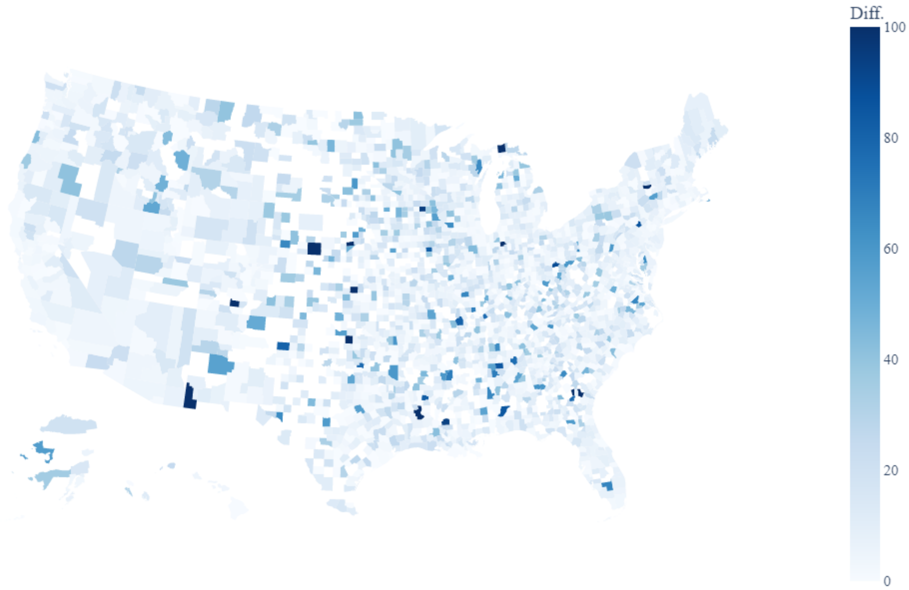
Figure A.2 plots the deviation in the average yearly interest rate between my estimated and the GCCP stock rate at the county level in the year 2016. The difference is calculated as $abs(R - R_{stock})/R_{stock})$. For example, if a county has an average interest rate of 450 bps according to GCCP and an average interest rate of 500 bps according to HMDA, the deviation would be $|450\% - 500\%|/500\% = 10\%$. The figure shows that the deviation between the GCCP and HMDA is below 10% for the majority of the counties.

Figure A.1: National Average of 30-year Conforming Mortgage Interest Rate



Note: This figure plots the average interest rate of 30-year conforming mortgage interest rate in GCCP and PMMS. The national average calculated by PMMS only includes fixed-rate mortgages with LTV equal to or lower than 80%, while the GCCP average includes all conventional mortgages that meet the conforming loan amount limit.

Figure A.2: National Average of 30-year Conforming Mortgage Interest Rate



Note: This figure plots the county-level difference between the interest rate estimation from the GCCP and the stock interest rate provided by Experian (GCCP stock rate). The difference is calculated as $abs(R - R_{stock})/R_{stock}$.

A.2 Potential Measurement Errors

For clarity, I use the italicized term *lender_key* to refer to the variable name in the GCCP database in this section. I use the *lender_key* of the first snapshot of each mortgage loan in the GCCP dataset to identify the lender of each mortgage. Next, I discuss two potential sources of measurement errors.

One potential source of measurement error is caused by the fact that information on a mortgage is only available when its first snapshot is recorded in the GCCP database. Thus, it is possible that the original lender sells a mortgage it originated to another lender before the first snapshot date of the mortgage[26]. In this case, the *lender_key* might identify the second-hand purchaser, but not the originator.

Figure A.3 shows that around 93% of the mortgages have their first snapshot dates within two years after their origination dates, while Figure A.4 that only a small fraction of mortgages have different *lender_key*s during the first two years of the sample. Thus, the total number of missing *lender_key* transactions is unlikely to be large for the 7% of the mortgages whose first snapshot dates are more than two years after the origination date.

Another possibility is that the *lender_key* in the GCCP data sometimes identifies the mortgage servicers instead of the mortgage originator. This could cause misclassification when the mortgage originator sells the mortgage servicing rights (MSR) separately from the cash flow rights prior to the first snapshot date. The most common scenario when such transactions occur is when an originator securitizes its mortgages through a service-release sale via a GSE[27]. In a service-release sale, a lender sells the cash flow rights of the mortgages to a GSE and sells the MSR to one or more transferee servicers, which are often nonbank lenders.

Figure A.5 graphically illustrates the likelihood of the second type of measurement error.

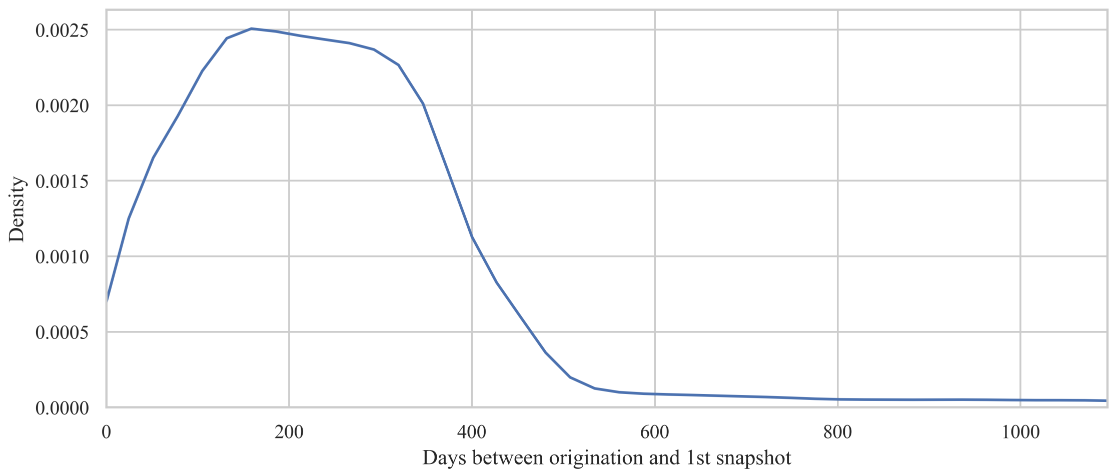
[26] Stanton et al. (2014) has a detailed discussion about the market structure of wholesale lending.

[27] It is also possible for an originator to sell the MSR to another lender while continuing to finance the mortgage on its balance sheet. While a secondary MSR market exists, transactions often occur in a large lump sum when lenders want to adjust their exposure to MSR. It is unlikely that the frequency of such transactions will systematically bias the sample.

First, misclassifications between mortgages from portfolio lenders and mortgages from OTD lenders are unlikely to occur when OTD lenders release MSR, because the purchasers of MSR are most likely also OTD lenders[28]. Misclassifications are more likely to occur when portfolio lender release their MSR to OTD lenders. However, the portfolio lenders, which are almost certainly banks, typically keep the MSR of the mortgages they originate. About 95% of the mortgages securitized by the banks retained their servicing rights (Federal Reserve (2016)). While data on servicing rights release of mortgages from portfolio lenders is not available, servicing rights are not likely to be released much more often than the securitized mortgages within the first snapshot dates. A simple calculation using the baseline portfolio lender classification shows that only 2.4% of the mortgages are misclassified as OTD mortgages. Thus, it is unlikely that this measurement error will significantly distort my estimation.

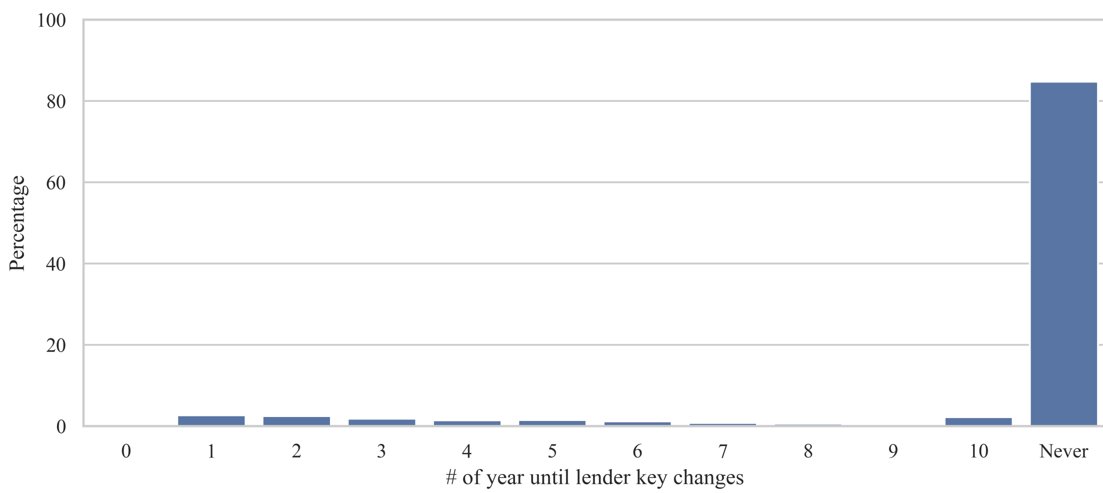
[28]It is also possible that some specialized mortgage servicing companies purchase MSR from lenders. But since specialized mortgage servicing companies are not portfolio lenders, the existence of these institutions is not going to bias the number of mortgages from portfolio lenders either.

Figure A.3: Distribution of days before 1st snapshot date



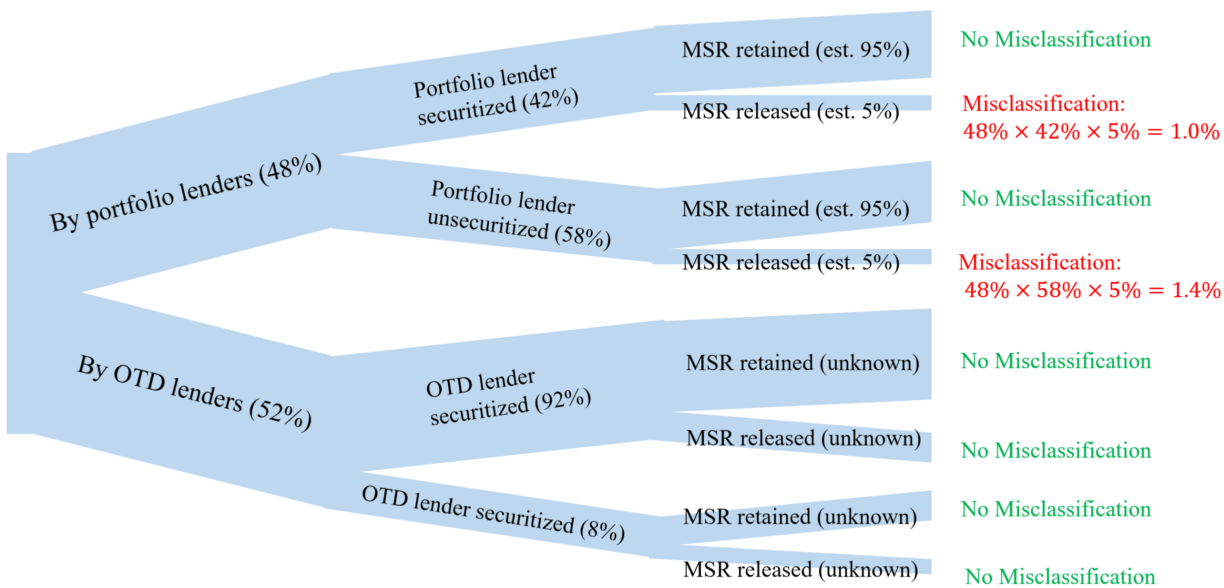
Note: This figure plots the distribution of the number of days between the mortgage origination date and the first snapshot date.

Figure A.4: Year of first *lender_key* change



Note: This figure plots the percentage of mortgages whose *lender_key*s have changed over the sample period. The horizontal axis is the number of years that have passed until the first *lender_key* change occurs. The rightmost bar plots the percentage of mortgages whose *lender_key* has never changed over the sample years.

Figure A.5: National Average of 30-year Conforming Mortgage Interest Rate



Note: This figure shows situations in which mortgages from portfolio lenders are misclassified as OTD mortgages and estimates the percentage of misclassification in each situation. The estimates in this figure are based on the baseline portfolio lender classification in Section IV.1. About 48% of the mortgages in the final conforming mortgage sample is from the portfolio lenders. Of the 48% of the mortgages from portfolio lenders, 42% are securitized and 58% are retained on the balance sheet of the lenders. I choose 5% as the estimated percentage of servicing release mortgages. The estimated misclassification due to MSR release of securitized and balance-sheet-financed mortgages are 1% and 1.4%, respectively.

A.3 Classification of Bank Lenders

In this paper, I classify lenders into two categories: bank lenders and nonbank lenders. I define a lender as a bank lender if it finances a large proportion of mortgages on its balance sheet or has consideration operations of credit products other than mortgage loans. Formally, I classify a *lender_key* as a bank if it meets one of the following criteria:

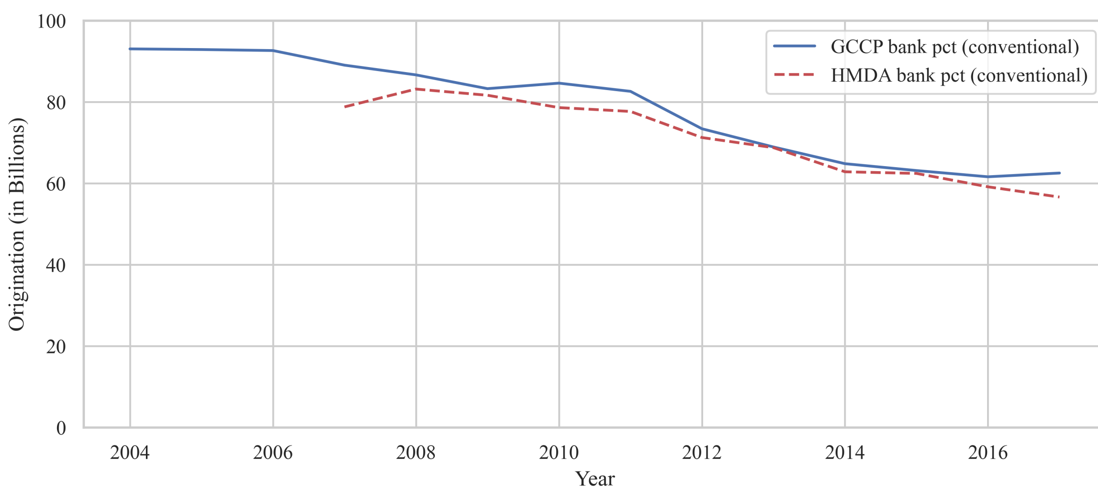
- Less than 80% of all conforming mortgages under its name are securitized throughout the whole sample period
- Both credit card balance AND auto loan balance account for more than 1% of its total outstanding consumer loan balance throughout the whole sample period
- Both credit card balance AND student loan balance account for more than 1% of its total outstanding consumer loan balance throughout the whole sample period
- Both auto loan balance AND student loan balance account for more than 1% of its total outstanding consumer loan balance throughout the whole sample period

The justification for this *lender_key* classification is that the nonbank lenders, who mostly engage in the originate-to-distribute business, do not hold mortgages for a long period of time. Typically, mortgages originated by nonbanks are either securitized via the GSEs or sold to other lenders within two months after origination. Thus, it is unlikely that the proportion mortgage nonbanks hold on their balance sheet exceeds $2/12 \approx 16.7\%$ at a given time of the year. Using the percentage of the non-securitized mortgages alone will still result in misclassification, as many banks operate in the originate-to-distribute model too. Hence, I use the next three criteria to include the lenders who specialize in originate-to-distribute business but also have other consumer credit products. Nonbank mortgage lenders typically specialize in mortgage lending alone, so it will be very unlikely for these lenders to operate other types of lending alone side mortgage lending business.

Figure A.6 and Figure A.7 show the comparison between the volume of bank mortgages in GCCP and HMDA along the time-series dimension and geographical dimension, respectively.

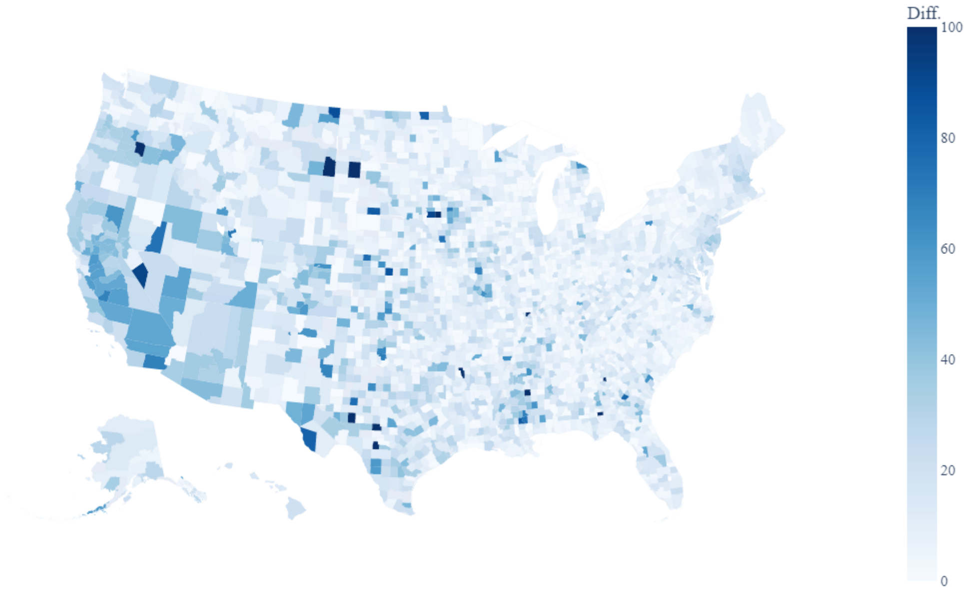
Note that since *lender_key* could identify the mortgage servicers and while HMDA only identifies the originators, some discrepancy is expected even if the bank classification method can perfectly identify the banks in the GCCP sample. On the time-series dimension, the percentage of bank loans classified using the new method match pretty well with the percentage of bank loans in HMDA data. On the geographical dimension, the deviation between GCCP and HMDA at each county is calculated as the absolute value of the percentage difference between GCCP bank percentage and HMDA bank percentage, i.e. $abs(pct(GCCP)_c - pct(HMDA)_c)/pct(HMDA)_c$, where $pct(sample)_c$ equals the dollar amount of bank loans in all sample year divided the dollar amount of all loans in all sample in county c. For example, if a county has 40% bank mortgages according to GCCP and 50% bank mortgages according to HMDA, the deviation would be $|40\% - 50\%|/50\% = 20\%$. Figure A.7 shows that the deviation between the GCCP and HMDA is below 20% for the majority of the counties.

Figure A.6: Share of Bank-Originated Mortgages on Time-Series Dimension



Note: This figure plots the shares of mortgages originated by bank lenders in GCCP and HMDA. The shares are calculated as the percentage of the total dollar amount of origination volume.

Figure A.7: Share of Bank-Originated Mortgages on Geographical Dimension



Note: The figure plots the county-level difference between the percentage of bank-originated mortgages in GCCP and HMDA data. The difference is calculated as the absolute percentage difference between GCCP bank percentage and HMDA bank percentage, i.e. $abs(pct(GCCP)_c - pct(HMDA)_c)/pct(HMDA)_c$.

A.4 PL premium with Alternative Cutoffs

Table A.1 reports the results from Equation (6) using alternative cutoffs of 60% and 80% in the portfolio lender classification. Panel A reports the result using a more lenient 80% cutoff, while Panel B reports the result using a stricter 60% cutoff. The premiums with 80% and 60% cutoffs are 3.56 to 7.74 bps and 4.45 to 18.01 bps, respectively. The magnitude of the estimation exhibits a negative relationship with the cutoff value, which is reasonable as the stricter cutoff likely captures the lenders whose mortgage financing methods are more dominated by portfolio lending.

Table A.1: Loan-Level Evidence of PL premium with Alternative Cutoffs

	Dependent Variable: Interest Rate (bps)					
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A Portfolio Lender Cutoff: <=80% Securitized					
Portfolio Lender	10.04***	7.62***	7.74***	5.72***	5.17***	3.56***
	(0.64)	(0.64)	(0.64)	(0.80)	(1.12)	(1.18)
Observations	507571	506250	498125	492459	491639	461126
R-Squared	0.35	0.37	0.37	0.39	0.44	0.49
	Panel B Portfolio Lender Cutoff: <=60% Securitized					
Portfolio Lender	14.44***	10.90***	10.81***	18.01***	6.83***	4.45***
	(0.63)	(0.62)	(0.62)	(0.89)	(1.54)	(1.69)
Observations	507571	506250	498125	492459	491639	461126
R-Squared	0.35	0.37	0.38	0.39	0.44	0.49
Loan Controls	Y	Y	Y	Y	Y	Y
Credit Risk Controls		Y	Y	Y	Y	Y
Credit Risk High-Orders			Y	Y	Y	Y
Lender Controls				Y	Y	Y
Quarter-County FEs	Y	Y	Y	Y	Y	Y
Lender FEs					Y	
Lender-County FEs						Y

Note: This table reports the results from Equation (6). The observations are at the loan level. Panels A and B report the estimations using the portfolio lender classifications with the 80% and 60% cutoffs, respectively. ”Loan Controls” include the loan amount, and loan purposes (refinance or new purchase). ”Credit Risk Controls” include borrower credit score, borrower debt-to-income ratio, and borrower income. ”Credit Risk High-Orders” include the second, the third, and the fourth power of all controls within ”Credit Risk Controls”. ”lender Controls” include the total amount of loan balance, the total number of credit card accounts, and the total number of auto loan accounts at the lender-quarter level. All standard errors are clustered at the county-quarter level and reported in parentheses.

A.5 Model Derivations

Let q be the quantity chosen by portfolio lender i, q' be the quantity chosen by portfolio lenders $j \neq i$. The maximization for lender i is:

$$
\begin{aligned}
&\max_{q_i} R_i q_i - r_i q_i \Rightarrow \\
&\max_{q_i} (a - bQ_i - b_{ij}Q_j) q_i - r_i q_i \Rightarrow \\
&\max_{q_i} (a - b(N_i - 1)q'_i - q_i - b_{i,j}Q_j) q_i - r_i q_i \Rightarrow \\
&\max_{q_i} (a - b(N-1)q'_i) q_i - bq^2 - b_{i,j}Q_j q_i - r_i q_i
\end{aligned}
\tag{A.1}
$$

The F.O.C. is:

$$
\begin{aligned}
&a - b(N-1)q'_i - 2bq_i - b_{ij}Q_j - r_i = 0 \Rightarrow \\
&a - b_{ij}Q_j - r_i - b(N+1)q_i = 0
\end{aligned}
\tag{A.2}
$$

Each type-i lender takes the total number of lenders of each type as given. By doing so, a type-i lender also takes the optimal aggregate quantity originated by the type-j lenders, Q_j, as given. Imposing a symmetric Nash equilibrium on each type of lender yields the optimal quantity of supply by each type of lender. Solving the F.O.C. yields the optimal quantity for each type of lender. The equilibrium quantities are:

$$
\begin{aligned}
q_1^* &= \frac{a - b_{12}Q_2 - r_1}{(N_1 + 1)b} \\
q_2^* &= \frac{a - b_{21}Q_1 - r_2}{(N_2 + 1)b}
\end{aligned}
\tag{A.3}
$$

where $Q_2 = N_2 q_2^*$ and $Q_1 = N_1 q_1^*$. Thus, we have:

$$
\begin{aligned}
Q_1^* &= \frac{N_1(a(b + bN_2 - b_{12}N_2) + b_{12}N_2 r_2 - b(1 + N_2)r_1)}{-b_{21}b_{12}N_2N_1 + b^2(1 + N_2)(1 + N_1)} \\
Q_2^* &= \frac{N_2(a(b + bN_1 - b_{21}N_1) + b_{21}N_1 r_1 - b(1 + N_1)r_2)}{-b_{21}b_{12}N_2N_1 + b^2(1 + N_2)(1 + N_1)}
\end{aligned}
\tag{A.4}
$$

The equilibrium prices are:

$$
\begin{aligned}
R_1^* &= -\frac{ab(b + bN_2 - b_{12}N_2 + bb_{12}N_2r_2 - b_{21}b_{12}N_2N_1r_1 + b^2(1+N_2)N_1r_1)}{b_{12}b_{21}N_1N_2 - b^2(1+N_1)(1+N_2)} \\
R_2^* &= -\frac{ab(b + bN_1 - b_{21}N_1 + bb_{21}N_1r_1 - b_{21}b_{12}N_2N_1r_2 + b^2(1+N_1)N_rr_2)}{b_{12}b_{21}N_1N_2 - b^2(1+N_1)(1+N_2)}
\end{aligned}
\tag{A.5}
$$

A.6 Testing Corollary 1 with Alternative Cutoffs

Table A.2 reports the result from Equation (8). Panels A, B, and C report the results using the 80%, 70%, and 60% cutoffs for portfolio lending classifications, respectively. The estimated coefficient of $\mathbb{1}^{PTF}_{l(i)} \times HHI_{c(i),q(i)}$ has a significant and positive loading across all specifications. In the baseline estimation in Panel B, one standard deviation increase in market concentration increases the PL premium by 5.47 to 6.54 bps (15.62 to 18.68 $\times$ 0.35). This result is consistent with the prediction in Corollary 1 that the PL premium increases when the overall market concentration increases.

Table A.2: PL premium and Market Concentration

	Dependent Variable: Interest Rate (bps)								
	Panel A: <=80% Securitized Cutoff			Panel B: <=70% Securitized Cutoff			Panel C: <=60% Securitized Cutoff		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Portfolio Lender	2.26*	2.24*	2.00	2.26*	2.24*	2.00	5.23***	4.77***	14.66***
	(1.24)	(1.25)	(1.30)	(1.24)	(1.25)	(1.30)	(1.25)	(1.25)	(1.41)
Portfolio Lender	18.68***	19.26***	15.62***	18.68***	19.26***	15.62***	20.66***	22.28***	15.09***
× HHI	(4.46)	(4.50)	(4.52)	(4.46)	(4.50)	(4.52)	(4.51)	(4.53)	(4.63)
Loan Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Credit Risk Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Credit Risk High-Orders		Y	Y		Y	Y		Y	Y
Lender Controls			Y			Y			Y
Quarter-County FEs	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	464610	457586	452632	464610	457586	452632	464610	457586	452632
R-Squared	0.36	0.36	0.37	0.36	0.36	0.37	0.36	0.36	0.37

Note: This table reports the results from Equation (8), where the measure of market concentration is defined in Equation (7). The observations are at the loan level. Panels A, B, and C report the results using the 80%, 70%, and 60% cutoffs for portfolio lending classifications, respectively. ”Loan Controls” include the loan amount, and loan purposes (refinance or new purchase). ”Credit Risk Controls” include borrower credit score, borrower debt-to-income ratio, and borrower income. ”Credit Risk High-Orders” include the second, the third, and the fourth power of all controls within ”Credit Risk Controls”. ”lender Controls” include the total amount of loan balance, the total number of credit card accounts, and the total number of auto loan accounts at the lender-quarter level. All standard errors are clustered at the county-quarter level and reported in parentheses.

A.7 Testing Corollary 2 with County-Year Level Panel Data

I employ the following fist-stage specification:

$$\Delta ln(Inq)_{c,y} = \sigma_c + \eta_y + \beta \Delta Bartik_{c,y} + \mu' \boldsymbol{X}_{c,y} + \varepsilon_{c,y} \tag{A.6}$$

where $\Delta ln(Inq)_{c,y}$ is the log change in the total number of mortgage inquiries in county c and year y. σ_c and η_y are county and year fixed effects. I also include a vector of county-level controls, $\boldsymbol{X}_{c,y}$, which includes the log changes in the following variables: population, average wage, average credit score, and average DTI.

I estimate the differential response to mortgage demand between portfolio lenders and OTD lending using the following second-stage specification:

$$\Delta Y^{PTF}_{c,y} - \Delta Y^{OTD}_{c,y} = \sigma_c + \eta_y + \beta \widehat{\Delta Inq}_{c,y} + \mu' \boldsymbol{X}_{c,y} + \varepsilon_{c,y} \tag{A.7}$$

The outcome variable $\Delta Y^{PTF}_{c,y} - \Delta Y^{OTD}_{c,y}$ is the difference in the changes of outcome Y between portfolio lenders and OTD lenders. $\widehat{\Delta Inq}_{c,y}$ is the instrumented number of mortgage inquiries. The second-stage estimation includes the same controls and fixed effects as the first-stage estimation. When Y is the average interest rate, the outcome variable is equal to the changes in PL premium.

The response to the mortgage demand shocks of each market segment can be estimated using the following specification:

$$\Delta Y_{c,y} = \sigma_c + \eta_y + \beta \widehat{\Delta Inq}_{c,y} + \mu' \boldsymbol{X}_{c,y} + \varepsilon_{c,y} \tag{A.8}$$

where the outcome variable $\Delta Y_{c,y}$ is the change in the average interest rate or origination volume. The other variables are defined the same as in Equation (A.7).

Table A.3 reports the results on interest rates. Columns (1), (2), and (3) report the results from Equation (A.7) where the outcome variable is the change in the PL premium, which equals the change in the difference between interest rates of portfolio lenders and OTD lenders, i.e. $\Delta R_{c,y}^{PTF} - \Delta R_{c,y}^{OTD}$. Columns (4), (5), and (6) report the results from Equation (A.8) where the outcome variable is the change in the interest rate of the portfolio lenders, while Columns (7), (8), and (9) report the results from Equation (A.8) where the outcome variable is the change in the interest rate of the OTD lenders. The results show that an increase in mortgage demand increases the interest rate for both portfolio lenders and OTD lenders and that the relative difference between the two groups of lenders, the PL premium, widens when demand increases. A ten percentage points increase in mortgage demand increases the interest rates charged by the portfolio lenders and OTD lenders by 10.31 and 4.46 bps, respectively, while the PL premium widens by 6.20 bps.

Table A.4 reports the results on origination volume. Columns (1), (2), and (3) report the results from Equation (A.7) where the outcome variable is the change in the difference between log origination volume of portfolio lenders and OTD lenders, i.e. $\Delta Vol_{c,y}^{PTF} - \Delta Vol_{c,y}^{OTD}$. Here, origination volume is measured by the log of the number of mortgage origination, i.e. $Vol = ln(\#ofOrigination)$. Columns (4), (5), and (6) report the results from Equation (A.8) where the outcome variable is the change in the origination volume of the portfolio lenders, while Columns (7), (8), and (9) report the results from Equation (A.8) where the outcome variable is the change in the origination volume of the OTD lenders. The results show that while the origination volumes for both the portfolio lenders and the OTD lenders increase, the magnitudes of their respective increases do not seem to differ significantly.

Table A.3: Response of PL premium to Demand Shocks

	Dependent Variable: Δ Interest Rate (bps)								
	Panel A: Δ PL premium			Panel B: Δ Avg. Rates of Portfolio Lenders			Panel C: Δ Avg. Rates of OTD Lenders		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\widehat{\Delta \# Inq_{c,y}}$	65.41**	63.94**	61.95**	107.09***	105.55***	104.13***	43.94**	44.31**	44.62**
	(26.77)	(26.64)	(26.41)	(23.02)	(22.84)	(23.01)	(20.22)	(20.07)	(19.84)
Δln(Population)		56.73	33.14		39.65	23.74		-12.21	-8.33
		(55.61)	(56.79)		(27.84)	(29.44)		(32.80)	(33.35)
Δ(Wage)		-0.74	-0.71		-0.52	-0.51		0.22	0.21
		(0.95)	(0.94)		(0.67)	(0.67)		(0.84)	(0.83)
Δ(Credit Score)			-0.58*			-0.35			0.11
			(0.32)			(0.22)			(0.20)
Δln(DTI)			50.46			51.35**			-1.92
			(36.41)			(23.52)			(23.55)
County FEs	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	12633	12623	12623	16399	16386	16386	15831	15819	15819
F-statistics	135.632	136.37	133.782	146.473	147.22	144.363	146.34	146.73	144.07

Note: This table reports the results from Equations (A.7) and (A.8) when the outcome variable is the change in average interest rate. All regressions are weighted by the number of mortgages in each county-year cell. All standard errors are clustered at the county level and reported in parentheses.

Table A.4: Response of Portfolio Lending Volume to Demand Shocks

	Dependent Variable: Δ Ln(# of Origination)								
	Panel A: $\Delta\frac{Ln(\#byPortfolioLenders)}{Ln(\#byOTDLenders)}$			Panel B: $\Delta Ln(\#byPortfolioLenders)$			Panel C: $\Delta Ln(\#byOTDLenders$		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\widehat{\Delta\#Inq_{c,y}}$	0.28*	0.29*	0.28*	1.91***	1.90***	1.89***	1.65***	1.63***	1.63***
	(0.16)	(0.15)	(0.15)	(0.15)	(0.15)	(0.14)	(0.16)	(0.16)	(0.16)
Δln(Population)		-0.75***	-0.81***		-0.45***	-0.51***		0.28**	0.28**
		(0.18)	(0.19)		(0.13)	(0.12)		(0.12)	(0.13)
Δ(Wage)		0.00	0.00		-0.01***	-0.01***		-0.01***	-0.01***
		(0.01)	(0.01)		(0.00)	(0.00)		(0.00)	(0.00)
Δ(Credit Score)			0.00			-0.00**			0.00
			(0.00)			(0.00)			(0.00)
Δln(DTI)			0.16			-0.34***			-0.48***
			(0.17)			(0.12)			(0.11)
County FEs	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	13604	13592	13592	17551	17537	17537	16692	16679	16679
F-statistics	139.754	140.591	137.844	151.889	152.57	149.599	148.418	148.849	146.164

Note: Note: This table reports the results from Equations (A.7) and (A.8) when the outcome variable is the change in the log of the total number of mortgage origination. All regressions are weighted by the number of mortgages in each county-year cell. All standard errors are clustered at the county level and reported in parentheses.

A.7.1 Test for correlates

I test whether income, debt-to-income ratio, and age can predict changes in the outcome variables by running the following regression:

$$\Delta Y_{c,y} = \sigma_c + \eta_y + \beta_1 \Delta Inc_{c,y-1} + \beta_2 \Delta DTI_{c,y-1} + \beta_3 \Delta age_{c,y-1} + \varepsilon_{c,y} \tag{A.9}$$

where the outcome variable, $\Delta Y_{c,y}$, is the change in the average interest rate and log mortgage origination amount of all lenders in county c and year y. $\Delta Inc_{c,y}$, $\Delta DTI_{c,y}$, and $\Delta age_{c,y}$. σ_c and η_y are county and year fixed effects, respectively. Table A.5 shows the results from Equation (A.9). The results show that the confounding factors do impact the two types of lenders separately. This is expected since these factors are presumably correlated with borrowers' credit riskiness, thereby should impact mortgage origination outcomes. However, these factors do not pose threats to the identification assumption as long as they do not impact the two types of lenders differentially.

A.8 A Quasi-Natural Experiment

Up to this point, the analysis in this paper has maintained that the difference in the cross-elasticity of demand between lenders is driven by the difference in the branch network. An alternative theory is that the financing method could drive the difference by itself. For example, financing methods could have a signaling effect. In the wake of the 2008 subprime mortgage crisis, the irresponsibility of many lenders to provide mortgage loans to consumers with no ability to pay back the loans has received much attention from the public. It is possible that consumers may view the act of financing mortgages on the balance sheet as a sign of responsible lending, as holding the loans on lenders' own balance sheets gives the lenders a stake in the performance of the loans. As a result, the borrowers may feel more confident that the lenders' screening is effective such that conditioning on loan approval, they are likely to be able to pay back the loans.

I use the passing of the Dodd-Frank Act as a setting for a quasi-natural experiment to test the

Table A.5: Robustness Test for the Effects of Potential Confounding Factors

	Portfolio Lenders Mortgages				OTD Lenders Mortgages			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: Dependent Variable: ΔInterest rate							
Δ Age	0.02			-0.30	-0.03			-0.24
	(0.29)			(0.30)	(0.22)			(0.23)
Δ Income		0.11		0.08		0.14**		0.13**
		(0.09)		(0.09)		(0.06)		(0.06)
Δ DTI			-1.33***	-1.32***			-0.53**	-0.55**
			(0.28)	(0.30)			(0.23)	(0.24)
Number of Obs	13950	14067	14061	13945	12930	13031	13029	12928
R-Squared	0.144	0.143	0.146	0.147	0.168	0.169	0.169	0.169
	Panel B: Dependent Variable: Δln(Vol)							
Δ Age	0.08			0.21***	0.20***			0.35***
	(0.06)			(0.06)	(0.07)			(0.07)
Δ Income		-0.20***		-0.25***		-0.23***		-0.27***
		(0.02)		(0.02)		(0.02)		(0.02)
Δ DTI			-0.37***	-0.45***			-0.35***	-0.37***
			(0.06)	(0.06)			(0.07)	(0.07)
Number of Obs	14899	15056	15047	14889	13566	13692	13689	13563
R-Squared	0.064	0.081	0.067	0.095	0.059	0.083	0.064	0.09
County FEs	Y	Y	Y	Y	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y	Y	Y	Y	Y

Note: This table reports the results from Equation (A.9). The observations are at the county-year level. Panel A reports the results when the outcome variable the change in average interest rate at lender-county level. Panel B reports the results when the outcome variable the change in log of dollar amount of originated mortgage at lender-county level.All regressions are weighted by the number of mortgages in each county-year cell. Columns (1) to (3) use the observations that are classified as portfolio lenders by the baseline classification. Columns (4) to (6) use the observations that are classified as OTD lenders by the baseline classification. All standard errors are clustered at the county level and reported in parentheses.

effect of portfolio lending on the mortgage interest rate. I use a difference-in-difference specification to test my hypothesis. The treatment group is the bank lenders, while the control group is the nonbank lender. The rationale is that bank lenders are more likely to securitize mortgages due to capital limit regulation imposed by the Dodd-Frank Act. Since Dodd-Frank Act is unlikely to impact the lenders' branch network in a few-year window. On the other hand, the effect caused by higher capital requirements has a much higher immediate impact. Thus, the differential effect of the Dodd-Frank Act estimated is likely driven mostly by method-specific factors, instead of lender-specific factors.

Figure A.8 shows the time series trend of the percentage of balance-sheet-financed mortgages originated by the bank lenders from 2009 to 2012. Right before Dodd-Frank Act take into effect, about 30-35% percent of the mortgages originated by bank lenders are financed on the balance sheets. After the Act took effect on July 21st, 2010, the percentage dropped to 26-28%. Thus, the Dodd-Frank Act can be regarded as a quasi-exogenous shock to the bank lender's ability to finance mortgages on their balance sheets. After the shock, we should expect the interest rate gap between banks and nonbanks to drop after the shock.

Formally, I use the following specification:

$$\begin{aligned} R_i = \sigma_{c(i),q(i)} + \quad & \beta_1 Bank_{l(i)} + \beta_2 Bank_{l(i)} \times PostDobbFrank_{q(i)} + \\ & \eta' \mathbf{X}_i + \gamma' \mathbf{X}_{j(i),q(i)} + \lambda' \mathbf{X}_{l(i),q(i)} + \varepsilon_i \end{aligned} \tag{A.10}$$

$Bank_{l(i)}$ is a dummy variable that equals to one if lender l is a bank lender. $PostDobbFrank_{q(i)}$ is a dummy variable that equals one if the quarter of origination of mortgage i takes place after the third quarter of 2010 when the Dobb-Frank Act takes effect. The coefficient of interest is β_2, which captures the marginal effect of the Dobb-Frank Act on the interest rate of bank lenders compared to nonbank lenders. The fixed effects and controls are all defined the same as in Equation (6).

Table A.6 reports the results from Equation (A.10). Across all specifications and sample periods, the effect of the Dodd-Frank Act on bank interest rates is 24.2 to 27.1 bps higher than that

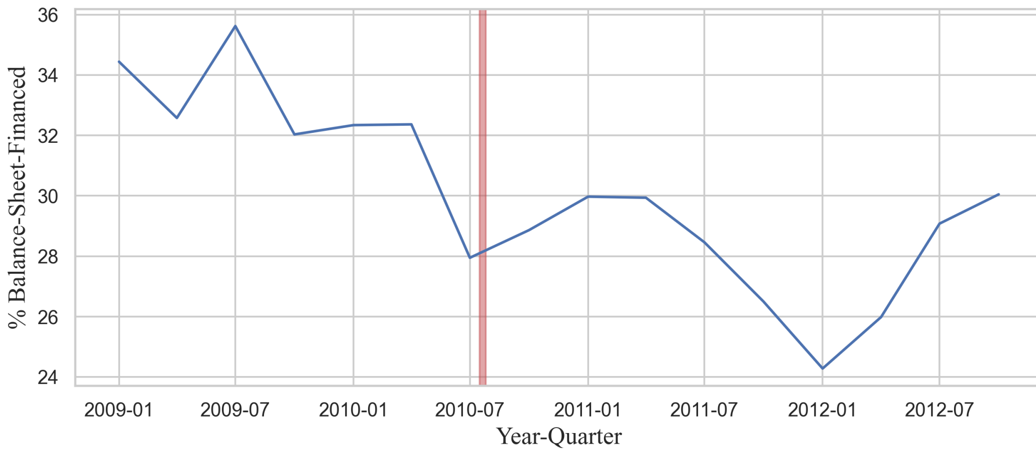
Table A.6: Quasi-Natural Experiment for the PL premium

	Dependent Variable: Interest Rate (bps)					
	(1)	(2)	(3)	(4)	(5)	(6)
Bank	2.25	0.87	1.02	4.63**	3.36	3.71
	(1.98)	(1.96)	(1.96)	(2.30)	(2.28)	(2.29)
Bank × PostDoddFrank	-27.75***	-27.05***	-26.70***	-24.20***	-24.04***	-23.73***
	(2.18)	(2.15)	(2.15)	(2.62)	(2.60)	(2.60)
Quarter-County FEs	Y	Y	Y	Y	Y	Y
Lender Controls	Y	Y	Y	Y	Y	Y
Loan Controls	Y	Y	Y	Y	Y	Y
Credit Risk Control		Y	Y		Y	Y
Credit Risk High-Order			Y			Y
Observations	281631	281007	277457	189799	189422	187229
R-Squared	0.29	0.31	0.31	0.21	0.23	0.23
Sample	2008-2013	2008-2013	2008-2013	2009-2012	2009-2012	2009-2012

Note: This table shows the results from Equation (A.10). ”Loan Controls” include the loan amount, and loan purposes (refinance or new purchase). ”Credit Risk Controls” include borrower credit score, borrower debt-to-income ratio, and borrower income. ”Credit Risk High-Orders” include the second, the third, and the fourth power of all controls within ”Credit Risk Controls”. ”lender Controls” include the total amount of loan balance, the total number of credit card accounts, and the total number of auto loan accounts at the lender-quarter level. All standard errors are clustered at the county-quarter level and reported in parentheses.

on the nonbank interest rate. One possible threat to the identification of this specification is that Dodd-Frank Act could impact the banks and nonbanks through channels other than the capital requirements. For example, the Act also regulates securitization standards, which could potentially raise the cost of securitization for bank lenders more than for nonbank lenders. However, this hypothesis is inconsistent with the drop in securitization rate for the bank lenders as shown in Figure A.8. In summary, the results of the quasi-natural experiment show that method-specific factors can at least contribute to part of the PL premium. However, in this paper, I remain conservative in claiming the causal relationship between the financing method and the portfolio lender premium.

Figure A.8: Drop in Portfolio Lending After Dodd-Frank Act



Note: This figure shows the time series of the percentage of mortgages that are financed on balance sheets by the bank lenders between 2009 and 2012. The vertical line indicates July 21st, 2010, the date on which Dodd-Frank Act came into effect.