The Portfolio Lending Premium in the Mortgage Market
*

Peter Han †

March 15, 2023

Abstract

This paper studies the effect of financing strategies on mortgage interest rates in the local markets using a novel loan-level dataset. I find that the lenders that finance a considerable amount of mortgage via portfolio lending charge a positive spread for mortgages of similar ex-ante risk profiles. I refer to this spread as the portfolio lending premium and estimate it to be about 9.12 to 12.77 bps. I show that this premium can be justified by a simple model where two types of lenders engage in a Cournot competition in a partially-segmented market. Furthermore, I find that the portfolio lending premium is 5.47 bps higher per one standard deviation increase in local market concentration and is 6.20 bps higher per ten percentage points increase in the demand for mortgages in the local market. These two results provide evidence that the two necessary conditions to generate a positive premium hold in the data.

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

---

*I am grateful to George Pennacchi, Charles Kahn, Jialan Wang, Julia Fonseca, Yufeng Wu, Joshua Pollet, Tatyana Deryugina, Qiping Xu, Heitor Almeida, Jaewon Choi, Don Fullerton, and participants at the Gies College of Business brownbag for valuable suggestions.

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

I Introduction

About sixty percent of the conforming mortgage loans originated between 2007 to 2017 in the United States are securitized. While the other forty percent, around seven trillion dollars in face value, are unsecuritized and directly financed on the balance sheet of the lenders, even though these loans meet the securitization standards posted by the Federal Housing Finance Agency (FHFA).

In this paper, I study the following question: How do mortgage financing strategies affect mortgage pricing? The methods of mortgage financing do not only affect the cost of origination (Fuster, Goodman, Lucca, Madar, Molloy and Willen (2013),Buchak, Matvos, Piskorski and Seru (2018b)) but also change the incentive of the lenders (Pennacchi (1988), Keys, Mukherjee, Seru and Vig (2010), Purnanandam (2011), Rajan, Seru and Vig (2015)), thus making the answer to this question theoretically ambiguous. On the one hand, the effective cost to finance mortgages via portfolio lending could be lower than securitization if the lender has access to cheap funding sources such as retail deposits. A lower cost of lending thus gives the portfolio lenders more leeway to undercut competitors. On the other hand, the portfolio lenders have more skin in the game since they internalize the default risk of the loans held on their balance sheets. More skin in the game may increase the incentive to bargain for a higher price to compensate for the risk. Such bargaining power may come in various forms, such as higher service quality[1] or expanding the economy of the scope of lending services[2].

To motivate my empirical analysis, I introduce a stylized model where two types of lenders with different funding costs engage in a Cournot competition in a partially-segmented market. I show that an equilibrium where the portfolio lenders charge a positive interest rate spread, henceforth referred to as the portfolio lending premium, than originate-to-distribute lenders (OTD lenders) can exist when (1) the lenders have market power in the local market, and (2) the portfolio lenders have a lower cross-elasticity of demand than the OTD lenders. In such an equilibrium, the

[1]For example, Citibank offers lender-paid assistance that reduces closing costs by up to $7500.

[2]The bank lenders typically offer a wide range of financial services. One incentive to create such a ”one-stop service” is to increase customer stickiness. In comparison, specialized mortgage lenders typically focus excluding on mortgage origination.

portfolio lending premium is higher when the market is more concentrated and when the demand for mortgages is higher.

I leverage a comprehensive dataset provided by Gies Consumer Credit Panel (GCCP) in my empirical analysis. GCCP contains yearly snapshots of about one million 30-year conforming mortgages originated between 2004 to 2017. Each snapshot contains information such as origination dates, mortgage terms (including interest rates), securitization status, borrower credit characteristics (including credit score, income, and debt-to-income ratio), and borrower demographic. The granularity of the data allows me to measure the size of the portfolio lending premium as well as its interaction with market concentration at the loan level. Additionally, GCCP also contains data on mortgage inquiries, which enables me to create a proxy for time-varying mortgage demand at the county level.

I begin my empirical analysis by showing that the local mortgage markets are consistent with an equilibrium with a positive portfolio lending premium. I estimate that the premium to be 9.12 to 12.77 bps during my sample period from 2007 to 2017. I choose a binary classification of lenders based on financing strategies: A lender is defined as a portfolio lender in a given quarter if the average securitization ratio of this lender in the past 12 quarters is lower than 70%. The choice of the binary classification is motivated by the polarization in the lenders' financing strategies observed from the data. More lenders choose to either securitize more than 80% or less than 20% of the mortgages than lenders that choose a mixed strategy across the sample years. My estimation of the portfolio lending premium is robust to a rich set of lender-level controls to account for the factors that could impact a lender's financing strategies, as well as different cut-off ratios used for lender classification.

Furthermore, I show two pieces of empirical evidence that are consistent with the two predictions about the equilibrium with a positive portfolio lending 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 the average interest rate. This result is of particular importance for policy-making as it suggests that a change in market concentration may change the relative balance of power between different segments in local markets.

Second, using a Bartik-type instrument for mortgage demand, I estimate that portfolio lending premium increases by 10.95 bps per one ten percentage points increase in the demand for mortgages in the local markets. 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. This result, along with the result on market concentration, provides additional support for the mechanism proposed in my theoretical framework.

The research question raised in this paper is highly related to but conceptually different from the literature on the comparison between bank and nonbank lenders in light of the raise of nonbank lending in the past decade. A rich body of research works has investigated the differences in areas such as funding cost (Buchak, Matvos, Piskorski and Seru (2018a)), regulatory burden (For example, Demyanyk and Loutskina (2016), Buchak et al. (2018b), Pennacchi (2019)), flexibility in the adoption of new technologies (For example, Gertler, Kiyotaki and Prestipino (2016), Fuster, Plosser, Schnabl and Vickery (2019), Bartlett, Morse, Stanton and Wallace (2022)), and etc. In my paper, I focus on the differences in financing strategies that affect bank and nonbank lenders at the same time, albeit to different degrees. 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)). Additionally, the difference in funding cost faced by portfolio lenders and OTD lenders cannot explain a positive portfolio lending premium, as the funding cost of portfolio lenders is typically lower than that of the OTD lenders.

This paper also contributes to the literature on the relevance of market concentration in the mortgage market. In particular, the current consensus in the literature is that the mortgage market is by and large a national market because of the existence of a secondary market. Fuster et al. (2013) argues that in a market with a large number of fringe firms, concentration should not lead to higher pricing. Hurst, Keys, Seru and Vavra (2016) finds that the interest of GSE-insured mortgage mortgages does not vary according to market concentration. Amel, Anenberg and Jorgensen (2018) finds that changes in MBS yields do not affect mortgage pricing differently in locations with different levels of concentration. 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. My paper contributes to this literature by providing a new angle from which market concentration affects mortgage pricing, which shed light on the competition within the mortgage market and provides empirical support for policy-making for the mortgage market.

This paper is also related to the literature on change 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 et al. (2010) finds that the subprime mortgages that are easier to be securitized are 10-25% more likely to default. While the previous literature focuses more on the moral hazard of securitization, my paper discusses the relationship between financing strategies and the lenders' product differentiation strategies, which result in lenders of different financing strategies facing different demand curves[3].

The rest of the paper is organized as follows: Section II lays out the institutional background and presents the theoretical framework that motivates my empirical analysis; Section III describes

[3]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.

the dataset; Section IV presents evidence of the portfolio lending premium; Section V discusses how portfolio lending is affected by local market concentration; Section VI discusses how the portfolio lending premium responds to local shocks to mortgage demand; Section VII concludes.

II Background and Conceptual 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 deposit-taking intuitions, such as commercial banks and thrift banks. In the recent twenty years, non-deposit-taking institutions have gained an increasingly larger share of the market in terms of origination volume. The GSEs are quasi-governmental entities that are established to provide liquidity to the housing market. There are two GSEs in the conforming mortgage market: Fannie Mae and Freddie Mac[4]. 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-

[4] 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 model. In this model, the lender sells the mortgages it originates to the GSEs or other purchasers of balance-sheet-financed mortgage loans. 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.

II.2 Theoritical Framework

In this section, I present a model of mortgage market competition to motivate my empirical analysis. The model is stylized and is only used as a conceptual framework to guide my following empirical analysis. The details of the model are as follows:

A local market (a county) has both portfolio lenders and OTD lenders. 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. In actuality, most bank lenders engage in portfolio lending and originate-to-distribute business at the same time, while the nonbank lenders still need to hold a small proportion of the mortgages on their balance sheet due to the time delay in securitizing the loans. Without a loss of generality, I make the simplification that all portfolio lenders participate in the portfolio lending business and all OTD lenders participate in the originate-to-distribute business. The funding cost of the portfolio lenders, r_1, equals the weighted average of debt and equity of the lenders and the funding cost of the OTD lenders, r_2, would be equal to the yield of the MBS. Though it is mostly the depository institutions that engage in portfolio lending, some depository institutions operate exclusively in the originate-to-distribute model. For example, Pennacchi (2019) shows that tax consideration can make it more profitable for a bank lender to securitize its loans. These lenders are essentially the same as nonbank lenders in the context of this model.

The market is partially segmented in the sense that the mortgages originated by portfolio lenders and OTD lenders are partial substitutes. The intuition is that the two types of lenders provide differentiated services due to the differences in their comparative advantages. For example, the lenders that finance mortgages on their balance sheets could potentially internalize more of the cost of the mortgage and as a result, are more diligent in the screening process. The borrowers perceive the more diligence in mortgage screening as a signal that the lenders are lending responsively. As a result, the borrowers could derive utility from the fact that their mortgages are being financed on the lenders' balance sheet.

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. In practice, the lenders' product differentiation strategies and the demand functions are endogenous. For simplicity, I assume that the lenders take their respective demand functions as given. 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 $j = 1$. The lenders earn the spread between the mortgage interest rate and its funding cost, $R_i - r_i$. Assume that a_i, b, and $b_{i,j}$ are positive constants. The term $b_{i,j}$ governs the cross-elasticity of demand faced by lenders of type i from lenders of type j. When $b_{ij} = b$, the mortgages offered by lenders of type i are perfect substitutes for the mortgages offered by lenders of type j. If $b_{ij} = 0$, the market becomes perfectly segmented. When $b_{ij} = b_{ji}$, the two types of lenders face the same level of cross-elasticity of demand. Based on the intuition mentioned earlier, I assume the two types of mortgages to be partial substitutes, i.e. $b > b_{ij}$. Also assume that $a > r_1$ and $a > r_2$, such that the equilibrium interest rates would be positive.

The lenders engage in a Cournot competition. Each lender of type i solves a maximization problem by choosing the optimal quantity of mortgage, q_i, to supply to the market:

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

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. Imposing a symmetric Nash equilibrium on each type of lender yields the optimal quantity of supply by each type of lender:

$$q_i^* = \frac{a - b_{ij} Q_j - r_i}{(N_i + 1) b} \tag{3}$$

where $q_i^* = N_i Q_i$. (3) is a set of two linear equation, from which Q_1 and Q_2 can be solved:

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

Plug (4) into (1) solves the equilibrium price for the two types of lenders:

$$R_i = \frac{a - b_{ij} Q_j + N_i r_i}{N_i + 1} \tag{5}$$

Upon solving the model, the following propositions can be made:

Proposition 1. *There exists $\hat{N} > 1$ and $\Delta > 0$, such that when $N < \hat{N}$ and $b_{12} - b_{21} < \Delta$, the portfolio lending premium, defined as the spread between the equilibrium interest rate charged by the portfolio lenders and the OTD lenders, is positive, i.e. $S := R_1 - R_2 > 0$. When $N \to \infty$ or $b_{12} - b_{21} \to 0$, the S approaches asymptotically to the difference between the funding cost, $r_1 - r_2$.*

When the supply of the portfolio lenders has a larger externality towards the OTD lenders, the portfolio lenders can charge a higher interest rate even when their funding cost is lower than that of the OTD lenders. The portfolio lenders, which internalize the risk and return of the loans more than the OTD lenders do, have a stronger incentive to bargain for a better price. The portfolio lenders can increase their bargaining power by providing better service to their borrowers and making their mortgage service less substitutable. An increase in service quality can be achieved in different ways, such as by improving loan processing services or integrating mortgage services into the other financial products the lenders offer. On the other hand, OTD lenders, whose profit is less dependent on the cash flow from the loans but rather the origination fee, are more likely to find it advantageous to pursue a quantity-over-quality strategy.

Corollary 1. *When the portfolio lending premium is positive, the portfolio lending premium becomes smaller when OTD lenders become more concentrated, i.e.* $\partial S/\partial N < 0$*, where* $N < \hat{N}$*.*

The intuition behind this proposition is that when market concentration increases, the type of lenders that receive a lower cross-elasticity of demand are able to leverage the increase in pricing power more effectively that the other type of lender.

Corollary 2. *When the portfolio lending premium is positive, the portfolio lending premium increases when there is a positive shock in the demand for mortgages in the local market, i.e.* $\partial S/\partial a > 0$ *and* $\partial Q_1/\partial a < \partial Q_2/\partial a$*.*

This proposition captures the intuition that mortgage interest rate only varies to demand shocks when the lenders internalize the change in equilibrium interest rate. If the mortgages offered by one type of lender are less substitutable than those offered by the OTD lenders, this type of lender can afford to raise their interest rate higher in response to demand shocks.

All the derivations and proofs are provided in Appendix A.1.

II.3 Numerical Example

I use a numerical example to illustrate the intuitions from the model. Panel A of Table 1 exhibits the parameter values of the example, while Panel B shows the solutions. A portfolio lending premium of 52.31 bps is generated from this example, even though the funding rate of the portfolio lenders is lower than that of the OTD lenders. Figure 1 graphically illustrates the comparative statistics in Proposition 1 and Corollaries 1 and 2. The Figure plots the relationship between the portfolio lending premium when one parameter changes while the other parameters take the same values as in Panel A of Table 1. Figure 1a shows that when the portfolio lending premium is positive when the externality from the OTD lenders is small enough relative to the externality imposed by the portfolio lenders, i.e. when b_{12} is smaller enough. Figure 1b shows that portfolio lending premium is higher when the market is less concentrated, i.e. when N increases. Figure 1c shows that the portfolio lending premium is higher when the demand functions for both types of mortgages make a parallel shift to the right, i.e. when a increases.

III Description of Data Sets

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 ID 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[5].

Besides mortgage data, GCCP also contains a data set of credit inquiry data. I use the credit inquiry for all mortgage loans[6]. 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 unique consumer ID.

Panel A of Table 2 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[7]. 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[8]. The average interest rate of the mortgages in the final sample is 5.50 percent. Appendix A.2 provides more details on the interest rate estimation and its validation. I identify the originator of a mortgage by its first

[5]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

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

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

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

owner recorded in the sample. Whether a mortgage is used for refinancing or new purchases is not directly observable from the data. Additionally, 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.

What is critical to the purpose of this paper is the GSE identifier. I classify all mortgages that are not insured by either Fannie Mae or Freddie Mac as balance-sheet-financed mortgages. I classify a mortgage as one financed through securitization if its GSE identifier identifies either Fannie Mae or Freddie Mac in the first snapshot date[9]. If a mortgage is not associated with a GSE, I classify the mortgage as one financed via portfolio lending. In the final sample, 57% of the mortgages are securitized through the two GSEs. Figure 2 shows that the balance-sheet-financed mortgages have higher interest rates compared to securitized mortgages throughout the sample years.

Panel B of Table 2 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 2 reports the summary statistics at the lender level by collapsing the mortgage level data by the lender key associated with each loan. There is a total of 4,488 lenders present in the sample, of which about 96% are bank lenders. Whether a lender is a bank lender is not directly observable in the GCCP data. I provide a detailed description of the methodology to identify bank lenders from the data in Appendix A.4. 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.

[9]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.

IV Evidence of Portfolio Lending Premium

This Section presents empirical evidence for Proposition 1. I construct a time-varying binary classification of portfolio lenders based on the observation of polarization in the financing strategies 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 measurement of financing strategies is critical to study the effect of financing strategies on mortgage pricing. Financing strategy affects both the cost of lending and the incentives of a lender at the institutional level. As a result, the financing strategies should also be measured at the lender level. Furthermore, 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 strategies amongst the lenders. Figure 3 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 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 4 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. Figure 5 shows the time series of the characteristics of the top 100 lenders. Prior to 2008, most of the portfolio lenders are large lenders that also have considerable lending in not only mortgage but also other credit product types. After 2008, the composition of portfolio lenders shifted towards smaller lenders that have smaller overall lending volume in all credit product types and operate in smaller geographic coverage.

IV.2 Main Results on Portfolio Lending Premium

In this section, I present evidence of a higher interest rate charged by portfolio lenders compared to OTD lenders at the aggregate level. I begin by presenting a series of graphical evidence. Figure 6 shows that the portfolio lenders charge a higher interest rate compared to the OTD lenders and that this difference is not explained simply by the differences between banks and nonbanks, as the premium still persists amongst the subsample of bank-originated mortgages.

I formally test the Proposition 1 by estimating the following loan-level specification:

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

i, j, l, c, and q are indices for loan, borrower, lender, county, and quarter, respectively, for all the specifications throughout this paper. 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. $\mathbf{X}_{l(i),q(i)}$ is the vector of lender controls, which include the number of operating states, number of operating counties, existing accounts for different credit product types, as well as the total amount of outstanding loan balance.

Importantly, I control for the set of lender controls that captures the time-varying heterogeneity between the bank and nonbank lenders and between banks that operate in different business models. In other words, 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. One concern of this result is the lack of control over the changes in the funding cost of the lenders. However, this concern is unlikely to impact the result for two reasons: First, the estimated economic magnitude is large enough that is unlikely to be explained by funding cost alone; Second, the cost to fund mortgages on the lender's balance sheet is typically lower, thus the difference in the effective funding cost between portfolio lenders and OTD lenders is likely to decrease the portfolio lending premium.

As lender IDs are anonymous in the GCCP dataset, I am not able to directly link lender IDs with Call Reports. I circumvent this problem by creating time-varying lender controls at the quarter level using the loan-level data of other major credit products in the GCCP dataset. More specifically, I calculate the total number of mortgage accounts, credit card accounts, and auto loan accounts. 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.

Table 3 reports the results from Equation (6). The estimated coefficient of the portfolio lender dummy captures the magnitude of the portfolio lending 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 Columns (1) to (4), the portfolio lending 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 portfolio lending premium is around 9.12 to 12.77 bps. Panel B reports the results using the sample of mortgages originated by bank lenders[10]. The estimations of the portfolio lending premium remain to be significant and the estimated magnitudes, if anything, become even larger than the ones in the pooled sample.

Table 4 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 strategies are more dominated by portfolio lending.

[10]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.4 provides more details on my method to identify bank lenders from the data.

IV.3 Quasi-Natural Experiment

I use the passing of the Dodd-Frank Act as a setting for a quasi-natural experiment to test the 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. Figure 7 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:

$$\begin{aligned} R_i = \sigma_{c(i),q(i)} + \quad & \beta_1 Bank_{l(i)} + \beta_2 Bank_{l(i)} \times PostDobbFrank_{q(i)} + \\ & \eta' \mathbf{X}_i + \gamma' \mathbf{X}_{j(i),q(i)} + \lambda' \mathbf{X}_{l(i),q(i)} + \varepsilon_i \end{aligned} \tag{7}$$

$Bank_{l(i)}$ 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 5 reports the results from Equation (7). 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 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 7. Overall, the results of the quasi-natural experiment provide strong additional support for the effect of the financing strategies on mortgage pricing.

IV.4 Discussion on the Portfolio Lending Premium

Proposition 1 predicts that under certain conditions, an equilibrium in which the portfolio lenders charge a higher interest rate despite a lower funding cost can exist. The results in Section IV.2 confirm that such an equilibrium is present in the local mortgage markets during the sample period. The higher interest rate charged by the portfolio lenders might process considerable pricing power in the local markets. Notably, the fact that there is a significant premium charged by the portfolio lenders amongst the bank lenders demonstrates that the findings in this Section cannot be explained by the regulatory and institutional differences between the banks and nonbanks alone.

The results from the quasi-natural experiment in Section IV.3 provide additional support for the relevance of the financing method in mortgage pricing at origination. Even so, I want to take a conservative stance and remain cautious about making a causal claim in this paper via the analysis in this Section alone. To provide further support for the effect of financing strategies on mortgage pricing, I focus on analyzing whether empirical evidence supports the mechanism proposed in Section II.2. In the mechanism proposed in Proposition 1, the keys to generating a positive portfolio lending premium are two necessary conditions: (1) lenders face a different demand curve when changing financing strategies, and (2) the market is not perfectly competitive. In the next two Sections, I go on to investigate the two testable predictions made in Corollary 1 and 2, which, if true, suggest that the two necessary conditions in Proposition 1 likely hold.

The results in this Section can be viewed along with the findings in Buchak et al. (2018b), where the authors find a positive Fintech premium, and the findings in Fuster et al. (2019), where the authors find a small but negative Fintech premium. Whereas Fintech lenders have high overlap with the OTD lenders defined in my paper, there are also many bank lenders who operate in the originate-to-distribute model[11]. Thus, the existence of a positive portfolio lending premium is not inconsistent with the previous findings, but rather explains a new dimension of heterogeneity amongst mortgage lenders.

V Portfolio Lending Premium and Market Concentration

One of the necessary conditions to generate a positive portfolio lending premium in Proposition 1 is that the lenders process market power. Corollary 1 predicts that the magnitude of the portfolio lending premium should also have a positive relationship with lender market power. The existence of such a positive relationship indicates that the necessary condition must hold. This Section provides empirical evidence that supports this relationship.

V.1 Measure of Market Concentration

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:

$$HHI_{c,q} = \sum_{\tau=1}^{12} \sum_{l} \left(\frac{Vol_{l,c,q-\tau}}{Vol_{c,q-\tau}} \right)^2 / 12 \tag{8}$$

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 $q-\tau$.

[11] I take as given the classification that all Fintech firms are nonbanks following the analysis in Fuster et al. (2019), where the authors find no major traditional deposit-taking lenders meet their criteria for Fintech firms through 2010 to 2016.

Figure 8 presents the geographic distribution of the baseline HHI. Consistent with the distribution of market concentrated in a number of other papers (Stanton, Walden, Wallace 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 north-eastern states and California. Figure 9 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.

V.2 Main Results on Market Concentration

I present the empirical analysis of the relationship between market concentration and the portfolio lending premium in this section. Figure 10 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 portfolio lending premium.

$$R_i = \sigma_{c(i),q(i)} + \beta_1 \mathbb{1}^{PTF}_{l(i),q(i)} + \beta_2 \mathbb{1}^{PTF}_{l(i),i(q)} \times HHI_{c(i),q(i)} + \eta' \mathbf{X}_i + \gamma' \mathbf{X}_{j(i),q(i)} + \lambda' \mathbf{X}_{l(i),q(i)} + \varepsilon_i \quad (9)$$

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 portfolio lending 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[12]. 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 variations at the local markets. 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 (9) 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.

Table 6 reports the result from Equation (9). 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 portfolio lending 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 portfolio lending premium increases when the overall market concentration increases.

These results are consistent with Corollary 1 and suggest that the mortgage market concentration has an important impact on the relative pricing power between portfolio lenders and OTD lenders. These results are not in conflict with previous findings that show little relationship between mortgage interest rates and market concentration. Rather, my findings complement the existing literature by highlighting the relevance of competition with and across segments on mortgage pricing. These competition dynamics has important policy implication. At the moment, regulators do not consider the effect of market concentration when evaluating the impact of lender mergers in the

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

local markets, even though market concentration is an important consideration for the evaluation of concentration in the deposit market (Buchak and Jørring (2021)). 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 Portfolio Lending Premium Respond to Demand Shocks

Besides the presence of lender market power, the second necessary condition for a positive portfolio lending premium in Proposition 1 is a lower cross-elasticity of demand faced by the portfolio lenders. Corollary 2 further shows that when the portfolio lenders face a lower cross-elasticity of demand, the portfolio lending premium has a positive relationship with the increase in the demand for mortgages in a local market. While it is challenging to directly measure the level of cross-elasticity of demand from the data, proofs for the relationship predicted in Corollary 2 may indicate that the necessary condition most likely holds. This Section provides support for the predictions made in Corollary 2 by employing a two-stage least squares specification with a Bartik-type instrument for shifts in mortgage demand.

VI.1 Instrumenting Demand Shocks

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:

$$\Delta Bartik_{c,y} = \sum_{b=1}^{11} w_{b,c,3} \Delta ln(Inq)_{b,\underline{c},y} \tag{10}$$

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)).

VI.2 Main Results on Response to Demand Shock

I employ the following fist-stage specification:

$$\Delta ln(Inq)_{c,y} = \sigma_c + \eta_y + \beta \Delta Bartik_{c,y} + \mu' \boldsymbol{X}_{c,y} + \varepsilon_{c,y} \tag{11}$$

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:

$$\Delta Y^{PTF}_{c,y} - \Delta Y^{OTD}_{c,y} = \sigma_c + \eta_y + \beta \widehat{\Delta Inq}_{c,y} + \mu' \boldsymbol{X}_{c,y} + \varepsilon_{c,y} \tag{12}$$

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. $\widehat{\Delta App}_{c,y}$ 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 portfolio lending premium.

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

$$\Delta Y_{c,y} = \sigma_c + \eta_y + \beta \widehat{\Delta Inq}_{c,y} + \mu' \boldsymbol{X}_{c,y} + \varepsilon_{c,y} \tag{13}$$

where the outcome variable $DeltaY_{c,y}$ is the change in the average interest rate or origination volume. The other variables are defined the same as in Equation (12).

Table 7 reports the results on interest rates. Columns (1), (2), and (3) report the results from Equation (12) where the outcome variable is the change in the portfolio lending premium, which equals the change in the difference between interest rates of portfolio lenders and OTD lenders, i.e. $\Delta R^{PTF}_{c,y} - \Delta R^{OTD}_{c,y}$. Columns (4), (5), and (6) report the results from Equation (13) 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 (13) 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 portfolio lending 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 portfolio lending premium widens by 6.20 bps.

Table 8 reports the results on origination volume. Columns (1), (2), and (3) report the results from Equation (12) 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 (13) 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 (13) 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.

The fact that increases in the instrumented mortgage demand increase both 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. The results are highly consistent with the predictions in Corollary 2. Intuitively, if the mortgages offered by the portfolio lenders are less substitutable than those offered by the OTD lenders, the portfolio lenders can afford to raise their interest rate higher in response to demand shock without losing market share. The results shown in this Section, together with the results in V, provide strong support that the two necessary conditions that generate a positive portfolio lending premium hold in the local markets.

VI.3 Identification Discussions

One concern in the interpretation of the estimates in Equation (13) is that credit demand shocks could change credit demand through channels other than the number of mortgage applicants. To examine whether the change in borrower riskiness violates the exclusion restriction, I run Equation 11 with the outcome variable being the change in the average borrower characteristics of the approved mortgages. I test the change in three characteristics: credit score, income, and DTI. Table 9 displays the result. While the estimates in Columns (1) to (3) confirm that the higher credit demand is indeed associated with a riskier borrower group, the estimates in Columns (4) to (6) show that the Bartik instrument does not impact the change in the riskiness of borrowers from the portfolio and OTD lenders differently. In an additional robustness check[13], I add the controls for the change in income, credit score, and DTI of mortgage applicants in both the first-stage and second-stage specifications to find that the results are robust to these additional controls.

Besides credit riskiness, it is also possible that the Bartik instrument can affect the equilibrium interest rate through changes in the demographic of loan applicants. 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. I run Equation (11) using the change in the demographics of different subgroups of the borrower population. Table 10 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. Overall, my robustness check does not find strong evidence of a violation of exclusion restriction in my two-stage least-square specification.

VII Conclusion

This paper studies the effect of financing strategies on mortgage pricing. To motivate my empirical analysis, I show that an equilibrium where portfolio lenders charge a higher interest rate than originate-to-distribute lenders can exist using a simple model of a partially segmented market. In my empirical analysis, I find an economically significant positive spread between the mortgages originated by the portfolio lenders and originate-to-distribute lenders does exist. I refer to this

[13] Available upon request.

premium as the portfolio lending premium and estimate its magnitude to be 9.12 to 12.77 bps. Furthermore, I show that the portfolio lending premium is 5.47 to 6.54 bps higher per one standard deviation increase in local market concentration and that I show that the portfolio lending premium is 6.20 bps higher per ten percentage points increase in mortgage demand. These two additional results provide evidence that the necessary conditions that generate a positive portfolio lending premium hold in the data. Overall, the findings in this paper highlight that in addition to the level of funding cost, the source of funding also has a first-order impact on mortgage interest rates.

References

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.

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.

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.

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.

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.

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.

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

Rajan, Uday, Amit Seru, and Vikrant Vig, "The failure of models that predict failure: Distance, incentives, and defaults," *Journal of financial economics*, 2015, *115* (2), 237–260.

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

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: Numerical Example for the Conceptual Framework

Panel A: Parameter Values

a	b	b_{12}	b_{21}	r_1	r_2	N_1	N_2
10	0.03	0.01	0.02	0.02	0.05	50	50

Panel B: Model Solution

	Q_1	Q_2	P_1	P_2	R_1	R_2	$R_1 - R_2$
	279.60	142.42	47.84	12.41	0.19	0.14	0.05

Table 2: 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 3: Loan-Level Evidence of Portfolio Lending 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-Square	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-Square	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). 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 4: Loan-Level Evidence of Portfolio Lending 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-Square	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-Square	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). 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.

Table 5: Quasi-Natural Experiment for the Portfolio Lending 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-Square	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 (7). ”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 6: Portfolio Lending 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-Square	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 (9), where the measure of market concentration is defined in Equation (8). 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.

Table 7: Response of Portfolio Lending Premium to Demand Shocks

	Dependent Variable: Δ Interest Rate (bps)								
	Panel A: Δ Portfolio Lending 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 (12) and (13) 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 8: 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 (12) and (13) 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.

Table 9: Robustness Test for Exclusion Restriction Violation through Credit Risk

	Outcome Variable: ΔY			Outcome Variable $\Delta Y^{PTF} - \Delta Y^{OTD}$		
	Y: Credit	Y: Income	Y: DTI	Y: Credit	Y: Income	Y: DTI
	(1)	(2)	(3)	(4)	(5)	(6)
Δ # of inquiries	-59.983***	-25.507***	3.369*	-24.915*	-19.264	0.642
	(7.443)	(7.138)	(1.791)	(14.456)	(15.963)	(3.479)
Demographic Controls	Y	Y	Y	Y	Y	Y
County FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
# of Obs	20533	20529	20477	20477	12413	6312
R Square	0.227	0.193	0.199	0.199	0.046	0.057

Note: 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.

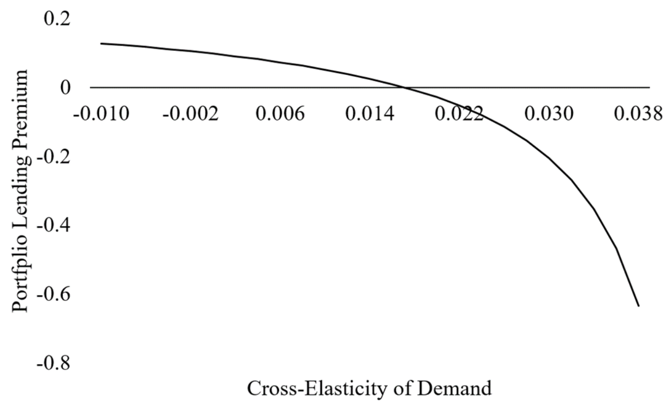
Table 10: Robustness Test for Exclusion Restriction Violation through Demographics

Variable	Coef.	SE
Δ % female of all approved mort.	-0.085	(0.121)
Δ % female of all approved portfolio lenders mort.	-0.128	(0.244)
Δ % female of all approved OTD lenders mort.	-0.025	(0.143)
Δ % collage of all approved mort.	0.052	(0.046)
Δ % collage of all approved portfolio lenders mort.	0.018	(0.080)
Δ % collage of all approved OTD lenders mort.	0.047	(0.054)
Δ # of adults in hosuehold of all approved mort.	-0.704*	(0.361)
Δ # of adults in hosuehold of all approved portfolio lenders mort.	-0.193	(0.642)
Δ # of adults in hosuehold of all approved OTD lenders mort.	-0.971**	(0.472)
Δ # of children in hosuehold of all approved mort.	-0.060	(0.264)
Δ # of children in hosuehold of all approved portfolio lenders mort.	0.293	(0.461)
Δ # of children in hosuehold of all approved OTD lenders mort.	0.031	(0.356)
Δ % homeowner of all approved mort.	-0.197	(0.127)
Δ % homeowner of all approved portfolio lenders mort.	-0.476*	(0.263)
Δ % homeowner of all approved OTD lenders mort.	-0.064	(0.156)
Δ diff. between portfolio lenders and OTD lenders % of female	-0.203	(0.294)
Δ diff. between portfolio lenders and OTD lenders % of college education	-0.059	(0.088)
Δ diff. between portfolio lenders and OTD lenders # of adults in household	0.761	(0.857)
Δ diff. between portfolio lenders and OTD lenders # of children in household	0.805	(0.593)
Δ diff. between portfolio lenders and OTD lenders % of homeowner in household	-0.399	(0.310)
Δ # of adults in loan applicants	0.106	(0.116)
Δ # of children in loan applicants	0.284***	(0.083)
Δ % of homeowners in loan applicants	0.098**	(0.041)
Δ % of female in loan applicants	0.000	(0.037)
Δ % of colleage and above in loan applicants	0.036	(0.039)
Δ average age in loan applicants	-0.517	(1.071)

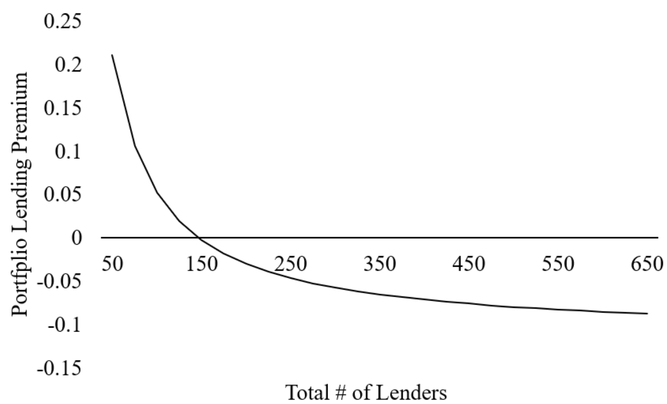
Note: This table reports the results from Equation (11) with different outcome variables. The outcome variables are reported in the ”Variable” Column. All of the regressions reported in this table have the same set of fixed effects (including county fixed effect, year fixed effect) and controls (including 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: Numerical Example

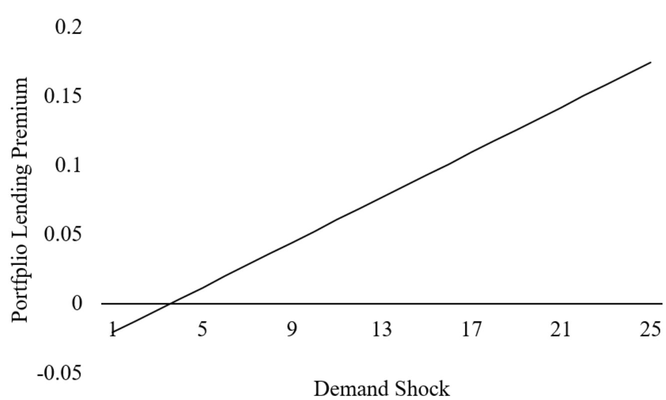
(a) Relationship between portfolio lending premium and $b_{1,2}$



(b) Relationship between portfolio lending premium and the total number of lenders



(c) Relationship between portfolio lending premium and demand shock

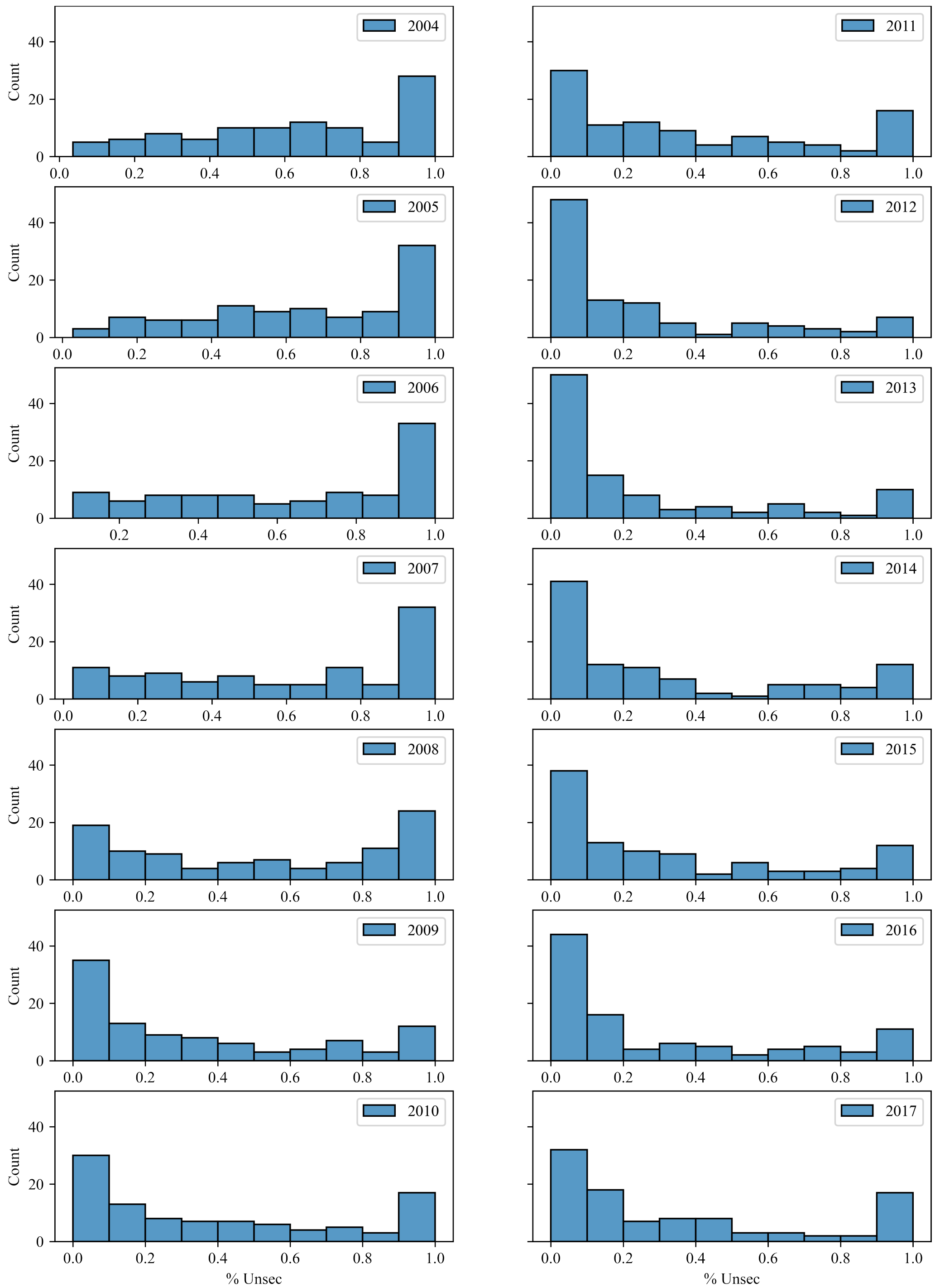


Note: This figure illustrates the relationship between the portfolio lending premium and the three parameters in the model discussed in Section II.2. 1a shows the relationship between the portfolio lending premium and b_{12}. 1b shows the relationship between the portfolio lending premium and N, where an increase in N means increases in N_1 and N_2 of the same proportion. 1c shows the relationship between the portfolio lending premium and a.

Figure 2: 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 2 shows the time series of the national average interest rates.

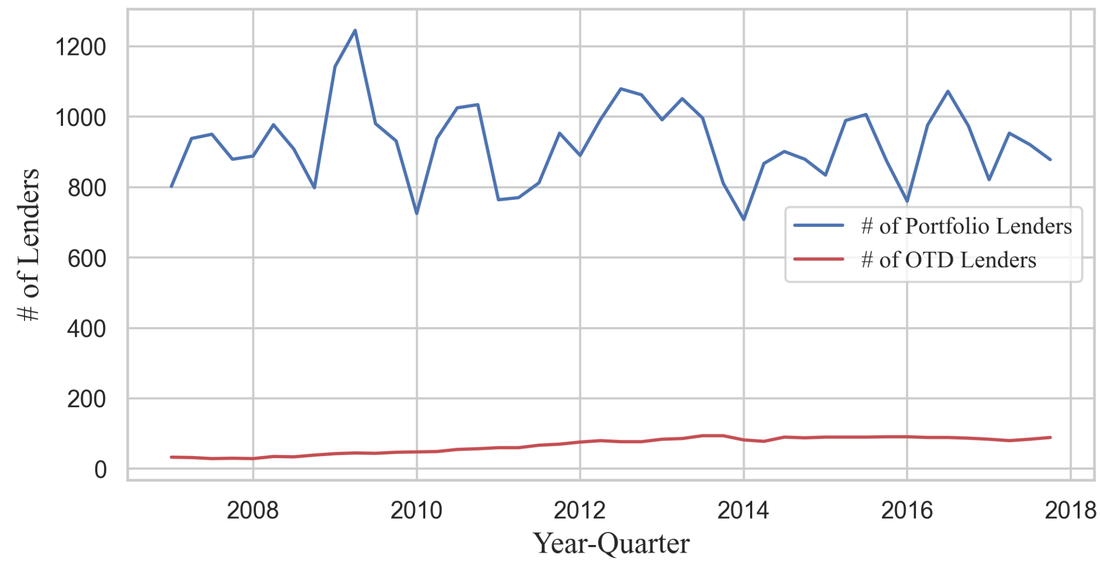
Figure 3: Lender Distribution by % of Balance-Sheet-Financed Mortgages



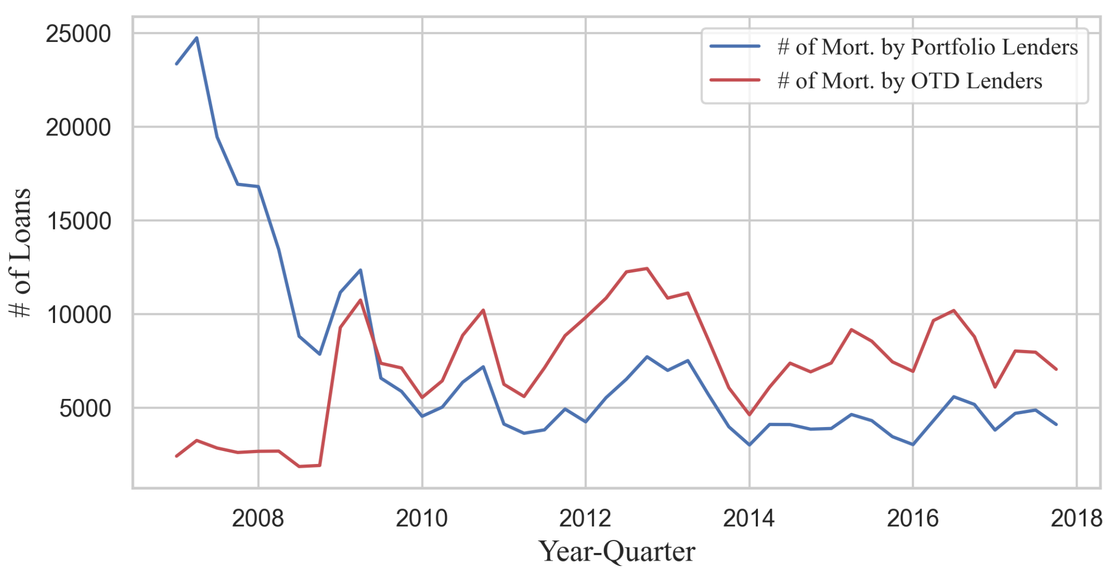
Note: This figure plots the distribution of the top 100 lenders by the percentage of balance-sheet-financed mortgages. The top 100 lenders are selected based on the total number of mortgages the lenders originated from 2004 to 2017. The percentage of balance-sheet-financed mortgages for each of the top 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 4: Number of Lenders by financing Strategies

(a) Number of Lenders by Financing Strategies



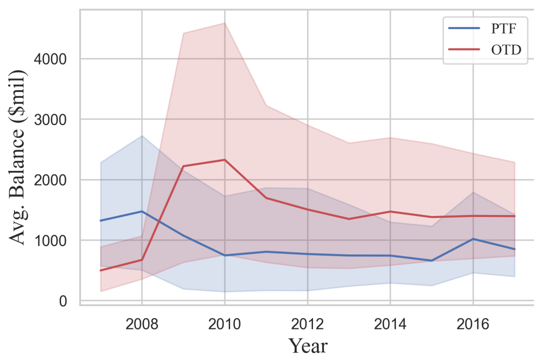
(b) Number of Originated Mortgages by Lender Types



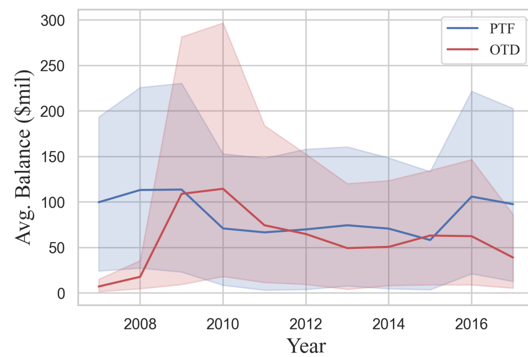
Note: This figure shows the fraction of lenders that are classified as portfolio lenders as well as the fraction of mortgages that are originated by portfolio lenders. Panel 4a shows the number of portfolio lenders and the number of OTD lenders from 2007 to 2017. The classification method is the baseline classification described in Section IV.1. Panel 4b shows the number of mortgages originated by portfolio lenders and the number of mortgages originated by OTD lenders from 2007 to 2017.

Figure 5: Lender Characteristics by Financing Strategies

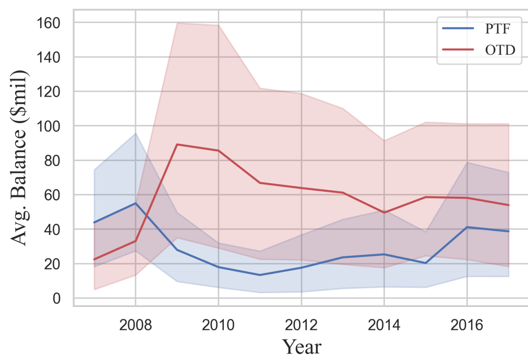
(a) Avg. Mortgage Balance



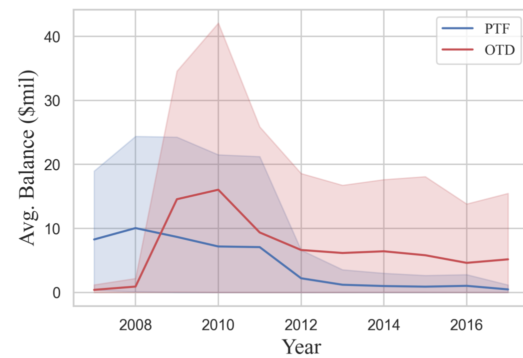
(b) Avg. Credit Card Balance



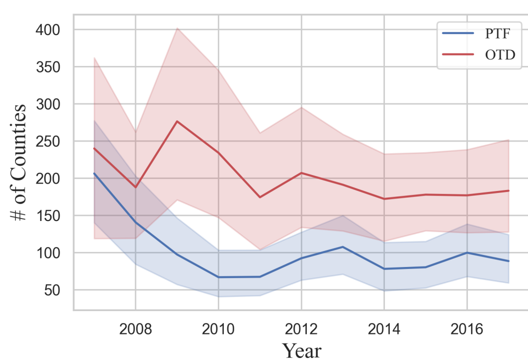
(c) Avg. Auto Loan Balance



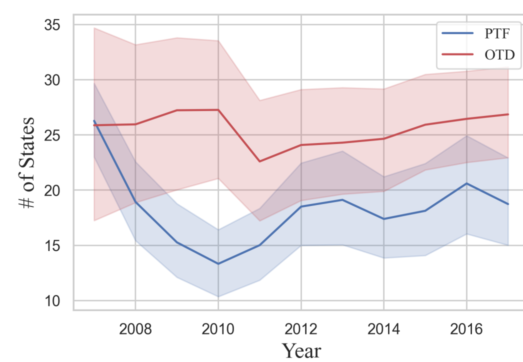
(d) Avg. Student Loan Balance



(e) Avg. # of Operating Counties



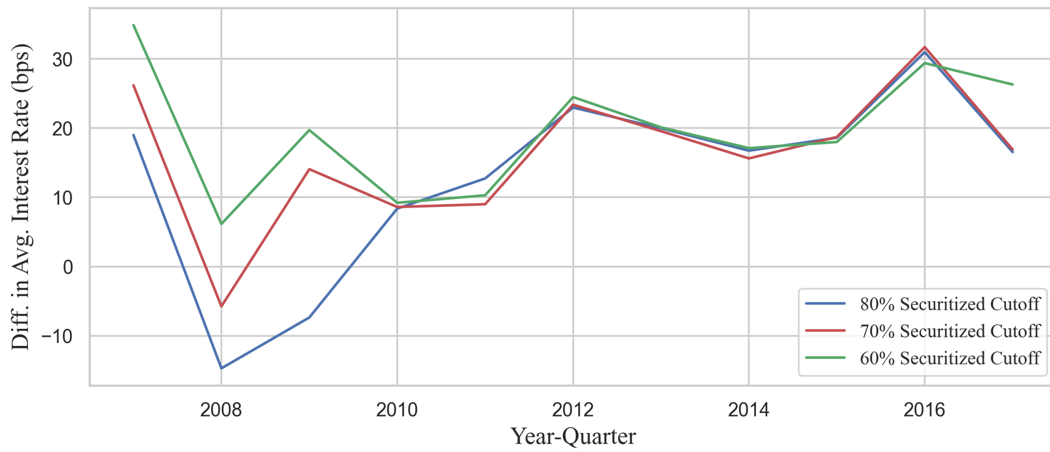
(f) Avg. # of Operating States



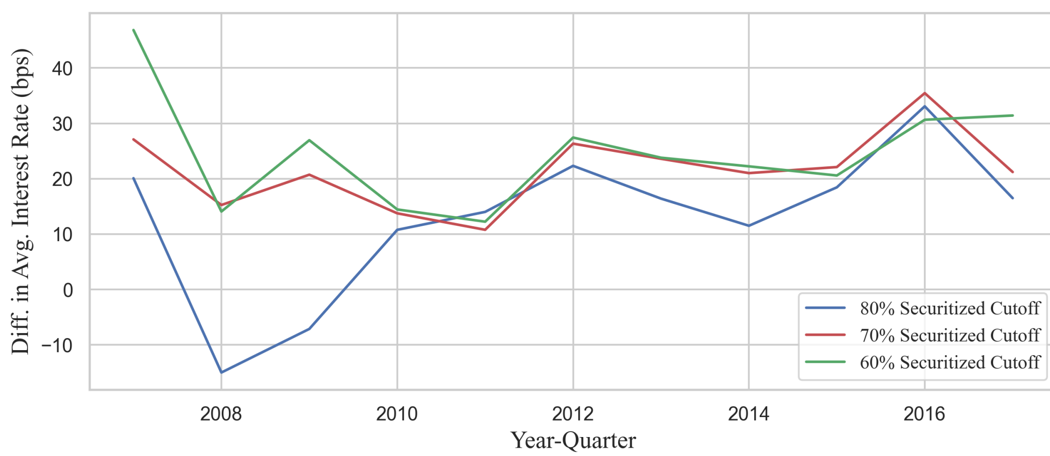
Note: This figure shows the time series of the characteristics of the top 100 lenders. The top 100 lenders each year are selected based on the total number of mortgages originated during the year. The lenders are classified into portfolio lenders and OTD lenders using the baseline classification. The blue lines plot the time series of portfolio lenders and the red lines plot the time series of the OTD lenders. The shaded areas plot the one standard deviation of a given characteristic.

Figure 6: Avg. Portfolio Lenders Interest Rate minus Avg. OTD Lenders Interest Rate

(a) All Mortgages

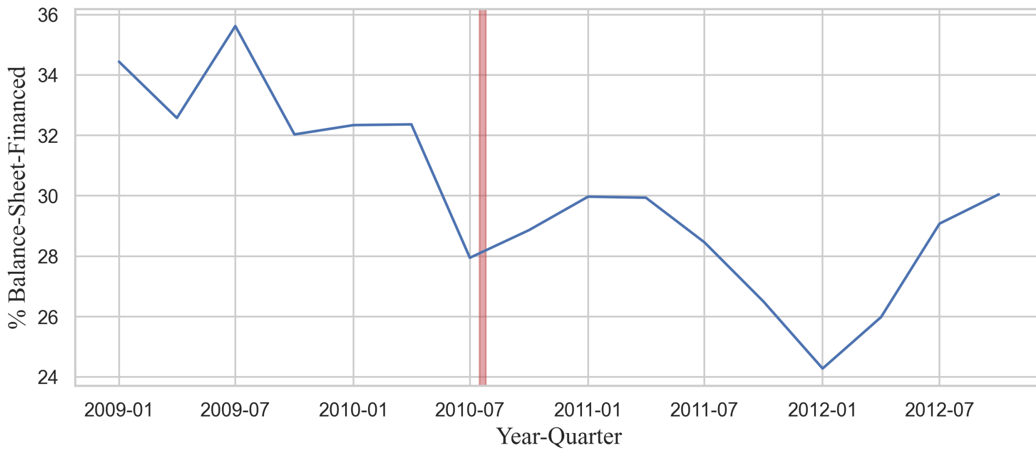


(b) Bank Mortgages



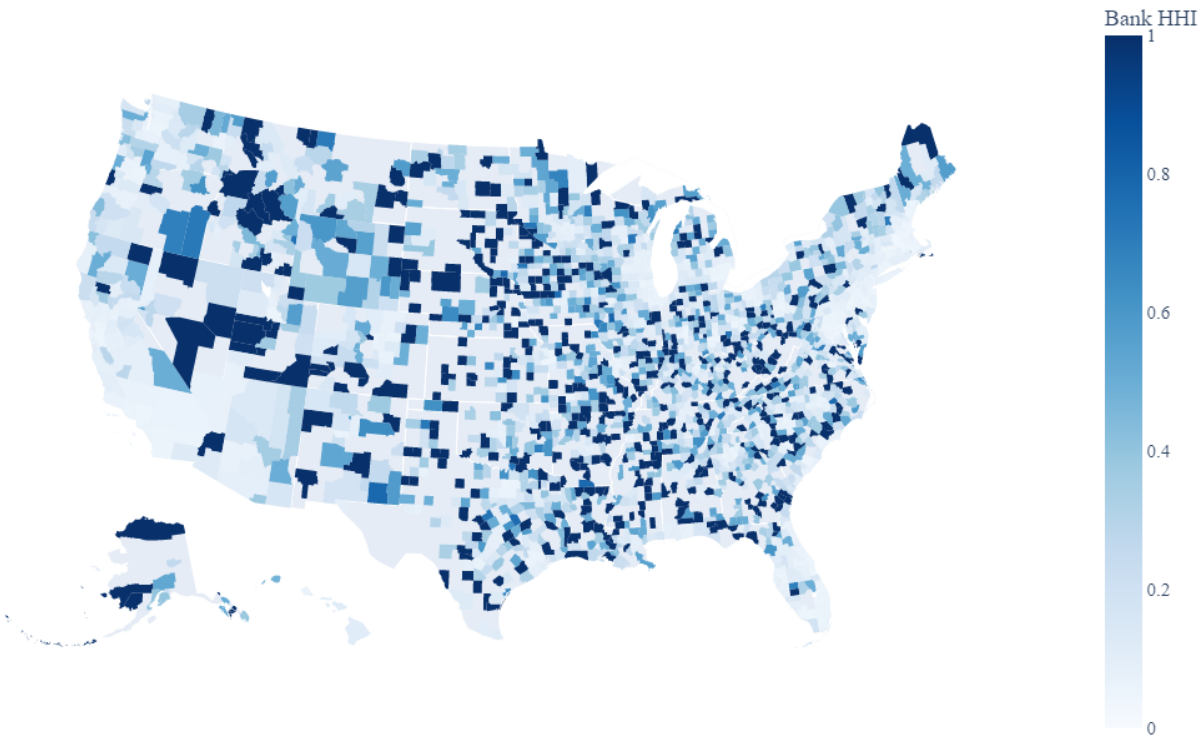
Note: This figure shows the gap between the average interest rate of mortgages originated by portfolio lenders and the average interest rate of mortgages originated by the OTD lenders, i.e. the portfolio lending premium. The blue, red, and green lines plot the premium calculated using the 80%, 70%, and 60% cutoffs, respectively. Panel 6a shows the difference in average interest rate in the full sample that includes both bank-originated and nonbank-originated mortgages. Panel 6b 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 7: 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.

Figure 8: Geographic Distribution of Market Concentration

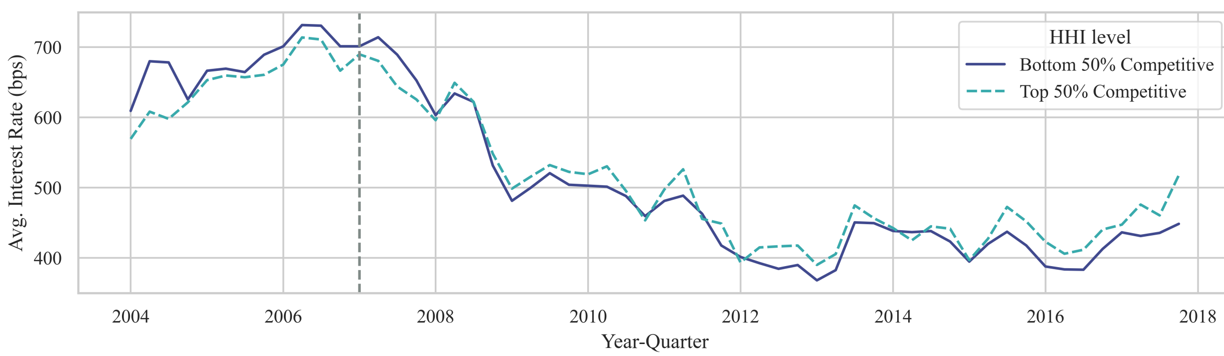


Note: This figure shows the geographic distribution of the county-level average of the baseline HHI as defined in Equation (8) in the year 2015.

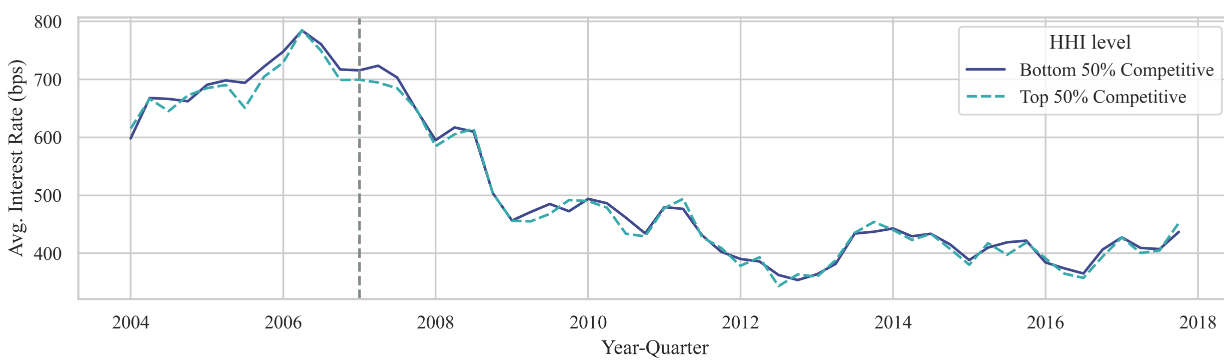
Figure 9: Parallel Trends of Avg. Interest Rate by HHI levels

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

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

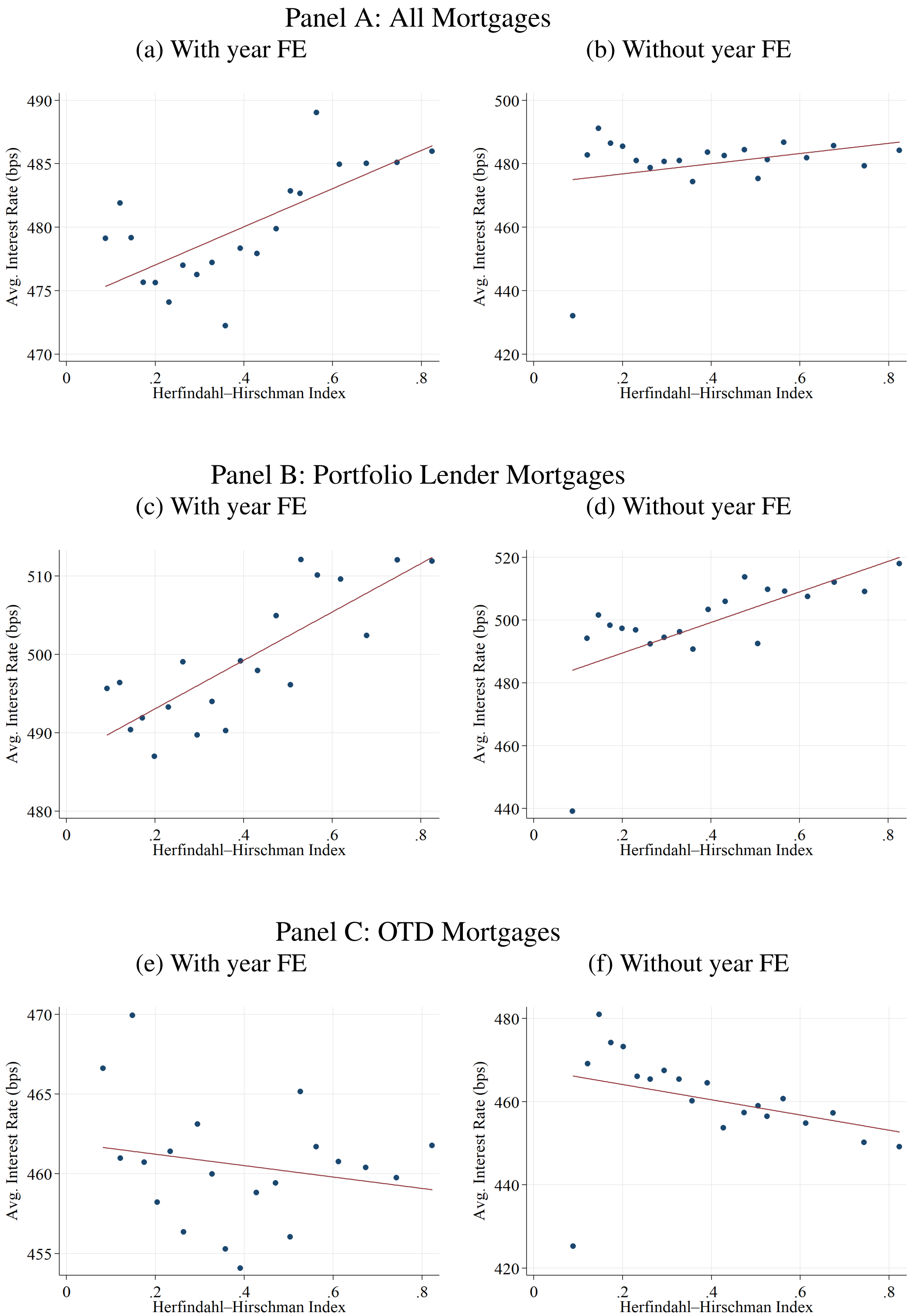


(c) Avg. Interest Rate of Nonbank Mortgages by HHI levels



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 (8).

Figure 10: Binscatter plots of average interest rate against market concentration



Note: This figure shows the binscatter plots of county-level average interest rate against county-level Herfindahl–Hirschman Index. Figures 10a, 10c, and 10e are demeaned by yearly average interest rate. Figures 10b, 10d, and 10f are not demeaned.

A Appendix

A.1 Model Derivation and Proofs

A.1.1 Model Derivation

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:

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

The F.O.C. is:

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

Solving the F.O.C. yields the optimal quantity for each type of lenders. Denote the expressions in terms of u and s, 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:

$$
\begin{aligned}
Q_1^* &= \frac{N_1(a(b + bN_2 - b_{12}N_2) + b_{12}N_2r_2 - b(1+N_2)r_1)}{-b_{21}b_{12}N_2N_1 + b^2(1+N_2)(1+N_1)} \\
Q_2^* &= \frac{N_2(a(b + bN_1 - b_{21}N_1) + b_{21}N_1r_1 - b(1+N_1)r_2)}{-b_{21}b_{12}N_2N_1 + b^2(1+N_2)(1+N_1)}
\end{aligned}
\tag{A.4}
$$

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.1.2 Proofs

Proof of Proposition 1

Proof. The spread between the interest rate of mortgages originated by portfolio lenders and OTD lenders, S, is:

$$
\begin{aligned}
S = R_1^* - R_2^* = \quad & \frac{ab(b_{12}N_2 - b_{21}N_1 + b(-N_2 + N_1))+}{b_{21}b_{12}N_2N_1 - b^2(1+N_2)(1+N_1)} \\
& \frac{b_{21}b_{12}N_2N_1(-r_2 + r_1)+}{b_{21}b_{12}N_2N_1 - b^2(1+N_2)(1+N_1)} \\
& \frac{b(-b_{12}N_2r_2 + b_{21}N_1r_1)+}{b_{21}b_{12}N_2N_1 - b^2(1+N_2)(1+N_1)} \\
& \frac{b^2(-N_1r_1 + N_2(r_2 + N_1r_2 - N_1r_1))}{b_{21}b_{12}N_2N_1 - b^2(1+N_2)(1+N_1)}
\end{aligned}
\tag{A.6}
$$

when $S = 0$

$$
\hat{b_{12}} = \frac{b(abN_2 - abN_1 + ab_{21}N_1 - bN_2N_1r_2 + bN_2N_1r_1 - bN_2r_2 + bN_1r_1 - b_{21}N_1r_1)}{N_2(ab - br_2 - b_{21}N_1r_2 + b_{21}N_1r_1)}
\tag{A.7}
$$

□

A.2 Validation of Interest Rate

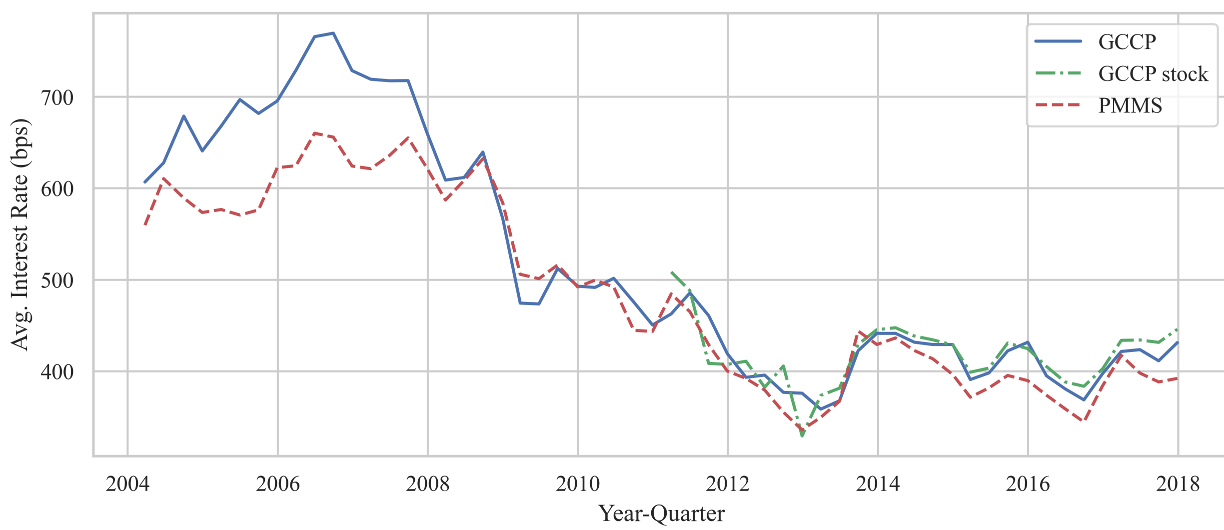
I calculate mortgage interest rates using a root-solving algorithm. Figure A.1 shows the comparison between the quarterly average interest from GCCP and the national average interest rate of the 30-year fixed rate conforming mortgage obtained from Freddie Mac's Primary Mortgage Market Survey (PMMS). 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 $abs(R -$

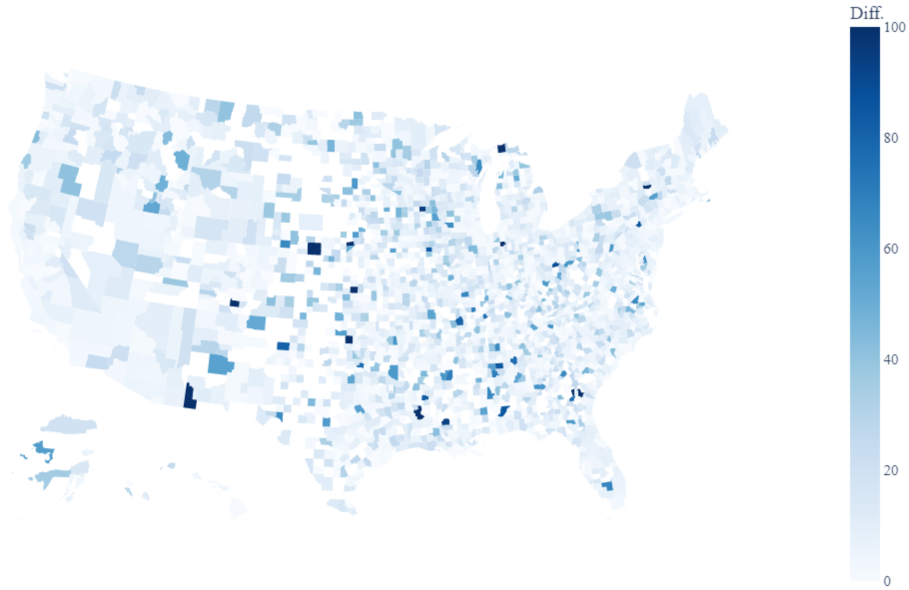
$R_{stock})/R_{stock})$. For example, if a county has an average interest rate of 450 bps according to GCCP and an average interest rate of 500 bps according to HMDA, the deviation would be $|450\% - 500\%|/500\% = 10\%$. The figure shows that the deviation between the GCCP and HMDA is below 10% for 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.3 Discussion about 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[14]. 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[15]. 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.

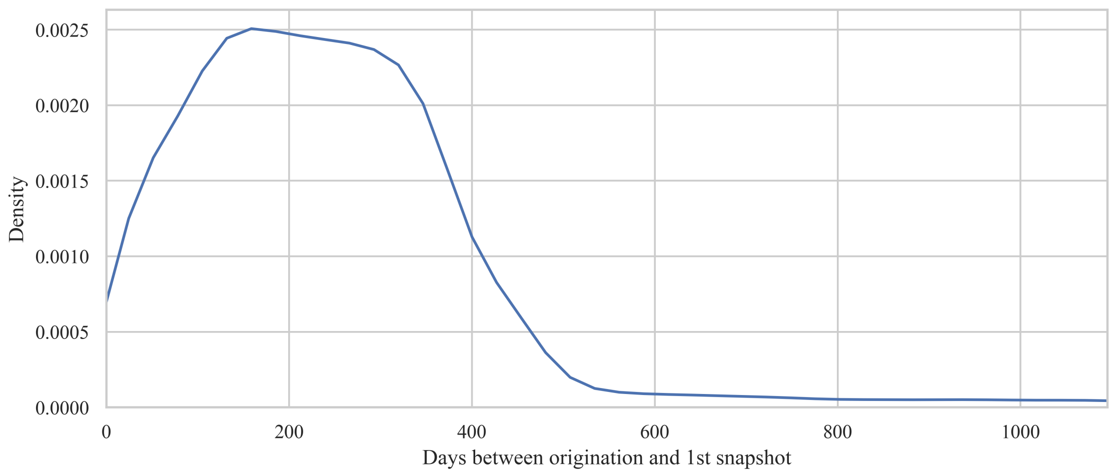
[14] Stanton et al. (2014) has a detailed discussion about the market structure of wholesale lending.

[15] 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[16]. 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.

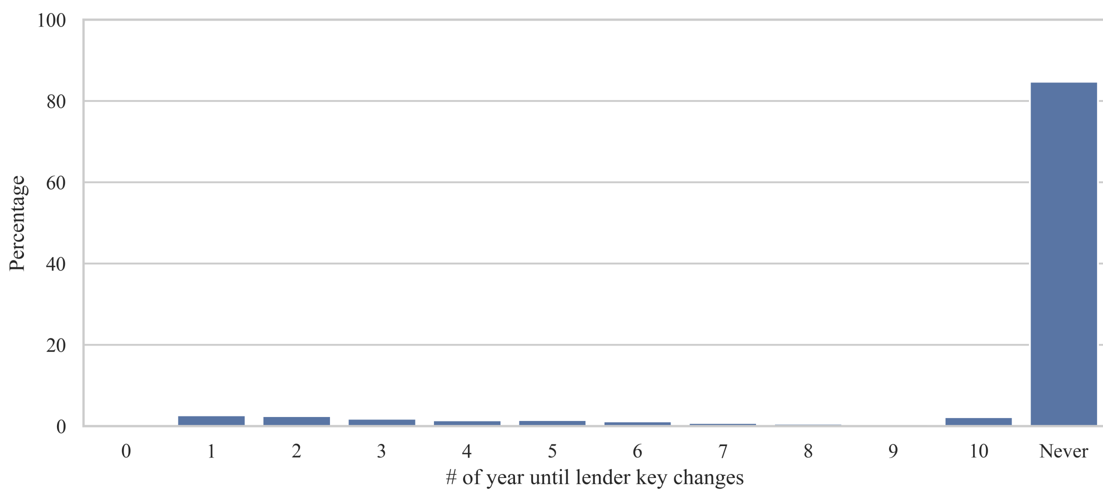
[16]It is also possible that some specialized mortgage servicing companies purchase MSR from lenders. But since specialized mortgage servicing companies are not portfolio lenders, the existence of these institutions is not going to bias the number of mortgages from portfolio lenders either.

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



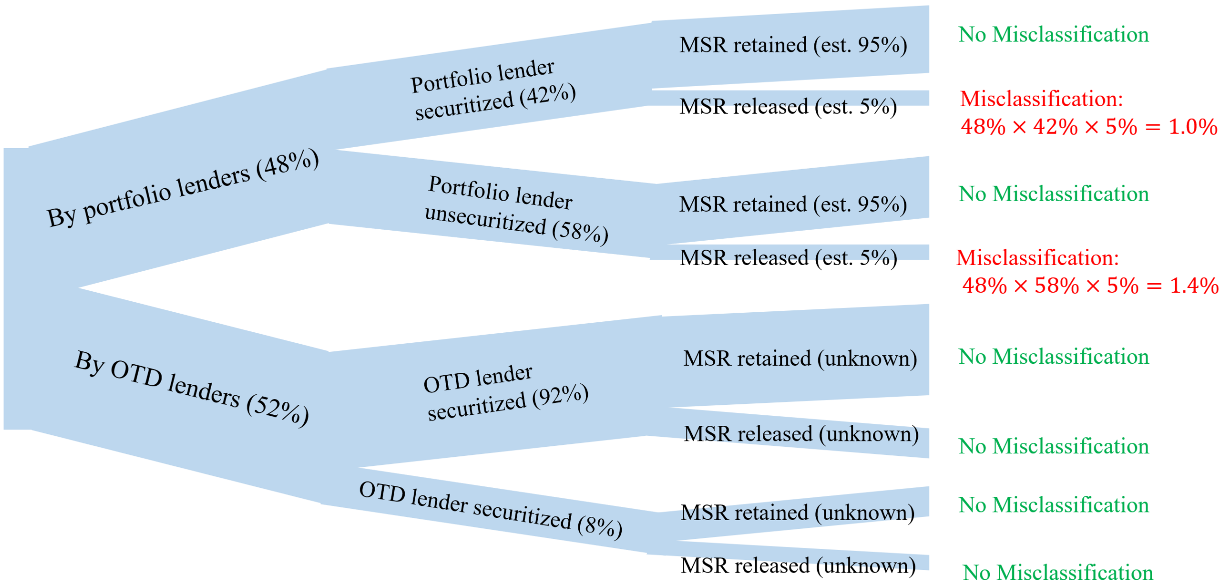
Note: This figure plots the distribution of the number of days between the mortgage origination date and the first snapshot date.

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.4 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:

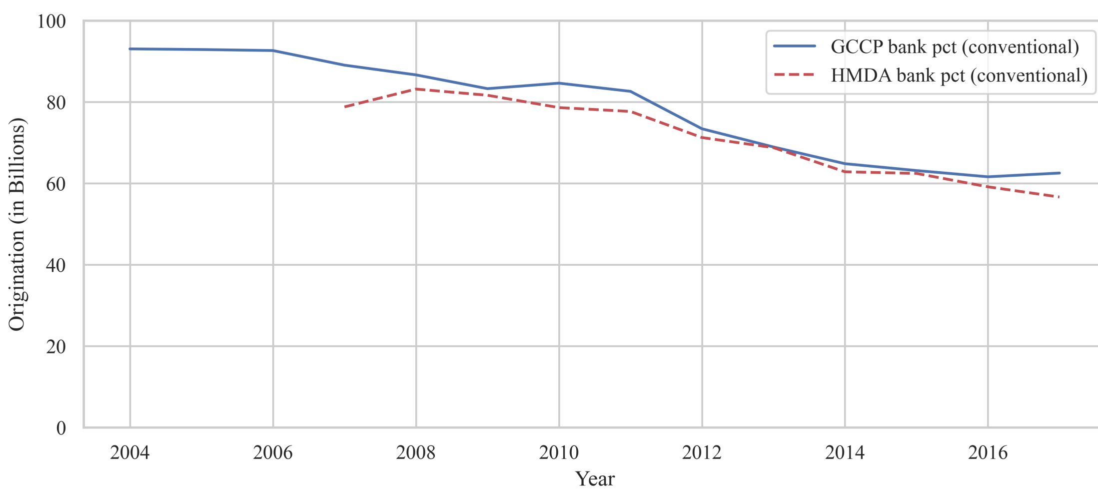
- Less than 80% of all conforming mortgages under its name are securitized throughout the whole sample period
- Both credit card balance AND auto loan balance account for more than 1% of its total outstanding consumer loan balance throughout the whole sample period
- Both credit card balance AND student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

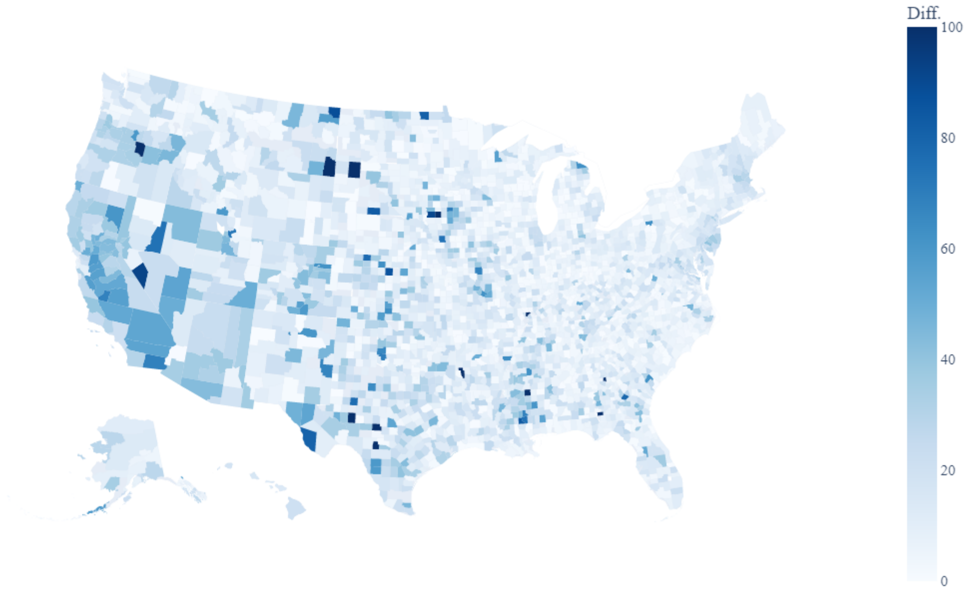
Note that since *lender_key* could identify the mortgage servicers and while HMDA only identifies the originators, some discrepancy is expected even if the bank classification method can perfectly identify the banks in the GCCP sample. On the time-series dimension, the percentage of bank loans classified 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. $abs(pct(GCCP)_c - pct(HMDA)_c)/pct(HMDA)_c$.