Shadow Banking and Branch Networks: An Analysis of Lending Patterns and Strategic Branching Decisions in the U.S. Mortgage Market

Link to the Latest Version

Long Bui

Queen's University

Abstract

This paper explores the effects of cannibalization and business-stealing within the lenders' branch networks in the U.S. shadow banking mortgage market. Given the difficulty in obtaining data related to the shadow bank branch network, a novel dataset, the Your-Economy Time Series dataset, is introduced to provide details on shadow bank branch locations from 2012 to 2017. Implementing a two-stage least squares approach shows that having a branch increases the lender's market share. However, the marginal effect of an additional branch on market share decreases as the total number of branches within a market increases, indicating the presence of market share cannibalization effects. In addition to market share cannibalization, there are business-stealing impacts from the presence of rival lenders' branches. Then, using an ordered probit with a control function approach, results show that when lenders decide to expand their branch network, they consider potential cannibalization and business-stealing effects. Finally, a structural model of entry and exit is estimated to decompose the impact of cannibalization and business-stealing effects on new branch openings. Counterfactual simulations show that muting the cannibalization and business-stealing effects leads to an increase in the establishment of new branches.

1 Introduction

This study explores the role of branch networks on the lending patterns and branching decisions of non-deposit-taking mortgage lenders, commonly referred to as “shadow banks”. Brick-and-mortar branches are important for the deposit-taking and lending functions of traditional banks.[1] However, the importance of brick-and-mortar branches for shadow banks is unclear.

Shadow banks do not accept deposits and operate based on an "originate-to-distribute" business model, i.e. instead of holding their originated loans on their balance sheet, loans are sold to investors on the secondary market (Vickery and Wright (2013)). Shadow banks also utilize lending technologies that eliminate the need for a physical brick-and-mortar branch. This innovation allows borrowers to secure mortgages over the phone or obtain pre-approved loans via online digital platforms. While traditional banks also employ these non-branch lending technologies, the effects on shadow banks may differ due to their lack of deposit-taking function.

Despite this, the presence of branch networks seems to be important for shadow banks. The ten largest shadow banks, excluding fintech companies, have an average of 10.8 branches across the United States compared to the average of 15.7 branches across all traditional banks insured by the Federal Deposit Insurance Corporation (“FDIC”).[2] A brick-and-mortar branch can result in various advantages, including increased visibility, strengthened brand recognition, provision of in-person customer service, and access to localized resources such as real estate agents and appraisers.

Given the presence of branches for shadow banks, this analysis aims to study various economic drivers that contribute to the opening of new branches, including market share cannibalization and competition (business-stealing by rivals). These determinants impact traditional branch networks and their deposit-taking and lending operations (Berger and Dick (2007) and Rysman, Townsend, and Walsh (2023)), thus suggesting that these factors also influence the decision-making strategies shadow banks use when establishing new branches.

First, the study examines the relationship between a lender’s branch network and its effect on their market share. Second, the analysis seeks to identify whether cannibalization exists within their branch network and whether business-stealing effects exist from their

[1] See, for example, Degryse and Ongena (2005), Agarwal and Hauswald (2010), Ergungor (2010), Gowrisankaran and Krainer (2011), Ho and Ishii (2011), Gilje, Loutskina, and Strahan (2016), Cortés and Strahan (2017), Nguyen (2019) and Aguirregabiria, Clark, and Wang (2020).

[2] Following Fuster et al. (2019), a lender is classified as a fintech lender if a borrower can initiate and obtain a pre-approved loan online via the lender’s website without speaking to a loan officer. The top 20 lenders as of December 2016, according to Fuster et al. (2019), contain eight non-fintech shadow banks, three fintech shadow banks, and the remaining nine are banks.

rivals' branch networks. Third, the study investigates how a shadow bank's branch network and its rivals' branch networks affect their branching decisions. Finally, a structural model with a flexible specification of fixed costs is estimated to account for factors such as entry cost, fixed operating cost, regulatory costs, and economies of density to study the effects of cannibalization and business-stealing on branching decisions.

A key challenge in studying shadow bank branch networks is the limited data. Unlike traditional bank branches, which are recorded in the Summary of Deposits ("SOD") dataset, there exists no publicly accessible, systematic data for shadow bank branches. Hence, a novel dataset, the Your-Economy Time Series ("YTS") dataset, is introduced. The YTS dataset is an annual establishment-level time series database that follows companies at their unique locations across the United States.

The YTS dataset is compiled from Infogroup, which collects the data from various sources, including industry reports and utility connects, and calls each establishment at its unique locations multiple times a year to verify and collect additional data. Infogroup's establishment data has been used in various studies to examine production networks (Barrot and Sauvagnat (2016)), entrepreneurs (Guzman and Stern (2020)), and grocery outlets (Clark, Horstmann, and Houde (2021)). Combining this YTS dataset with the Home Mortgage Disclosure Act dataset, which contains a list of shadow banks and transacted loan level data, creates a unique dataset containing shadow banks, their branch locations, and lending activity across the United States. Since the shadow bank branch locations from the YTS dataset have not been utilized before, in a companion paper, Bui (2023) shows the YTS dataset for traditional banks insured by the FDIC does well in identifying the branch locations registered in the Summary of Deposits Bank Call Reports. Using this analysis, our underlying assumption is that the YTS dataset does just as well at recording branch locations for shadow banks.

A challenge in investigating branching patterns lies in delineating markets, so lenders' decisions remain unaffected by events in other markets. Such events include potential demand spillover effects. To address this issue, events beyond a distance threshold are assumed not to affect a lender's branching decision, which allows the clustering of branching locations into separate markets that do not affect one another.

To quantify an additional branch's impact on a lender's market share, an ordinary least squares ("OLS") regression specification shows that having an additional branch is associated with a 0.16% increase in the lender's market share. However, endogeneity issues are a concern since unobserved mortgage demand determinants may correlate with the lender's branching decisions. To address this issue, a two-stage least squares ("TSLS") regression approach is used to analyze a branch's impact on a lender's market share. The first stage of the TSLS utilizes the distance between branches within a lender's branch network as instruments. The instruments in the first stage used are motivated by cost synergies. Lenders can spread costs, such as advertising and management, across their branch network, similar to Relihan (2018). Results indicate that the OLS regression understates the effect of an additional branch. The TSLS results show that when a lender opens their first branch in a market, the average lender's market share increases by 2.03%. This increase is more than six times the average market share of 0.32% across lenders, with or without a branch, who have at least some lending activity. However, the effect of an additional branch on a lender's market share gets smaller as the number of a lender's branches within a market increases. This result shows the cannibalization effects of having multiple branches in a market. Finally, by using the rival lenders' branch characteristics as instruments, the two-stage least squares analysis shows that there are also business-stealing effects from rival branches within a market. Specifically, a 1% increase in the number of rival branches decreases the lender's market share by 0.23%.

Next, an ordered probit model, similar to Igami and Yang (2016) and Rysman, Townsend, and Walsh (2023), is used to examine the branching decisions of shadow banks. The empirical findings reveal a significant negative correlation between a lender's choice to open a new branch and the presence of both their own and rivals' branches. A control function approach is used to mitigate potential endogeneity issues, such as lenders' propensity to open branches in markets with unobservable positive demand shocks. In the first stage of this method, the distances between branches within the lenders' branch network serve as instruments. The results show that lenders are less likely to open a branch with an existing branch in the market. However, though negative, the estimated parameter associated with the presence of competitors' branches lacks statistical significance.

The reduced-form regressions show the presence of both cannibalization and business-stealing effects. Consistent with these results, the ordered probit model also indicates that shadow banks exhibit a decreasing likelihood of opening a new branch when the shadow bank has a pre-existing branch presence or heightened competition from rival branches. However, the cannibalization and business-stealing effects cannot fully rationalize the observed branching patterns. In particular, despite the cannibalization effect, branching patterns in the data show that shadow banks with one or two branches are more likely to open a new branch than those with no pre-existing branch. This observation could be due to fixed costs such as entry costs, fixed operating costs, and regulatory costs or economies of density. The empirical analysis cannot separately identify these factors. Hence, to quantify the importance of these factors, the paper develops a structural model of entry and exit, wherein fixed costs are permitted to adjust as a firm's pre-existing branch presence changes. To identify the fixed costs, the estimation uses a revealed preference approach based on firms' observed branch opening decisions and the extent to which these cannot be explained by cannibalization or business-stealing effects alone.

The structural model of entry and exit follows the framework formulated by Seim (2006). The model is an entry game involving many heterogeneous lenders making simultaneous moves under incomplete information. In this model, lenders are involved in loan origination (buying loans from borrowers) and selling them on the secondary market to investors, thereby emulating the originate-to-distribute business model. The margin realized through this transaction is denoted as the intermediation fee. The lender's market share is determined by the density of their own and competitors' branch networks and various demand factors. The intermediation fee and the lender's market share make up the variable profit component of the model. For the other components of the profit equation, lenders bear a fixed cost of operating an additional branch they open. They are also subject to a private cost shock, visible to the lender but unobserved to all other lenders.

To estimate the model, loan prices (both buying and selling) and loan volume are considered exogenous, while the market share is estimated through the TSLS analysis. Additionally, it is assumed that the cost shock drawn by the lender follows an Extreme Value Type 1 distribution. With the cost shock, the fixed cost is estimated by maximizing a log-likelihood function with a nested fixed-point problem.

The counterfactual analysis shows that the number of new branches nationally would rise by 4.4% in an environment without cannibalization effects. In comparison, the absence of business-stealing effects would result in a 6.3% increase. These effects vary across different U.S. regions, with the Northeast and Midwest experiencing a larger business-stealing effect than cannibalization. In contrast, the South and West regions exhibit comparable impacts from both effects. When undertaking the counterfactuals across lenders grouped by the number of branches they have in a market, suppressing the cannibalization channel has a large impact, particularly for lenders with a pre-existing high number of branches in a market.

Previous research has explored various aspects of the mortgage market, including the reasons for the shift from traditional banks to shadow banks (Buchak, Matvos, et al. (2018)), the effects of traditional bank branch consolidation (Nguyen (2019)), the differences between traditional banking and fintech lending (Fuster et al. (2019)), and the shadow bank's funding sources (Jiang (2019)). This study is the first to analyze the effect of shadow banks' branch networks on lending patterns and their strategic branching decisions.

Berger and Dick (2007) examines the first mover advantage for bank branch entry concerning gaining depositors and Gowrisankaran and Krainer (2011) study how automatic teller machine ("ATM") entry decisions are made in response to regulating ATM surcharges. More recently, Aguirregabiria, Clark, and Wang (2016) analyzed bank branch networks for geographical diversification purposes, and Rysman, Townsend, and Walsh (2023) investigated how the effect of bank branching decisions after the 1997 financial crisis affected access to credit. These papers show that branches are important for depositors and the transformation of deposits into loans. Since shadow banks do not take deposits, this study can isolate the effect of branches on lending activities without the deposit side of the business.

Local competition is an important factor in influencing market outcomes. As evidenced by Buchak and Jørring (2021), market concentration impacts non-interest fees, with the top decile of the most concentrated markets incurring fees 35 basis points higher than their least concentrated counterparts. Furthermore, Scharfstein and Sunderam (2016) shows that monetary policy transmission to mortgage rates is dampened in markets with a high concentration level. Our analysis concentrates on one specific element contributing to local competition - the branch network of the shadow banking sector.

Few papers have examined the shadow bank branch network. Relihan (2018) studied the shadow banks' branch network, using the Nationwide Mortgage Licensing System dataset, and their effects on low socioeconomic-status borrowers and the credit quality of local applicant pools before the 2008 financial crisis. Different from Relihan (2018), this paper examines the effect of having a branch on lending, the cannibalization effect, the business-stealing effect, and their branching decisions.

Finally, this paper contributes to the literature on cannibalization and pre-emptive entry literature. Igami and Yang (2016) studies cannibalization and pre-emptive entry in the context of the burger industry. Holmes (2011) studies the trade-off between economies of density and cannibalization in the Wal-Mart network of retail stores. This paper analyzes the cannibalization and business-stealing effects in the U.S. mortgage market for shadow banks.

The rest of the paper proceeds as follows. The next Section describes the U.S. mortgage market and the lending landscape. Section 3 describes the data used, Section 4 contains the empirical analyses, Section 5 outlines the model, and Section 6 shows the estimates and describes the counterfactual experiments. Finally, Section 7 concludes.

2 Industry Details

2.1 Lenders in the U.S. Mortgage Market

The landscape of lenders in the U.S. mortgage market comprises two types of institutions: traditional banks, which are depository institutions, and shadow banks, which are non-depository institutions. They mainly differ in their funding sources and what they do with the loans they originate.

The main funding source for traditional banks, or depository institutions, comes from customer deposits, which include checking, savings, and time deposits. Traditional banks use these funds to initially extend mortgage loans to borrowers. Once a loan is originated, the bank can either hold it on its balance sheet, add it to its portfolio of assets, or sell it to investors in the secondary market, often in mortgage-backed securities. By retaining loans on their balance sheets, traditional banks are exposed to the credit risk associated with them. However, when they sell the originated loans to investors, they transfer some or all of the associated credit risk to the purchaser. This option is gaining increasing popularity among banks.

In comparison, shadow banks', or non-depository institutions', funding sources differ from traditional banks, as they do not rely on customer deposits. Instead, they obtain funds through lines of credit, typically from traditional banks or other financial institutions (Jiang (2019)). The primary business model of shadow banks is to originate loans and then sell them to investors on the secondary market (originate-to-distribute business model). This approach allows shadow banks to generate income from loan origination fees and servicing fees, while simultaneously transferring the credit risk associated with the loans to investors in the secondary market. The originate-to-distribute business model is particularly attractive to shadow banks, enabling them to maintain lower capital requirements than traditional banks.[3]

In conclusion, the U.S. mortgage market is characterized by a diverse landscape of lenders, including traditional and shadow banks. While traditional banks have customer deposits as an option for their funding source and can choose to hold originated loans on their balance sheets or sell them to investors, shadow banks use lines of credit to fund their mortgage lending activities and primarily follow the originate-to-distribute business model.

2.2 Lending Technologies

In the U.S. mortgage market, lending technologies have evolved over the years, offering lenders various ways to interact with borrowers. These interactions can take place through three primary channels: (i) in-person branch meetings, (ii) phone calls, and (iii) online platforms.

In-person branch meetings have been the traditional method of interaction between lenders and borrowers. Borrowers visit the lender's brick-and-mortar branch to discuss their mortgage needs, submit loan applications, and review loan terms with a loan officer.

[3]Vickery and Wright (2013) discuss this originate-to-distribute business model in detail.

Phone-based interactions give borrowers the convenience of discussing their mortgage needs with a loan officer without visiting a brick-and-mortar branch. This method of communication has gained popularity as it offers greater flexibility and accessibility for borrowers who prefer not to visit a branch in person.

In line with the fintech definition provided by Fuster et al. (2019), a lender is classified as a fintech lender if a borrower can apply for and obtain a pre-approved loan online without speaking to a loan officer. This digital process often involves automated underwriting systems that assess the borrower's creditworthiness and provide a loan decision quicker than traditional methods.

In contrast, non-fintech lenders require interaction with borrowers through in-person branch meetings or phone calls to obtain pre-approval for a loan. While these lenders may utilize the same technology to streamline the application process, the key difference is the requirement for human interaction in the pre-approval stage. For this study, fintech lenders are removed from the set of shadow banks considered.

2.3 Traditional Bank vs. Shadow Bank Branches

Traditional bank branches play a multifaceted role in providing financial services. These branches often serve as the primary point of contact between the bank and its customers. They offer a variety of services, including, but not limited to, account opening, loan application processing, customer service, and safe deposit boxes. The physical presence of these branches contributes to relationship building with customers and overall customer experience. In response to the evolving technologies, traditional banks increasingly use non-branch methods to interact with customers. Consumers can now increasingly do much of their banking online without visiting a brick-and-mortar branch. As of 2019, some traditional banks (e.g. Bank of America) have fintech lending technologies as defined by Buchak, Matvos, et al. (2018).[4]

While shadow bank branches do not provide the full spectrum of services as traditional bank branches, they accommodate face-to-face interactions with borrowers. Shadow banks facilitate these interactions through their brick-and-mortar branches, which house their loan officers. The physical premises of these branches can vary. For instance, some adopt the form of conventional storefronts. For example, Bay Equity Home Loans maintains a traditional walk-in storefront at "*5610 Scotts Valley Dr, Scotts Valley, CA 95066*". Conversely, other lenders operate within corporate buildings that prospective borrowers can visit. An example is Bayshore Mortgage Inc., whose office headquarters is at "*1920 Greenspring Drive,*

[4] An updated list of fintech lenders for 2019 as defined by Buchak, Matvos, et al. (2018) can be downloaded at https://sites.google.com/view/fintech-and-shadow-banks.

Timonium, MD 21093". While the external view of this address via Google Maps shows a corporate building, in-person consultations with a loan officer are available upon request.

Comparatively, the most significant similarity between shadow bank branches and traditional bank branches lies in their role as customer contact points, providing in-person customer services. However, the extent and nature of services provided differ. Traditional bank branches provide a broad array of financial services, whereas shadow bank branches operate with the narrow aim of originating mortgages and mortgage-related services.

2.4 Retail Market, Secondary Market and the Intermediation Fee

Lenders in the structural model are intermediaries who match borrowers in the retail market to investors in the secondary market. More formally, the U.S. mortgage market is comprised of two main components: the retail market and the secondary market. These markets play distinct roles in the mortgage lending process. The retail market is where mortgage loans are originated, involving direct transactions between lenders and borrowers. In this market, lenders, such as traditional banks and shadow bank institutions, extend mortgage loans to borrowers. The retail market encompasses the entire process, from the borrower's loan application to the final approval and funding of the mortgage loan. As discussed in Section 2.2, loan originators may use various lending channels, including in-person branch meetings, phone, and online platforms, to interact with borrowers during this stage.

Originated mortgage loans in the retail market can then be sold in the secondary market. In the secondary market, existing mortgage loans are packaged into mortgage-backed securities ("MBS") and are bought, sold, and traded among investors. Major players in the secondary market include government-sponsored enterprises ("GSEs") such as Fannie Mae, Freddie Mac, and Ginnie Mae, as well as private financial institutions and investors. By purchasing loans from the retail market, the secondary market provides liquidity to the mortgage market, allowing lenders to originate more loans and maintain a steady flow of funds.

The intermediation fee refers to the cost associated with connecting borrowers and lenders in the mortgage market. This fee compensates shadow bank loan originators. Intermediation fees can include various components, such as loan origination fees, points, and other costs borrowers pay to obtain a mortgage loan. The fee is expressed as a percentage of the loan amount. It varies depending on factors such as the borrower's creditworthiness and the interest rate-rebate combination that the borrower chooses.

3 Data

3.1 Your-Economy Time Series Data

The Your-Economy Time Series ("YTS") dataset is an annual establishment-level time series database tracking businesses at their specific geographic locations across the United States. The YTS dataset is assembled by the Business Dynamics Research Consortium ("BDRC"), which sources business data files from Infogroup (currently Data Axle), where the data is collected and verified.

Infogroup obtains data from various sources, including annual reports, government documents (e.g., 10-Ks and SEC filings), internet searches, utility connects/disconnects, and other relevant channels. To further ensure the data's accuracy, Infogroup contacts each establishment at least once a year. This direct communication allows Infogroup to verify the data's authenticity and gather any additional information about the establishment.

The YTS dataset provides information about each establishment at its unique location, including details such as (i) the name of the business or lender, (ii) the specific address of the establishment, including street address, city, county, state, and ZIP code, (iii) the range of years during which the establishment was operational, and (iv) the North American Industry Classification System ("NAICS") code, which indicates the primary industry in which the establishment operates. The dataset is utilized to obtain the geographical location of shadow bank branches across the United States.

The Infogroup dataset has been used in various academic studies. Barrot and Sauvagnat (2016) used the dataset to study production networks, Guzman and Stern (2020) used the dataset to study entrepreneurs and Clark, Horstmann, and Houde (2021) used the dataset to study grocery outlets. This paper uses the dataset in a novel context, investigating the branch network of shadow banks. In a companion paper, Bui (2023) verifies the quality of the YTS dataset by comparing the branch locations from the publicly available SOD dataset with those in the YTS dataset. The SOD dataset contains branch locations for all traditional banks insured by the FDIC.

This comparison shows that at least 82% of all traditional banks, as documented in the SOD dataset, have at least one of their branches recorded in the YTS dataset. Furthermore, when a traditional bank is included in the YTS dataset, we observe that, on average, 89-92% of a bank's county presence is recorded in the YTS data over the 2012-2017 period. Furthermore, over the 2012-2017 period, for a given traditional bank and county combination represented in the SOD dataset, on average, between 85-91% of branches within this traditional bank and county combination are accounted for in the YTS dataset.[5]

Given the high match rate between the YTS and SOD datasets, it is assumed that the YTS dataset is similarly reliable in recording the locations of shadow bank branches. This assumption guides our confidence in using this dataset to examine the branch networks in the shadow banking sector.

3.2 The Home Mortgage Disclosure Act Data

The Home Mortgage Disclosure Act ("HMDA") dataset aids in understanding and addressing concerns related to the provision of credit to borrowers in the United States. Enacted by Congress in 1975 and implemented by the Federal Reserve Board's Regulation C, the HMDA's primary purpose is to provide insight into the mortgage lending practices of financial institutions, focusing on addressing concerns about unfair credit terms being provided to borrowers.

The HMDA dataset encompasses most mortgages originating in the United States. If a financial institution exceeds a total asset threshold and originated over a pre-determined loan amount, then it must report various types of information related to its mortgage lending activities. Financial institutions, including traditional banks, credit unions, and savings associations, must report to the Home Mortgage Disclosure Act under specific conditions. The obligations apply to institutions that (i) possess total assets exceeding $44 million, (ii) operate a branch in a Metropolitan Statistical Area ("MSA"), (iii) originate at least one loan, and (iv) originate a minimum of 25 home loans within the preceding two calendar years.

Shadow banks have different prerequisites for reporting to HMDA. These entities are required to report if they (i) operate with for-profit status, (ii) purchase loans constituting more than 10% of the loans they originated or surpassing $25 million in total, (iii) maintain a branch or receive a mortgage application within an MSA or metropolitan division in the preceding calendar year, and (iv) have assets exceeding $10 million or originate at least 100 mortgages in the preceding calendar year.

The information lenders must report includes specific details about each loan origination, such as the loan amount, loan type, loan purpose (e.g., home purchase, home improvement, or refinancing), and property location. Moreover, the HMDA dataset includes some borrower characteristics, such as race, income, and other demographic information. The primary use of this dataset is to calculate lenders' market share across geographic areas.

[5] For more details, refer to the companion paper Bui (2023).

3.3 Optimal Blue Data

Optimal Blue is a technology platform loan officers use to obtain interest rate and rebate combinations from different lenders, which can then be offered to borrowers. By entering a borrower's characteristics into the platform, loan officers can access various available loan options, enabling borrowers to find the most suitable terms for their financial needs.

To use the platform, loan officers follow a three-step process: (i) the loan officer enters the borrower's information (e.g. credit score, loan-to-value ratio, loan amount, etc.) and the property and loan characteristics (e.g. property location, loan purpose etc.) into the platform, (ii) Optimal Blue's platform then provides a list of lenders, interest rates, and rebate combinations that can be offered to the borrower based on their characteristics, and (iii) once the borrower and loan officer agree on the loan terms, the loan is locked, obligating the lender to fund the loan at the agreed-upon terms. This final step secures the loan and ensures both parties are committed to the transaction.

To calculate the intermediation fee, the price of the loan minus the cost of the loan, for the structural model, this dataset is used to obtain the cost at which the lender buys a loan from the borrower. The variables used from this dataset are the Fannie Mae and Freddie Mac product identifier, the date at which a loan is locked, the price of the loan net of loan-level price adjustments, and the property location.

First, the dataset is restricted to loans that are Fannie Mae and Freddie Mac loan products, and the price net of loan-level price adjustments is used as the price at which the lender buys the loan from the borrower.

To obtain the associated price which the lender sells the loan for, the interest rate at which the borrower has agreed to pay over the life of the loan, and the date from when the loan is locked is used to match with the To-Be-Announced Mortgage-Backed Securities ("TBA-MBS") trades in the Trade Reporting and Compliance Engine ("TRACE") dataset.

3.4 Trade Reporting and Compliance Engine Data

The TRACE dataset is a comprehensive database managed by the Financial Industry Regulatory Authority ("FINRA") that provides detailed information on transactions in the U.S. fixed-income market, including corporate bonds, agency debt, and securitized products. Established in 2002, TRACE has become a tool for regulators, market participants, and researchers seeking to monitor and analyze the fixed-income market. The platform offers insights into trading activity, pricing, and market trends, promoting transparency.

FINRA requires member firms to report all eligible transactions in TRACE-eligible securities to the platform within a specified time frame. The reporting process ensures that the data remains up-to-date and accurately reflects the current state of the fixed-income market.

The TBA-MBS dataset obtained from TRACE provides detailed information on transactions in the TBA-MBS market. TBA-MBS are forward-settling agency mortgage-backed securities traded on a "to-be-announced" basis, wherein the exact pool of mortgages to be delivered at settlement is not specified at the time of the trade. The TRACE TBA-MBS dataset covers these securities' trading activity, pricing, and transaction volume. It also contains the near-universe of TBA-MBS transactions (Gao, Schultz, and Song (2017)).

This paper uses the closing price of the TBA-MBS at the end of the day as the price for which the lender sells the originated loan in the secondary market. The calculated margin earned on a loan is the closing price of the TBA-MBS trade minus the loan price net of loan-level price adjustments from the Optimal Blue dataset. This margin is averaged and aggregated to the market-year level to calculate the variable profit the lender earns in the structural model.

4 Empirical Evidence

For the empirical investigation, markets are constructed with the assumption that these markets are independent from one another and that lenders do not implement geographic diversification strategies. In the U.S. mortgage market literature, analyses are commonly undertaken at the county level (for example, Scharfstein and Sunderam (2016) and Buchak and Jørring (2021)). However, to analyze branching decisions, there could be spillover effects from the county's neighboring counties. That is, there could be significant demand spillovers from neighboring counties, violating the assumption of independent markets.

Past papers have approached the construction of independent markets differently. Zheng (2016) integrates observed demand with machine learning techniques to discern independent markets. Rysman, Townsend, and Walsh (2023) characterizes independent markets in rural Thailand based on the distance between branches. Given the limited detail in the data regarding shadow banks and loan originations and with the focus on the most populated counties, this paper cannot use those methods. Hence, the objective is to develop an algorithm to define independent markets, so lenders with branches in a given market are in close competition while reducing the likelihood of demand spillover among these markets.

To construct these markets (shown in Figure 1), the process is first initiated with the list of the top 500 U.S. counties selected based on their mean population between 2012-2017.[6] The list of the top 500 counties ensures that the centroid counties chosen are similar. We use

[6] For consistency across time, all geographic locations are mapped to 2023 counties. This means Connecticut is removed from the set of considered states as their defined county borders changed in 2022.

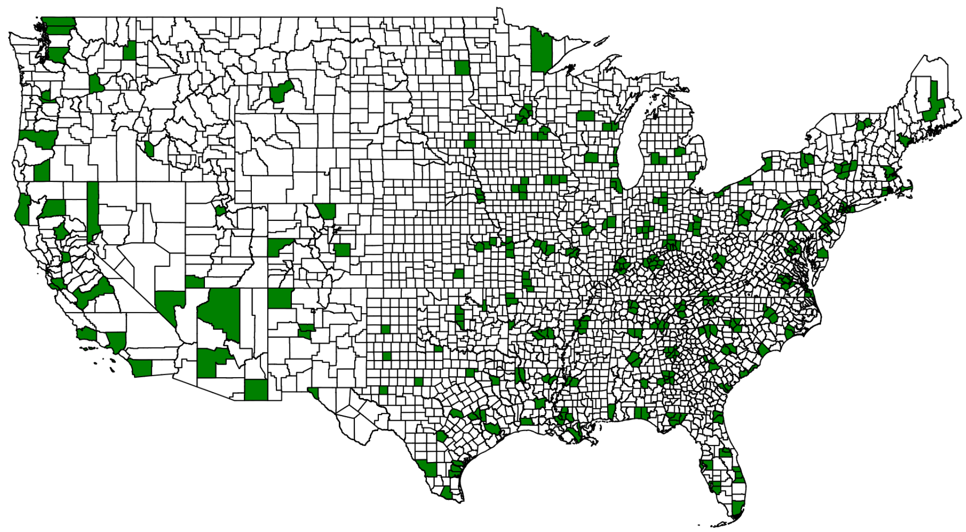

Figure 1: Constructed markets and their counties

an algorithm with the following steps: (i) select the county with the highest population from our list of top 500 counties, defining this as a 'centroid' county; (ii) include every county within a 25-mile radius from the center of the centroid county and label this group of counties as a single market; (iii) remove all neighboring counties to the centroid county from the list of top 500 counties; and (iv) choose the next most populated county from the modified list of top 500 counties, ensuring it is at least 75 miles from any other centroid county. Then, repeat steps (ii)-(iv).

This algorithm generates a collection of markets where each of the centroid counties ranks among the top 500 in terms of population, and all counties situated within a 25-mile radius of the centroid county are geographically clustered to their corresponding centroid county, resulting in a set of independent markets. By maintaining a minimum distance of 75 miles between each centroid county, the algorithm effectively eliminates any geographic overlap among the constructed markets. This distance serves as a buffer, ensuring that branches located on the periphery of our markets are at least 25 miles from the nearest market, thereby minimizing the potential for demand spillover effects.

The application of this algorithm leads to the formation of 138 distinct markets, collectively containing 328 counties. Several of these counties are among the initial top 500 counties, ranked by population from 2012 to 2017. Specifically, 219 of these initial top 500 counties are contained within the 138 constructed markets.[7] Figure 1 shows the geographical distribution of these markets, with each county included in the market highlighted in green.

[7]Note that by construction, all 138 centroid counties are from the list of the top 500 counties. Hence, 81 of the 219 counties are attached to a centroid county.

Variable	Obs	Mean	Std. Dev.	Min	Max
Market share (%)	298,080	0.07	0.43	0	18.51
Strictly positive market share (%)	63,244	0.32	0.90	5.77e-05	18.51
Number of own branches	298,080	0.06	0.45	0	27
(At least 1 branch) Number of own branches	10,496	1.64	1.78	1	27
Number of rival shadow bank branches	298,080	32.71	44.10	0	238
Number of rival trad. bank branches	298,080	265.23	395.40	19	3400
Total number of rival branches	298,080	297.94	434.48	21	3631
Mean population	298,080	558,083	1,066,141	50,732	1.01e+07
Mean household income	298,080	202,424	358,986	18,776	3,295,198
Number of counties per market	138	2.38	1.91	1	11
Positive change in the number of branches	1,208	1.15	0.48	1	5
Negative change in the number of branches	1,144	-1.12	0.44	-5	-1

Table 1: Summary statistics

Table 1 provides summary statistics concerning the branch network for a balanced panel of 360 shadow bank lenders in the HMDA and YTS datasets for the entire 2012-2017 year period. There are 298,080 year, market, and lender observations. Conditional on a shadow bank lender having at least one branch in a market, the shadow bank maintains 1.64 branches on average. The average shadow bank lender's market share of total loans originated by all types of lenders across all markets is 0.07%. However, when restricting the sample to where the shadow bank lender has a strictly positive market share, their mean market share is 0.32%. On average, the number of rival shadow and traditional bank branches in a market is 32.7 and 265.2, respectively. Furthermore, from year to year, when shadow banks expand their branch network in a market, on average, they increase their branch network by 1.15 branches. In contrast, when shadow banks decrease their branch network, they reduce their branch network by 1.12 branches.

4.1 Number of Branches and Market Share

This section examines the relationship between the branch presence in a defined market (as above) and the lender's market share. The ordinary least squares ("OLS") regression specification is

$$s_{imt} = \alpha_0 + \alpha_1 n_{imt} + \alpha_2 n_{imt}^2 + \alpha_3 n_{-imt} + \Gamma_1' \mathbf{X}_{1mt} + \epsilon_{imt} \quad (1)$$

where the dependent variable s_{imt} is lender i's market share for market m and year t. For the explanatory variables, the regression contains the number of branches lender i has in market m and year t, denoted by n_{imt}, the squared number of branches in market m and year t, denoted by n_{imt}^2, and the number of rival (traditional bank and shadow bank) branches

(000's) in market m and year t, denoted by n_{-imt}. The variable $\mathbf{X}_{1mt}$ contains the mean market population (000,000s), mean market household income ($000,000s), total market loan amount ($000,000s), lender-market fixed effect, and year fixed effect.

There are endogeneity concerns when using the OLS regression specification. One concern is omitted variable bias due to unobserved demand and cost shocks. For example, in lending markets with unobserved adverse demand shocks and for cost efficiency reasons (e.g. cheap funding options), lenders still decide to operate a branch. In this case, the estimate of the lender's branch on market share will be biased downward.

To address the endogeneity issue, seven instruments are used for the three variables: 1) the number of own branches more than 100 miles from the centroid county, adjusted by distance, 2) the number of own branches 50-100 miles away from the centroid county, and the squared term, 3) the number of own branches 100-150 miles away from the centroid county, 4) the number of own branches more than 150 miles away from the centroid county and within the same state, 5) the number of all rival branches that is 50-100 miles away from the centroid county, and 6) the number of all rival branches that is 100-150 miles away from the centroid county.[8]

These instruments are motivated by Holmes (2011), Goolsbee and Syverson (2008), and Relihan (2018), who find economies of density in Walmart store locations, the airline industry, and branch locations, respectively. Following Relihan (2018), the idea behind the economies of density with branches arises from sharing advertising and management costs, which incentivize new branches to be close to their pre-existing branch network (Berger, Leusner, and Mingo (1997), Bos and Kool (2006) and Felici and Pagnini (2008)).

Table 2 shows the OLS and TSLS results for full sample, Columns (1) and (2), and the sample restricted to markets, Columns (3) and (4), where the lender either has a branch in the same state as the market or has a strictly positive market share in the same state as the market, labeled as "Active states". After controlling for the year fixed effect and the lender-market fixed effect, the estimate of interest ($\hat{\alpha}_1$) in Column (1) shows an additional branch corresponds with a 0.16% higher market share. To address the endogeneity issues discussed above, Column (2) uses the characteristics of the lender's own and rival branch networks as instruments. The TSLS results show that an additional branch within a market increases the lender's market share by 2.03%, an effect larger than the OLS estimate. These results are consistent with the restricted sample in Columns (3) and (4).

[8]For an example of the instrument (1) where the number of branches is adjusted by distance, consider a county that is x miles away from the centroid of the market, denoted as c_x, with n_{ic_xt} branches. We multiply n_{ic_xt} by the reciprocal of the distance between the centroid of market m and county c_x, denoted by $d(m, c_x)$, i.e., $n_{ic_xt} \times \frac{1}{d(m,c_x)}$. These calculations are performed for each county more than 100 miles from the centroid county of market m. The results are then summed up.

	Full sample		Active states	
	(1)	(2)	(3)	(4)
	OLS	IV	OLS	IV
Own branches (n_{imt})	0.155***	2.032***	0.155***	1.954***
	(0.021)	(0.313)	(0.021)	(0.300)
Own branches sq. (n_{imt})	-0.005***	-0.096**	-0.005***	-0.090**
	(0.001)	(0.044)	(0.001)	(0.042)
Rival branches in 000's (n_{-imt})	-0.082**	-0.478**	-0.118	-0.792*
	(0.036)	(0.227)	(0.072)	(0.441)
Market loan amount ($000,000,000,000s)	0.649***	0.972*	1.216***	1.556*
	(0.223)	(0.506)	(0.454)	(0.931)
Mean population (000,000s)	0.181**	0.232	0.098	0.163
	(0.079)	(0.239)	(0.130)	(0.396)
Mean household income ($000,000s)	-0.533**	-1.393*	-0.479	-1.885
	(0.233)	(0.796)	(0.374)	(1.335)
Constant	0.088***		0.214***	
	(0.015)		(0.032)	
Kleibergen-Paap rk LM statistic		10.647		10.691
Cragg-Donald Wald F statistic		244.519		123.103
Hansen J statistic		3.816		3.433
Observations	298080	298080	132504	132504
R^2	0.853		0.850	

Note: Standard errors are given in parentheses. The symbols ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. The standard errors are clustered at the market level where there are 138 clusters. All regressions include lender-market fixed effect and year fixed effect. There are 7 instruments used in the first stage of the TSLS regressions.

Table 2: Market OLS and TSLS regression results

The second term of interest is the estimated coefficient for the squared number of branches within a market ($\hat{\alpha}_2$). If the estimate $\hat{\alpha}_2$ is negative, then each additional branch present in the market cannibalizes its own market share. If $\hat{\alpha}_2$ is positive, then the estimate indicates economies of density. Table 2, Column (2), shows the estimated α_2 coefficient is -0.095. The negative estimated coefficient shows that an additional branch cannibalizes market share. This effect gets stronger when the sample is restricted to active states for lenders.

Finally, the estimate of coefficient α_3 captures the business-stealing effect. Estimates using the TSLS in Column (2) show that a one standard deviation increase in the number of rival branches decreases the lender's market share by 0.479%. This effect increases to 0.559% on the restricted sample in Column (4).

To summarize, branches are important for shadow bank lenders to gain market share in local markets despite the evolving technology. Lenders with a strictly positive market share on average have a 0.32% market share while having a branch on average increases their market share to 2.03%. However, there are cannibalization effects from having multiple branches in the same market, and business-stealing effects from rival branches exist.

4.2 Branching Decisions

Next, the branching decisions in relation to a lender's branch network are examined similarly to Igami and Yang (2016) and Rysman, Townsend, and Walsh (2023). The ordered probit specification is

$$a_{i,m,t} = \gamma_1 n_{i,m,t-1} + \gamma_2 n_{-i,m,t-1} + \Gamma_2^{'} \mathbf{X}_{2,m,t-1} + \nu_{i,m,t-1} \tag{2}$$

where the branch opening and closing decisions, i.e. the change in the number of branches in market m from year $t-1$ to year t, are

$$a_{i,m,t} = \begin{cases} -1 & \text{if } \Delta n_{i,m,t} < 0 \\ 0 & \text{if } \Delta n_{i,m,t} = 0 \\ +1 & \text{if } \Delta n_{i,m,t} > 0 \end{cases} . \tag{3}$$

For the explanatory variables, the ordered probit contains the number of branches lender i has in market m and year t, denoted by $n_{i,m,t}$ and the number of rival (shadow bank and traditional bank) branches in 000's in market m and year t, denoted by $n_{-i,m,t}$. The variable $\mathbf{X}_{2,m,t-1}$ contains the mean market population (000,000s), mean market household income ($000,000s), total market loan amount ($000,000s), lender fixed effect and market fixed effect.

	(1)	(2)	(3)	(4)
Dependent variable: $a_{imt} \in \{-1, 0, +1\}$				
Lagged own branches (n_{imt-1})	-0.202***	-0.688**	-0.202***	-0.687**
	(0.044)	(0.323)	(0.043)	(0.323)
Lagged rival branches in 000's (n_{-imt-1})	-1.920*	-0.879	-1.830	-0.855
	(1.2918)	(1.229)	(1.444)	(1.414)
Lagged market loan amount ($000,000s)	27.840***	27.728***	26.813***	26.320***
	(4.294)	(4.245)	(4.560)	(4.479)
Lagged mean market population (000,000s)			-1.498	-1.652
			(1.088)	(1.140)
Lagged mean household income ($000,000s)			4.109	5.647
			(3.398)	(3.872)
Lagged own branch residual		0.490		0.489
		(0.330)		(0.329)
Lagged rival branch residual		-0.003		-0.002
		(0.002)		(0.002)
Cutoff 1	-3.152***	-2.671***	-3.140***	-2.549***
	(0.315)	(0.426)	(0.482)	(0.670)
Cutoff 2	2.199***	2.683***	2.212***	2.805***
	(0.314)	(0.438)	(0.484)	(0.686)
Observations	248400	248400	248400	248400
Pseudo R^2	0.063	0.063	0.063	0.063

Note: Standard errors are given in parentheses. The symbols ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. The standard errors are clustered at the market level where there are 138 clusters. All regressions include lender fixed effects and market fixed effects.

Table 3: Ordered probit regression results

Endogeneity concerns arise from omitted variable bias in this specification. For example, if lenders choose to open a new branch in regions with higher unobserved demand shocks, then the estimated coefficients of opening a new branch would have a positive bias.

To address these endogeneity concerns, a control function approach is used to supplement the analysis. The two instruments used in the first stage of the estimation are: 1) the number of own branches more than 50 miles from the centroid county adjusted by distance, and 2) the number of rival branches between 50-100 miles away from the centroid county.

Columns (2) and (4) of Table 3 estimate the specification using the control function approach. Columns (3) and (4) control for additional demand factors: mean market population and mean market income. Table 3 shows the estimate of a lender's own branch presence γ_1 on their choice to open a new branch or close an existing one is negative. Column (3) shows an estimate of -0.202 compared to -0.687 in Column (4), showing the upward bias from the omitted variable bias. The coefficient estimates suggest lenders are concerned about their branch presence when deciding to open a new branch. This finding is consistent with the studies conducted by Igami and Yang (2016) and Rysman, Townsend, and Walsh (2023), who explored the dynamics of burger chains in Canada and bank networks in Thailand, respectively.

The second coefficient of interest is the effect of the rivals' branch presence on a lender's decision to open a new branch. Since the coefficient estimate is negative, lenders are also concerned about their rivals' branch presence. However, this estimate is statistically insignificant. Finally, examining the demand coefficient estimates in Column (4) of Table 3, the market size (Lagged market loan amount in $000,000s) is positively correlated with new branches. To summarize, lenders are concerned about their own branch network and the market size in which they choose to open a new branch.

5 Model

The model is based on Seim (2006). Consider an environment with M geographic markets, indexed by $m \in \mathcal{M} = \{1, 2, ..., M\}$, T years, indexed by $t \in \{1, 2, ..., T\}$, and I shadow banks, indexed by $i \in \{1, 2, ..., I\}$. For every market-year combination (m, t), each lender i observes the state variables: the number of existing branches n_{imt}, the number of rival (shadow and traditional bank) branches n_{-imt}, the market size q_{mt}, population, household income, average loan price $\overline{p}_t$ per $100, and the average loan cost $\overline{c}_{mt}$ per $100. Given an exogenous probability λ of an existing branch closing and an idiosyncratic cost shock ϵ_{imt}, each lender i independently chooses whether to open a new branch or not $a \in \mathcal{A} = \{0, +1\}$ in each market m for the year t. Lenders' branching decisions factor in the prospective moves of their rivals in relation to their own strategic choices.

As discussed in Section 2.4, shadow banks earn profits by purchasing and selling loans in the retail and secondary markets, respectively. The fee accrued from these transactions constitutes their profit. Lenders procure a portion s_{imt} of the entire market size q_{mt}, denoted as q_{imt}. They then buy loans amounting to q_{imt} from borrowers at an average cost $\overline{c}_{mt}$. With this loan volume q_{imt}, lenders enter into the secondary market to sell the procured loan amount q_{imt} to investors at an average price $\overline{p}_t$. Hence, the margin earned, $\overline{p}_t$ minus $\overline{c}_{mt}$ is denoted as $\overline{\phi}_{mt}$. This paper assumes lenders operate as price takers in the retail and secondary markets.

Considering the effects of market share cannibalization and business-stealing effects from their own and rivals' branch presence, muting these effects will increase the likelihood of shadow bank lenders opening a new branch. This study aims to quantify the impact of these two channels.

The profit of lender i in market m and year t is equal to

$$\Pi_{imt}\left(a_{imt}, \boldsymbol{a}_{-imt}, \mathbf{X}_{mt}\right) = VP_{imt}\left(a_{imt}, \boldsymbol{a}_{-imt}, \mathbf{X}_{mt}\right) - \alpha_k^{FC}\left(n_{imt}\right) \times \mathbf{1}_{\{a_{imt}=+1\}} + \boldsymbol{\epsilon}_{imt}\left(a_{imt}\right) \tag{4}$$

where $\mathbf{X}$ contains the state variables: margin $\overline{\boldsymbol{\phi}}$, quantity $\boldsymbol{q}$, number of own and rival branches $\boldsymbol{n}$ and local demand factors: market size, population, and household income. There are K market types where the net fixed cost of opening a new branch in market type k is captured by $\alpha_k^{FC}\left(n_{mt}\right)$, a function of the current number of branches the lender has in market m and year t. The error term $\boldsymbol{\epsilon}_{imt} = \left(\epsilon_{imt}^{(0)}, \epsilon_{imt}^{(+1)}\right)$ is a private cost shock for each action $a_{imt} \in \mathcal{A}$ that is observed by the lender but unobserved to all other lenders. It is assumed to be drawn independently from an Extreme Value Type I distribution.

The expected variable profit is defined as

$$\mathbb{E}VP_{imt}\left(\cdot\right) = \overline{\phi}_{mt} \times q_{mt} \times \mathbb{E}s_{imt} \tag{5}$$

where the margin $\overline{\phi}_{mt}$ and quantity q_{mt} are exogenously determined. The expected market share is defined as

$$\begin{aligned}\mathbb{E}s_{imt} = {} & \alpha_1\left(\tilde{n}_{imt} + a_{imt}\right) + \alpha_2\left(\tilde{n}_{imt} + a_{imt}\right)^2 \\ & + \alpha_3\left[\tilde{n}_{-imt} + \sum_{-i}\sum_{a' \in \mathcal{A}} a' \times \mathbf{1}_{\left\{a_{-imt}=a'\right\}} \times \tilde{\Pr}_{-imt}\left(a_{-imt} = a'\right)\right] \\ & + \alpha_4 \text{Market size}_{mt} + \alpha_5 \text{Population}_{mt} + \alpha_6 \text{Household income}_{mt} + \xi_{imt}\end{aligned} \tag{6}$$

where $\tilde{n}$ is the expected number of branches a lender has in market m and year t when they open a new branch. It is defined as

$$\tilde{n} = \lambda \max\left\{n - 1, 0\right\} + (1 - \lambda)n \tag{7}$$

where λ is the probability that an existing branch closes down, and the probability $\tilde{\Pr}$ denotes the belief that lender i has about the probabilities of rival lenders doing action $a' \in \mathcal{A}$. Parameter α_1 measures the per-branch effect on market share. The parameter α_2 measures the cannibalization or economies of density effect from the lender's own branch network. If α_2 is positive, then the lender benefits from having more than one branch in a given market m and year t. This is the economies of density effect. However, if α_2 is negative, then having an additional branch reduces the lender's per-branch market share.

This is the market share cannibalization effect. The parameter α_3 measures the competitive effect of their rivals' branch presence on their own market share. The parameters α_4, α_5, and α_6 are for local demand factors: market size (total loan amount in the market q_{mt}), population, and household income, respectively. Finally, ξ_{imt} is the lender-market-year fixed effect.

The centers of the constructed markets are at least 75 miles away from one another to minimize potential demand spillover effects. Hence, similar to Rysman, Townsend, and Walsh (2023), this model assumes each market m and year t combination is independent. The shadow bank's market profit depends only on its own branch network, rivals' branch network, and local demand factors. Furthermore, lenders do not undertake geographic risk diversification strategies. Since this model is static, these assumptions allow for the lender's branching problem for each year and market to be solved independently.

5.1 Equilibrium

For each market m and year t, every lender i chooses their action a_{imt}, which maximizes their profit given their private cost shock and the state variables: margin, quantity, own branch presence, rival branch presence, and demand factors. Their value function is

$$v_{imt}\left(\boldsymbol{a}_i\left(\boldsymbol{\epsilon}_i\right), \boldsymbol{a}_{-i}\left(\boldsymbol{\epsilon}_{-i}\right), \mathbf{X}\right) \quad = \quad \max_{\boldsymbol{a}_i \in \mathcal{A}} \mathbb{E}\Pi_i\left(\boldsymbol{a}_i\left(\boldsymbol{\epsilon}_i\right), \boldsymbol{a}_{-i}\left(\boldsymbol{\epsilon}_{-i}\right), \mathbf{X}\right) \tag{8}$$

where the lender is taking expectations over their rivals' expected actions, and the cost shock is assumed to be drawn from an independent and identically distributed Extreme Value Type I distribution. Hence, the best-response probability mapping for each market and year combination is given by

$$\Pr_{imt}\left(a_{imt} = a\right) = \Psi\left(\tilde{\mathbf{Pr}}_{-imt}\right) = \frac{\exp\left\{\mathbb{E}\Pi_{imt}\left(a_{imt} = a, \tilde{\mathbf{Pr}}_{-imt}, \cdot\right)\right\}}{\sum_{a' \in \mathcal{A}} \exp\left\{\mathbb{E}\Pi_{imt}\left(a_{imt} = a', \tilde{\mathbf{Pr}}_{-imt}, \cdot\right)\right\}} \tag{9}$$

for every lender i, market m, and year t. The function Ψ is a mapping of the lenders' probabilities of opening a branch onto the probability space. Hence, a symmetric perfect-Nash equilibrium is a fixed point of the probability mapping

$$\Pr^*_{imt} = \Psi\left(\mathbf{Pr}^*_{-imt}\right) \tag{10}$$

for each lender i, market m, and year t.

5.2 Estimation

The vector of structural parameters that need to be estimated is

$$\Theta = \left(\left\{\alpha_k^{FC}(n=0), \alpha_k^{FC}(n=1), \alpha_k^{FC}(n=2), \alpha_k^{FC}(n=3), \alpha_k^{FC}(n=4+)\right\}_{k=1}^{4}, \{\alpha_j\}_{j=1}^{6}, \boldsymbol{\xi}\right). \tag{11}$$

The fixed cost is estimated for four bins and by U.S. region. The bins contain lenders with 0, 1, 2, 3, or 4+ branches in a market-year combination. For each of these bins, a separate fixed cost is estimated for those opening a branch in the Midwest, Northeast, South, and West regions. Since the fixed cost captures other factors, in addition to entry cost and fixed operating costs, such as economies of density, the fixed cost is allowed to vary based on the number of existing branches in a market-year combination.

First, the probability of a branch closing down is taken from the data, equaling 0.46%. Second, the market share parameter estimates from the TSLS in Section 4.1 are used for the structural model parameters $\{\alpha_j\}_{j=1}^{6}$. Using the estimates $(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\alpha}_4, \hat{\alpha}_5, \hat{\alpha}_6)$, ξ_{imt} is calculated for each observation as

$$\begin{aligned} \hat{\xi}_{imt} \quad = \quad & s_{imt} - \hat{\alpha}_1 n_{imt} - \hat{\alpha}_2 n_{imt}^2 + \hat{\alpha}_3 n_{-imt} \\ & - \hat{\alpha}_4 \text{Market size}_{mt} + \hat{\alpha}_5 \text{Population}_{mt} + \hat{\alpha}_6 \text{Household income}_{mt}. \end{aligned} \tag{12}$$

Finally, to estimate the remaining structural parameters, an approach similar to Igami and Yang (2016) and Rysman, Townsend, and Walsh (2023) is used where maximum likelihood is used to estimate the parameters

$$\boldsymbol{\theta} = \left\{\alpha_k^{FC}(n=0), \alpha_k^{FC}(n=1), \alpha_k^{FC}(n=2), \alpha_k^{FC}(n=3), \alpha_k^{FC}(n=4+)\right\}_{k=1}^{4}.$$

The maximum likelihood estimate of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}} = \arg\max_{\theta} \sum_i \sum_m \sum_t \log \Pr_{imt}^*\left(a_{imt} \middle| \mathbf{X}_{mt}, \boldsymbol{\theta}\right)$$

where $\Pr_{imt}^*$ is the equilibrium conditional choice probability for lender i in market m and year t given state variables $\mathbf{X}_{mt}$ and parameters $\boldsymbol{\theta}$.

6 Model Estimates and Counterfactuals

Table 4 presents the estimated fixed costs derived from the maximum likelihood estimation. Given the absence of any pre-existing branches, it shows an estimated fixed cost of $35.9

	Estimate	Std. Error
Fixed cost ($000,000's) for region Midwest		
0 branches	35.906	(0.125)
1 branch	29.823	(0.174)
2 branches	22.948	(0.258)
3 branches	21.044	(0.378)
4 or more branches	11.640	(0.236)
Fixed cost ($000,000's) for region Northeast		
0 branches	52.344	(0.160)
1 branch	39.295	(0.186)
2 branches	38.038	(0.250)
3 branches	21.138	(0.409)
4 or more branches	14.979	(0.267)
Fixed cost ($000,000's) for region South		
0 branches	40.462	(0.066)
1 branch	31.898	(0.110)
2 branches	15.516	(0.158)
3 branches	9.993	(0.224)
4 or more branches	8.198	(0.204)
Fixed cost ($000,000's) for region West		
0 branches	88.476	(0.085)
1 branch	41.078	(0.127)
2 branches	30.651	(0.151)
3 branches	26.210	(0.289)
4 or more branches	19.045	(0.196)
Other parameters		
Own branch (α_1)	2.032	(0.313)
Own branches sq. (α_2)	-0.096	(0.044)
Rival branches in 000's (α_3)	-0.478	(0.227)
Market loan amount in trillions (α_4)	0.972	(0.506)
Mean population in millions (α_5)	0.232	(0.239)
Mean household income in millions (α_6)	-1.393	(0.796)
	Mean	**Std. Dev.**
Lender-Market-Year fixed effect (ξ)	0.496	0.932

Note: The mean market size in the Midwest, Northeast, South, and West regions are $3.9bn, $6.1bn, $3.6bn, and $7.9bn respectively.

Table 4: Parameter estimates

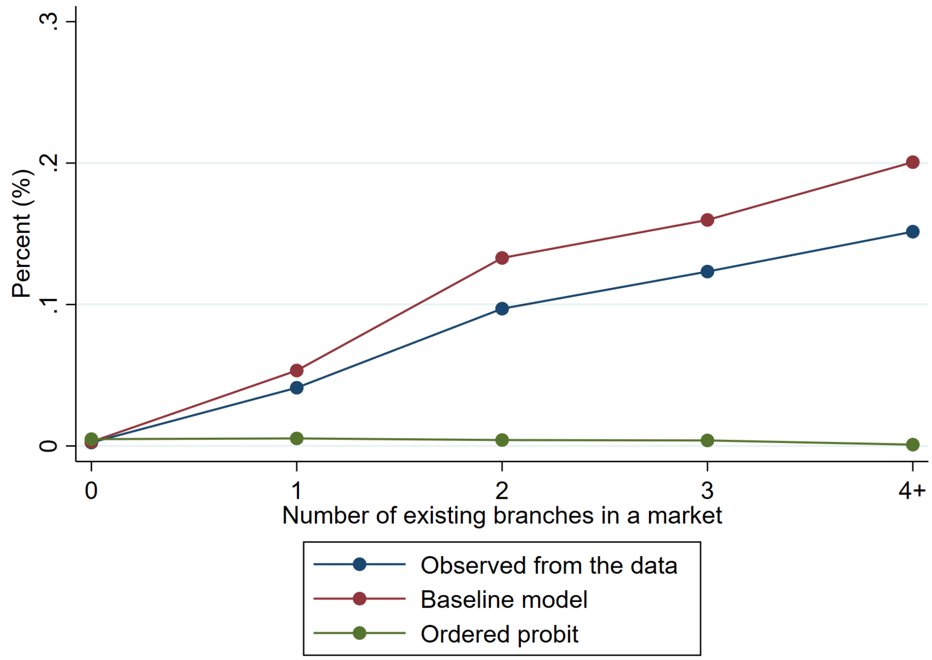

Figure 2: Model fit

million for lenders opening a new branch in a market in the Midwest. This fixed cost decreases to \$11.6 million when the lender already has a footprint of four or more branches in the market. Furthermore, the fixed cost is highest in the West and Northeast regions. However, across all regions, the fixed cost for lenders with 0 branches is 1% of the average market size in the region. This fixed cost decreases to 0.25% of the average market size for lenders with four or more pre-existing branches in a single market. This scaling of the fixed cost to market size shows that the fixed costs estimated include factors other than fixed costs. It includes other fixed costs such as fixed operating costs and regulatory costs. Moreover, the declining trend in fixed costs across all regions underscores the role of economies of density when lenders venture to establish multiple branches.

Economies of density can manifest in various forms, such as the expenditures associated with advertising and management. These expenditures may have cost efficiencies when branches within a network are nearby. Furthermore, economies of density for shadow banks can arise when there are fixed costs associated with orchestrating funding options, such as the capital requisites for opening a line of credit. Additionally, specific lender-investor relationships may be important for selling originated loans. For example, correspondent lenders must abide by a stipulated underwriting standard with their wholesale counterparts.

Using the ordered probit and estimated structural model, we compare the predicted

	U.S. wide	Midwest	Northeast	South	West
Counterfactual: Total number of branches					
Data	12112	2163	1520	4743	3686
Baseline model	12324	2181	1533	4800	3809
Model w/o cannibalization	12868	2277	1660	4939	3990
Model w/o business-stealing	13105	2344	1839	4940	3981
Model w/o cannibalization or business-stealing	13602	2426	1955	5059	4160
Percentage change in the number of branches from the baseline model					
Model w/o cannibalization	4.4%	4.4%	8.3%	2.9%	4.8%
Model w/o business-stealing	6.3%	7.5%	20.0%	2.9%	4.5%
Model w/o cannibalization or business-stealing	10.4%	11.2%	27.5%	5.4%	9.2%

Table 5: Number of total shadow bank branches by counterfactual for U.S. wide and regions

probability of a lender opening a new branch, given a certain number of pre-existing branches in a market, with the observed probability for lenders with the same number of branches. Figure 2 shows the estimated average predicted probabilities for the ordered probit and the structural model. For comparison, we include the observed probability of a lender opening a new branch in a market based on their pre-existing number of branches. We calculate the observed probability by identifying lenders with a specified number of pre-existing branches in a market, counting the number of newly opened branches within this group, and then dividing that count by the total number of lenders in this group.

Results from Figure 2 show that the ordered probit model encounters challenges in accurately capturing the data, especially when lenders have one or more pre-existing branches in the market. Furthermore, the ordered probit indicates a diminishing likelihood of opening a new branch as the count of existing branches increases. In contrast, the structural model's predicted probabilities more closely align with the observed data patterns. Beyond numerical proximity, this estimated structural model shows an increasing trend in predicted probabilities as a lender's number of pre-existing branches in the market increases.

6.1 Counterfactual analysis: cannibalization and business-stealing decomposition

To undertake the counterfactuals, wherein cannibalization and business-stealing effects are nullified, we re-estimate the model, assigning a zero value to the appropriate parameter. Specifically, the market share equation incorporates both cannibalization and business-stealing effects. Hence, to neutralize the cannibalization channel, we assign a value of zero to the parameter α_2, and to neutralize the business-stealing effect, we assign the parameter α_3 a value of zero. After the parameter modification, we re-estimate the profits and probability of a lender opening a new branch in a market. By omitting the associated effect, the model effectively adds the reduction in profit due to the presence of the respective channel.

The counterfactual analysis undertaken nationally, as shown in Table 5, shows that relative to the baseline model, an environment without cannibalization effects, i.e. the parameter α_2 is set to zero, the branch count increases by 4.4% (equivalent to 544 branches). In comparison, in an environment without business-stealing effects, i.e. the parameter α_3 is set to zero, the increase in the branch count is 6.3% (781 branches). This finding shows a larger business-stealing impact on the number of new branch openings relative to the cannibalization effect. Additionally, eliminating both effects increases the establishment of new branches by 10.4%.

The effect of cannibalization and business-stealing effects is dependent upon market structure. As presented in Table 5, counterfactuals undertaken at the U.S. regional level show these effects are largest in the Northeast region. In the Midwest and Northeast regions, the business-stealing effect is larger than the cannibalization effect in magnitude. In comparison, in the South and West regions, the impact of both effects on branch establishments appears comparable.

To understand the cannibalization effect, first note that the probability of opening a branch diminishes when a lender already has pre-existing branches in the market. Thus, in the counterfactual, where the cannibalization effect is muted compared to the baseline model, lenders with many branches in the market exhibit a higher difference in the probability of opening a new branch. Consequently, the emergent market structure, post the muting of the cannibalization effect, depends on the number of lenders having a high branch presence. For instance, the Northeast region has the highest ratio of such lenders. 9.5% of lenders with at least one branch in the Northeast region operate four or more branches within the market. In this environment, the business-stealing effect from rival branches persists, leading to a feedback mechanism where a lender opens a branch based on their belief that their rivals will also open a branch.

Removing the business-stealing channel sets the coefficient α_3 equal to zero in the profit equation. Given that α_3 was previously negative, this exclusion increases variable profit. The overall number of market rival (shadow and traditional bank) branches affects the magnitude of this increase. For instance, the Northeast region averages 1439 rival (shadow and traditional bank) branches, whereas the South region averages 467 rival (shadow and traditional bank) branches.

From the data, Figure 3 shows the calculated probability of a lender opening a new branch, grouped by lenders having a certain pre-existing number of branches within a market. It also includes the average predicted probability of a lender opening a branch by the

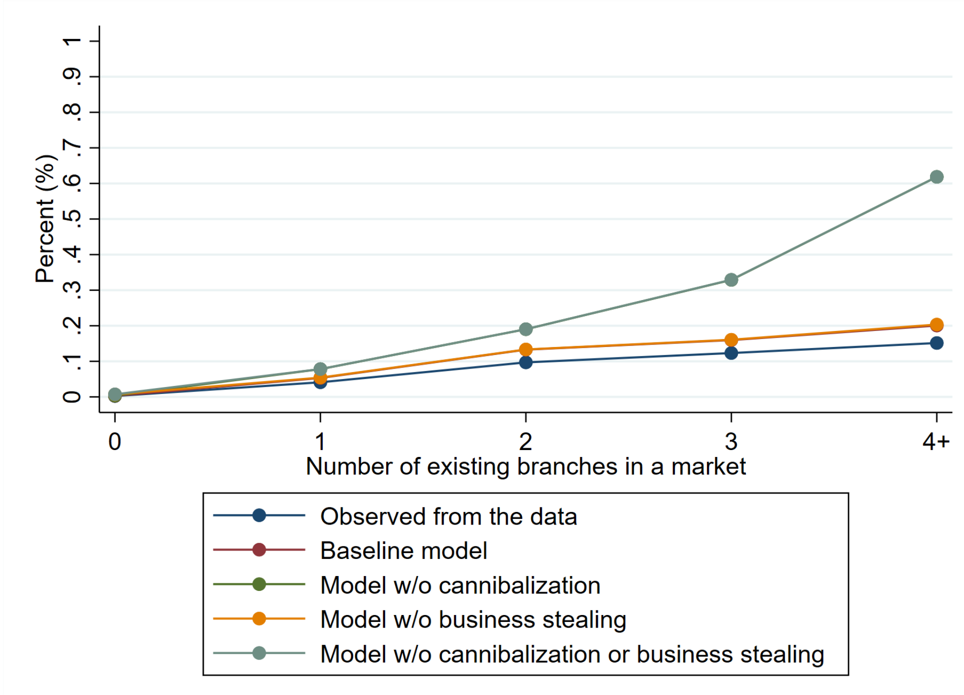

Figure 3: Probability of opening a branch in a market, by the number of branches a lender has in a market

Number of own branches in a market	0	1	2	3	4+
Observation count	190960	5030	1185	365	460
Probability of opening a branch					
Data	0.0025	0.0412	0.0970	0.1233	0.1515
Baseline model	0.0029	0.0533	0.1329	0.1598	0.2007
Model w/o cannibalization	0.0034	0.0781	0.1902	0.3289	0.6186
Model w/o business-stealing	0.0070	0.0543	0.1334	0.1607	0.2034
Model w/o cannibalization or business stealing	0.0072	0.0783	0.1904	0.3293	0.6186
Total number of branches opened					
Data	468	207	115	45	70
Baseline model	553	268	158	58	92
Model w/o cannibalization	650	393	225	120	285
Model w/o business-stealing	1328	273	158	59	94
Model w/o cannibalization or business stealing	1384	394	226	120	285

Table 6: Probability of opening a new branch and the number of new branches, by the number of branches a lender has in a market

groupings in a market across various counterfactuals: the baseline model (where both cannibalization and business-stealing effects are active), a scenario without the cannibalization effect, a scenario without the business-stealing effect, and a scenario without both effects. Table 6 presents the probability values for the data, baseline model, and counterfactuals, as shown in Figure 3. The lower section of Table 6 calculates the number of new branches for each group of lenders. To compute these values, we multiply the observation count by the probability that a lender will open a branch.

The analysis reveals that muting the business-stealing channel has comparable impacts on the probability of lenders opening a new branch across the different groups where lenders are categorized. In comparison, the muting of the cannibalization channel displays a larger impact depending on the number of pre-existing branches. For instance, relative to the baseline model, lenders with a single branch in a market exhibit an increase in the probability of opening an additional branch by 2.48%. This result translates to 125 new branches. In comparison, the likelihood of opening a new branch is higher for lenders with four or more branches in a market. This increase, quantified at 41.9%, indicates the opening of 193 new branches across the United States. These results are qualitatively similar when this analysis is undertaken by U.S. regions, as shown in Appendix 7.

7 Conclusion

Despite technological advancements that seemingly diminish the necessity of physical branches, data shows that shadow banks have a tangible branch presence in the United States. The empirical findings show that an additional branch increases market share, albeit with diminishing returns due to cannibalization effects. The lender's market share is also negatively impacted by their rivals' branch presence. An ordered probit model shows that a lender's branch presence in a market negatively impacts their probability of opening a new branch, and their rivals' branch presence has a negative but statistically insignificant impact.

The empirical findings show that while cannibalization and business-stealing effects exist, they do not account for observed branching patterns, particularly in scenarios where shadow banks have pre-existing branches in a market. The paper constructs a structural model based on the framework of a many-player game with simultaneous moves under incomplete information. This structural model estimates a flexible specification of fixed costs to match branching patterns in the data more accurately. This model is then used to study the cannibalization and business-stealing effects on branching decisions. The counterfactual analyses reveal the sensitivity of branch expansion to the presence of cannibalization and business-stealing effects, both nationally and regionally.

References

Agarwal, Sumit and Robert Hauswald (2010). “Distance and Private Information in Lending”. In: *The Review of Financial Studies*.

Aguirregabiria, Victor, Robert Clark, and Hui Wang (2016). “Diversification of Geographic Risk in Retail Bank Networks: Evidence from Bank Expansion after the Riegle-Neal Act”. In: *RAND Journal of Economics*.

— (2020). “The Geographic Flow of Bank Funding and Access to Credit: Branch Networks, Local Synergies and Competition”. Working paper.

Barrot, Jean-Noël and Julien Sauvagnat (2016). “Input specificity and the propagation of idiosyncratic shocks in production networks”. In: *The Quarterly Journal of Economics*.

Berger, Allen N. and Astrid A. Dick (2007). “Entry into Banking Markets and the Early-Mover Advantage”. In: *Journal of Money, Credit and Banking*.

Berger, Allen N., John H. Leusner, and John J. Mingo (1997). “The efficiency of bank branches”. In: *The Journal of Monetary Economics*.

Bos, Jaap WB and Clemens JM Kool (2006). “Bank efficiency: The role of bank strategy and local market conditions”. In: *Journal of Banking and Finance*.

Buchak, Greg and Adam Jørring (2021). “Do Mortgage Lenders Compete Locally? Implications for Credit Access”. Working paper.

Buchak, Greg, Gregor Matvos, et al. (2018). “Fintech, regulatory arbitrage, and the rise of shadow banks”. In: *Journal of Financial Economics*.

Bui, Long (2023). “Shadow bank branch networks”. Working Paper.

Clark, Robert, Ignatius Horstmann, and Jean-FranÃ§ois Houde (2021). “Hub and spoke cartels: Theory and evidence from the grocery industry”. Working Paper.

Cortés, Kristle Romero and Philip E Strahan (2017). “Tracing out capital flows: How financially integrated banks respond to natural disasters”. In: *Journal of Financial Economics*.

Degryse, Hans and Steven Ongena (2005). “Distance, Lending Relationships,and Competition”. In: *The Journal of Finance*.

Ergungor, Ozgur Emre (2010). “Bank Branch Presence and Access to Credit in Low- to Moderate-Income Neighborhoods”. In: *Journal of Money, Credit and Banking*.

Felici, Roberto and Marcello Pagnini (2008). “Distance, Bank Heterogeneity and Entry in Local Banking Markets”. In: *The Journal of Industrial Economics*.

Fuster, Andreas et al. (2019). “The Role of Technology in Mortgage Lending”. In: *The Review of Financial Studies*.

Gao, Pengjie, Paul Schultz, and Zhaogang Song (2017). "Liquidity in a Market for Unique Assets: Specified Pool and To-Be-Announced Trading in the Mortgage-Backed Securities Market". In: *The Journal of Finance*.

Gilje, Erik P, Elena Loutskina, and Philip E Strahan (2016). "Exporting liquidity: Branch banking and financial integration". In: *The Journal of Finance*.

Goolsbee, Austan and Chad Syverson (2008). "How Do Incumbents Respond to the Threat of Entry? Evidence from the Major Airlines". In: *The Quarterly Journal of Economics*.

Gowrisankaran, Gautam and John Krainer (2011). "Entry and pricing in a differentiated products industry: evidence from the ATM market". In: *The RAND Journal of Economics*.

Guzman, Jorge and Scott Stern (2020). "The state of American entrepreneurship: New estimates of the quantity and quality of entrepreneurship for 32 US States, 1988-2014". In: *American Economic Journal: Economic Policy*.

Ho, Katherine and Joy Ishii (2011). "Location and competition in retail banking". In: *International Journal of Industrial Organization*.

Holmes, Thomas J (2011). "The Diffusion of Wal-Mart and Economies of Density". In: *Eonometrica*.

Igami, Mitsuru and Nathan Yang (2016). "Unobserved heterogeneity in dynamic games: Cannibalization and preemptive entry of hamburger chains in Canada". In: *Quantitative Economics*.

Jiang, Erica Xuewei (2019). "Financing Competitors: Shadow Banks' Funding and Mortgage Market Competition". Working paper.

Nguyen, Hoai-Luu Q. (2019). "Are Credit Markets Still Local? Evidence from Bank Branch Closings". In: *American Economic Journal: Applied Economics*.

Relihan, Lindsay (2018). "Branches in Local Mortgage Markets". Working paper.

Rysman, Marc, Robert M. Townsend, and Christoph Walsh (2023). "Bank Branching Strategies in the 1997 Thai Financial Crisis and Local Access to Credit". Working paper.

Scharfstein, David and Adi Sunderam (2016). "Market Power in Mortgage Lending and the Transmission of Monetary Policy". Working paper.

Seim, Katja (2006). "An Empirical Model of Firm Entry with Endogenous Product-Type Choices". In: *The RAND Journal of Economics*.

Vickery, James and Joshua Wright (2013). "TBA trading and liquidity in the agency MBS market". In: *Economic Policy Review*.

Zheng, Fanyin (2016). "Spatial Competition and Preemptive Entry in the Discount Retail Industry". Working Paper.

Appendix A

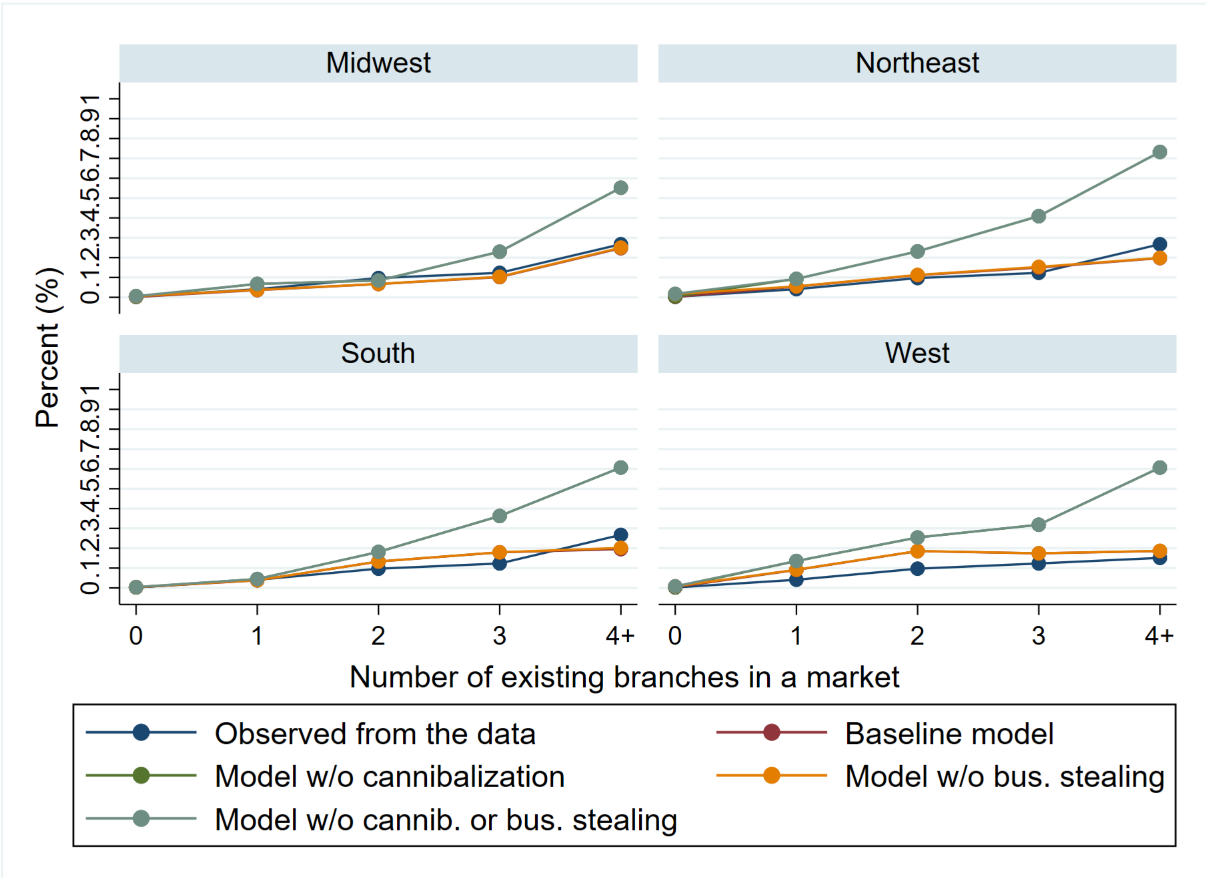

Figure 4: Probability of opening a branch in a market, by the number of branches a lender has in a market and region

Number of own branches	0	1	2	3	4
Region: Midwest					
Observation count	41879	952	225	70	74
Data	0.0025	0.0412	0.0970	0.1233	0.1515
Baseline model	0.0019	0.0362	0.0669	0.1018	0.2475
Model w/o cannibalization	0.0027	0.0675	0.0846	0.2294	0.5523
Model w/o business-stealing	0.0058	0.0376	0.0673	0.1037	0.2504
Model w/o cannibalization or business stealing	0.0062	0.0677	0.0849	0.2301	0.5523
Region: Northeast					
Observation count	21882	535	147	40	76
Data	0.0025	0.0412	0.0970	0.1233	0.1667
Baseline model	0.0023	0.0543	0.1114	0.1499	0.1975
Model w/o cannibalization	0.0041	0.0929	0.2313	0.4086	0.7325
Model w/o business-stealing	0.0163	0.0552	0.1129	0.1533	0.1989
Model w/o cannibalization or business stealing	0.0175	0.0930	0.2314	0.4091	0.7325
Region: South					
Observation count	84867	2222	472	146	133
Data	0.0025	0.0412	0.0970	0.1233	0.1515
Baseline model	0.0030	0.0382	0.1330	0.1795	0.1957
Model w/o cannibalization	0.0032	0.0450	0.1815	0.3624	0.6067
Model w/o business-stealing	0.0046	0.0387	0.1332	0.1797	0.2019
Model w/o cannibalization or business stealing	0.0046	0.0450	0.1817	0.3628	0.6068
Region: West					
Observation count	42332	1321	341	109	177
Data	0.0025	0.0412	0.0970	0.1233	0.2673
Baseline model	0.0040	0.0908	0.1856	0.1743	0.1862
Model w/o cannibalization	0.0042	0.1356	0.2542	0.3185	0.6064
Model w/o business-stealing	0.0080	0.0919	0.1861	0.1745	0.1868
Model w/o cannibalization or business stealing	0.0082	0.1359	0.2543	0.3187	0.6064

Table 7: Probability of opening a new branch, by the number of branches a lender has in a market and region

Number of own branches	0	1	2	3	4
Region: Midwest					
Observation count	41879	952	225	70	74
Data	103	39	22	9	11
Baseline model	80	34	15	7	18
Model w/o cannibalization	111	64	19	16	41
Model w/o business-stealing	241	36	15	7	19
Model w/o cannibalization or business stealing	260	64	19	16	41
Region: Northeast					
Observation count	21882	535	147	40	76
Data	54	22	14	5	13
Baseline model	51	29	16	6	15
Model w/o cannibalization	89	50	34	16	56
Model w/o business-stealing	356	30	17	6	15
Model w/o cannibalization or business stealing	384	50	34	16	56
Region: South					
Observation count	84867	2222	472	146	133
Data	208	91	46	18	20
Baseline model	252	85	63	26	26
Model w/o cannibalization	271	100	86	53	81
Model w/o business-stealing	389	86	63	26	27
Model w/o cannibalization or business stealing	391	100	86	53	81
Region: West					
Observation count	42332	1321	341	109	177
Data	104	54	33	13	47
Baseline model	170	120	63	19	33
Model w/o cannibalization	179	179	87	35	107
Model w/o business-stealing	341	121	63	19	33
Model w/o cannibalization or business stealing	348	180	87	35	107

Table 8: Total number of new branches, by the number of branches a lender has in a market and region