Locational Rents and Deposit Franchise Value: Uncovering the Role of Distance in Deposit Pricing

ABSTRACT

Using novel foot traffic data from millions of cell phone devices across the U.S., I study the extent to which the distance between a bank and its customers affects the pricing (interest rates) of its deposit products. Instrumenting the distance of the customers with regional broadband access status, I find substantial evidence for spatial price discrimination in the deposit market. The distance of the customers from a branch negatively affects the price of its deposit products; this price-distance relationship is stronger in a highly concentrated market, consistent with the exercise of market power. Cross-sectional analysis reveals that this negative effect of the distance is present for time deposits, but not for transactional deposits. This effect is more pronounced for small banks and intensifies with the maturity period of the deposit products. Furthermore, paying lower rates for deposits sourced from distant customers translates into higher bank profitability. These results provide evidence of the presence of locational rents in the deposit markets that contribute to a bank's deposit franchise value.

JEL classification: E30, E44, G21, G51, L11

Keywords: Distance, Deposit Pricing, Market Power, Multi-Product Pricing, Cross-Selling, Big Data.

1 Introduction

There is a widespread trend of consolidation in the banking industry for the last couple of decades (Berger et al. [2000]). Accordingly, the number of brick-and-mortar bank branches in the U.S. has declined steadily for more than a decade[1] (Keil and Ongena [2020], Di and Pattison [2020], Petersen and Rajan [2002]). The number of bank branches in a region is negatively related to the physical distance from banks' customers[2]. Though the number of physical branches is decreasing, trends in the banking sector and findings of recent research[3] indicate that distance still plays an important role in the banking market (Drexler et al. [2019]). The impact of geographic proximity to the customer on lending has been examined extensively in the literature (Degryse and Ongena [2005], Di and Pattison [2020], Nguyen [2019], and Herpfer, Mjøs, and Schmidt [2022] etc.). Though most of the bank value comes from the liability or deposit side rather than the credit side (Egan, Lewellen, and Sunderam [2022]), so far there is no empirical work that studies the effect of the distance in the deposit market.

Retail deposits are important for banks as they provide a low-cost, stable source of funds and generate fee income (Clark et al. [2007]). Retail deposits constitute more than 70% of bank liabilities (Drechsler et al. [2017]) and a large portion of a bank's cost of capital is its retail deposit interest rate (Granja et al. [2022]). Therefore, it is important to see how the distance impacts the retail deposit price as the cost of capital is one of the key factors in determining bank's profitability. The impact of the physical distance on the deposit market is left mostly unexplored in the academic research because of the

[1]The number of banks in the United States has been declining for more than twenty years. But, the number of branches had been increasing till 2009, then it started to fall. It continues to decline till today. In appendix table-A1, the number of banks and branches in the U.S. for last couple of decades are reported.

[2]When branch banks shutter, people sometimes must travel farther to conduct their banking business. In some cases, customers must drive significant distances or forgo financial transactions. (Source: https://www.stlouisfed.org/)

[3]Though number of banks is decreasing for more than 10 years, physical offices remain a vital channel through which FDIC-insured institutions deliver financial services to their customers and per capita density of banking offices also remain high (Breitenstein and McGee [2015]).

lack of data availability. My paper aims to fill this gap in literature, and I circumvent this data problem by using foot traffic data for the customers of the bank branches obtained through tracking millions of cell-phone devices. So, exploiting this extensive geolocation data, I systematically examine the impact of distance on deposit pricing. My paper asks the following research questions: What is the impact of distance on deposit price? Does the effect differ between transactional deposit and savings deposit? Why do some banks choose to increase their branch network? Does the branch network help them to increase profitability through reducing distance?

To get the distance of a branch's customers, this study makes use of the bank branch customers' footprint data across U.S. from January 2018 to January 2021, sourced from around forty-five million smartphone devices. Around 70% bank branches of FDIC's Summary of Deposit (SOD) data are observed in this geolocation data. Besides that, the comparative graphs show that the observed branches for foot traffic in the geolocation data represents the branches of the Summary of Deposit data quite well. Using this granular data, I am able to get the distance for the customers of bank branches for my empirical analysis. The summary statistics of the distance variable provide some rare unique perspectives about the banking in the U.S. It reveals that on an average a customer is 10.183 KM away from his/her corresponding banking branch. It also discovers the time variation in the distance of the customers as we can observe that the distance declines gradually from 10.586 KM to 9.495 KM during our sample periods, even during the non-covid periods. Another interesting point is that the distance of the customers is higher for large banks[4] and rural branches relative to small banks[5] and urban branches respectively. Moreover, it also points out that large branch network helps banks to attract distant customers.

My first baseline tests examine the effect of the distance of the customers of a branch

[4] The banks with total assets of $100 billion or more. (Source:https://www.federalreserve.gov/supervisionreg/large-financial-institutions.htm)

[5] The banks with total assets of less than $1.322 billion. (Source:https://www.federalreserve.gov/newsevents/pressreleases/bcreg20201217a.htm)

on the price of the deposit products of that branch. The results of my panel regressions show that the distance of the customers of a branch has a negative effect on the price of the deposit products of that branch. I find that an increase of one standard deviation in the distance of the customers of a branch reduces the prices for that branch's deposit products by 0.54 basis points. These results are statistically significant and robust to controlling for different measures of competition (numbers of branches in a county, distance to the nearest competing branch and deposit concentration in the county). These aggregate results provide evidence for the spatial price discrimination of the banks in the deposit market.

Then, I conduct a series of robustness tests to show that my baseline results are robust under different scenarios. First, there were some disruptions in the financial market because of the Covid and the supply of Stimulus by the government. So, I conduct the tests for non-covid years to ensure that my results are not affected by these disruptions. Second, county boundaries are not always adequate confines for a local economy and often reflect political boundaries rather than an area's local economy. People may live in one county and go to another county for job and maintain bank account with the banks near their working place. To address this concern, I rerun my tests for Commuting Zones (CZs)[6] instead of for counties. To better delineate local economies, these commuting zones are used in many Economics literature. (e.g., Foote, Kutzbach, and Vilhuber [2021]; Atkin [2016] and Dorn, Hanson, Majlesi, et al. [2020]). Finally, the location of bank branches which shapes the distance of the customers and the prices set by the banks for those branches may be endogenously determined by the banks. To address this concern for endogeneity, I take advantage of the existing digital divide in U.S. and run IV regressions using high speed internet access status of a region as an instrumental variable for the distance of the customers. The assumption of using this IV is that the access to high-speed internet reduces

[6] Commuting Zones reflect the local economy where people live and work. Commuting zones are bigger than the counties but smaller than the states. In appendix-B, I provide some examples of commuting zones that illustrate how counties are situated across different commuting zones and how one commuting zone can lay in two states.

the demand for physical banking thus decrease the observed distance between customers and bank branches but it has no direct effect on the deposit pricing. The results of all the robustness tests and IV regressions are similar like our baseline regressions and provide support for my initial findings.

Next, I perform the cross-sectional analysis for regions with different levels of competition, for various types of banks and for different deposit products. Regression results for the distance and local competition reveal that the spatial price discrimination in the deposit market exists when banks have market power. Analyzing for types of banks, I find that the effect is more pronounced for small banks and the branch network size helps to augment the impact. Then, I find that the effect of the distance of the customers on deposit pricing is present for time deposits, but not for transactional deposits. This finding exhibits the importance of the distance in creating banks' value as Egan et al. [2022] observes that the main driver of bank value is time deposits compared to transactional deposits.

In addition, I also study the effect of the distance on the deposit volume and profitability. Results from both branch and bank level regressions show that the distance has a negative impact on deposit volume. Distance plays a dual role in the deposit market. On the one hand, it reduces cost of acquiring deposits through providing lower rates to the distant customers which increases profitability. On the other hand, the lower rate decreases overall deposit volume which has a negative impact on profitability. Which of these channels dominates is ultimately an interesting question. The result of the regression on Net Interest Margin (NIM) suggests that banks actually increase net profitability through offering lower price to the distant depositors.

This paper contributes to three strands of the literature. First, it adds to the literature on the impact of physical proximity in banking. Large numbers of theoretical and empirical works explore the effect of geographical distance on the lending. Previous papers have confirmed that the distance between borrowers and their banks is increasing (Petersen and Rajan [2002] and DeYoung et al. [2011]); there is a cyclical component in the lending distance (Granja, Leuz, and Rajan [2022]); the distance decreases during the crisis period as bank reduces distant loan during the downturn (Degryse, Matthews, and Zhao [2018] and Presbitero, Udell, and Zazzaro [2014]) and the distance of the borrowers is higher for the large banks (Berger et al. [2005]). The impact of distance on credit outcome is also documented in previous empirical works which find that distance is negatively related with loan rate (Herpfer et al. [2022] and Degryse and Ongena [2005]) and overall credit quantity (Nguyen [2019]); positively related with default probabilities (DeYoung, Glennon, and Nigro [2008] and Loutskina and Strahan [2011]) and borrower proximity reduces transaction and monitoring costs through facilitating the collection of soft information (De Haas, Ongena, Qi, and Straetmans [2021] and Agarwal and Hauswald [2010]), which is mainly collected through close interactions between bankers and borrowers (Liberti and Petersen [2019]). Though relative to lending, the liabilities/deposits are more strongly associated with bank market value (Egan et al. [2022]), there is a lack of research that empirically examines the impact of the distance in the deposit market. My paper fills that gap in the literature through revealing that the distance of the customers is negatively associated with deposit prices.

Second, my paper contributes to the broad literature of price discrimination and market power. As a microeconomic pricing strategy, price discrimination is explored extensively in the economics literature (Phlips [1983], Varian [1985], and Bonatti and Cisternas [2020] etc.). It is also thoroughly studied in the banking literature specifically in the credit market. Banks differentiate the price in the credit market based on different factors such as regional characteristics (Meyer [1967]); the distance of the borrower from the branch (Degryse and Ongena [2005]); the credit risk of the borrower (Magri and Pico [2011]); whether borrower is experienced and confident (Woodward and Hall [2012]); bargaining power of the borrower (Allen, Clark, and Houde [2014]) and whether a borrower is a new or existing borrower

(Ongena, Paraschiv, and Reite [2021]). Price discrimination in banking is related with market power since banks usually offer differentiated price to gain some premium when they have reasonable market power (Gary-Bobo and Larribeau [2004]) and Drechsler et al. [2017] shows that banks also exercise market power in the deposit market. My paper uncovers the existence of the spatial price discrimination in the deposit market and shows that this discrimination exists when banks have market power.

Finally, my findings add to the bank's branch network literature. Previous works in this area show how branch network changes because of interstate and intrastate banking and bank branching regulatory modifications (Jayaratne and Strahan [1996], Huang [2008], Jayaratne and Strahan [1998], Rice and Strahan [2010], Beck et al. [2010], and Keil and Müller [2020]). Branch network facilitates information sharing within banks (De Haas, Ongena, Qi, and Straetmans [2021]); stabilizes financial system through enabling portfolio diversification (Hubbard [2001]); facilitates local market competition (Carlson and Mitchener [2005], Puri and Rocholl [2008], Carlson and Mitchener [2009], Kuehn [2018]) and plays an important role in integrating the lending market and ensuring access to credit (Gilje, Loutskina, and Strahan [2016] and Nguyen [2019]). My paper contributes to this literature through showing that branch network helps banks to become more profitable through providing less interest to the distant depositors.

The remainder of this paper is structured as follows. In section-2, I develop the main testable hypotheses for the empirical analysis. Section-3 of the paper gives a brief description of the data. In this part, I also present summary statistics of the key variables of my analysis. Section-4 examines the effect of the distance of the customers on the price of deposit products. Then, section-5 explores the relationship between distance and deposit price for markets with different levels of concentrations. This part also includes cross sectional analysis for different types of deposit products. Section-6 includes the robustness tests for my analysis and the instrument variable regressions to address the concern for endogeneity. In section-7, I explore the relationship between distance, deposit volume and profitability. Finally, a brief conclusion of the findings of this paper is given in section-8.

2 Hypothesis Development

Distance plays an important role in the banking as it is strongly associated with relationship banking. However, the expected effect of the distance on pricing of the banking products is not conclusive from earlier works. Radecki [1998] shows that many banks set uniform rates for both deposits and retail loans across an entire state or broad regions of a large state. Examining the Bank Rate Monitor data, Heitfield [1999] and Park and Pennacchi [2008] confirms Radecki's finding that larger banks often set uniform rates across cities. Banks price uniformly if they cannot observe customers' locations or are prevented from charging different prices to different customers (Degryse and Ongena [2005]). So, based on uniform pricing theory, the distance of the customers should not have any effect on the price of the deposit products.

Alternatively, according to transportation cost theory, closer customers will bear higher transportation costs if they try to do banking with a competing bank as they are located further away relative to the existing bank. In other words, as customers usually bear the transportation expense, the opportunity cost of doing banking with other competing banks rather than the nearest bank is higher for the closer customers (Lederer and Hurter Jr [1986]). So, banks can capitalize on the increased opportunity cost for the closer customers and can charge some rents based on that. Empirically, Degryse and Ongena [2005] and Herpfer, Mjøs, and Schmidt [2022] show that banks take advantage of the higher transportation cost through charging higher prices to the nearest borrowers. Accordingly, in the deposit market, banks need to compensate distant customers for the extra transportation cost for maintaining business with them. So, transportation cost theory posits that banks can engage in price discrimination based on the location of the depositors and can provide lower rate to the nearest depositors.

On the other hand, Elliehausen and Wolken [1997] and Boyd, Leonard, and White [1994] reveals from two different survey results that customers can choose a bank for a lot of other factors besides proximity, which is broadly related with transaction cost, such as: ease of conducting business, services offered, reputation, quick service, friendliness of employees, and even based on some personal relationship etc. In the deposit market, banks set the deposit spread through exercising market power to maximize profit (Drechsler, Savov, and Schnabl [2017]). Now, when a bank observes that depositors come to a branch for some other factors rather than proximity, it indicates that the specific branch has some sort of market power and banks can take advantage of it by offering a lower deposit rate to the distant customers.

Besides that, banks gain from not only selling both deposit and loan products but from selling both to the same customer (Basten and Juelsrud [2022] and Petersen and Rajan [1994]). Herpfer, Mjøs, and Schmidt [2022] shows that lower travel time increases the likelihood of initiating a new banking relationship and this motivates banks to accept lower deposit spreads from likely future borrowers (Basten and Juelsrud [2022]). So as like as marker power theory, cross selling motive suggests that banks should offer higher price to the closest depositors for the prospect of getting future borrowers.

H_0: (Uniform Pricing Hypothesis): The distance of the customers has no effect on the price of the deposit products as banks set price uniformly.

H_1a: (Transportation Cost Hypothesis): Banks may offer higher price to the distant depositors to compensate them for the higher transportation cost.

H_1b: (Market Power/ Cross Selling Hypothesis): Banks provide lower price to the distant depositors to increase their profit margin through exercising market power.

I test these hypotheses using a combination of datasets including novel foot traffic data with an appropriate methodology. From the results of my empirical analysis, I find evidence for the second alternative hypothesis. It means banks price differentiate in the deposit market through offering lower price to the distant customers. Further analysis also reveals that this discrimination is the result of banks' exercising market power. In the next section, I describe my sample selection procedures, the variable definition and present the summary statistics for my key variables.

3 Data Description and Variable Formation

In this section, I describe the major data sources and the definition of the key variables used in my empirical analysis.

3.1 Sample Selection & Data Sources

My data set combines information from a variety of sources including Foot-traffic Data, RateWatch Data, Federal Deposit Insurance Corporation's (FDIC's) Summary of Deposit (SOD) Data, FDIC's Call Report Data, FCC's Broadband Access Data, Rural-Urban Continuum Data, U.S. Census Data and U.S. Bureau of Labor Statistics Data. I get the price of different deposit products from Rate Watch Data, the market share of different banks from SOD data, bank-related variables, and controls such as ROA and asset size from FDIC's Call Report Data, rural urban classification data from Rural-Urban Continuum and other regional characteristics data from U.S. Census Data and U.S. Bureau of Labor Statistics Data. My sample period is determined by the joint availability of data in Foot-Traffic Data and RateWatch Data. The sample period starts in 2018 as Foot-Traffic Data is available form 2018 and ends in 2021 as the Rate Watch data is available only till 2021.

3.1.1 Foot-traffic Data

I observe foot traffic of the customers of different bank branches from beginning in January 2018 to January 2021. To observe the foot traffic pattern of the customers of different banks, I use anonymized geolocation data derived from approximately forty-five million smartphones for around 10 million Point-of-Interest (POI) locations in the U.S., provided by a commercial company that provides Point-of-Interest (POI) and location-based services data in the U.S., United Kingdom, Canada and 30 other countries. The firm observes human mobility patterns by partnering with smartphone apps that get opt-in consent from the users to record their location. However, the dataset does not contain any personal identifiable information, such as usernames, race, home address, the IP address of the mobile devices to ensure the privacy of the users.

[Insert Table-1 Here]

The geolocation data that I use in this paper has two parts: Point-of-Interest (POI) dataset and Foot-traffic dataset. Point-of-Interest (POI) dataset provides different attributes for different Point-of-Interests (in my case, bank branches), for example geospatial coordinates, addresses, brand affiliation, open hours, and locational categories etc. In panel A of table-1, I show the most common banks in the summary of deposits (SOD) data. In panel B, I provide the number of Point-of-Interest (POI) or bank branches of the most 10 common brands or banks that are observed in the geolocation data. We can find that the most observed banks are quite similar in the two datasets. In figure-1, I show the major banks that are observed in the geolocation data. This figure shows that the most frequent words among the word corpus of banks name covered in this dataset and as expected, the large banks are observed more as they have more branches across united states. In figure-2(a) and 2(b), I point out the locations of all bank branches of FDIC's Summary of Deposit (SOD) data and the locations of all bank branches observed in the foot traffic data in 2019 respectively. The relative graphs show that the observed branches in the geolocation data represents the branches of the Summary of Deposit data very well.

[Insert Figure-1, 2(a) and 2(b) Here]

To examine it further, in figure- 3(a), I illustrate the location of all bank branches in a sample county (Shasta County, CA) from summary of deposits (SOD) dataset for the year 2019. In 3(b), I point out the location of all bank branches in the Shasta County, CA for which customer footprint is observed through cellphone tracking in the year 2019. The figures visually demonstrate that the observed branches in the geolocation data is indeed a good representative of the summary of deposits (SOD) data. This geolocation data is compiled by the data provider firm through crawling open store locators on the web, using public APIs, and working in partnership with a third-party on additional data sources to fill in gaps. After collecting the data, a rigorous de-duping and merging process is used to make sure the dataset is clean for usage.

[Insert Figure-3(a) and 3(b) Here]

On the other hand, foot-traffic dataset provides the customers' mobility data. It provides insights into where people travel from to get to the specific place, and where else they go. The dataset is built by licensing aggregated and anonymized mobility data that has been sourced from mobile applications of the smartphone, and one smartphone is treated as one visitor/customer. A visit is counted if a mobile device in the data panel dwells for at least 4 minutes at a given POI. This dataset also provides visitor home census block group. To determine people's home census block group, the provider analyzes six weeks of data during nighttime hours (treating 6 p.m. to 7 a.m. as common nighttime) and assigns a home location for a mobile device. The foot-traffic data is vast as it is pulled from about forty-five million unique devices. In this paper, I use foot-traffic data on weekly basis as the deposit price data (Ratewatch) is also available on weekly basis. Foot-traffic dataset covers one week starting Monday and ending end of day on Sunday.

I merge the two geolocation datasets (Point-of-Interest (POI) dataset and Foot-traffic dataset) using Placekey. Placekey is a free, universal standard identifier for any physical place or location. It replaces a location's physical address and latitude-longitude data with a unique standard identifier. As it uses a unique identifier for each physical place on the earth, it solves the problem of linking locations by addresses that are spelled differently (e.g., 3000 July Street, Suite 2212 vs. 3000 July St., #2212) or by latitude-longitude geocodes that differ for having different decimal digits but refer to the same place. Each Placekey has two parts "What" and "Where". When both parts are combined, the result reads as What@Where (For example: zzw-222@8dx-6xt-52k is the placekey for Chase Branch @ 250 W State St, Baton Rouge, LA 70802).

The first three characters in the What Part (in our example: zzw) indicate the physical address. An address at "3000 July Street, Suite 2212" will have a different first three characters than "3000 July Street, Suite 2213". But "3000 July Street, Suite 2212" will have the same address encoding as "3000 July St., #2212" to consider common address formats. The next three characters in the What Part (in our example: 222) refers to the specific Point-of-Interest (POI). It changes if a new business opens at the same address of a previous business that closed. For example, if the ownership of a bank branch changes because of Merger & Acquisition, the two branches (old and new branch) will have the same address /first three character (in our example: zzw), but different points-of-interest/ next three characters (in our example: 234 instead of 222). So, they both will have unique placekeys. The Where Part is made up of three unique character sequences (in our example: 8dx-6xt-52k). Using Uber's open source H3 grid system, this part encodes a hexagonal region on the surface of the earth based on the latitude-longitude of the centroid of the Point-of-Interest (POI).

Through merging these two geolocation datasets, I can track the foot traffic of the customers of the different bank branches. In 4(a), I illustrate the customers' footprints of a specific branch [Chase Bank Branch (Dana Drive in Shasta County, CA)] in a random week (1st Week of February, 2019). We can find from which census block groups customers came to that branch during that week. This merged dataset also help us to see the competition between bank branches on weekly basis. In 4(b), I show the market network for the same Chase Bank Branch in the same week. The figure portrays that the branch shares territory with four other bank branches (Plumas Bank at Hilltop Dr., Cornerstone Community Bank at E Cypress Ave, Bank of America at East St. and Tri Counties Bank at Hilltop Dr.) in that specific week.

[Insert Figure-4(a) and 4(b) Here]

Using this merged dataset, I get the average distance of the branch customers from a branch. However, as the data does not capture every customer that visits a bank branch, one might argue that the findings can be attributed to the smartphone owners only. To address this concern, it can be noted that as of 2020 about 85% of American adults owned a smartphone[7]. Besides that, according to the analysis of user characteristics, the data provider firm posits that its data sample is representative of the U.S. population based on income characteristics, age, and demographics etc. Another concern about the average distance measure of the bank customers might be that most of the customers do not go to the physical branch as they conduct their banking business through online or mobile banking. But, FDIC's Survey of Household Use of Banking and Financial Services finds that branch visits is still prevalent even among people that used online or mobile banking as their primary method of account access[8]. According to Jefferies retail banking survey,

[7]The vast majority (97%) of Americans now own a cellphone of some kind. The share of Americans that own a smartphone is now 85%, up from just 35% in 2011. (Source: https://www.pewresearch.org/internet/fact-sheet/mobile/)

[8]In 2019, 79.9 percent of banked households that used mobile banking as their primary method visited

physical banking locations, which are still considered as important particularly by younger people, are a top factor in picking a new bank[9]. So, it shows that customers still prefer and do visit physical branches in person.

3.1.2 Deposit Product Price Data

For deposit pricing, I use a panel data set of interest rate of different deposit products from January 2018 to January 2021 at the branch level provided by RateWatch. RateWatch was established in 1989 and it collects thousands of data points from nearly 100,000 financial institution locations across the United States every week. RateWatch accumulates deposit rate from banks' branches all over the country via telephone, fax, e-mail, and also by scraping banks' web sites on weekly basis. Market research specialists also verify the last change date when calling contacts and the effective dates of faxes, emails, and websites[10]. Since both banks and regulators such as the FDIC use this dataset, it ensures that the quality of the data is great. The data are available on a weekly frequency from 2001 to 2021. I compare RateWatch's branch list to the list of branches in the FDIC's Summary of Deposits (SOD) data and find that its data set covers about 75%–85% of the market depending on the period. The dataset also reports whether a branch sets its own deposit rates (Rate-setter) or whether the branch uses rates that are set by another branch. For robustness check, I also run my baseline regressions only for rate-setter branches.

I merge the RateWatch data with the FDIC's Summary of Deposits (SOD) data using the FDIC's branch identifier. In this paper, I use mainly two types of deposit products: Transactional Deposits and Saving Deposits. Among these categories, I choose the products that are most commonly offered across all bank branches. For transactional deposit, I

a branch and 18.8 percent visited ten or more times. (Source:How America Banks: Household Use of Banking and Financial Services)

[9]Physical banking locations "are still viewed as important" and are still a top factor in picking a new bank, according to a new Jefferies retail banking survey. Respondents are still relying on physical locations, especially those age 18 to 34. (Source: CNBC)

[10]Source: https://www.spglobal.com/marketintelligence/en/campaigns/ratewatch

use checking deposit with minimum account size of $2500 ($2.5K Checking Accounts); for savings deposit, I use money market deposit accounts with minimum account size of $25,000 ($25K Money Market Accounts); and for time deposits, I use 12-month certificates of deposit with minimum account size of $10,000 ($10K 12-month CDs) following the existing literature (Drechsler, Savov, and Schnabl [2017], Cave et al. [2022] and Azar et al. [2022]). For uninsured time deposits, I use 12-month certificates of deposit with minimum account size of $250,000 ($250K 12-month CDs). For observing the impact of products' maturity period on the relationship of deposit price and the distance of the customers, I also consider 03-month / 12-month/ 24-month/ 36-month certificates of deposit with minimum account size of $10,000.

3.1.3 Branch Deposit data

I get the deposit volume and the geographic characteristics of all branches of depository institutions operating in the United States between 2018 and 2021 from the FDIC's Summary of Deposits (SOD) data [an annual survey of branch office deposits as of June 30 for all FDIC-insured institutions, including insured U.S. branches of foreign banks][11]. This dataset links each branch to its parent bank and contains information on the latitude-longitude, location, name, and deposit volume of all bank branches in the United States. I use this dataset to obtain data on the total number of banks' branches, to calculate the distance of the nearest competing branch from a specific branch using Haversine Formula[12] and to measure existing bank concentration in each county and in each commuting zones during my sample period following the existing literature (Cetorelli and Strahan [2006]; Drechsler et al. [2021] and Azar et al. [2022]). To be more specific, I calculate each county's Herfindahl-Hirschman Index (HHI_County) and each commuting zones' Herfindahl-Hirschman Index ($HHI\_CommutingZone$) based on the deposits held by each bank in a region in a specific

[11]Source: https://www.fdic.gov

[12]Source: https://en.wikipedia.org/wiki/Haversine_formula

year.

To merge geolocation data with FDIC's Summary of deposit (SOD) data, at first, I use Placekey's free Application Programming Interface (API) to generate the Placekey for all the bank branches in FDIC's Summary of deposit (SOD) dataset using geographical coordinates, location information. Then I merge both the dataset using Placekey and I am able to match 68.74% of the branches in the SOD data following this process. This is how, I construct a weekly panel dataset of physical branch visits for the bank customers at the branch level. Then, I merged this panel data with deposit product price data using the FDIC's branch identifier.

3.1.4 Bank level Data

For bank level data, I use the bank Reports of Condition and Income, also known as Call Reports, provided by the Federal Reserve Bank of Chicago. All balance sheet data for the banks in my sample come from this dataset, which banks file on quarterly basis. I use this dataset to categorize my sample into large banks vs small banks based on assets size, as well as to create some bank-related control variables such as Deposit Diversity, Liquidity, Return on Assets and Z-score. The merged dataset of Foot-traffic data, RateWatch data and FDIC's SOD data is at branch level. I match this bank-level call reports data to previously merged branch-level data using the FDIC Certificate ID, a unique number assigned to each depository institution by the Federal Deposit Insurance Corporation (FDIC).

3.1.5 Broadband Access Data

I use broadband access data from Federal Communications Commission (FCC)'s census tract data on internet access services. This dataset provides accurate data pinpointing where broadband service is available, and where it is not available. This dataset is reliable as different broadband service providers and even the government use this dataset to make decisions about where service is needed and how to expand the broadband services. FCC collects this data on internet access connections in the United States twice a year and all broadband service providers are required to submit the data on where they offer internet access service. The dataset is available at census tract level, and it includes data on residential fixed internet access connections per 1,000 households for service at least 10Mbps down / 1 Mbps up. Based upon the number of connections per 1,000 households FCC classifies regions into six categories [0: 0 connection per 1,000 households; 1: 1 to 200 connections per 1,000 households; 2: 201 to 400 connections per 1,000 households; 3: 401 to 600 connections per 1,000 households; 4: 601 to 800 connections per 1,000 households; 5: More than 800 connections per 1,000 households]. There is variation regarding broadband connectivity status from one period to another (time variation). Besides that, there are also regional variations in accessing internet through broadband connection which create digital divide in U.S. for using financial services through internet.

Apart from that, I use Rural-Urban Continuum Dataset[13] to categorize a region as rural or urban. Based on population density, urbanization, and daily commuting patterns, the rural-urban commuting area (RUCA) codes categorize U.S. census tracts into 10 categories. The value "1" indicates the most urbanized areas. On the other hand, value "10" means the extremist rural areas. This indicator is a reliable indicator for rural areas as it is also used by the Federal Office of Rural Health Policy. Then, I collect weekly fed rate data from Federal Reserve's Economic Data website[14]. I use weekly fed rate data as the foot-traffic data and the ratewatch data are available on weekly basis. Finally, I get my regional control variables from U.S. Census Data and U.S. Bureau of Labor Statistics Data. I get the population of different counties and commuting zones from U.S. Census Data[15] and the average income of different counties and commuting zones from U.S. Bureau of Labor

[13]Source: https://www.ers.usda.gov/data-products/rural-urban-continuum-codes.aspx
[14]Source: https://fred.stlouisfed.org/series/FEDFUNDS
[15]Source: https://www.census.gov

Statistics Data[16].

3.2 Summary Statistics

In table-2(a), I provide the summary statistics for the key variables that I employ in my empirical analysis. The variables are defined in Appendix C.

[Insert Table-2(a) Here]

For observing the effect of customers' distance on pricing, my key explanatory variable is $Distance\_Customer(KM)$. I use the foot-traffic data to get this variable. It is defined as the average distance of the customer in kilometer from the centroid of a specific branch in a given week. In panel-A of table-2(a), I place the summary stat of this variable. It shows that the average distance of the customers home from a bank branch in U.S. is 10.183 KM. In the empirical analysis, I use the logarithmic transformed form of this key variable.

In panel-B, I present the price or annualized interest rate for different deposit products. This data is collected from ratewatch data and available at weekly basis. Here, rate means annualized interest rate of the all the deposit products (INTCK2.5K, MM25K, 03MCD10K, 06MCD10K, 12MCD10K, 24MCD10K, 36MCD10K , and 12MCD250K) used in our analysis. Then in the following rows, I show the price of different types of deposit products. The average price of deposit products varies between 24.553 and 80.331 basis points. It is also observed that the price of uninsured products is higher than the price of the insured products, which is quite rational.

In panel-C, the control variables are described. The control variables include fedrate, alternative measures of local market concentration (distance to nearest competing bank, HHI), some other bank specific and regional characteristics variables. In panel-D, I show

[16]Source: https://www.bls.gov

the instrument variable *BroadbandAccess*. This variable indicates the high-speed internet access status of a region from Federal Communications Commission (FCC)'s region classification. The value ranges from 0 to 5. The mean value is 3.306 means out of 1000 households 461 households have highspeed broadband connection. I also describe some other variables used in the analysis for deposit volume and profitability under this panel-D.

[Insert Table-2(b) and 2(c)Here]

In table-2(a), the variable $Distance\_Customer(KM)$ is observed across different years. It reports that the average distance of bank customers decreases to 9.495KM from 10.586 during our sample period. The declining trend in the distance of the bank customers from their respective branches is observed even during the no-covid periods. Then, the distance variable is reported for different bank types and regions In table-2(c). The average distance of the customers for large bank is 10.474 KM where for small banks it is 9.954KM. It is also noted that branch network helps banks to attract distant customers. From the regional perspective, the average distance of customers of a rural bank branch (10.529 KM) is higher than that of an urban bank branch (10.353 KM).

4 Results for Baseline Regression

4.1 Methodological Remarks

In this section, I run my panel regressions to find the answers of the following research questions: What is the impact of the customers' distance on deposit price? Do banks price discriminate based on the geolocation of the depositors? To get the answers of these questions, I examine the data using the following Ordinary Least Square(OLS) model specifications:

$$Price_{i,j,t} = \beta_0 + \beta_0 Dist_{j,t} + \beta_1 Comp_{l(j),t} + \beta_2 Fed_t + \gamma . Z_{l(j),t} + \delta_t + \theta_{k(j)} + \zeta_i + \alpha_j + \epsilon_{i,j,t}. \quad (1)$$

Here, the dependent variable, $Price_{i,j,t}$, is the price or annualized interest rate of the deposit product i for a given bank branch j in a given period t. The main explanatory variable, $Dist_{j,t}$, is the logarithmic transformation of the average distance in kilometer of the customers of a branch j in a given period t. $Comp_{j,t}$ is the different alternative measures of local market concentration of the region (county/commuting zones) l where branch j is located, Fed_t is the fedrate at period t, $Z_{l(j),t}$ is a vector of other local market characteristics of the region $l(j)$. Following the literature, I control for this vector of local market characteristics which could confound the relationship between average distance of customer and price of different deposit products. This vector includes Log Wage (the logarithmic transformation of the average wage of the people of a given region $l(j)$) and Log Population (the logarithmic transformation of the total population of a specific region $l(j)$). Even though I control for a set of local market characteristics, the price of deposit products might be affected due to some unobserved heterogeneity in the banking industry or due to changes in overall macroeconomic environment. So, to set control for the changes in macroeconomic environment over time, I use time fixed effect δ_t. To control for any unobserved time-invariant bank specific heterogeneity, I use Bank Fixed Effects $\theta_{k(j)}$ for bank k which owns branch j; to control for any unobserved product specific heterogeneity that is time-invariant, I also use Product Fixed Effect ζ_i; and to control for any unobserved branch specific heterogeneity that is time-invariant, I include branch Fixed Effect α_j. Finally, I cluster the standard error at the region level to remove the concern about the shared component in the variation of price data for a specific region.

4.2 Average Price and Distance

Table-3 presents the results from the regressions of the distance of the depositors on the price of the deposit products using the model in equation-1. The table contains four specifications of the same model. In specification (1), county fixed effect is used instead of the region-specific control variables and in specification (2)-(4), alternative measures of local market competition and other regional control variables are used replacing county fixed effect. The model specifications ensure that it captures within bank variation and the results show that the distance of the customers of a branch has a negative impact on the pricing of the deposit products of that branch. The coefficients of the distance variable vary between 0.317 and 0.330 and the results are robust to different measures of competition (number of competitors in the county, local market concentration and distance to the nearest competing branch). Economically speaking, an increase of one standard deviation in the distance of the customers of a branch (around 16 KM increase in the distance from a branch), the prices for that branch's deposit products reduce by 0.54 basis points. From all the specifications, it is observed that banks price discriminate in the deposit market based on the distance of the customers.

[Insert Table-3 Here]

5 Results for Cross-sectional Analysis

Here, I conduct the cross-sectional analysis. Specifically, I try to answer the following questions in this section: Does local market concentration have any impact on the relationship between the distance and price in the deposit market? Does the branch network help banks to increase the impact through reducing distance? Does the effect of the distance differ between transactional deposit and savings deposit? How does the price-distance relationship change with the change in the maturity of the deposit products?

5.1 Average Price, Distance and Local Competition

In this part, the relationship between deposit price and the distance is examined from the perspectives of local market concentrations. The results of the regressions are shown in table-4(a). In column (1), the coefficient of the interaction term between HHI and the Distance is -1.249, meaning that the negative impact of the distance on the deposit price is not monotonic and intensifies with increase in local market concentration. In column (2) and column (3), I explore the relationship between price and distance for areas with low and high market concentration respectively. The higher negative coefficient for the high concentrated market relative to low concentrated market also indicates that banks exercise market power through offering lower price to distant customers only when market competition is low. I also run the regressions for the commuting zones instead of the counties[17] and the impact of the market power on the price-distance relationship is also observed for the commuting zones.

[Insert Table-4(a) Here]

5.2 Average Price for Different Products and Distance

The effect of the distance on the prices of different deposit products is observed in table-4(b). In column (1) and column (2), the regressions are conducted for insured products (the three most common insured products: Interest checking account with minimum amount \$2.5K, Money market account with minimum amount \$25K and 12-month certificate of deposit product with minimum amount \$10K) and uninsured product (the most common uninsured product: 12-month certificate of deposit product with minimum amount \$250K). The negative effect is observed for both types of deposit products, but it is higher and more significant for uninsured deposit product. In column (3), (4) and column (5), the dependent

[17]The results for the commuting zones are exhibited in appendix table-A2.

variables are price of the most common checking account product, saving product, and time deposit product respectively. The results show that the effect exists for time deposit, but not for transactional deposit. The effect is especially stronger for time-deposit and the coefficient -0.421 means for one standard deviation increase in the customers distance the price of small CD products increases by 0.69 basis points. The findings of this part are really interesting since time deposits are more important for banks' value creation relative to transactional deposits (Egan et al. [2022]).

[Insert Table-4(b) Here]

5.3 Average Price for Products with Different Maturity and Distance

I examine how the effect of the distance on deposit price changes with the change in the maturity of the deposit products in this section. To observe the effect of maturity, I use the prices of certificate of deposit products with minimum amount $10K for different maturity periods and the results are presented in table-4(c). In column (1) to column (4), the dependent variables are rates of 03-month, 12-month, 24-month and 36-month CD with minimum amount $10K respectively. All the coefficients are negative and significant. The coefficients vary between 0.270 and 0.696 and the magnitude increases with the increase in the maturity periods of the deposit products.

[Insert Table-4(c) Here]

5.4 Average Price, Distance and Different Types of Banks

Next, I observe whether banks' branch networks help them to exercise market power through offering lower price to distant customers. In column (1) of table-4(b), the coefficient of the interaction term between the distance and branch network variables is negative (-0.197) and significant. It indicates that branch network indeed magnifies the impact of distance on the deposit pricing. In column (2) and column (3), the regressions are run for banks with lower 25% branch network and with higher 25% branch networks respectively. The higher and statistically significant coefficient for the banks in highest quartile branch network group provide supports for my findings in column (1). In column (4) and column (5), the impact of distance is examined for small and large banks. This negative effect of the distance of the customers is more pronounced for small banks compared to large banks.

[Insert Table-4(d) Here]

There are two important implications from the findings of this table. They are:

- The branch networks help banks to offer lower rate to the distant depositors through providing banks market power.
- Second, the distance to the depositors is more important for small banks as the impact of the distance is larger for smaller banks relative to large banks.

6 Robustness Tests and IV Regression

This part includes a battery of robustness tests to show that my results are robust to different scenarios. It also contains the instrument variable regressions to alleviate the concern for endogeneity.

6.1 Average Price and Distance for Non-covid Years

There might be a concern regarding my sample period is that it includes the covid period. There was a huge positive shock to the deposit volume during the covid period because of the increasing precautionary motives of saving among the people during the pandemic and the supply of stimulus money by the government. The disruptions in the deposit market during the covid period is depicted in the google search trends for "Deposit" in figure- 5. To ensure that, my results are not affected by these disruptions, I rerun my baseline regressions for non-covid periods and present the results in table-5(a). We can see that the results are similar like our baseline regressions which illustrate that my previous results are not driven by covid disruptions in the deposit market.

[Insert Figure-5 and Table-5(a) Here]

6.2 Average Price and Distance for Commuting Zones

Another concern regarding my empirical methodology might be regarding the market area as people might live in one county but commute to another county for work or business and maintain bank account in that county. It is a justified concern as county boundaries are not always adequate confines for a local economy and often reflect political boundaries rather than an area's local economy. To address this concern, I run the baseline regressions for commuting zones[18], which better delineate the local economies, instead of counties. The results are analogous to that of my baseline regressions, and they show that my results are robust to different market definitions.

[Insert Table-5(b) Here]

6.3 Average Price and Distance for Rate-Setter Location Only

Here, I want to examine whether my results are robust to the organizational distance. There are two types of bank branches in the ratewatch data. They are rate-setters and rate-

[18] Commuting Zones reflect the commuting nature of the people. Commuting zones are bigger than the counties but smaller than the states. In appendix-B, I provide detailed examples of commuting zones.

followers. In other words, not all bank branches set the price of their products themselves. Rate-setter are the bank branches that set their own prices and set price for other branches under their control. To ensure that my findings in the baseline regressions are robust to the organizational distance, I run the regressions only for the rate-setter branches. The notion behind using only rate-setter branches is as the rate-setter branches set their own prices, they can change the prices based upon the variation of the distance of their depositors. The results of these regressions are presented in table-5(c). The findings are similar to the results of baseline regressions. So, the findings of the baseline regressions are indeed robust to organizational distance.

[Insert Table-5(c) Here]

6.4 IV (Broadband Access) Regressions

In the previous parts, it is observed that the findings of the primary regressions are robust to different specifications and scenarios. But banks' decision to set prices for deposit products and locate their branches may be correlated with some unobservable factors. To address such concern, I use treatment regression analysis using an instrumental variable (IV) that affects the distance of the customers of the bank branches (relevance condition) but has no direct impact on the pricing of the deposit product directly (exclusion condition) exploiting the fact that the access to high-speed internet reduces the demand for physical banking thus decrease the observed distance between customers and bank branches but it has no direct effect on the deposit pricing. In this regard, I take advantage of the existing digital divide[19] in the U.S. and in figure- 6 I illustrate these regional variations in accessing high speed internet for June, 2019.

[19]The idea of the "digital divide" refers to the growing gap between the underprivileged members of society, especially the poor, rural, elderly, and handicapped portion of the population who do not have access to computers or the internet; and the wealthy, middle-class, and young Americans living in urban and suburban areas who have access. (source: https://cs.stanford.edu/people/eroberts/cs181/projects/digital-divide/start.html)

[Insert Figure-6 and Table-5(d) Here]

In table-5(d), I report the results of the treatment regressions using the broadband access status of a region as instrument variable for the distance. In the Panel-A of table-5(d), I show the results of the first stage IV regressions. We see that all the coefficients are negative and significant which portray that our IV, the broadband access status of a region, is negatively related with the physical distance of the customer as expected. The F-stat varies between 90.23 and 95.41, which suggests that our instrument variable is not a weak instrument. Panel-B of table-5(d)reports the results for second stage regressions. We find that all the coefficients of the second stage are negative and significant. The magnitudes of the coefficients for the IV regressions are higher relative to the OLS regressions as IV regressions are estimating the local average treatment effect. In fine, the findings of the IV regressions also support the findings of the panel regressions and show that banks do offer lower price to the distant depositors.

7 Tests for Deposit Volume and Profitability

In this section, I test the effect of the distance of the depositors on the deposit volume through running the regressions both at branch level and bank level. I specifically try to answer the following questions here: What is the impact of the distance on the deposit volume? Does it increase banks' deposit productivity? How does the distance of the customers affect banks' profitability?

7.1 Deposit Volume (Branch) and Distance

Here, I conduct the analysis on the deposit volume at branch level. As the deposit volume for branches is available only on annual basis, I transform the weekly distance variable to the annual variable by calculating the average distance of the customers for the bank branches in a year. The model specifications are like the one specified in equation-1except that the dependent variable here is the logarithmic transformation of the branch deposit volume. The results of the regressions are presented in table-6(a).

[Insert Table-6(a) Here]

The results show that distance of the depositors has a negative effect on the deposit volume of the branch. The coefficient of the column (1) is -0.0063 which means for one standard deviation increase in the average distance of the customers of a bank branch, the deposit volume of that branch decreases by 1.029% (around 1.496 million). In the earlier parts, we find that banks offer lower price to the distant customers. Here, we observe the consequences of offering lower rate to the distant customer as it decreases the overall volume of the deposits for the bank branches.

7.2 Deposit Volume (Bank), Deposit Productivity, NIM and Distance

Lastly, I run the analysis at bank level. As the deposit volume and other control variables for the banks are available only on quarterly basis, I transform the weekly distance variable to the annual variable for the banks. At first, I calculate the average distance of the customers for the branches in a quarter. Then, I get the deposit weighted average distance for all the branches of a bank in a quarter and use this value as the distance of the customers of a bank in that specific quarter. I use the following control variables in the model: Bank Asset (total amount of assets in billions of a specific bank), Deposit Diversity (the concentration of demand, time and saving deposits for a bank in a specific quarter), Return On Assets (net interest income over total assets for a bank), Liquidity (cash over total assets for a bank), and Z-score (the sum of ROA and the equity ratio over the three year standard deviation of ROA).

[Insert Table-6(b) Here]

In table-6(b), I report the results of the regressions at the bank level. In column (1) and (2), the dependent variable is the logarithmic transformation of the banks' deposit volume. I find that the distance also negatively affects the deposit volume of the banks. The column (1) coefficient -.0051 implies for one standard deviation increase in the average distance of the customers of a bank, the deposit volume of that bank decreases by 0.83% (around 29.538 million). In column (3), the dependent variable is the logarithmic transformation of the deposit productivity of a bank and the result shows that the banks' deposit productivity declines with the increase in distance.

So far, from the previous results, we find distance has two different implications in the deposit market. On the one side, it helps banks to reduce the cost of acquiring deposit through providing lower rate to the distant customers which boosts profitability. On the flip side, the lower rate to the distant customers decreases overall deposit volume which shrinks profitability. It is interesting to observe which of these channels dominates ultimately. In column (4), the result of the regression on Net Interest Margin (NIM) discloses that the distance of the customers positively affects NIM. It means that banks actually contribute to their deposit franchise value through offering lower price to the distant depositors.

8 Conclusion

The ability of banks to exercise market power is related with the physical distance from their customers. Existing literature shows that banks derive monopoly rents from proximate customers in the credit market. What remains unexplained is the impact of the distance in the deposit market. Exploiting cell-phone tracking data for the customers of the bank branches, I find substantial evidence for spatial price discrimination in the deposit market. The results of the baseline regressions show that the distance of the customers of a branch has a negative effect on the price of the deposit products of that branch. I reveal that for a one standard deviation increase in the distance of the customers of a branch, the prices for that branch's deposit products reduce by 0.54 basis points and the deposit volume decreases by around 1.496 million.

Moreover, the results find that the impact of the distance is exists only in highly concentrated market which provides banks ideal conditions to exercise the market power. The price-distance relationship is more pronounced for small banks relative to large banks. Apart from that, large branch network also helps banks to charge locational rent through offering lower price to the distant depositors. Finally, my findings shed light on how banks price discriminate in the deposit market to increase their profitability.

References

Agarwal, Sumit, and Robert Hauswald, 2010, Distance and private information in lending, *The Review of Financial Studies* 23, 2757–2788.

Allen, Jason, Robert Clark, and Jean-François Houde, 2014, Price dispersion in mortgage markets, *The Journal of Industrial Economics* 62, 377–416.

Atkin, David, 2016, Endogenous skill acquisition and export manufacturing in mexico, *American Economic Review* 106, 2046–85.

Azar, José, Sahil Raina, and Martin Schmalz, 2022, Ultimate ownership and bank competition, *Financial Management* 51, 227–269.

Basten, Christoph, and Ragnar Juelsrud, 2022, Deposit pricing with cross-selling considerations. a new micro foundation for the deposit channel of monetary policy, *Working Paper* .

Beck, Thorsten, Ross Levine, and Alexey Levkov, 2010, Big bad banks? the winners and losers from bank deregulation in the united states, *The Journal of Finance* 65, 1637–1667.

Berger, Allen N, Robert DeYoung, Hesna Genay, and Gregory F Udell, 2000, Globalization of financial institutions: Evidence from cross-border banking performance, *Brookings-Wharton Papers on Financial Services* 2000, 23–120.

Berger, Allen N, Nathan H Miller, Mitchell A Petersen, Raghuram G Rajan, and Jeremy C Stein, 2005, Does function follow organizational form? evidence from the lending practices of large and small banks, *Journal of Financial Economics* 76, 237–269.

Bonatti, Alessandro, and Gonzalo Cisternas, 2020, Consumer scores and price discrimination, *The Review of Economic Studies* 87, 750–791.

Boyd, William L, Myron Leonard, and Charles White, 1994, Customer preferences for financial services: an analysis, *International Journal of Bank Marketing* .

Breitenstein, Eric C, and John M McGee, 2015, Brick-and-mortar banking remains prevalent in an increasingly virtual world, *FDIC Quarterly* 9, 37–51.

Carlson, Mark, and Kris James Mitchener, 2005, Branch banking, bank competition, and financial stability, *Journal of Money, Credit and Banking* 38, 1293–1328.

Carlson, Mark, and Kris James Mitchener, 2009, Branch banking as a device for discipline: Competition and bank survivorship during the great depression, *Journal of Political Economy* 117, 165–210.

Cave, Joshua, Kevin Keasey, and Danilo V Mascia, 2022, Bank branch deposit competitiveness and deposit growth: Granular data and a new measure of competitiveness, *Economics Letters* 210, 110174.

Cetorelli, Nicola, and Philip E Strahan, 2006, Finance as a barrier to entry: Bank competition and industry structure in local us markets, *The Journal of Finance* 61, 437–461.

Clark, Timothy, Astrid Andrea Dick, Beverly Hirtle, Kevin J Stiroh, and Robard Williams, 2007, The role of retail banking in the us banking industry: risk, return, and industry structure, *Economic Policy Review* 13.

De Haas, Ralph, Steven Ongena, Shusen Qi, and Stefan Straetmans, 2021, Move a little closer? information sharing and the spatial clustering of bank branches, *CEPR Discussion Paper No. DP15829* .

Degryse, Hans, Kent Matthews, and Tianshu Zhao, 2018, Smes and access to bank credit: Evidence on the regional propagation of the financial crisis in the uk, *Journal of Financial Stability* 38, 53–70.

Degryse, Hans, and Steven Ongena, 2005, Distance, lending relationships, and competition, *The Journal of Finance* 60, 231–266.

DeYoung, Robert, W Scott Frame, Dennis Glennon, and Peter Nigro, 2011, The information revolution and small business lending: The missing evidence, *Journal of Financial Services Research* 39, 19–33.

DeYoung, Robert, Dennis Glennon, and Peter Nigro, 2008, Borrower–lender distance, credit scoring, and loan performance: Evidence from informational-opaque small business borrowers, *Journal of Financial Intermediation* 17, 113–143.

Di, Wenhua, and Nathaniel Pattison, 2020, Distant lending, specialization, and access to credit, *FRB of Dallas Working Paper* .

Dorn, David, Gordon Hanson, Kaveh Majlesi, et al., 2020, Importing political polarization? the electoral consequences of rising trade exposure, *American Economic Review* 110, 3139–83.

Drechsler, Itamar, Alexi Savov, and Philipp Schnabl, 2017, The deposits channel of monetary policy, *The Quarterly Journal of Economics* 132, 1819–1876.

Drechsler, Itamar, Alexi Savov, and Philipp Schnabl, 2021, Banking on deposits: Maturity transformation without interest rate risk, *The Journal of Finance* 76, 1091–1143.

Drexler, Alejandro, Andre Guettler, Daniel Paravisini, and Ahmet Ali Taskin, 2019, Competition between arm's length and relational lenders: Who wins the contest?, *Working Paper* .

Egan, Mark, Stefan Lewellen, and Adi Sunderam, 2022, The cross-section of bank value, *The Review of Financial Studies* 35, 2101–2143.

Elliehausen, Gregory E, and John D Wolken, 1997, Banking markets and the use of financial services by small and medium-sized businesses, *J. Reprints Antitrust L. & Econ.* 27, 289.

Foote, Andrew, Mark J Kutzbach, and Lars Vilhuber, 2021, Recalculating...: How uncertainty in local labour market definitions affects empirical findings, *Applied Economics* 53, 1598–1612.

Gary-Bobo, Robert J, and Sophie Larribeau, 2004, A structural econometric model of price discrimination in the french mortgage lending industry, *International Journal of Industrial Organization* 22, 101–134.

Gilje, Erik P, Elena Loutskina, and Philip E Strahan, 2016, Exporting liquidity: Branch banking and financial integration, *The Journal of Finance* 71, 1159–1184.

Granja, João, Christian Leuz, and Raghuram G Rajan, 2022, Going the extra mile: Distant lending and credit cycles, *The Journal of Finance* 77, 1259–1324.

Heitfield, Erik A, 1999, What do interest rate data say about the geography of retail banking markets?, *The Antitrust Bulletin* 44, 333–347.

Herpfer, Christoph, Aksel Mjøs, and Cornelius Schmidt, 2022, The causal impact of distance on bank lending, *Management Science* .

Huang, Rocco R, 2008, Evaluating the real effect of bank branching deregulation: Comparing contiguous counties across us state borders, *Journal of Financial Economics* 87, 678–705.

Hubbard, R Glenn, 2001, Money, the financial system, and the economy, *American Enterprise Institute for Public Policy Research* .

Jayaratne, Jith, and Philip E Strahan, 1996, The finance-growth nexus: Evidence from bank branch deregulation, *The Quarterly Journal of Economics* 111, 639–670.

Jayaratne, Jith, and Philip E Strahan, 1998, Entry restrictions, industry evolution, and dynamic efficiency: Evidence from commercial banking, *The Journal of Law and Economics* 41, 239–274.

Keil, Jan, and Karsten Müller, 2020, Bank branching deregulation and the syndicated loan market, *Journal of Financial and Quantitative Analysis* 55, 1269–1303.

Keil, Jan, and Steven Ongena, 2020, The demise of branch banking technology, consolidation, bank fragility, *Swiss Finance Institute Research Paper* .

Kuehn, Joseph, 2018, Spillovers from entry: the impact of bank branch network expansion, *The RAND Journal of Economics* 49, 964–994.

Lederer, Phillip J, and Arthur P Hurter Jr, 1986, Competition of firms: Discriminatory pricing and location, *Econometrica: Journal of the Econometric Society* 623–640.

Liberti, José María, and Mitchell A Petersen, 2019, Information: Hard and soft, *Review of Corporate Finance Studies* 8, 1–41.

Loutskina, Elena, and Philip E Strahan, 2011, Informed and uninformed investment in housing: The downside of diversification, *The Review of Financial Studies* 24, 1447–1480.

Magri, Silvia, and Raffaella Pico, 2011, The rise of risk-based pricing of mortgage interest rates in italy, *Journal of Banking & Finance* 35, 1277–1290.

Meyer, Paul A, 1967, Price discrimination, regional loan rates, and the structure of the banking industry, *The Journal of Finance* 22, 37–48.

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

Ongena, Steven, Florentina Paraschiv, and Endre J Reite, 2021, Determinants of price discrimination and switching mortgage provider in times of regulation and digitalization, *Swiss Finance Institute Research Paper* .

Park, Kwangwoo, and George Pennacchi, 2008, Harming depositors and helping borrowers: The disparate impact of bank consolidation, *The Review of Financial Studies* 22, 1–40.

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

Petersen, Mitchell A, and Raghuram G Rajan, 2002, Does distance still matter? the information revolution in small business lending, *The Journal of Finance* 57, 2533–2570.

Phlips, Louis, 1983, *The Economics of Price Discrimination* (Cambridge University Press).

Presbitero, Andrea F, Gregory F Udell, and Alberto Zazzaro, 2014, The home bias and the credit crunch: A regional perspective, *Journal of Money, Credit and Banking* 46, 53–85.

Puri, Manju, and Jörg Rocholl, 2008, On the importance of retail banking relationships, *Journal of Financial Economics* 89, 253–267.

Radecki, Lawrence J, 1998, The expanding geographic reach of retail banking markets, *Economic Policy Review* 4.

Rice, Tara, and Philip E. Strahan, 2010, Does credit competition affect small-firm finance?, *The Journal of Finance* 65, 861–889.

Varian, Hal R, 1985, Price discrimination and social welfare, *The American Economic Review* 75, 870–875.

Woodward, Susan E, and Robert E Hall, 2012, Diagnosing consumer confusion and sub-optimal shopping effort: Theory and mortgage-market evidence, *American Economic Review* 102, 3249–76.

Figure 1. Banks Covered in Geolocation Data

This figure shows the most frequent words among the word corpus of banks' name covered by the cellphone foot traffic data. From the figure, we can see that the large banks are observed more as they have more branches across united states.

Figure 2(a). Bank Branches in FDIC's Summary of Deposits (SOD) data
This figure points out the location of all bank branches across united states that are covered in FDIC's Summary of Deposits (SOD) data in the year 2019.

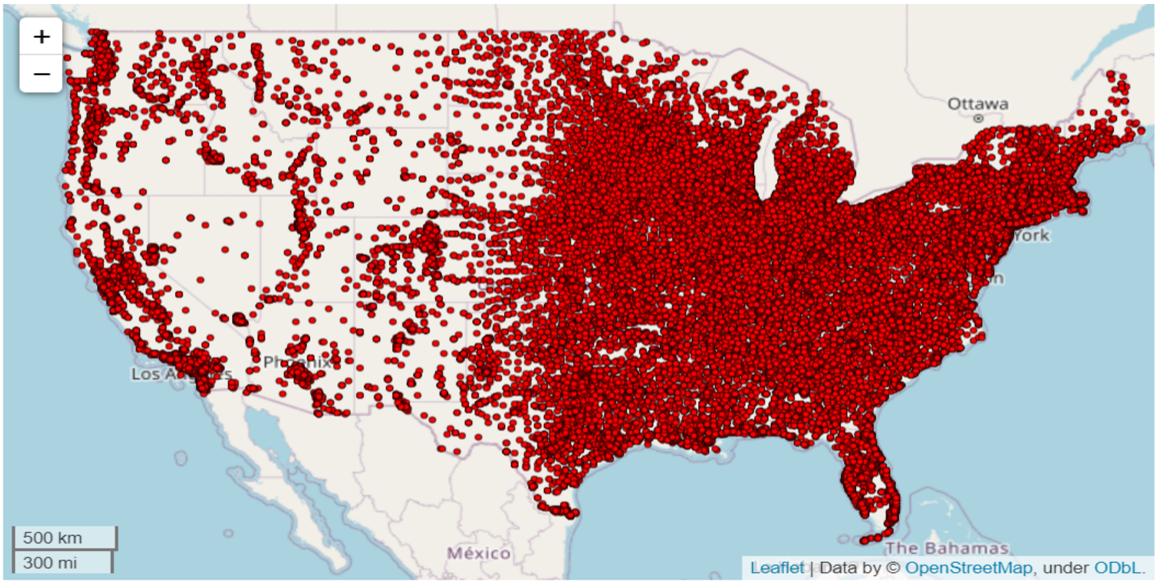

Figure 2(b). Bank Branches Observed in Geolocation Data
Here, I show all the branches for which customers' footprints are observed through cellphone tracking in the year 2019. It shows that the observed branches are good representative of the population/SOD Data.

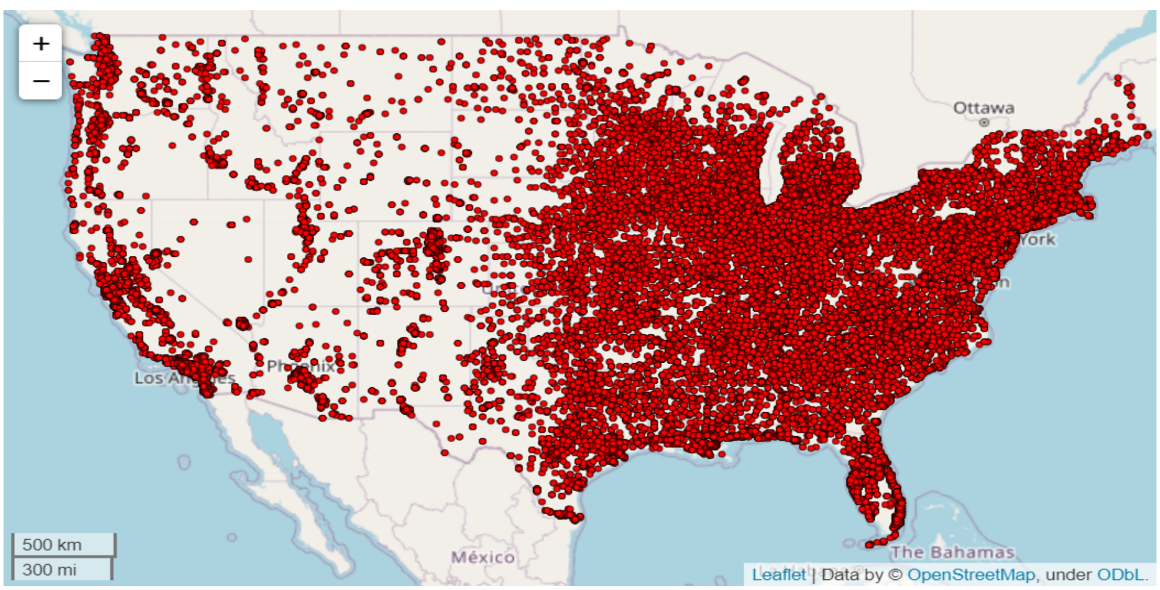

Figure 3(a). Bank Branches in Summary of Deposits (SOD) data for a Sample County (Shasta County, CA)

This figure illustrates the location of all bank branches in Shasta County, CA from Summary of Deposits (SOD) dataset for the year 2019. It shows that some branches are quite close to one another whereas some others are quite far from their competitor branches.

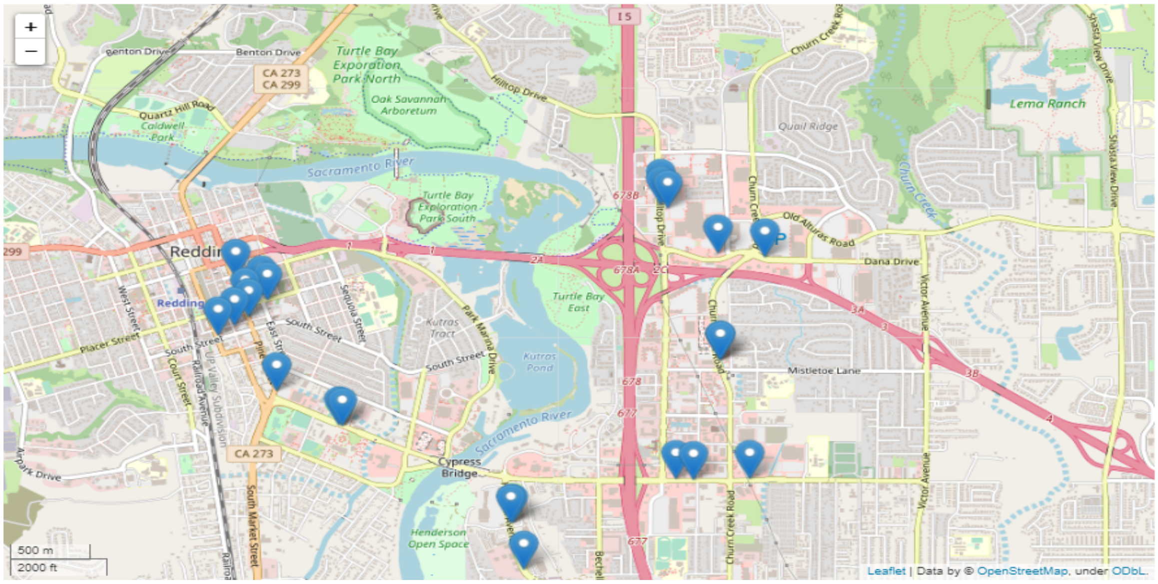

Figure 3(b). Bank Branches Observed in Geolocation Data for a Sample County (Shasta County, CA)

This figure points out the location of all bank branches in Shasta County, CA for which customer footprints are observed through cellphone tracking in the year 2019. We can see that the observed branches in the geolocation data are good representative of SOD branches.

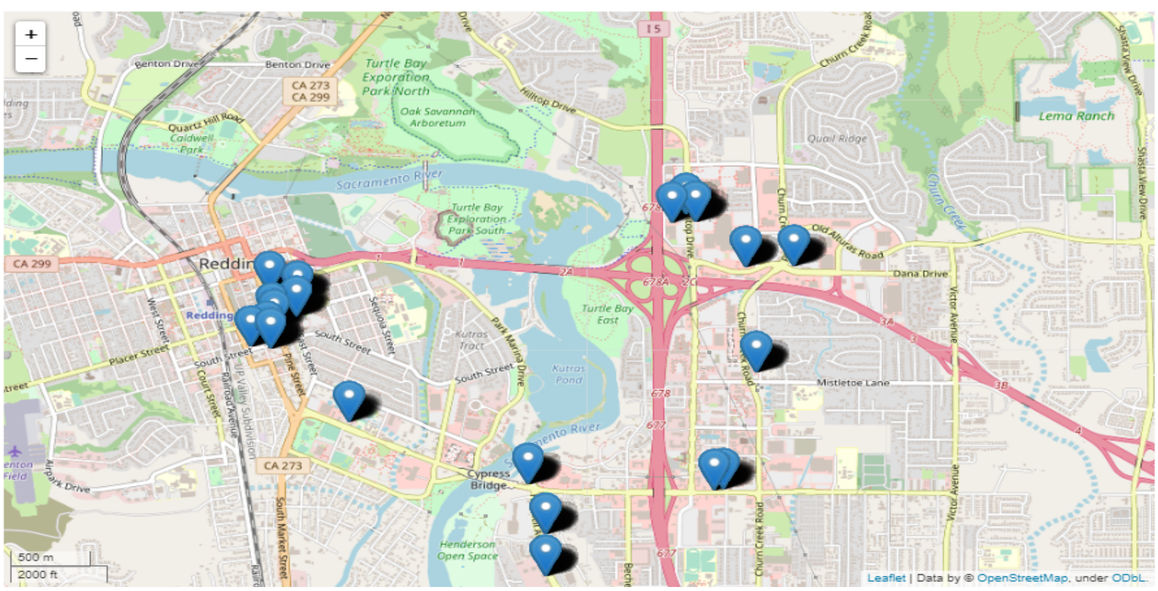

Figure 4(a). Customers' Footprints of a Bank Branch in a Week

Here, we observe the customers' footprints of a specific branch [Chase Bank Branch (Dana Drive in Shasta County, CA)] in a random week (1st Week of February, 2019). We can see, from which census block groups, customers came to that branch during that week.

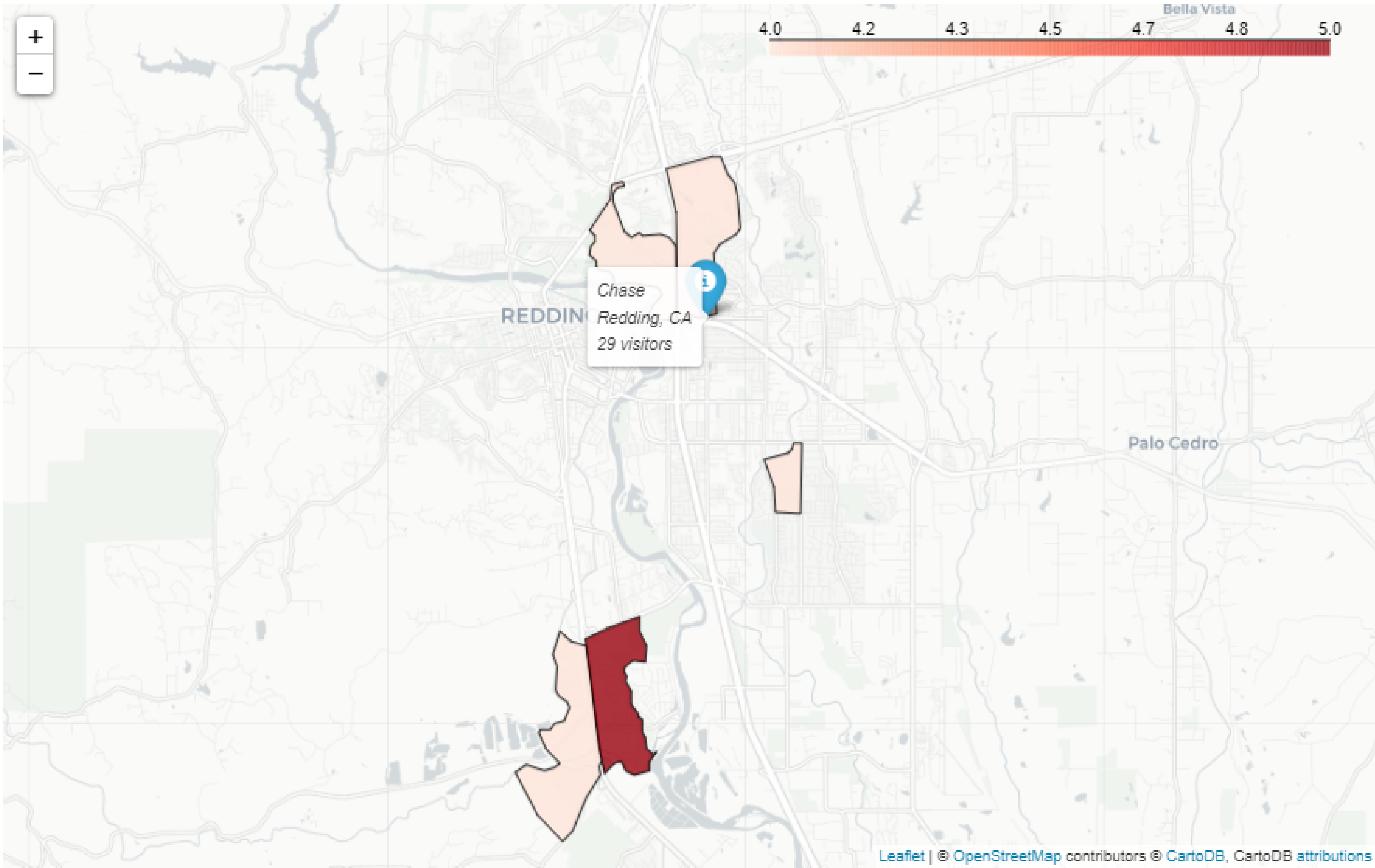

Figure 4(b). Market Network for a Branch in a Week

This figure portrays the market network for a branch (the same Chase Bank Branch at Dana Drive in Shasta County, CA) in a random week (1st Week of February, 2019). We can see that it shares territory with four other bank branches [Plumas Bank at Hilltop Dr., Cornerstone Community Bank at E Cypress Ave, Bank of America at East St. and Tri Counties Bank at Hilltop Dr.] in that specific week.

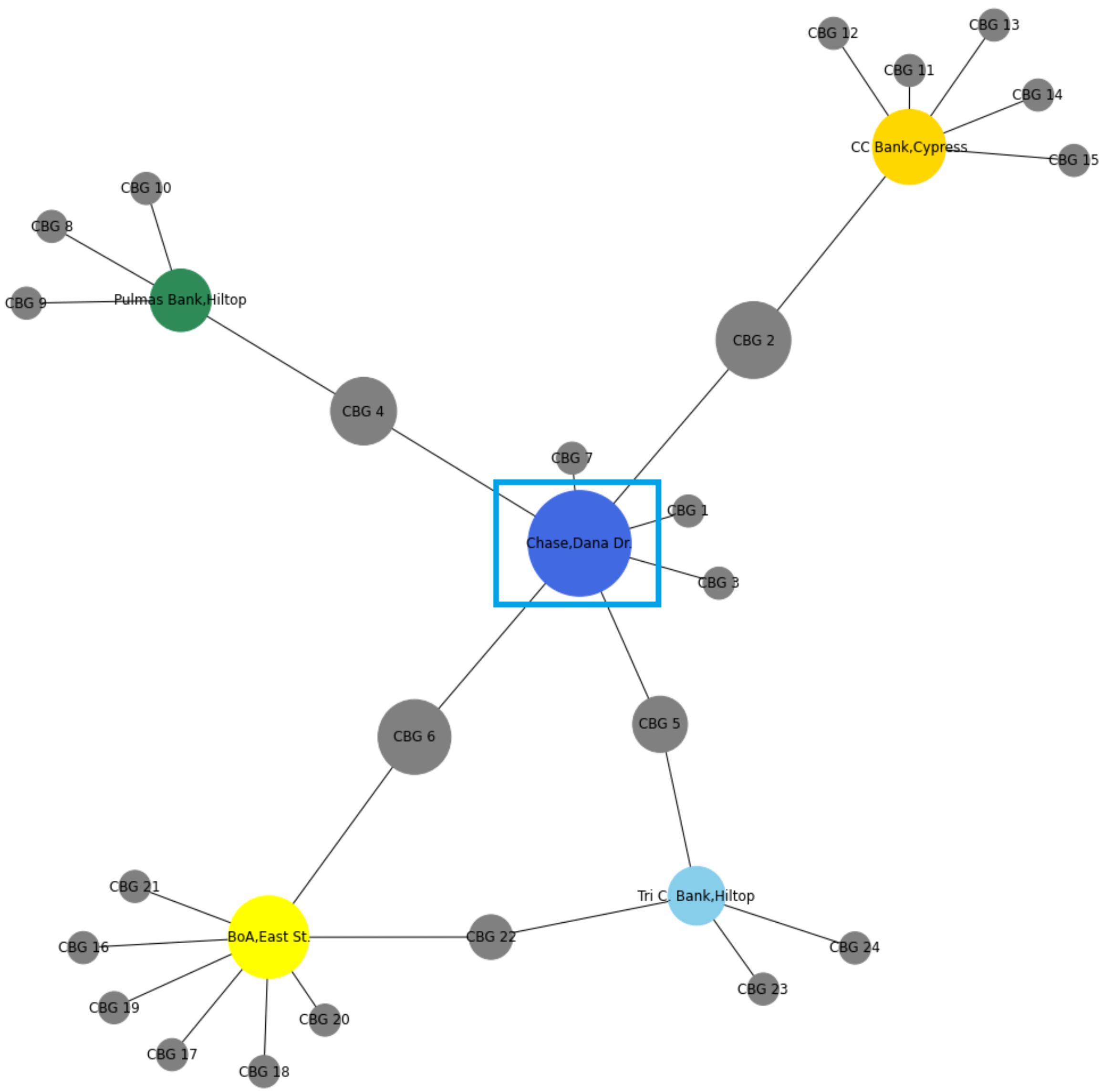

Figure 5. Google Search Interest for "Deposit" over Sample Period.
There were some disruptions in the financial market because of the Covid and the supply of Stimulus by the government. This figure depicts these disruptions in the deposit market. To ensure that my result is not affected by these disruptions, as a robustness check, I also run my analysis only for pre-covid periods.

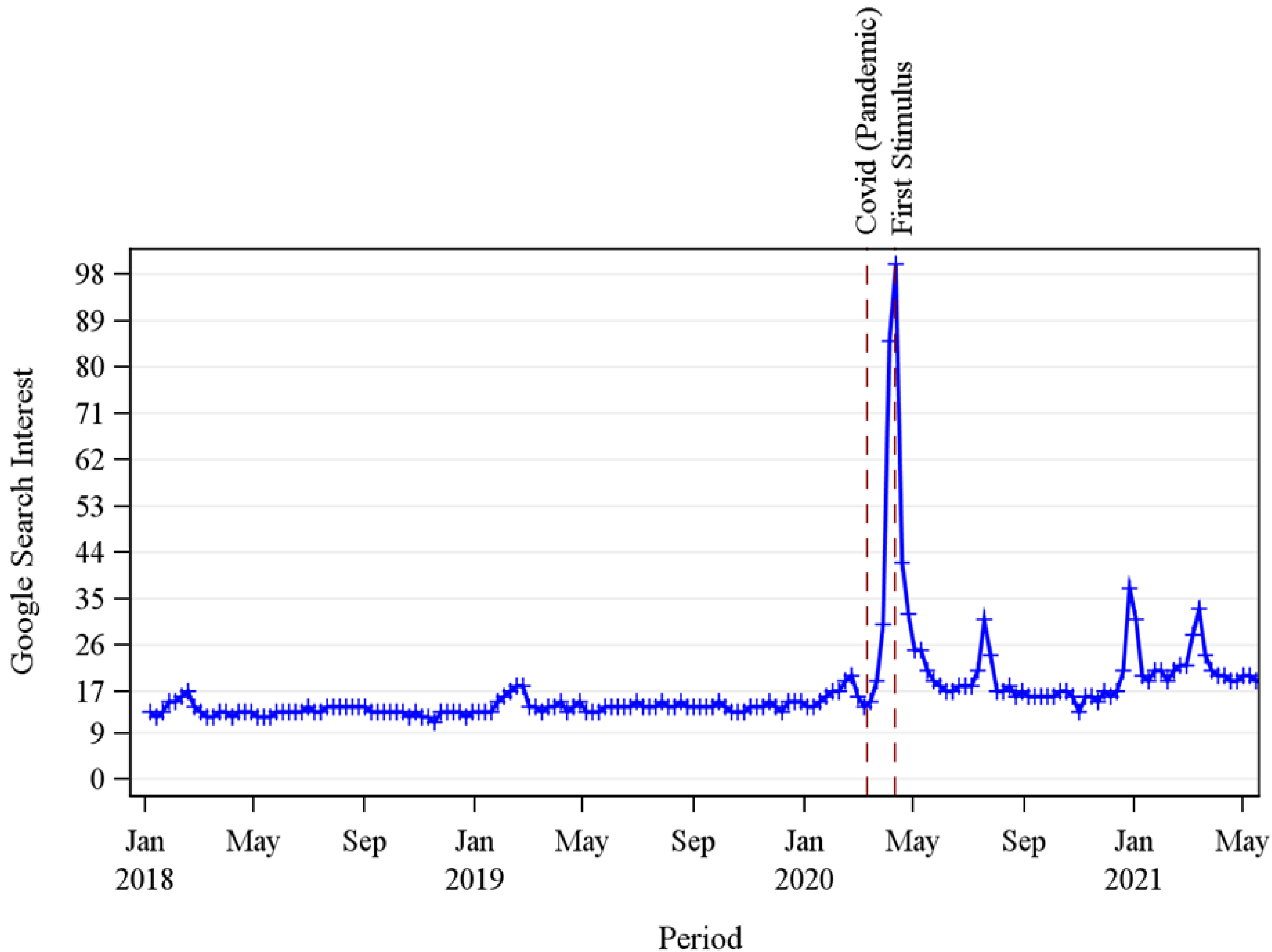

Figure 6. Broadband Connections per 1000 Households by Census Tract for June 2019

This figure depicts the number of residential fixed Internet access service connections per 1,000 households in June 2019 from Federal Communications Commission (FCC)'s broadband subscribership data. FCC classifies regions into six categories [0: 0 connection per 1,000 households; 1: 1 to 200 connections per 1,000 households; 2: 201 to 400 connections per 1,000 households; 3: 401 to 600 connections per 1,000 households; 4: 601 to 800 connections per 1,000 households; 5: More than 800 connections per 1,000 households]. It shows that there are a lot of regional variation besides time variation from the point of internet access through broadband connection.

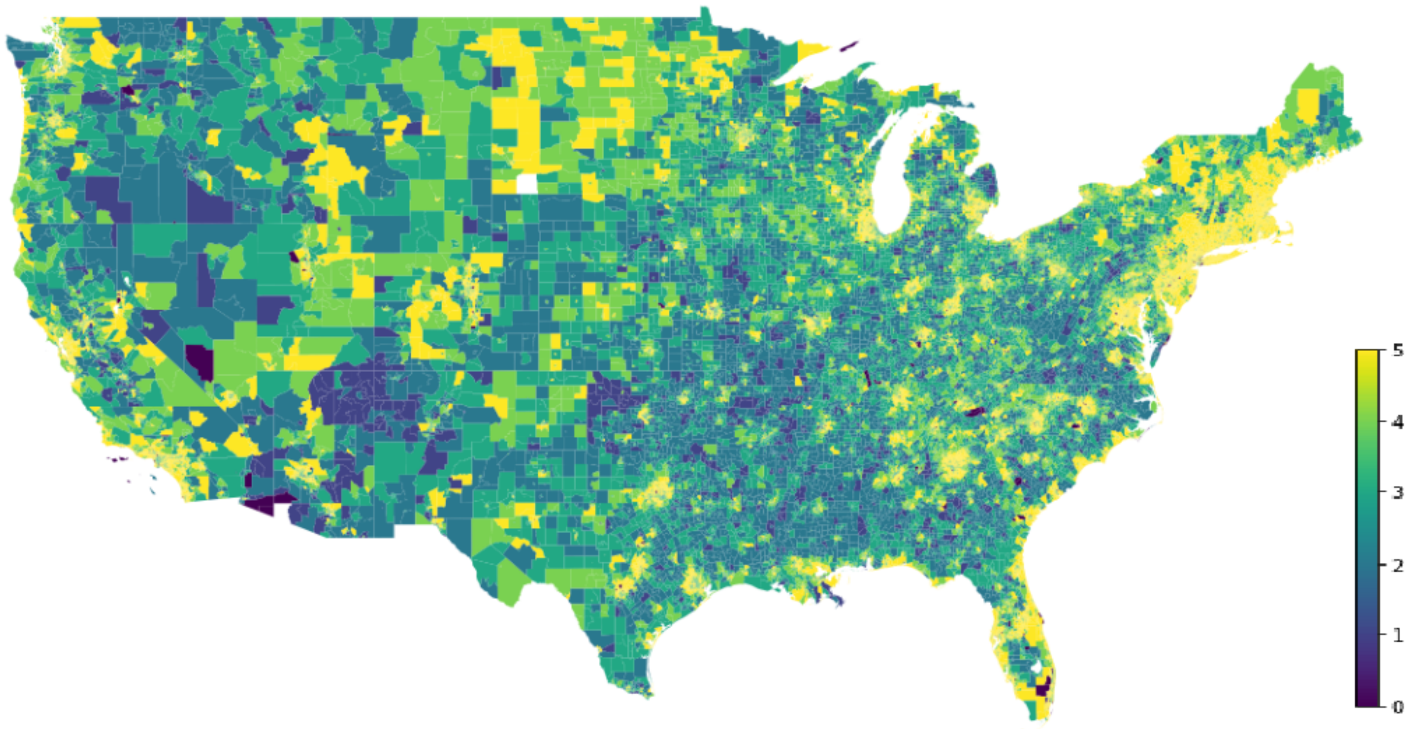

Table 1: The 10 Most Common Point of Interest (POI) Brands/ Banks
In Panel-A, this table reports the most common Banks in the Summary of Deposits (SOD) data. In Panel-B, I show the most common Point of Interest (POI) Brands in the geolocation data. Around 70% of the SOD bank branches are covered by the Geolocation Data.

Panel A: Summary of Deposits (SOD) Data	
Banks	Number of POI
Wells Fargo	5914
Chase	5646
Bank of America	4590
U.S. Bank	3155
BB&T (Branch Banking and Trust)	3150
PNC Financial Services	2998
Regions Bank	1545
The Huntington National Bank	1508
Fifth Third Bank	1299
TD Bank	1282
Panel B: Geolocation Data	
Banks	Number of Branches
Chase	5347
Wells Fargo	4656
Bank of America	4198
PNC Financial Services	2988
U.S. Bank	2786
BB&T (Branch Banking and Trust)	1668
Regions Bank	1392
M&T Bank	1362
Fifth Third Bank	1187
SunTrust Banks	1165

Table 2(a): Summary Statistics of the Key Variables

This table reports summary statistics of the key variables used in the study. In panel A, it shows the key independent variable ($Distance\_Customer(KM)$). The next six rows in panel B show the interest rate for different deposit products. In panel C, I present the regional and bank or branch level control variables. The instrument variable and some other variables are reported in panel D. All the variables are defined in Appendix C.

Variables:	Mean	SD	Min	P25%	Median	P75%	Max	N
Panel A: Distance								
$Distance\_Customer(KM)$	10.183	16.632	0.03	5.02	8.23	12.40	1,407.02	1,662,176
Panel B: Interest Rate								
$Rate$	54.895	57.464	0.10	10.00	32.00	80.00	350.00	3,158,724
$Rate(InsuredProducts)$	34.502	44.028	0.10	5.00	15.00	40.00	300.00	1,226,042
$Rate(UninsuredProducts)$	80.331	59.791	1.00	30.00	65.00	121.00	304.00	313,636
$Rate(Saving)$	24.553	25.047	0.10	10.00	15.00	30.00	250.00	405,422
$Rate(SmallCD/12MCD10K)$	66.614	55.774	1.00	25.00	50.00	100.00	300.00	429,998
$Rate(03MCD10K)$	26.833	29.756	0.50	6.00	15.00	35.00	284.00	371,991
Panel C: Controls								
$Fedrate$	1.485	0.869	0.05	0.47	1.70	2.20	2.44	3,158,724
$Distance\_Competitor(KM)$	1.932	20.965	0.00	0.11	0.25	0.82	1,734.78	3,148,690
HHI_County	0.216	0.132	0.05	0.13	0.18	0.26	1.00	3,158,724
$HHI\_CommutingZone$	0.128	0.086	0.04	0.08	0.11	0.15	1.00	3,158,724
$No.ofBranches/Bank$	427.945	1,069.168	1.00	3.00	8.00	79.00	5,871.00	3,158,724
$Branch\_Network$	2.977	2.484	0	1.10	2.08	4.37	8.68	3,158,724
$RuralIndicator$	3.439	2.297	1.00	1.00	3.00	6.00	9.00	3,158,724
Panel D: Other Variables								
$BroadbandAccess$	3.306	0.907	1.00	3.00	3.00	4.00	5.00	2,134,236
$BranchDeposit(Million)$	145.384	1,315.090	0.00	40.58	72.54	128.89	161,906.19	161,289
$BankDeposit(Billion)$	3.546	47.338	0.00	0.12	0.27	0.65	2,011.34	47,807
$BankAsset(Billion)$	4.942	70.897	0.01	0.15	0.32	0.78	3,207.52	47,807
$NetInterestMargin$	0.007	0.002	-0.02	0.01	0.01	0.01	0.05	47,805

Table 2(b): Summary Statistics of the Distance for Different Years

This table reports summary statistics of the key independent variables $Distance\_Customer(KM)$ (the average distance of a branch customers' home from the specific branch in a given week) for different years [2018-2021]. We can observe a gradual decline in the distance across years.

Variables:	Mean	SD	Min	P25%	Median	P75%	Max	N
$Distance\_Customer(KM)$ Year 2018	10.586	17.959	0.07	5.12	8.38	12.76	1,407.02	500,915
$Distance\_Customer(KM)$ Year 2019	10.321	17.794	0.03	5.09	8.28	12.49	1,196.57	588,053
$Distance\_Customer(KM)$ Year 2020	9.707	13.748	0.06	4.89	8.05	12.03	1,034.28	524,944
$Distance\_Customer(KM)$ Year 2021	9.495	16.159	0.15	4.85	7.92	11.82	1,047.30	48,264

Table 2(c): Univariate Analysis for the Distance

This table reports univariate mean value of the key independent variables $Distance\_Customer(KM)$ for different banks' and regions. The univariate analysis is done from the perspective of large vs small banks, banks with large branch network vs banks with small branch network and urban vs rural branches. SD are reported in parentheses below the mean value. All other variables are defined in Appendix C.

Variables:		$Distance\_Customer(KM)$			
$LargeBanks$	10.474	$Branch\_Network(HighstQrtl)$	10.596	$RuralBranch$	10.529
	(17.718)		(20.037)		(21.914)
$SmallBanks$	9.954	$Branch\_Network(LowstQrtl)$	9.749	$UrbanBranch$	10.353
	(15.406)		(14.639)		(16.278)
$diff$	0.520		0.847		0.176
$t-stat$	15.457		22.028		4.131

Table 3: Average Price and Distance for All Years.

This table reports the regression on Average Price (*Rate*). Here, the independent variable "Distance_Customer (KM)" is the average distance of a branch customers' home from the specific branch in a given week. And the dependent variables are: *Rate*(*AllProducts*) [Rate of all deposit products in the analysis]. Fixed effects (f.e.) are denoted at the bottom of the table and robust standard errors are reported in parentheses below the coefficients. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. All other variables are defined in Appendix C.

Variables	(1) *Rate* (*AllProducts*)	(2) *Rate* (*AllProducts*)	(3) *Rate* (*AllProducts*)	(4) *Rate* (*AllProducts*)
$Log(1 + Distance\_Customer)$	-0.330***	-0.320***	-0.323***	-0.317***
	(0.051)	(0.051)	(0.051)	(0.051)
$No.ofCompetitors$		3.015***		
		(0.938)		
$Log(1 + Distance\_Competitor)$			-3.360***	
			(0.401)	
HHI_County				-9.426***
				(1.226)
$LogPopulation$		-4.024***	-0.786***	-2.351***
		(0.862)	(0.304)	(0.314)
$LogWage$		92.057***	91.436***	91.823***
		(1.532)	(1.531)	(1.528)
$Fedrate$	16.322***	16.258***	16.252***	16.256***
	(0.129)	(0.129)	(0.129)	(0.129)
Constant	32.105***	-921.456***	-939.752***	-925.414***
	(0.220)	(16.916)	(16.232)	(16.180)
Observations	1,662,176	1,662,176	1,657,259	1,662,176
Adjusted R-squared	0.700	0.701	0.701	0.701
Bank FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
Product FE	Yes	Yes	Yes	Yes
Branch FE	Yes	Yes	Yes	Yes
County FE	Yes	No	No	No
SE	Robust	Robust	Robust	Robust

Table 4(a): Average Price, Distance and Local Competition

I report the results of the relation between average price (*Rate*) and the distance of the customers for high and low local market concentration areas. Here, the independent variable "Distance_Customer (KM)" is the average distance of a branch customers' home from the specific branch in a given week. Here, the dependent variables are: $Rate(AllProducts)$ [Rate of all deposit products in the analysis]. In column (1), the key coefficient is the coefficient of the interaction term between HHI and the Distance. In column (2) and column (3) I observe the relationship between price and distance for low and high market concentration areas. Fixed effects (f.e.) are denoted at the bottom of the table and robust standard errors are reported in parentheses below the coefficients. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. All other variables are defined in Appendix C.

	$Rate(AllProducts)$		
Variables	(1)	(2)	(3)
	$AllRegion$	$LowHHI$	$HighHHI$
$Log(1 + Distance\_Customer)$	-0.027	-0.096	-0.370***
	(0.095)	(0.110)	(0.090)
$Log(1 + Distance\_Customer)$*$HHI\_County$	-1.249***		
	(0.334)		
HHI_County	-6.435***	-396.417***	13.254***
	(1.468)	(14.281)	(2.036)
$LogPopulation$	-2.396***	-80.348***	0.340
	(0.312)	(7.403)	(0.363)
$LogWage$	91.870***	8.586*	77.938***
	(1.528)	(4.695)	(2.239)
$Fedrate$	16.255***	14.810***	18.515***
	(0.129)	(0.274)	(0.256)
Constant	-926.088***	1,003.390***	-806.698***
	(16.179)	(103.689)	(23.234)
Observations	1,662,176	430,118	403,494
Adjusted R-squared	0.701	0.694	0.708
Bank FE	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes
Product FE	Yes	Yes	Yes
Branch FE	Yes	Yes	Yes
SE	Robust	Robust	Robust

Table 4(b): Average Price for Different Products and Distance

This table reports the regression on average price of different products. Here, the independent variable "Distance_Customer (KM)" is the average distance of a branch customers' home from the specific branch. Here, the dependent variables are: $Rate(Insured)$ [Rate of three most common Insured products (INTCK2.5K, MM25K, 12MCD10K)]; $Rate(Un-Insured)$ [Rate of the most common Un-insured product (12MCD250K)]; $Rate(Checking)$ [Rate of the most common checking product (INTCK2.5K)]; $Rate(Saving)$ [Rate of the most common checking product (MM25K)]; $Rate(SmallCD)$ [Rate of the most common CD (12MCD10K)]. Fixed effects (f.e.) are denoted at the bottom of the table and robust standard errors are reported in parentheses below the coefficients. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. All other variables are defined in Appendix C.

	(1)	(2)	(3)	(4)	(5)
Variables	$Rate(Insured)$	$Rate(Uninsured)$	$Rate(Checking)$	$Rate(Saving)$	$Rate(SmallCD)$
$Log(1+Distance\_Customer)$	-0.148**	-0.337***	0.017	-0.064	-0.421***
	(0.069)	(0.125)	(0.032)	(0.058)	(0.113)
HHI_County	-5.096***	7.146	-2.320***	4.396***	-17.980***
	(1.485)	(4.688)	(0.479)	(1.375)	(2.780)
$LogPopulation$	18.740***	62.973***	8.288***	12.071***	54.332***
	(3.201)	(9.105)	(1.628)	(3.528)	(6.367)
$LogWage$	63.146***	67.100***	21.442***	37.121***	112.407***
	(2.038)	(4.914)	(0.982)	(1.854)	(3.712)
$Fedrate$	10.156***	19.519***	1.733***	6.668***	21.017***
	(0.181)	(0.397)	(0.093)	(0.170)	(0.312)
Constant	-873.451***	-1,404.975***	-318.524***	-523.685***	-1,795.253***
	(41.693)	(115.780)	(19.921)	(44.375)	(81.338)
Observations	644,619	162,956	204,301	213,800	226,153
Adjusted R-squared	0.633	0.839	0.709	0.736	0.787
Bank FE	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes
Product FE	Yes	No	No	No	No
Branch FE	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes
SE	Robust	Robust	Robust	Robust	Robust

Table 4(c): Average Price for Products with Different Maturity and Distance

This table reports the regression on average price of products with different maturity period. Here, the independent variable "Distance_Customer (KM)" is the average distance of a branch customers' home from the specific branch. Here, the dependent variables are: $Rate(03MCD10K)$ [Rate of 03 month CD with minimum amount $10K]; $Rate(12MCD10K)$ [Rate of 12 month CD with minimum amount $10K]; $Rate(24MCD10K)$ [Rate of 24 month CD with minimum amount $10K]; $Rate(36MCD10K)$ [Rate of 36 month CD with minimum amount $10K]. Fixed effects (f.e.) are denoted at the bottom of the table and robust standard errors are reported in parentheses below the coefficients. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. All other variables are defined in Appendix C.

Variables	(1) *Rate* $(03MCD10K)$	(2) *Rate* $(12MCD10K)$	(3) *Rate* $(24MCD10K)$	(4) *Rate* $(36MCD10K)$
$Log(1 + Distance\_Customer)$	-0.270***	-0.421***	-0.696***	-0.327**
	(0.075)	(0.113)	(0.127)	(0.138)
HHI_County	-26.507***	-17.980***	-12.928***	-31.538***
	(1.778)	(2.780)	(2.691)	(3.178)
$LogPopulation$	19.191***	54.332***	45.392***	-5.094
	(4.134)	(6.367)	(7.117)	(7.697)
$LogWage$	104.909***	112.407***	114.218***	130.808***
	(2.868)	(3.712)	(4.009)	(4.440)
$Fedrate$	8.502***	21.017***	26.092***	29.171***
	(0.199)	(0.312)	(0.348)	(0.378)
Constant	-1,328.412***	-1,795.253***	-1,699.263***	-1,279.808***
	(54.585)	(81.338)	(90.034)	(96.799)
Observations	197,440	226,153	221,815	210,933
Adjusted R-squared	0.743	0.787	0.793	0.810
Bank FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
Branch FE	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes
SE	Robust	Robust	Robust	Robust

Table 4(d): Average Price, Distance and Different Types of Banks

This table reports the regression on average price for different types of banks. Here, the independent variable "Distance_Customer (KM)" is the average distance of a branch customers' home from the specific branch. Here, the dependent variables are: $Rate(AllProducts)$ [Rate of all deposit products in the analysis]. In column (1), (2) and (3), we see the analysis for banks with different branch network sizes. In column (4) and (5), we see the analysis for small vs large banks. Fixed effects (f.e.) are denoted at the bottom of the table and robust standard errors are reported in parentheses below the coefficients. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. All other variables are defined in Appendix C.

	$Rate(AllProducts)$				
Variables	(1)	(2) $Branch\_Network$ $(LowstQrtl)$	(3) $Branch\_Network$ $(HighstQrtl)$	(4) $SmallBanks$	(5) $LargeBanks$
$Log(1 + Distance\_Customer)$	0.240***	-0.089	-0.564***	-0.291***	-0.248**
	(0.077)	(0.091)	(0.088)	(0.060)	(0.098)
$Log(1 + Distance\_Customer)$	-0.197***				
$*Branch\_Network$	(0.020)				
HHI_County	-9.260***	-17.591***	19.681***	-13.395***	24.809***
	(1.221)	(4.817)	(1.448)	(1.679)	(1.472)
$LogPopulation$	-2.400***	-50.097***	-42.098***	-2.961***	-69.968***
	(0.313)	(4.410)	(4.900)	(0.388)	(5.998)
$LogWage$	92.530***	42.704***	176.817***	65.557***	19.891***
	(1.526)	(2.683)	(4.446)	(1.736)	(5.922)
$Fedrate$	16.262***	18.918***	9.864***	18.646***	7.450***
	(0.129)	(0.262)	(0.203)	(0.164)	(0.224)
Constant	-932.355***	141.098**	-1,380.633***	-625.667***	661.115***
	(16.161)	(56.063)	(77.764)	(18.448)	(96.399)
Observations	1,662,176	400,345	450,709	1,013,498	289,235
Adjusted R-squared	0.700	0.732	0.656	0.724	0.639
Quarter FE	Yes	Yes	Yes	Yes	Yes
Product FE	Yes	Yes	Yes	Yes	Yes
Branch FE	Yes	Yes	Yes	Yes	Yes
SE	Robust	Robust	Robust	Robust	Robust

Table 5(a): Average Price and Distance for Non-covid Years

This table reports the baseline regression on average price for non-covid years. Here, the independent variable "Distance_Customer (KM)" is the average distance of a branch customers' home from the specific branch. And the dependent variables are: $Rate(AllProducts)$ [Rate of all deposit products in the analysis]. Fixed effects (f.e.) are denoted at the bottom of the table and robust standard errors are reported in parentheses below the coefficients. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. All other variables are defined in Appendix C.

Variables	(1) $Rate$ $(AllProducts)$	(2) $Rate$ $(AllProducts)$	(3) $Rate$ $(AllProducts)$	(4) $Rate$ $(AllProducts)$
$Log(1+Distance\_Customer)$	-0.179***	-0.179***	-0.179***	-0.181***
	(0.059)	(0.059)	(0.060)	(0.059)
$No.ofCompetitors$		18.982***		
		(1.315)		
$Log(1+Distance\_Competitor)$			0.123	
			(0.513)	
HHI_County				11.182***
				(2.062)
$LogPopulation$		-18.918***	-2.702***	-1.587***
		(1.172)	(0.309)	(0.347)
$LogWage$		18.210***	16.644***	17.608***
		(2.219)	(2.222)	(2.214)
$Fedrate$	11.606***	11.612***	11.619***	11.604***
	(0.319)	(0.319)	(0.320)	(0.319)
Constant	41.659***	-3.035	-105.644***	-131.261***
	(0.648)	(24.459)	(23.484)	(23.352)
Observations	1,189,750	1,189,750	1,185,620	1,189,750
Adjusted R-squared	0.752	0.752	0.752	0.752
Bank FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
Product FE	Yes	Yes	Yes	Yes
Branch FE	Yes	Yes	Yes	Yes
County FE	Yes	No	No	No
SE	Robust	Robust	Robust	Robust

Table 5(b): Average Price and Distance for Commuting Zones

This table reports the baseline regression on average price for commuting zones instead of counties. I use out10 regions as Commuting Zones which are bigger than counties. Here, the independent variable "Distance_Customer (KM)" is the average distance of a branch customers' home from the specific branch. And the dependent variables are: $Rate(AllProducts)$ [Rate of all deposit products in the analysis]. Fixed effects (f.e.) are denoted at the bottom of the table and robust standard errors are reported in parentheses below the coefficients. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. All other variables are defined in Appendix C.

Variables	(1) $Rate$ $(AllProducts)$	(2) $Rate$ $(AllProducts)$	(3) $Rate$ $(AllProducts)$	(4) $Rate$ $(AllProducts)$
$Log(1 + Distance\_Customer)$	-0.331***	-0.298***	-0.299***	-0.298***
	(0.051)	(0.051)	(0.051)	(0.051)
$No.ofCompetitors\_CZ$		-15.782***		
		(1.567)		
$Log(1 + Distance\_Competitor)$			-1.903***	
			(0.393)	
$HHI\_CommutingZone$				-10.661***
				(1.328)
$LogPopulation\_CZ$		11.362***	-0.297	-0.385
		(1.419)	(0.657)	(0.644)
$LogWage\_CZ$		175.196***	174.896***	176.107***
		(2.048)	(2.045)	(2.041)
$Fedrate$	16.322***	16.210***	16.210***	16.215***
	(0.129)	(0.128)	(0.128)	(0.128)
Constant	32.106***	-1,909.470***	-1,832.873***	-1,844.266***
	(0.220)	(24.144)	(23.416)	(23.354)
Observations	1,662,176	1,662,176	1,657,259	1,662,176
Adjusted R-squared	0.700	0.701	0.702	0.701
Bank FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
Product FE	Yes	Yes	Yes	Yes
Branch FE	Yes	Yes	Yes	Yes
Cmtng Zone FE	Yes	No	No	No
SE	Robust	Robust	Robust	Robust

Table 5(c): Average Price and Distance for Rate-Setter Location Only

This table reports the baseline regression on average price for rate setter locations only. Rate-setter are the bank branches that set their own prices and set price for other branches under their control. Here, the independent variable "Distance_Customer (KM)" is the average distance of a branch customers' home from the specific branch. And the dependent variables are: $Rate(AllProducts)$ [Rate of all deposit products in the analysis]. Fixed effects (f.e.) are denoted at the bottom of the table and robust standard errors are reported in parentheses below the coefficients. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. All other variables are defined in Appendix C.

Variables	(1) $Rate$ $(AllProducts)$	(2) $Rate$ $(AllProducts)$	(3) $Rate$ $(AllProducts)$	(4) $Rate$ $(AllProducts)$
$Log(1 + Distance\_Customer)$	-0.322***	-0.311***	-0.314***	-0.308***
	(0.051)	(0.051)	(0.051)	(0.051)
$No.ofCompetitors$		2.415**		
		(0.938)		
$Log(1 + Distance\_Competitor)$			-3.347***	
			(0.401)	
HHI_County				-10.633***
				(1.229)
$LogPopulation$		-3.543***	-0.818***	-2.494***
		(0.862)	(0.304)	(0.314)
$LogWage$		92.858***	92.277***	92.654***
		(1.532)	(1.531)	(1.528)
$Fedrate$	16.324***	16.259***	16.254***	16.257***
	(0.129)	(0.129)	(0.129)	(0.129)
Constant	32.127***	-933.446***	-948.392***	-932.412***
	(0.220)	(16.912)	(16.232)	(16.181)
Observations	1,656,376	1,656,376	1,651,459	1,656,376
Adjusted R-squared	0.700	0.701	0.701	0.701
Bank FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
Product FE	Yes	Yes	Yes	Yes
Branch FE	Yes	Yes	Yes	Yes
County FE	Yes	No	No	No
SE	Robust	Robust	Robust	Robust

Table 5(d): IV (Broadband Access) Regressions

This table reports the instrumental variable (IV) regression on average price. Panel-A: [It reports first-stage iv regression result, where the dependent variable "Distance_Customer (KM)" is the average distance of a branch customers' home from the specific branch. Here **Broadband Access Status** of a region is used as the instrumental variable]. Panel-B: [It reports second-stage treatment regression result. Here, the dependent variables are: $Rate(AllProducts)$ [Rate of all deposit products in the analysis]. Fixed effects (f.e.) are denoted at the bottom of the table and robust standard errors are reported in parentheses below the coefficients. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. All other variables are defined in Appendix C.

Panel-A: First-stage Regression

Variables	$Log(1 + Distance\_Customer)$			
	(1)	(2)	(3)	(4)
$BroadbandAccess$	-0.0197***	-0.0192***	-0.0197***	-0.0194***
	(0.0020)	(0.0020)	(0.0020)	(0.0020)
Control Variables	Yes	Yes	Yes	Yes
F-test (p-value)	0.000	0.000	0.000	0.000
F Stat	95.41	90.23	94.31	91.32

Panel-B: Second-stage Regression

	(1)	(2)	(3)	(4)
Variables	$Rate(AllProducts)$			
$Log(1 + Distance\_Customer)(fit)$	-0.8623***	-0.8500***	-0.9111***	-0.8368***
	(11.157)	(11.371)	(11.624)	(11.194)
$No.ofCompetitors$		22.297***		
		(2.668)		
$Log(1 + Distance\_Competitor)$			0.381	
			(0.833)	
HHI_County				30.436***
				(7.071)
Observations	1,088,968	1,088,968	1,085,031	1,088,968
Bank FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
Product FE	Yes	Yes	Yes	Yes
Branch FE	Yes	Yes	Yes	Yes
County FE	Yes	No	No	No
Wald chi2 test	0.000	0.000	0.000	0.000

Table 6(a): Deposit Volume (Branch) and Distance

This table reports the regression on logarithm transformation of deposit volume at branch level ($Log(BranchDeposit)$). Here, the independent variable "Distance_Customer (KM)" is the average distance of a branch customer's home from the specific branch in a year (the distance is converted to yearly observations as branch level deposit volume data is available on annual basis). And the dependent variable is $Log(BranchDeposit)$ [Log of deposit volume at branch level on yearly basis]. Fixed effects (f.e.) are denoted at the bottom of the table and robust standard errors are reported in parentheses below the coefficients. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. All other variables are defined in Appendix C.

	(1)	(2)	(3)	(4)
Variables		$Log(BranchDeposit)$		
$Log(1 + Distance\_Customer)$	-0.0063	-0.0352***	-0.0349***	-0.0353***
	(0.006)	(0.006)	(0.006)	(0.006)
$No.ofCompetitors$		-0.0471***		
		(0.010)		
$Log(1 + Distance\_Competitor)$			-0.2438***	
			(0.005)	
HHI_County				-0.1510***
				(0.028)
$LogPopulation$		0.0673***	0.0232***	0.0242***
		(0.008)	(0.003)	(0.003)
$LogWage$		0.5075***	0.4723***	0.4962***
		(0.018)	(0.017)	(0.018)
Constant	4.3177***	-1.7889***	-0.9525***	-1.3025***
	(0.013)	(0.204)	(0.172)	(0.174)
Observations	107,758	107,823	107,356	107,823
Adjusted R-squared	0.387	0.334	0.346	0.334
Bank FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
County FE	Yes	No	No	No
SE	Robust	Robust	Robust	Robust

Table 6(b): Deposit Volume (Bank), Deposit Productivity, NIM and Distance
This table reports the regression on logarithm transformation of deposit volume at bank level ($Log(BankDeposit)$). Here, the independent variable "Distance_Customer (KM)" is the average distance of a bank customers' home from the branches of that bank in a quarter (the distance is converted to quarterly observations as bank level deposit volume data is available on quarterly basis). In column (1) and (2) the dependent variables are: $Log(BankDeposit)$ [Log of deposit volume at bank level on yearly quarterly basis]; in column (3) the dependent variable is log of deposit productivity and in column (4) the dependent variable is Net Interest Margin. Fixed effects (f.e.) are denoted at the bottom of the table and robust standard errors are reported in parentheses below the coefficients. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. All other variables are defined in Appendix C.

Variables	(1) Log $(BankDeposit)$	(2) Log $(BankDeposit)$	(3) $Log(Deposit$ $Productivity)$	(4) $NetInterest$ $Margin$
$Log(1 + Distance\_Customer)$	-0.0051***	-0.6432***	-0.0311***	0.0002***
	(0.002)	(0.010)	(0.004)	(0.000)
$BankAsset$		0.0047***	-0.0001	-0.0000***
		(0.000)	(0.000)	(0.000)
$DepositDiversity$		-4.4202***	-1.5102***	-0.0008***
		(0.074)	(0.037)	(0.000)
$ReturnOnAssets$		-2.2920	-3.1364	0.1766***
		(4.228)	(2.431)	(0.018)
$Liquidity$		-3.5542***	2.0946***	0.0016***
		(0.113)	(0.058)	(0.000)
$Z - score$		-0.0000**	-0.0000***	-0.0000***
		(0.000)	(0.000)	(0.000)
Constant	-0.9117***	2.6639***	0.3224***	0.0064***
	(0.002)	(0.039)	(0.020)	(0.000)
Observations	36,366	30,718	30,701	30,718
Adjusted R-squared	0.996	0.482	0.282	0.111
Bank FE	Yes	No	No	No
Quarter FE	Yes	Yes	Yes	Yes
SE	Robust	Robust	Robust	Robust

Appendix A

Table A1: No of Banks and Bank Branches in U.S.

This table reports the number of banks and the number of branches in the U.S. from 1998 to 2021. There is a consistent decline in the number of banks throughout the period. However, the number of bank branches increases initially till 2009 and then declines steadily.

Year	Number of Banks	Number of Branches
1998	10738	67029
1999	10346	68743
2000	10116	70204
2001	9752	71082
2002	9474	72088
2003	9256	78560
2004	9066	80488
2005	8856	82074
2006	8767	84580
2007	8605	87258
2008	8441	89132
2009	8185	91874
2010	7821	91657
2011	7523	98193
2012	7255	97340
2013	6950	96339
2014	6669	94725
2015	6358	93272
2016	6068	91834
2017	5797	89857
2018	5551	88075
2019	5313	86392
2020	5076	85050
2021	4960	81818

Table A2: Summary Statistics of Distance and Time Spent for Different Types of Banks

This table reports summary statistics of two important variables $Distance\_Customer(KM)$ (the average distance of a branch customers' home from the specific branch) and $TimeSpent(Minutes)$ (the average time spent by a branch customers' at that specific branch) for different types of banks and bank branches.

Variables:	Mean	SD	Min	P25%	Median	P75%	Max	N
For Small Banks:								
$Distance\_Customer(KM)$	9.954	15.406	0.04	4.78	8.07	12.18	1,034.28	1,013,498
$TimeSpent(Minutes)$	107.962	152.240	4.00	8.00	30.00	165.00	1,437.00	2,006,571
For Large Banks:								
$Distance\_Customerr(KM)$	10.474	17.718	0.07	5.36	8.43	12.35	1,407.02	289,235
$TimeSpent(Minutes)$	75.724	140.868	4.00	6.50	9.00	52.00	1,423.00	498,950
For Small Branch Network:								
$Distance\_Customerr(KM)$	9.749	14.639	0.04	4.39	8.01	12.16	1,034.28	400,345
$TimeSpent(Minutes)$	98.479	146.250	4.00	8.00	25.50	143.00	1,437.00	841,645
For Large Branch Network:								
$Distance\_Customerr(KM)$	10.596	20.037	0.07	5.34	8.43	12.51	1,407.02	450,709
$TimeSpent(Minutes)$	84.994	147.183	4.00	6.50	10.00	87.00	1,423.00	795,121
For Single State Banks:								
$Distance\_Customer(KM)$	9.955	15.413	0.04	4.85	8.13	12.13	1,034.28	992,248
$TimeSpent(Minutes)$	108.653	153.426	4.00	8.00	30.00	165.00	1,437.00	1,959,866
For Multi State (GT5) Banks:								
$Distance\_Customer(KM)$	10.878	22.243	0.07	5.32	8.40	12.62	1,407.02	360,262
$TimeSpent(Minutes)$	83.705	147.991	4.00	6.50	10.00	75.00	1,423.00	639,394
For Urban Area:								
$Distance\_Customer(KM)$	10.353	16.278	0.03	5.40	8.45	12.62	1,407.02	1,055,994
$TimeSpent(Minutes)$	107.205	154.951	4.00	8.00	21.50	163.50	1,437.00	1,938,944
For Rural Area:								
$Distance\_Customer(KM)$	10.529	21.914	0.04	3.64	7.95	12.41	1,034.28	194,519
$TimeSpent(Minutes)$	101.194	150.721	4.00	8.00	25.00	148.00	1,435.00	435,712

Table A3: Average Price, Distance and Local Competition for Commuting Zones

I report the results of the relation between average price (*Rate*) and the distance of the customers for commuting zones (instead of counties) with high and low local market concentration. Here, the independent variable "Distance_Customer (KM)" is the average distance of a branch customers' home from the specific branch in a given week. Here, the dependent variables are: *Rate*(*AllProducts*) [Rate of all deposit products in the analysis]. In column (1), the key coefficient is the coefficient of interaction term between HHI of a commuting zone and the Distance. In column (2) and column (3) I observe the relationship between price and distance for low and high market concentration areas. Fixed effects (f.e.) are denoted at the bottom of the table and robust standard errors are reported in parentheses below the coefficients. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. All other variables are defined in Appendix C.

Variables	*Rate*(*AllProducts*)		
	(1)	(2)	(3)
	AllRegion	*LowHHI*	*HighHHI*
$Log(1 + Distance\_Customer)$	-0.163*	-0.260***	-0.492***
	(0.084)	(0.092)	(0.104)
$Log(1 + Distance\_Customer)$*$HHI\_CommutingZone$	-1.051**		
	(0.522)		
$HHI\_CommutingZone$	-8.258***	87.845***	-32.803***
	(1.766)	(23.738)	(2.882)
$LogPopulation\_CZ$	-0.390	41.908***	-62.013***
	(0.644)	(8.309)	(7.035)
$LogWage\_CZ$	176.081***	193.772***	127.516***
	(2.041)	(4.902)	(4.314)
$Fedrate$	16.217***	16.257***	17.536***
	(0.128)	(0.274)	(0.251)
Constant	-1,844.226***	-2,572.087***	-489.013***
	(23.353)	(117.573)	(102.180)
Observations	1,662,176	418,459	409,764
Adjusted R-squared	0.701	0.731	0.695
Bank FE	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes
Product FE	Yes	Yes	Yes
Branch FE	Yes	Yes	Yes
SE	Robust	Robust	Robust

Appendix B Commuting Zones

I use Out10 regions as Commuting Zones (CZs) in this paper. They are bigger than the counties but smaller than the states. There are 625 out10 regions whereas the number of counties are 3141[20]. County boundaries are not always adequate confines for a local economy and often reflect political boundaries rather than an area's local economy. So, to better delineate local economies, these Commuting Zones (CZs) were first developed in the 1980s. In short, Out10 regions are geographic units of analysis intended to more closely reflect the local economy where people live and work. Commuting Zones (CZs) are used in many economics literature to better understand the local economic activities of a region.

An out10 region consists of multiple counties but a county cannot be in two out10 regions. However, the counties under an out10 regions can be in two states. These distinctions will be clear from the following examples:

Out10 Region 01:

It consists of nine counties. The FIPS codes for the counties are: 1001, 1011, 1013, 1039, 1041, 1041, 1051, 1085, 1101, and 1109. All these counties are situated in Alabama.

Out10 Region 02:

It consists of ten counties. The FIPS codes for the counties are: 1003, 1097, 1129, 28039, 28041, 28045, 28047, 28059, 28109, and 28131. The first three counties (1003, 1097, and 1129) are situated in Alabama and the last seven counties (28039, 28041, 28045, 28047, 28059, 28109, and 28131) are situated in Mississippi.

[20] Source:https://www.usgs.gov

Appendix C Variable Definition

Variable	Description of Variable
BankAsset	Total amount of assets in billions of a specific bank. This data is available on quarterly basis.
BankDeposit	Total amount of deposits in billions of a specific bank. This data is available on quarterly basis.
BranchDeposit	Total amount of deposits in millions of a specific bank branch. This data is available on annual basis.
Branch_Network	Logarithmic transformation of the number of total branches of a specific bank observed in the summary of deposit (sod) data.
BroadbandAccess	This variable indicates the high speed internet access status of a region from Federal Communications Commission (FCC)'s region classification. The value ranges from 0 to 5 [0: 0 connection; 1: 1 to 200 connections; 2: 201 to 400 connections; 3: 401 to 600 connections; 4: 601 to 800 connections; 5: More than 800 connections per 1,000 households].
CommutingZones	I use Out10 regions as Commuting Zones (CZs) which are bigger than the counties but smaller than the states. An out10 region consists of multiple counties but a county cannot be in two out10 regions.
DepositDiversity	The concentration of demand, time and saving deposits for a given bank in a specific quarter.
DepositProductivity	Total deposit amount over total interest expenses for a given bank in a specific quarter.
Distance_Customer	The average distance of the customers' homes in kilometer from a bank branch in a specific week.
Distance_Competitor	The distance of the nearest competing branch in kilometer from a bank branch.
FedRate	Interest rate at which depository institutions trade balances held at Federal Reserve Banks with each other overnight in a given week.
HHI_County	Herfindahl-Hirschman Index (HHI) is the sum of the squared deposit-market shares of all banks that have branch(es) in a given county in a specific year.
HHI_CommutingZones	Herfindahl-Hirschman Index (HHI) is the sum of the squared deposit-market shares of all banks that have branch(es) in a given commuting zone area (out10 region) in a specific year.
$INTCK2.5K$	Interest Checking Account with Minimum Amount \$2.5K.

Variable	Description of Variable
$LargeBanks$	The banks with total assets of $100 billion or more.
$Liquidity$	Cash over total assets for a given bank.
Log_Wage	Logarithmic transformation of the average wage of the people of a county in a given year.
$LogWage\_CZ$	Logarithmic transformation of the average wage of the people of a given commuting zone area (out10 region) in a specific year.
$Log\_Population$	Logarithmic transformation of the total population of a county in a given year.
$LogPopulation\_CZ$	Logarithmic transformation of the total population of a given commuting zone area (out10 region) in a specific year.
$MM25K$	Money Market Account with Minimum Amount $25K.
$MultiStateBanks$	The banks that operate in more than five states.
$NetInterestMargin$	The difference between the interest income and the interest paid, relative to the total assets for a given bank.
$No.ofBranches/Bank$	The number of total branches of a specific bank observed in the summary of deposit (sod) data.
$No.ofCompetitors$	Total number of bank branches in each county.
$No.ofCompetitors\_CZ$	Total number of bank branches in a given commuting zone area (out10 region) in a specific year.
$Rate$	Annualized interest rate of the all the deposit products (INTCK2.5K, MM25K, 03MCD10K, 06MCD10K, 12MCD10K, 24MCD10K, 36MCD10K , and 12MCD250K) used in our analysis.
$Rate(Checking)$	Annualized interest rate of the most common checking product "Interest Checking Account with Minimum Amount $2.5K (INTCK2.5K)".
$Rate(InsuredProducts)$	Annualized interest rate of three most common insured products (INTCK2.5K, MM25K, 12MCD10K).
$Rate(Saving)$	Annualized interest rate of the most common saving product "Money Market Account with Minimum Amount $25K (MM25K)".
$Rate(SmallCD)$	Annualized interest rate of the most common CD "12 Month Certificate of Deposit (CD) product with Minimum Amount $10K (12MCD10K)"
$Rate(UninsuredProducts)$	Annualized interest rate of the most common uninsured product "12 Month Certificate of Deposit (CD) product with Minimum Amount $250K (12MCD250K)".
$Rate(03MCD10K)$	Annualized interest rate of 03 Month Certificate of Deposit (CD) product with Minimum Amount $10K

Variable	Description of Variable
$Rate(12MCD10K)$	Annualized interest rate of 12 Month Certificate of Deposit (CD) product with Minimum Amount $10K
$Rate(24MCD10K)$	Annualized interest rate of 24 Month Certificate of Deposit (CD) product with Minimum Amount $10K
$Rate(36MCD10K)$	Annualized interest rate of 36 Month Certificate of Deposit (CD) product with Minimum Amount $10K
$ReturnOnAssets$	Net interest income over total assets for a given bank.
$RuralBranch$	The branch that is located in a rural area based on $RuralIndicator$ variable.
$RuralIndicator$	The regions are classified into 10 groups based on the rural-urban commuting area (RUCA) codes. The value ranges from 1 to 10. The value 1 indicates the most urbanized areas. On the other hand, value 10 means the extremist rural areas.
$SingleStateBanks$	The banks that operate only in a single state.
$SmallBanks$	The banks which have total assets of less than $1.322 billion.
$UrbanBranch$	The branch that is located in an urban area based on $RuralIndicator$ variable.
$Z - score(SolvencyScore)$	The sum of ROA and the equity ratio over the three year standard deviation of ROA for a given bank.