Consumer Valuation of Network Convenience: Evidence from the Banking Industry*

Hui Wang
Guanghua School of Management
Peking University
Email: jackie.wang@gsm.pku.edu.cn

Andrew T. Ching
Carey Business School
Johns Hopkins University
Email: andrew.ching@jhu.edu

Xinyu Zhao
Department of Economics
Johns Hopkins University
Email: xzhao47@jhu.edu

First draft: December 21st, 2016

This draft: November 18th, 2021

Abstract

This paper develops and estimates a structural demand model which incorporates consumer preferences for geographical convenience of retailers' networks. Unlike the previous literature, which assumes consumers only care about retailer's outlet located closest to their home locations, our model explicitly allows consumers to consider two outlets of a retailer located nearest their home and work locations, respectively. We apply our model to the U.S. retail banking industry. To estimate our model, we use the U.S. Journey to Work dataset to obtain consumers' home and work locations in the population, and combine it with a dataset that details each branch's deposit and location of 132 isolated cities. The results show that consumers value the proximity of outlets to home and work locations roughly equally. In the counterfactual experiments, we evaluate the impact of various Work From Home policies on how consumers value a bank's branch network. In the baseline counterfactual experiment, where everyone is assumed to work from home, the market shares of banks specializing in serving work locations drop by 0.7 percentage point (which is 8.6% of the mean market share in the data).

Keywords: Spatial Competition; Retail Banking Network; Demand Estimation; Post COVID-19 Pandemic; Work From Home Policy

JEL Codes: C25, D14, G21, L11, L14, L89, M21

* Please direct all correspondence to Andrew T. Ching, andrew.ching@jhu.edu.

1. Introduction

It has long been recognized that the location of retail outlets plays a crucial role in product differentiation and spatial competition. The empirical literature of spatial demand has emphasized its importance in a wide range of industries.[1] As far as we know, all existing papers extend the model pioneered by Hotelling (1929) and assume consumers prefer retail outlets that are closer to a *single* location, often times their home location. We have learned a lot from this literature. However, a potentially important factor is ignored: many consumers also spend a significant amount of their time at other locations, in particular their workplaces. Hence, an empirical model that fails to incorporate consumers' both home and work locations could cause misleading inference in consumer preferences, the value of retailers' networks, and the nature of spatial competition.

In this paper, we contribute to this literature by developing and estimating a structural multi-location demand model for the U.S. retail banking industry. To estimate our model, we make use of data from the U.S. Department of Transportation's Journey to Work survey, data on branch locations and deposits, as well as average household income at the tract level. In particular, the Journey to Work dataset, which provides the number of workers who commute from one tract to another tract to work, allows us to obtain the share of workers' home and work locations at the tract level. Compared with previous research, which only considers consumers' home locations, our framework can better capture the *effective* distance between consumers and banks' branch networks, because our estimation results show that consumers value the proximity to branches from both home and work locations almost equally.

When the recent COVID-19 pandemic ends, many workers will likely be allowed to work remotely on a permanent basis, or in a hybrid format.[2] An interesting question is how such a change in the company policy would affect the value of retail networks to consumers. We use our demand model to conduct counterfactual experiments to answer this question. We implement various Work From Home policies by varying how it would apply to different income segments of the population (we use income thresholds as a proxy for the types of jobs that allow workers to work from home). In the baseline experiment, where we assume everyone works from home, our model predicts that (i) for banks specializing in serving work locations,

[1] This includes car dealers (Albuquerque and Bronnenberg, 2012), hospitals (Gaynor and Vogt, 2003), nursing homes (Ching et al., 2015), gasoline stations (Chan et al., 2007; Houde, 2012; Manuszak, 2010), supermarkets (Orhun, 2013; Smith, 2004), discount stores (Zhu and Singh, 2009), fast-food restaurants (Thomadsen, 2005; 2007), movie theatres (Davis, 2006), and retail banking (Bao and Ni, 2017; Dick, 2008; Ho and Ishii, 2011).

[2] https://www.forbes.com/sites/ashleystahl/2021/02/01/5-lasting-changes-to-expect-in-the-workplace-post-covid/?sh=28cd0f5a213d (assessed on August 22nd, 2021).

their market shares would reduce by 0.7 percentage point (8.6% of the mean market share in the data); (ii) by contrast, for banks specializing in serving home locations, their market shares would increase by 1.5 percentage points (30.6% of the mean market share in the data). When assuming only workers above certain income cutoffs would work from home, the impacts of this policy become smaller in magnitudes, but the qualitative results remain unchanged.

We should also highlight that our paper has significantly extended the retail banking demand estimation literature. In the previous works, Dick (2008) and Bao and Ni (2017) use branch density as a proxy for a bank's network size instead of explicitly taking distance into account; Ishii (2008) and Ho and Ishii (2011) incorporate the distance from consumers' homes to the closest and the second closest bank branches in the consumers' indirect utility function. Unlike our study, these previous works have ignored the possibility that consumers may want to use branches close to their work locations.[3]

We should point out that Houde (2012) also uses data similar to the Journey to Work dataset to study spatial competition in the retail gasoline market in Quebec City, Canada. He assumes that the distance between a gas station and consumer is measured by its shortest deviation from the commuting path between his/her home and work locations. Hence, in Houde's framework, each consumer has only one most preferred gas station per retail chain. In contrast, we assume that each consumer is characterized by two locations - home and work locations. Hence, each consumer can have up to two most preferred branches per bank: one for each location. Moreover, we allow consumers to value their home and workplace locations differently, whereas Houde's approach assumes that consumers treat all locations along the commuting path equal. While Houde's approach is appropriate for some retail markets such as retail gasoline market, we believe our two-location approach is more suitable for some other retail markets, including retail banking. In the case of banking services, consumers may want to visit branches during work breaks with minimum need of driving.

The remainder of the paper is organized as follows. Section 2 provides background information on the U.S. retail banking industry. Section 3 describes the data used in this paper. Section 4 presents the empirical model, the estimation algorithm, and the identification issues. Section 5 shows the estimation results and the counterfactual experiments. Section 6 concludes and discusses future directions.

[3] For supply side spatial competition models, Kuehn (2018, 2021) and Kumar (2018) consider firms' branch entry decisions in different cities, but their demand models also do not incorporate consumer work locations. The literature on firms' spatial competition and location choices needs to simplify the demand side. Often times, researchers assume a reduced form latent profit function and assume the demand side is embedded in the profit function without explicitly modeling it (e.g., Zhu and Singh, 2009; Holmes, 2011; Orhun, 2013; Yang, 2020).

2. Industry Background

Because of its long tradition against a nationwide banking system, the U.S. banking industry has a large number of independent and localized banking institutions.[4] After several waves of deregulation, the U.S. banking industry has undergone significant consolidation, marked by a 29 percent decrease in the number of banking institutions over the past two decades. However, in 2015 there are still about 6,350 banking institutions, which is much higher than the international norm.[5] During the same period, the number of physical branches steadily increased by 17 percent from 81,300 in 1994 to 94,720 in 2014. On average, the number of branches per bank had expanded considerably from 6.3 in 1994 to 14.2 in 2014 based on the data from FDIC Summary of Deposit.

The continued expansion of branches may seem surprising given that it happened in a period of rapid technological advancements in the U.S. banking industry. One may expect that the widely used automatic teller machines (ATMs) and the development of online banking through the internet or mobile phones should herald the demise of the branches. However, consumers still heavily rely on branches to access some important banking services. Using the Federal Deposit Insurance Corporation (FDIC) 2013 household survey, Burhouse et al. (2014) find that about a third of all bank consumers used bank tellers as their most common banking channel.[6] Even though the COVID-19 pandemic has pushed many consumers to use online banking, the retail banking industry still expects branches to remain an important channel to do business with customers when the economy re-opens. In particular, a recent survey conducted by Financial Brand suggests that closing a branch could lead to roughly 50% of the customers switching to another bank with a local branch that provides easy access (Cocheo, 2020). A physical branch has its advantages over other distribution channels in services that require more interactive communications, e.g., applying for personal loans, mortgages, opening

[4] Regulations imposing geographic restrictions on branching activities include the McFadden Act of 1927 and the Banking Act of 1933, both of which prohibited branching across state lines.

[5] The U.S. banking industry is very different from the rest of the world. In most countries (such as Canada, Germany, France, Japan, UK, etc.) a small number of large banks dominate the banking industry with nationwide branches. In 1994, the U.S. banking industry had about 10,000 commercial banks, 93% of which had less than $500 million in assets. The top ten largest banks had only 36.6% of the total assets, compared to Canada or UK, where 4 or 5 banks held about 80-90% of the total assets (Mishkin and Eakins, 2009).

[6] A survey conducted by Javelin in August 2012 finds the same trend. The survey shows that 40% of mobile bankers chose to deposit funds in person at a branch as their most preferred channel, whereas only 6% of mobile bankers chose mobile banking as their most preferred channel. https://www.javelinstrategy.com/coverage-area/leveraging-omnichannel-approach-drive-15b-mobile-banking-cost-savings (accessed on November 18, 2021).

saving accounts, etc. (Dick, 2006, 2008). For these reasons, branches remain the best place for banks to conduct complicated transactions (Burhouse et al., 2014).

As in other retail businesses, location convenience remains an important factor for consumers in the banking industry. As shown in the Federal Reserve Board's Survey of Consumer Finance from 1992 to 2013, around 45% of the consumers cited "location of branch office" as the primary reason for consumers to select their banks. Moreover, many survey studies report that consumers often cite convenient branch locations near where they live and work as the top reason for their bank choices (e.g., Kiser, 2002; Burhouse et al., 2014).

When consumers need banking services other than withdrawing cash, they need to visit branches of their own banks. Although it is possible to withdraw cash from other banks' ATMs, consumers need to incur a surcharge fee.[7] Survey evidence shows that most consumers avoid using ATMs of other banks. According to the 2003 consumer survey conducted by the American Bankers Association, nearly 70% of people who used ATMs regularly did not pay any surcharge fee at all, and about 10% paid less than $3 per month.

It is worth pointing out that even though a bank's ATM network may affect consumers' bank choices, branch network should play the first order importance because the vast majority of branches (if not all) have ATMs on site (Dick, 2007; Yang and Ching, 2014).

3. Data

We construct a cross-sectional dataset from various sources in 2000. Table 1 reports the summary statistics of our dataset.

3.1 Markets

To construct a set of markets for our analysis, we first select a set of middle-sized U.S. cities identified by the Census Bureau, with populations ranging from 40,000 to 180,000. Our market definition is the same as the definition of "city group" in Seim (2006), where she assigns neighboring cities to the same market if they lie within 10 miles of each other, and either share boundaries or consist of neighborhoods that partially overlap with both cities.[8] Markets with less than 40,000 population are excluded because the population in these cities tends to be concentrated and display limited variations in consumers' locations and commuting patterns. Markets with more than 180,000 population are also excluded because they may have more

[7] The average surcharge per transaction is $2.86 in 2001 according to the Public Interest Research Groups (PIRGs) national survey.
[8] We thank Katia Seim for sharing her city group sample based on the 1990 Census. We update the sample using the 2000 data.

public transportation options, which could affect consumers' preferences for geographic proximity.

We further restrict our attention to "isolated markets" by follow the criteria in Seim (2006): (i) the largest neighboring market within a 10-mile radius of the focal market's centroid has less than 15,000 population, and (ii) the largest neighboring market within a 20-mile radius of the focal market's centroid has less than 30,000 population. The distance between two markets is calculated using their population-weighted centroids with the Haversine formula (Sinnott, 1984).

If we select markets that are not isolated, there could be a non-trivial portion of the population commuting in between two adjacent markets, making it hard to determine the market size. By contrast, for isolated markets, one's home and work locations will be mostly confined within the market; we also do not need to worry about how to deal with branches located at the outskirts of markets because there are hardly any.

Our final dataset includes 132 isolated markets, with an average population of 97,883. As shown in Figure 1, the sample covers almost every state in the U.S. Each market is covered by a set of Census tracts. Census tracts are small and non-overlapping statistical subdivisions of counties, covering the entire U.S. The average population per tract is about 4,200. Since the boundaries of the markets do not necessarily coincide with the boundaries of the tracts, we select the minimum number of tracts that cover the entire city without significantly changing the market size in terms of either geographic area or population.[9]

We choose Census tracts as the basic consumer location unit within a market because (i) it is the finest geographic level at which Journey to Work data is available; (ii) Census tracts have been extensively used as the basic geographic unit in the empirical literature of retail banking (Chang et al., 1997; Ho and Ishii, 2011).

In our analysis, we assume that consumers live or work at the population-weighted centroid of each tract. The distance between a consumer and a branch is defined by the distance between the tract centroid where the consumer is located and the exact location of the branch. The centroid coordinates of tracts are available from the Census Bureau's geographic correspondence engine MABLE/Geocorr.[10] We believe centroids of tracts are reasonable

[9] We first identify tracts that have overlapping regions with the cities of interest. Then, we drop those tracts whose overlapping parts with the city contain either (i) less than 1 percent of the tract's population or (ii) less than 3 percent of the tract's area. Last, the cities' boundaries are expanded (if necessary) to include all of the remaining tracts.

[10] Available at https://mcdc.missouri.edu/geography.

proxies for consumer locations because tracts are usually very small, with 70% of the tracts having areas less than 10 square miles, equivalent to a circle with a radius of 1.8 miles.

Using the 2000 Census, we obtain demographic information (including population and income) at the tract level. We assume that the average income at the tract level applies to everyone who lives in that tract. The number of tracts per market ranges from 9 to 51, with an average of 23. Panel A of Figure 2 provides an example, Brownsville, Texas. The shaded area shows the original shape of the market, and the solid lines delineate the selected tracts.

3.2 Workers' Home and Work Locations

Workers' home and work location information is obtained from the Journey to Work (JTW) tables in the Census Transportation Planning Package (CTPP), from the U.S. Department of Transportation. The Census defines "workers" as people 16 years and older who were employed and at work, full-time or part-time, during the Census 2000 reference week. The dataset provides the number of workers who live in tract A (i.e., home location) and work in tract B (i.e., work location) for all markets in our sample.[11]

Panel C of Table 1 describes the commuting pattern of the workers at the market level. The first row calculates the percentage of workers who work in a tract different from their homes. Among the 132 markets in our sample, an average market has 85.40% of the workforce working in tracts different from where they live. The second row calculates the average commute distance at the market level, weighted by workers' income. On average, consumers travel 4.97 miles from their home to work, which is about 10 minutes of driving with ordinary traffic.

Panel B of Figure 2 illustrates the variations in the home/work locations for Brownsville, Texas. We calculate the tract-level population distribution when people are at home and work separately. We use the degree of darkness to indicate its population share (darker colors mean higher population shares). Panel B(i) shows the population distribution in Brownsville, Texas when consumers stay at home; Panel B(ii) shows the distribution when consumers go to work. It can be seen from these two sub-figures that many consumers commute from suburban areas to the downtown area, where tracts are usually zoned for commercial use.

Extending this analysis to all cities in our sample, we draw a scatter plot of tracts in all cities in Figure 3, where the x-axis (y-axis) shows the tract-level income-weighted population

[11] We assume banks' market shares among workers are the same as banks' market shares among the whole population. In Appendix A, we explain the conditions under which this assumption is valid.

distribution when people are at work (home).[12] If a tract has a higher income-weighted work population share, that means it has more people going there to work than living there, and vice versa. Figure 3 displays a large variation in tracts' specialization in residential vs. work purposes. This is consistent with Table 1, which shows that most workers (85.4%) live and work in different tracts. We will further discuss how banks spatially position themselves to serve different types of tracts in Section 3.6.

3.3 Branch Locations

For commercial banks and saving associations, branch level addresses (street, city, ZIP code) come from the 2000 Summary of Deposits (SOD), which are collected by the Federal Deposit Insurance Corporation (FDIC) on June 30th of each year. The corresponding information for credit unions comes from the Financial Performance Reports of the National Credit Union Association.

To assign banks and branches to a market, we geocode their addresses by coordinates. On average, there are 32 branches and 11 banks per market. Moreover, all markets are served by at least one credit union. The average market share of credit unions in the sample markets is 27 percent.

3.4 Deposits, Market Shares, and Bank Level Characteristics

For commercial banks and saving associations, we obtain branch-level deposit amounts and bank affiliations from SOD. We follow the literature (such as Dick, 2008 and Ho and Ishii, 2011) and study the demand for dollar deposits. The market size in this study is defined as the total amount of deposits in commercial banks, savings associations and credit unions. Market shares are calculated in terms of the *amounts of deposits* as well.

Detailed bank-level characteristics are derived from balance sheets and income statements in banks' quarterly reports for regulation agencies: the Federal Reserve Board (FRB)'s Report on Condition and Income (CALL Reports) for commercial banks, and the Office of Thrift Supervision (OTS)'s Thrift Financial Report (TFR) for savings associations.

Annual deposit interest rates are calculated as the ratio of annual interest expenses to deposits for each bank. By construction, there is no variation in the interest rate within an institution across markets. This clearly holds for banks with a single branch and banks which operate in one market. In addition, studies based on directly-measured interest rates found that

[12] For a more detailed definition and a numerical example, refer to Appendix C.

multi-market banks generally offer the same interest rate within one state and very similar interest rates across states (e.g., Radecki, 1998; Heitfield, 1999). We therefore believe our interest rate measure is reasonable.

Our bank characteristics include (i) the age of the bank (measured by the number of years since established), (ii) the size of the bank (measured by the total institution assets),[13] (iii) the average number of employees per branch, (iv) the number of states in which the bank operates, and (v) a dummy variable indicating if a bank has only one branch in the market. The age of the bank may reflect its experience, expertise, and perceived reputation in the market. The total assets of the bank serve as a proxy for the range of benefits consumers may receive, e.g., better facilities, products, and services. The number of employees per branch serves as a proxy for service quality, as more employees per branch normally means shorter waiting time and more personal interactions. The number of states that a bank has presence provides a measure of the geographic diversification of the bank. Finally, the dummy for a single-branch bank captures the possibility that such banks may specialize in services which do not require geographic convenience.

3.5 Distance

The literature typically uses the spherical distance (also known as the air distance) to measure how far away two locations are. It is calculated by the Haversine formula (e.g., Sinnott, 1984). However, when measuring the distance between the branches and Census tracts, the spherical distance tends to underestimate the actual travel distance because it ignores the exact road network and conditions between the destinations. To address this concern, we obtain the driving distance for all the 312,253 tract-branch combinations in our sample using the Google Maps Application Programmable Interface (API). Appendix B details how the driving distance is constructed and compares it with the spherical distance.

3.6 Preliminary Evidence for the Importance of Including Work Locations

To gain some insights about how banks position themselves, we will show to what extent banks serve tracts that are primarily home locations vs. work locations. We first classify tracts based on the income-weighted home and work population shares. Intuitively, we classify a tract as *home* (*work*) tract if it is primarily for residential (work) purpose. If a tract falls somewhere in

[13] Bank sizes are classified it into three categories: (i) small banks with assets less than $100 million, (ii) mid-sized banks with assets between $100 million and $300 million, (iii) large banks with assets more than $300 million.

between, we classify it as *other* tract. The details of our definition and classification procedure are in Appendix C. Our classification leads to 474 home tracts, 676 work tracts, and 1929 other tracts.

We next document how banks in our sample position themselves in the market. For any given market, we classify a bank as *home*, *work*, or *other* bank based on how many home tracts and work tracts it serves.[14] A *home* (*work*) bank is a bank that serves at least one home (work) tract and no work (home) tract. If a bank is neither a home bank nor a work bank, we classify it as *other* bank. Table C1 in Appendix C shows that there are 42 home banks, 743 work banks, and 703 other banks based on our classification. This classification reveals that banks are much more likely to set up branches that target work areas. Hence, we expect the recent Work From Home policies during the COVID-19 pandemic can lead to significant changes in how consumers value branch networks. We will use our counterfactual experiments to shed light on this issue in Section 5.2.

4. The Model

4.1 The Demand Model

We now turn to our structural demand model. Consider a sample of markets $m = 1, \ldots, M$, each with I_m consumers and J_m banks. Throughout this section, vectors are represented by boldface letters, scalars are represented by regular lowercase letters, and matrices are represented by capital letters.

Each consumer $i \in I_m$ earns income y_i annually. Consumers consider two locations where they might access banking services: home (H) and work (W). Let $\boldsymbol{LC_i} = (LC_i^H, LC_i^W)$ denotes the geographic coordinates of these two locations for consumer i.[15] Bank $j \in J_m$ has B_{jm} branches in market m, and their locations are characterized by the coordinate vector $\boldsymbol{LB_{jm}} = \{LB_{jm}^b\}_{b=1}^{B_{jm}}$, where superscript b is the index for a branch. The indirect utility of consumer i from choosing bank j is,

$$u_{ijm} = \beta^r r_j + \boldsymbol{x_j'\beta} + \gamma_i \cdot D_{ijm} + \xi_{jm} + \xi_m + \epsilon_{ijm}, \quad (1)$$

where r_j is the deposit interest rate of bank j; $\boldsymbol{x_j}$ is a vector of observable bank-level characteristics that determine the quality of banking services; D_{ijm} denotes the *effective*

[14] This definition is bank-market specific. To put it another way, our definition allows for the possibility that a bank which operates in multiple cities is a home bank in one city, but a work bank in another.
[15] Empirically, these are the coordinates of the centroids of the tracts where the consumer live and work respectively.

distance between bank j and consumer i in market m (which is defined below); β^r, $\boldsymbol{\beta}$ and γ_i are consumer i's preferences on the characteristics; ξ_{jm} captures bank j's unobserved characteristics in market m; ξ_m is the market fixed effect, which captures the popularity of the outside option in market m; ϵ_{ijm} is the idiosyncratic random utility which is known to consumers but unobserved to researchers.

For notational simplicity, the market subscript m is suppressed throughout the remainder of this subsection. Readers should bear in mind that most variables, except for interest rate r_j and bank level characteristics x_j, are market specific. The indirect utility in Eq(1) can then be expressed as,

$$u_{ij} = \beta^r r_j + \boldsymbol{x_j'\beta} + \gamma_i D_{ij}(\boldsymbol{LC_i}; \boldsymbol{LB_j}) + \xi_j + \epsilon_{ij}. \tag{2}$$

Let $\omega(\cdot,\cdot)$ denote the distance between the two coordinates. Then,

$$D_{ij} = \rho \cdot \min_{b\in\{1,\dots,B_j\}} \omega(LC_i^H, LB_j^b) + (1-\rho) \cdot \min_{b\in\{1,\dots,B_j\}} \omega(LC_i^W, LB_j^b), \tag{3}$$

where $\rho \in [0,1]$ captures the relative importance of home and work locations. When $\rho = 1$, our model becomes the single-location model in the literature, where only consumers' home locations matter. We assume that ρ is homogeneous for all consumers.[16]

We assume that the individual specific marginal disutility in the effective travel distance is γ_i. We further assume that $\gamma_i = \gamma_0 + \gamma_1 \cdot (y_i - \bar{y})$, where y_i is the income of consumer i and $\bar{y}$ is the average income in the population; γ_0 represents the mean level of the marginal disutility from traveling; γ_1 captures the effect of an increase in income on this marginal disutility. We expect γ_0 and γ_1 to be both negative, as consumers with higher income tend to have higher opportunity cost of time.

The specification of the demand system is completed by defining an outside option. Following Dick (2008), we choose the credit unions to be the outside option. The indirect utility from choosing the outside option is:

$$u_{i0} = \xi_0 + \epsilon_{i0}\ , \tag{4}$$

where ξ_0 is the mean utility of choosing the outside option. We will normalize ξ_0 to zero for identification reasons.

Consumers in our model can be characterized by a vector $(\boldsymbol{LC_i}, y_i, \boldsymbol{\epsilon_i})$, where $\boldsymbol{\epsilon_i} = (\epsilon_{io}, \epsilon_{i1}, \dots, \epsilon_{iJ})$. Consumer i chooses bank j if

[16] We also consider heterogenous ρ_i by allowing it to depend on the income of individual i, i.e., $\rho_i = \rho_0 + \rho_1 y_i$. The estimated ρ_1 is not significantly different from zero, and the estimates on other coefficients are similar to the results presented in Section 5.1. The results from this alternative model are available upon request.

$$u_{ij} > u_{ik}, \forall k \in \{1, 2, \ldots, J, 0\} \text{ and } k \neq j. \quad (5)$$

To simplify notations, rewrite the utility function in Eq(2) as

$$u_{ij} = \delta_j + \gamma_i D_{ij} + \epsilon_{ij}, \quad (6)$$

where $\delta_j = \beta^r r_j + \boldsymbol{x}_j' \boldsymbol{\beta} + \xi_j$ is the common utility that consumers in the same market receive from patronizing bank j. Assuming that ϵ_{ij} follows an i.i.d. Type I Extreme Value distribution, the probability of choosing bank j can be expressed as

$$P_{ij} = \frac{\exp(\delta_j + \gamma_i D_{ij})}{1 + \sum_{k=1}^{J} \exp(\delta_k + \gamma_i D_{ik})}. \quad (7)$$

As mentioned in Section 3.4, we follow the literature and construct the market shares using the amounts of deposits in dollar terms. The predicted market share of deposits for bank j, denoted by s_j, is therefore

$$s_j = \frac{\int dep_i \cdot P_{ij} \, dG(\boldsymbol{LC_i}, y_i, dep_i)}{\int dep_i \, dG(\boldsymbol{LC_i}, y_i, dep_i)}, \quad (8)$$

where dep_i is the amount of deposits made by consumer i, and $G(\boldsymbol{LC_i}, y_i, dep_i)$ is the distribution of $(\boldsymbol{LC_i}, y_i, dep_i)$. However, there is one complication -- we do not observe dep_i in the data. We therefore assume that dep_i is a function of consumer i's income y_i. More specifically, as a first-order approximation, we assume that $dep_i = \alpha \cdot y_i$. Although this functional form may seem restrictive, it is a common approach taken by the existing demand studies for the retail banking industry (e.g., Ishii, 2008, and Ho and Ishii, 2011). An advantage of this approach is that α does not need to be estimated since it is canceled out in the calculation:

$$s_j = \frac{\int \alpha \cdot y_i \cdot P_{ij} \cdot dF(\boldsymbol{LC_i}, y_i)}{\int \alpha \cdot y_i \cdot dF(\boldsymbol{LC_i}, y_i)} = \frac{\int y_i \cdot P_{ij} \, dF(\boldsymbol{LC_i}, y_i)}{\int y_i \, dF(\boldsymbol{LC_i}, y_i)}, \quad (9)$$

where $F(\boldsymbol{LC_i}, y_i)$ is the distribution of $(\boldsymbol{LC_i}, y_i)$. Two assumptions that we mentioned earlier will simplify the integration in Eq(9): (i) the centroid of a tract represents the locations of all consumers in this tract; (ii) the average income of a tract applies to all consumers in this tract. With these two simplifications, we can then construct the distribution of consumers, $F(\boldsymbol{LC_i}, y_i)$, and the effective distance between consumer i and bank j, D_{ij}. Because there are a finite number of tracts, we can rewrite the integration in Eq(9) using summation. Appendix D provides the details.

In our demand model, the price elasticity of bank j is given by

$$\eta_{jk} = \frac{\partial s_j}{\partial r_k} \cdot \frac{r_k}{s_j} = \begin{cases} \beta^r \cdot \dfrac{r_j}{s_j} \cdot \dfrac{\int y_i P_{ij}\left(1 - P_{ij}\right) dF(\boldsymbol{LC_i}, y_i)}{\int y_i \, dF(\boldsymbol{LC_i}, y_i)}, & \text{if } j = k \\ -\beta^r \cdot \dfrac{r_k}{s_j} \cdot \dfrac{\int y_i P_{ij} P_{ik} \, dF(\boldsymbol{LC_i}, y_i)}{\int y_i \, dF(\boldsymbol{LC_i}, y_i)}, & \text{otherwise} \end{cases} \quad (10)$$

It should be emphasized that because we model the distances from consumers to branches from both their home and work locations, our model allows for more general substitution patterns compared with previous work. In particular, when one bank drops its interest rate, its customers may switch to other banks with branches not too far away from either their home or their work locations. In contrast, Ho and Ishii (2011) only models branches' distances to consumers' home locations. By incorporating the work locations, our model can capture heterogeneous consumer preferences for banks' network convenience in a more general way.

4.2 Estimation

4.2.1 Algorithm

To estimate the model in Section 4.1, we follow Berry, Levinsohn and Pakes (1995) (hereafter BLP). We use ξ_j to capture bank j's unobserved characteristics (e.g., brand recognition) observed by consumers but not by researchers.[17] We allow deposit interest rate to be endogenous, but assume branch locations are exogenous. We will explain our instruments in Section 4.2.2.

The set of parameters to be estimated is $(\beta^r, \boldsymbol{\beta}, \gamma_0, \gamma_1, \rho)$, denoted by $\boldsymbol{\theta}$. Partition $\boldsymbol{\theta}$ into two parts, i.e., $\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \boldsymbol{\theta_2})$, where $\boldsymbol{\theta_1} = (\beta^r, \boldsymbol{\beta})$ contains all the linear parameters, and $\boldsymbol{\theta_2} = (\gamma_0, \gamma_1, \rho)$ consists of the non-linear parameters that govern consumers' perception about distances. We stack the unobserved bank qualities ξ_{jm} in market m into a vector $\boldsymbol{\xi_m} = \left(\xi_{1,m}, \dots, \xi_{J_m,m}\right)'$, and further stack ξ_m in all markets into $\boldsymbol{\xi} = (\boldsymbol{\xi_1'}, \dots, \boldsymbol{\xi_M'})'$. There are 132 markets in the data and 1,488 market-bank combinations, so $\boldsymbol{\xi}$ is of dimension 1,488. All other market-bank level variables are stacked in the same way, including $\boldsymbol{\delta}$, X, and Z, where $\boldsymbol{\delta}$ is the common utility defined in Eq(6), X are the firm-level observed characteristics, and Z are the instrumental variables. We will discuss our choice of instruments in the next subsection.

To obtain the parameter estimates, we minimize the following GMM objective function:

[17] We do not include advertising in our study. For a study which explicitly estimates the effects of advertising on bank choices, see Honka et al., (2017).

$$\min_{\theta} \; Z'\xi(\theta) W Z'\xi(\theta), \tag{11}$$

where W is a positive semi-definite weighting matrix. We will start with an initial guess of the parameter vector and carry out the following steps in the estimation algorithm.

Step 1: For any given $\boldsymbol{\theta_2}$, find the mean bank qualities $\boldsymbol{\delta}$ such that for all markets m and all banks j in the market,

$$s_{jm}(\boldsymbol{\delta}, \boldsymbol{LB}; \theta_2) = S_{jm}, \tag{12}$$

where S_{jm} is the observed market share in the data, and s_{jm} is the market share predicted by Eq(9). Eq(12) can be inverted numerically to solve for $\boldsymbol{\delta}$ using the contraction mapping proposed by BLP. Denote the unique solution to Eq(12) by $\boldsymbol{\delta}(\boldsymbol{\theta_2})$.

Step 2: After obtaining the mean bank qualities $\boldsymbol{\delta}(\boldsymbol{\theta_2})$ from Step 1, $\boldsymbol{\theta_1} = (\beta^r, \boldsymbol{\beta})$ can be obtained from the orthogonality condition that $\mathrm{E}\left[z_j\left(\delta_j(\boldsymbol{\theta_2}) - \beta^r r_j - \boldsymbol{x_j}'\boldsymbol{\beta}\right)\right] = \boldsymbol{0}$ using the generalized method of moments (GMM) estimator, i.e.,

$$\min_{\beta^r, \boldsymbol{\beta}} Z'(\boldsymbol{\delta}(\boldsymbol{\theta_2}) - \beta^r \boldsymbol{r} - X\boldsymbol{\beta}) W Z'(\boldsymbol{\delta}(\boldsymbol{\theta_2}) - \beta^r \boldsymbol{r} - X\boldsymbol{\beta}), \tag{13}$$

Denote the value of Eq(13) by $J(\boldsymbol{\theta_2})$.

Step 3: Search over $\boldsymbol{\theta_2}$ to minimize $J(\boldsymbol{\theta_2})$. In this search process, each trial value of $\boldsymbol{\theta_2}$ requires to redo Steps 1 and 2. Hence, the calculations done in these two steps will need to be repeated many times until convergence.

Note that the effect of $\boldsymbol{\theta}$ on bank unobservables $\boldsymbol{\xi}$ can be decomposed into two parts, i.e., $\boldsymbol{\xi}(\boldsymbol{\theta}) = \boldsymbol{\xi}(\boldsymbol{\delta}(\boldsymbol{\theta_2}), \boldsymbol{\theta_1})$. The above procedure exploits this property to significantly reduce the computation time by searching only in the space of $\boldsymbol{\theta_2}$.

This procedure is repeated three times. In the first iteration, the weighting matrix W is taken to be $(Z'Z)^{-1}$, and the GMM estimator is equivalent to the 2-Stage Least Square estimator. In the second and third iterations, the weighting matrix W is updated to be an estimate of the efficient weighting matrix using the parameters from the previous round. The difference between the 2-step and 3-step estimators is small, indicating a fast convergence. The results from the 2-step estimator are presented in Section 5.

4.2.2 Instruments

We allow banks' interest rates to be endogenous, and hence we need to come up with instruments that are mean independent of ξ_j. Our first set of instruments are cost shifters, which include: (i) operating expenses, (ii) other expenses, (iii) non-performing loans, (iv) equity-asset ratio, and (v) an indicator of whether the bank belongs to a bank holding company (BHC).[18] The first four variables are normalized by bank assets. Operating expenses include premises and equipment expenses on utilities, building maintenance, and ordinary repairs. Other expenses include legal fees, postage, deposit insurance assessments, amortization of assets, directors' fees, etc. Non-performing loans provide a measure of the amount of bad debt carried by a bank. The equity-asset ratio measures a bank's financial stability. Lastly, banks that belong to a holding company may have better access to certain internal capital pooling arrangements, which can lower the costs of raising funds and improve their liquidity positions (Houston et al., 1997). All these instruments could potentially affect a bank's ability to offer a competitive interest rate. We obtain these variables from the Federal Reserve Board's Reports of Condition and Income (CALL Reports) and the Office of Thrift Supervision (OTS) Thrift Financial Report (TFR). Our identification assumption (exclusion restrictions) is that these five cost shifters only affect a bank's marginal cost, but do not enter a consumer's utility function.

If banks compete in a Bertrand pricing game, the characteristics of competitors in the local market could influence a bank's pricing decision, but are independent of the bank's unobserved product characteristics. As a result, these characteristics can also serve as instruments for prices, as shown in BLP. Hence, our second set of instruments are the average value of the same set of characteristics included in the utility function of other competitors in the market.

Our third set of instruments are the average distances of rival banks to consumers in each market as additional instruments for the interest rate of bank j. More specifically, the competitors' average distance to consumers' home locations is defined as,

$$\overline{D}_j^H = \frac{1}{J-1}\sum_{k\neq j} D_k^H, \tag{14}$$

where $D_k^H = \frac{1}{I}\sum_i \min_{b\in\{1,\ldots,B_j\}} \omega(LC_i^H, LB_k^b)$ is bank k's average distance to consumers' home locations, and the competitors' average distance to consumers' work locations is defined similarly as,

$$\overline{D}_j^W = \frac{1}{J-1}\sum_{k\neq j} D_k^W, \tag{15}$$

where $D_j^W = \frac{1}{I}\sum_i \min_{b\in\{1,\ldots,B_j\}} \omega(LC_i^W, LB_j^b)$ is bank k's average distance to consumers' work locations.

Note that we follow the literature (e.g., Dick, 2008; Ho and Ishii, 2011) and assume that branch locations and consumer locations are exogenous. Hence, the effective distances between

[18] The set of instruments chosen are similar to those used in Dick (2008) and Ho and Ishii (2011).

consumers and banks are also exogenous in our model. The assumption of exogenous branch locations in our static demand framework is motivated by the fact that, unlike setting the interest rate, entry and exit decisions require much longer-term planning (e.g., zoning restrictions and lease terms). The assumption of exogenous consumer locations (in the sense that they are uncorrelated with banks' ξ_j's) is motivated by our intuition that it seems unlikely for consumers to decide where to live and work mainly based on banks' branch networks. Hence, it follows that $E(\overline{D}_j^H \xi_j) = 0$ and $E(\overline{D}_j^W \xi_j) = 0$, which allow us to use $\overline{D}_j^H$ and $\overline{D}_j^W$ as instruments as well.

4.2.3 Identification of $\boldsymbol{\rho}$

The parameter ρ measures the relative importance of home and work locations. Its identification relies on the variation in the distribution of consumers' home-branch and work-branch distances across banks and across markets. Intuitively, ρ determines how the model maps these distributions to market shares, after controlling for other banks' characteristics and market fixed effects.

We should note that the correlation between the distance of the closest branch to home and that to work is 0.295, which indicates that they are only mildly correlated. Hence, we should be able to identify their relative importance in our model.

5. Estimation Results

5.1 Single and Multi-location Models

Table 2 reports the estimation results of the structural demand model, with Columns (1) and (2) for the single-location model (by imposing that consumers only care about the distance from home, i.e., $\rho = 1$), and Columns (3) and (4) for the multi-location model.

There are three main takeaways. First, for the multi-location model, the utility weight on the closeness to home (ρ) is 0.553, and is statistically significant. This suggests that an average consumer values the geographic convenience to home and work roughly the same, and ignoring the consumers' work locations could lead to misleading results.

Second, the coefficient on the mean level of utility from travel distance (γ_0) is negative and significant, as expected, but its magnitude is larger in the multi-location model than in the single-location model. Regarding the heterogeneity in the disutility from travel due to one's income level (γ_1), it is negative and significant in the multi-location model, indicating that consumers who have higher income dislike traveling more. This is intuitive because the opportunity cost of time typically increases with income. Interestingly, in the single-location model, γ_1 is very close to zero and insignificant, which is counterintuitive.

To put the magnitudes of γ_0 and γ_1 into perspective, in Column (3) of Table 2, the ratio of γ_0 to the deposit interest rate coefficient (β^r) is -1.715/1.565=-1.096, suggesting that an average consumer values having "one mile" closer to a bank to be equivalent to an increase in the annual interest rate by 1.096 percentage point. Moreover, for every $1,000 increase in annual income, consumers would need an additional 0.06 percentage point increase in interest rate to do an equivalent trade-off.

The coefficients on other bank-level characteristics also have the expected signs. Banks with more employees per branch, multiple branches per market, larger assets, and operating in more states tend to attract more deposits.

We also use Eq(10) to calculate the interest rate elasticity for each bank. The results are summarized in Table 3. The mean interest rate elasticity of demand (in terms of market share) is 2.75, which is consistent with the estimates in the literature. For example, the estimates from Dick (2008) range from 2 to 3, and the median estimate from Adam et al. (2007) is about 3.47. Table 3 also shows that banks with only one branch in the local market are more price sensitive compared to banks with multiple branches. This is consistent with Panel A of Figure 4, which shows that the interest rate elasticity is negatively correlated with the number of branches of a bank. This suggests that banks with more branches also tend to have shorter effective distance to consumers, allowing them to remain competitive even if their interest rates are lower than other competitors with fewer branches. However, this competitive advantage could change if consumers can work from home, making branches near work locations to become less relevant. Our counterfactual experiments will investigate this possibility.

With our structural model, we can also produce the average consumer characteristics of any given bank, X_j., weighted by consumer choice probabilities. This allows us to gain more insights about what types of customers a bank can attract. For each bank j, we calculate

$$X_j = \frac{\int x_i^c \cdot P_{ij} \cdot y_i \, dF(\boldsymbol{LC_i}, y_i)}{\int P_{ij} \cdot y_i \, dF(\boldsymbol{LC_i}, y_i)} \tag{16}$$

where x_i^c is any given consumer characteristic; P_{ij} is the choice probability that consumer i chooses bank j; y_i is the income of consumer i.

We report four (bank-level) average consumer characteristics in Table 3 and Figure 4. The first one is "the percentage of bank j's customers who work and live in different tracts." In this case, $x_i^c = I(LC_i^H \neq LC_i^W)$ in Eq(16), where $I(\cdot)$ is an indicator function for whether consumer

i lives and work in different tracts. As shown in Table 3, on average 82% of banks' customers work and live in different tracts, and this number is significantly higher for multi-branch banks (85%) than for the single-branch banks (78%). This is also consistent with Panel B of Figure 4. This reinforces the point made above -- multi-branch banks are better positioned to cater consumers who need to commute to a different tract for work.

It is possible that consumers who live and work in different tracts may only have one most preferred branch if it happens that the same branch is closest to both their home and work locations. If this is common, single-location and multi-location models will not make much difference. Hence, to show to what extent the multi-location model matters, the second statistic of interest is the "share of customers who patronize two *different* branches of the bank". In this case, we set,

$$x_i^c = I\{\underset{b\in\{1,\ldots,B_j\}}{\operatorname{argmin}}\ \omega(LC_i^H, LB_j^b) \neq \underset{b\in\{1,\ldots,B_j\}}{\operatorname{argmin}}\ \omega(LC_i^W, LB_j^b)\} \qquad (17)$$

in Eq(16). By construction, it is 0 for single-branch banks. For a multi-branch bank, the mean value is 57% (see the third row in Table 3). This implies that more than half of the consumers of a multi-branch bank use two branches in its network. This demonstrates the multi-branch banks' advantage by providing more access to commuting consumers.

The third statistic of interest is the "average distance of customers from home to work". We can see that single-branch banks' customers usually work closer to their homes, compared to multi-branch banks' customers. On average, the former travel 1.1 miles (or 36%) shorter from home to work than the latter (see the fourth row of Table 3).

The last statistic of interest is the "average effective distance from customers to the bank". We set $x_i^c = D_{ij}$ in Eq(16), where D_{ij} is as defined in Eq(3). On average, a consumer needs to travel 2.52 miles to access their bank's services. The average distance to a multi-branch bank is much shorter than that to a single-branch bank (2.04 vs. 3.37 miles, shown in the fourth row of Table 3). As shown in Panel E of Figure 4, the effective distance decreases with the number of branches.

5.2 Counterfactual: The Work From Home (WFH) Experiment

The recent COVID-19 pandemic has forced many workers to work from home. Within a short period, the vast majority of office workers have adopted video conferencing apps and worked remotely from home. Companies are now planning whether to allow their workers to work from home permanently or in some kinds of hybrid format (e.g., a few days per week). Such a change in work arrangements has important implications on the value of retailers' networks. In particular, branches located around office buildings may generate much less business. To shed some light on how this change in company policies could affect the value of retail banking networks, we conduct three counterfactual experiments: Experiment 1 (CF1) assumes everyone can work from home, i.e., setting $\rho = 1$ for all individuals; Experiment 2 (CF2) assumes only individuals with income above the 33rd percentile work from home; Experiment 3 (CF3) assumes only individuals with income above the 67th percentile work from home.[19] For (CF2) and (CF3), we use income as a proxy for jobs that are more likely office-based and try to depict the scenario where only those individuals can work from home. This is done by changing ρ to 1 for individuals whose income satisfies the threshold.

Which types of banks will benefit or suffer from such work from home arrangements? We summarize our counterfactual findings based on home banks, work banks, and other banks defined in Section 3.6. Table 4 and Figure 5 show the results of the three counterfactual experiments.

In sum, the findings are consistent across all three counterfactual experiments. We find that work banks tend to suffer and home banks tend to benefit from the Work From Home arrangements. When everyone works from home, work banks' market shares on average reduce by 0.7 percentage point, which is 8.6% of the sample mean. By contrast, home banks' market shares increase by 1.5 percentage points, which is 30.6% of the sample mean. The results are robust even when we consider only individuals with income above the 33rd or 67th percentile work from home. As expected, the magnitude of the changes shrinks because there are fewer consumers who are affected by the policy, but the directions of the changes remain unchanged. Interestingly, the average impact of this policy on *other* banks (recall that other banks are those which do not have any branch in either home or work tracts) is also positive. This is primarily because work banks lose their competitive advantage. Note that the null hypotheses that the three bank types are affected equally in the counterfactuals are all rejected in a t-test.

Besides comparisons on the mean effects, we also compare the distributions of changes in the counterfactuals by bank type. Figure 5 shows that the impact of Work From Home policies is more severe for work banks than on other types of banks. All Kilmogorov-Smirnov tests that any two distributions are the same (e.g., the distribution of changes in market shares for work banks and that for home banks when everyone works from home are the same) are

[19] Notice that the traditional single-location model, where consumers only enjoy the convenience to home locations, cannot answer such a question. We need to rely on our multi-location model to investigate how banks' market shares change under this counterfactual experiment.

rejected at the 5% significance level. Similar to the comparison on the mean values, the differences by bank type are larger when more people are allowed to work from home.

We also investigate how the changes in market shares vary with the number of branches under our counterfactuals. In figure 6, we plot the change in market share against the number of branches for different Work From Home scenarios. There is a significant negative relationship between change in market share and number of branches in the local market, indicating that banks with larger branch network tend to suffer more from the Work From Home polices. This is mainly because larger banks are more likely to have branches close to the work tracts. Specifically, when all workers work from home, the average marginal effect of having an additional branch is -0.29 percentage point in market share. This converts to roughly 2.9 million dollars in deposits on average, given that the average total deposits in each market is about one billion dollars. In particular, among the 21 banks with more than 10 branches in the local market, on average they lose 11 percent of their existing market share if all workers work from home. This pattern is robust for the counterfactuals which allow workers with different income levels to work from home. The only subtle difference is that the less workers are allowed to work from home, the smaller the negative impact on larger banks.

We should note that the main point of our analysis is to illustrate how one can use our structural model as a tool to evaluate the impact of various Work From Home company policies. There are certainly other possible Work From Home policies that are worth investigating, and one can change the parameters of our model in different ways to examine their impacts.

6. Concluding Remarks

In this paper, we develop and estimate the first structural multi-location consumer demand model for the retail banking industry. Our multi-location demand model recognizes that consumers can utilize different branches of the same bank for services, and it is this important feature that differentiates our model from the previous literature. The lack of data on consumers' most visited locations has long hindered the research on consumers' preferences for geographic convenience of retail networks. We address this issue by utilizing the U.S. Journey to Work data, which contains detailed information on the two main locations at which consumers spend most of their time: home and work. This crucial piece of information, along with information on branch locations and bank market shares across markets, reveal how consumers value their home and work locations, and improves our understanding about the role of banks' branch networks in spatial competition.

By estimating our model using data for a sample of isolated markets in the U.S., we find evidence that consumers value their home and work locations almost equally. Moreover, consumers are willing to tradeoff one mile of effective travel distance for about one percentage point increase in annual interest rate.

Motivated by the current COVID-19 pandemic situation, we conduct counterfactual experiments where firms implement various Work From Home policies, which apply to workers with income higher than certain levels. Comparing the simulated market share outcomes under the counterfactuals with those in the actual scenario, we find that banks with larger networks would likely suffer while banks with fewer branches may benefit. These counterfactual results illustrate how one can use our structural model as a tool to evaluate the impact of the recent Work From Home company policy. It should be highlighted that the previous research, which only considers consumers' home locations (similar to our single-location model), is not designed for these counterfactual experiments. This is because such models have always assumed that work locations do not matter in consumer's decision-making process in the first place.

One caveat of our Work From Home counterfactual experiments is that we use data from 2000 to estimate our model. Readers should view our counterfactual experiment as an exercise to illustrate what research questions our framework could answer. Future research can use more recent data to estimate this model and re-run our counterfactual experiments. Another direction for future research is to extend our framework to study consumer preferences for retail networks in other industries. Chain retailers, such as fast-food restaurants, coffee shops, supermarkets, department stores, etc., often have loyalty programs. Consumers can only redeem points at outlets of the same chain but not across chains. Hence, loyalty programs generate a similar type of network effects. By considering consumers' multiple locations and explicitly modeling how it interacts with the retail networks, we hope this research direction can help us improve our understanding about spatial competition and estimate the value of a retailer's network more accurately. This can be very useful for retailers when they consider expanding or contracting their networks, and for us to quantify its welfare implications.

Reference

Albuquerque, Paulo and Bart Bronnenberg (2012): "Measuring the Impact of Negative Demand Shocks on Car Dealer Networks," *Marketing Science*, 31(1), 4-23.

Amel, Dean F., and Martha Starr (2002): "Market Definition in Banking: Recent Evidence," *The Antitrust Bulletin*, 63, 63-89.

Bao, Weining and Jian Ni (2017): "Could Good Intentions Backfire? An Empirical Analysis of the Bank Deposit Insurance," *Marketing Science*, 36(2), 301-319.

Berry, Steven, James Levinsohn, and Ariel Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 841-890.

Burhouse, Susan, Matthew Homer, Yazmin Osaki, and Michael Bachman (2014): "Assessing the Economic Inclusion Potential of Mobile Financial Services," whitepaper, Division of Depositor and Consumer Protection, Federal Deposit Insurance Corporation.

Chan, Tat Y., V. Padmanabhan, and P.B. Seetharaman (2007): "An Econometric Model of Location and Pricing in the Gasoline Market," *Journal of Marketing Research*, 44(4), 622-635.

Chang, Angela, Shubham Chaudhuri, and Jith Jayaratne (1997): "Rational Herding and the Spatial Clustering of Bank Branches: An Empirical Analysis," Federal Reserve Bank of New York Research Paper, No. 9724.

Ching, Andrew T., Fumiko Hayashi, and Hui Wang (2015): "Quantifying the Impacts of Limited Supply: The Case of Nursing Homes," *International Economic Review*, 56(4), 1291-1322.

Cocheo, Steve (2020): "Branches Still Dominate, But You Won't Need as Many to Compete," Financial Brand. Available at:

https://thefinancialbrand.com/97734/covid-19-coronavirus-pandemic-branch-digital-competition-mobile-online/ [Accessed on July 5, 2020.]

Davis, Peter (2006): "Spatial Competition in Retail Markets: Movie Theaters," *The RAND Journal of Economics*, 37, 964-982.

Dick, Astrid A. (2006): "Nationwide Branching and Its Impact on Market Structure, Quality, and Bank Performance," *The Journal of Business*, 79, 567-592.

Dick, Astrid A. (2007): "Market Size, Service Quality, and Competition in Banking," *Journal of Money, Credit and Banking*, 39, 49-81.

Dick, Astrid A. (2008): “Demand estimation and consumer welfare in the banking industry,” *Journal of Banking & Finance*, 32, 1661-1676.

Gaynor, Martin and William B. Vogt (2003): “Competition among Hospitals,” The *RAND Journal of Economics*, 34(4), 764-785.

Heitfield, Erik A. (1999): “What Do Interest Rate Data Say about the Geography of Retail Banking Markets?” *The Antitrust Bulletin*, 44, 333-347.

Ho, Katherine and Joy Ishii (2011): “Location and Competition in Retail Banking,” *International Journal of Industrial Organization*, 29(5), 537-546.

Holmes, Thomas J. (2011): “The Diffusion of Wal-Mart and Economies of Density,” *Econometrica*, 79(1), 253-302.

Honka, Elisabeth, Ali Hortacsu, and Maria Ana Vitorino (2017): “Advertising, Consumer Awareness, and Choice: Evidence from the US Banking Industry,” *The RAND Journal of Economics*, 48(3), 611-646.

Hotelling, Harold (1929): “Stability in Competition,” *Economic Journal*, 39, 41-57.

Houde, Jean-Francois (2012): “Spatial Differentiation and Vertical Mergers in Retail Markets for Gasoline,” *American Economic Review*, 102(5), 2147-2182.

Houston, Joel, Christopher James, and David Marcus (1997): “Capital market frictions and the role of internal capital markets in banking,” *Journal of Financial Economics*, 46, 135-164.

Ishii, Joy (2008): “Compatibility, Competition, and Investment in Network Industries: ATM Networks in the Banking Industry,” working paper.

Kiser, Elizabeth K. (2002): “Household Switching Behavior at Depository Institutions: Evidence from Survey Data,” *The Antitrust Bulletin*, 47, 619-640.

Kuehn, Joseph (2018): “Spillovers from entry: the impact of bank branch network expansion,” *RAND Journal of Economics*, 49(4), 964-994.

Kuehn, Joseph (2021): “Strategic Complementarities in Bank Branching Decisions,” *Journal of Industrial Economics*, 68(4), 640-692.

Kumar, Pradeep (2018): “Market Power and Cost Efficiencies in Banking,” *International Journal of Industrial Organization*, 57(2), 175-223.

Manuszak, Mark (2010): “Predicting the impact of upstream mergers on downstream markets with an application to the retail gasoline industry,” *International Journal of Industrial Organization*, 28(1), 99-111.

Mishkin, Frederic S. and Stanley G. Eakins (2009): “Financial Markets and Institutions,” Pearson, Sixth Edition.

Orhun, Yesim (2013): “Spatial Differentiation in the Supermarket Industry: The Role of Common Information,” *Quantitative Marketing and Economics*, 11(3), 3-37.

Radecki, Larry J. (1998): “The Expanding Geographic Reach of Retail Banking Markets,” *Economic Policy Review*, 4, 15-33.

Seim, Katja (2006): “An Empirical Model of Firm Entry with Endogenous Product-Type Choices,” *The RAND Journal of Economics*, 37, 619-640.

Sinnott, Roger W. (1984): “Virtues of the Haversine,” *Sky & Telescope*, 68(2), 158.

Smith, Howard (2004): “Supermarket choice and supermarket Competition in Market Equilibrium,” *The Review of Economic Studies*, 71(1), 235-263.

Thomadsen, Raphael (2005): “The Effect of Ownership Structure on Prices in Geographically Differentiated Industries,” *The RAND Journal of Economics*, 36(4), 908-929.

Thomadsen, Raphael (2007): “Product Positioning and Competition: The Role of Location in the Fast Food Industry,” *Marketing Science*, 26(6), 792-804.

Yang, Botao and Andrew Ching (2014): “Dynamics of Consumer Adoption of Financial Innovation: The Case of ATM Cards,” *Management Science*, 60, 903-922.

Yang, Nathan (2020): “Learning in Retail Entry,” *International Journal of Research in Marketing*, 37 (2), 336-355.

Zhu, Ting and Vishal Singh (2009): “Spatial Competition with Endogenous Location Choices: An Application to Discount Retailing,” *Quantitative Marketing and Economics*, 7(1), 1-35.

Table 1. Summary Statistics

Panel A: Market Level (132 markets)	Mean	Std. Dev.	Q(1)	Q(50)	Q(99)
Population	97883	39300	45184	89380	228467
Largest incorporated place within 10 miles	2882	2541	0	2512	9830
Largest incorporated place within 20 miles	8735	6742	568	6624	29592
Number of tracts	23	10	9	21	51
Number of banks	11	4	4	11	24
Number of branches	33	14	10	30	75
Herfindahl–Hirschman Index (HHI)	1938	735	985	1842	4398
Panel B: Tract level (3077 tracts)					
Area, in squared miles	7.73	35.57	0.17	1.80	88.68
Population	4195	2125	561	3816	10548
Income per capita (annual)	17633	7231	1499	17115	38091
Panel C: Journey to Work Data Aggregated at the Market Level (132 markets)					
% workers who work and live in different tracts (weighted by income)	85.40%	3.34%	74.75%	85.49%	91.38%
Average commute distance (in miles, weighted by income)	4.97	1.26	1.93	4.98	8.07
Panel D: Bank-market level (1488 bank-markets)					
Deposit (in millions)	117.66	167.39	0.49	70.58	660.42
Deposit market share	6.50%	7.41%	.03%	3.80%	32.38%
Number of branches	2.90	2.29	1.00	2.00	11.00
Panel E: Bank level (957 banks)					
Interest rate (semi-annual, percent)	4.32	0.69	2.07	4.38	5.78
Employees per branch (in 10s)	2.11	13.13	0.40	1.31	7.45
Bank age: 0-20	0.22	0.41	0.00	0.00	1.00
Bank age: 20-60	0.22	0.41	0.00	0.00	1.00
Bank age: 60+	0.56	0.50	0.00	1.00	1.00
Dummy of single branch bank	0.44	0.50	0.00	0.00	1.00
Number of states that a bank operates in	1.40	1.66	1.00	1.00	8.00
Assets (in millions)	3983	25831	15	223	75908
Bank size: small (0-100M)	0.27	0.44	0.00	0.00	1.00
Bank size: medium (100M-300M)	0.31	0.46	0.00	0.00	1.00
Bank size: large (300M+)	0.42	0.49	0.00	0.00	1.00

Table 2. Estimation Results of Single and Multiple Location Models

	Single-location Model			Multi-location Model		
	(1)	(2)		(3)	(4)	
	Est.	SE.		Est.	SE.	
γ_0 (Coefficient on distance: mean)	-1.190	(0.705)	*	-1.715	(0.440)	***
γ_1 (Coefficient on distance: interaction with deviation from mean income in thousands)	0.022	(0.530)		-0.100	(0.027)	***
ρ (Weight on closeness to home)	-	-		0.553	(0.235)	**
Bank level characteristics						
Deposit interest rate	1.075	(0.437)	**	1.565	(0.516)	***
Employee/branch	0.023	(0.004)	***	0.022	(0.005)	***
Bank age 0-20	0.224	(0.115)	*	0.348	(0.127)	***
Bank age 20-60	-0.205	(0.103)	**	-0.168	(0.132)	
Small bank (assets < $100m)	-0.743	(0.119)	***	-0.624	(0.146)	***
Mid-size bank ($100m < assets < $300m)	-0.577	(0.100)	***	-0.425	(0.112)	***
Number of states	0.021	(0.008)	***	0.026	(0.009)	***
Bank-market level characteristics						
Single branch dummy	-0.803	(0.133)	***	-0.566	(0.180)	***
Constant	0.977	(0.894)		2.103	(1.089)	*
Market FE	Y			Y		

Note:
Columns 1 and 3 report parameter estimates. Columns 2 and 4 report standard errors. Significance at 10%, 5%, and 1% are indicated by *, **, and *** respectively.

Table 3. Inferred Bank-level Characteristics

	All banks (obs = 1488)	Single-branch (obs = 543)	Multi-branch (obs = 945)	Difference (Multi - Single)
Interest rate elasticity	2.75	2.98	2.62	-0.36 ***
	(0.64)	(0.70)	(0.56)	
% of customers who work and live in different tracts	0.82	0.78	0.85	0.07 ***
	(0.11)	(0.15)	(0.06)	
% of customers who patronize two branches of the bank	0.36	0.00	0.57	0.57 ***
	(0.30)		(0.16)	
Expected distance of customers from home to work	3.94	3.24	4.35	1.11 ***
	(1.40)	(1.19)	(1.36)	
Expected effective distance from customers to the bank	2.52	3.37	2.04	-1.33 ***
	(1.54)	(2.09)	(0.77)	

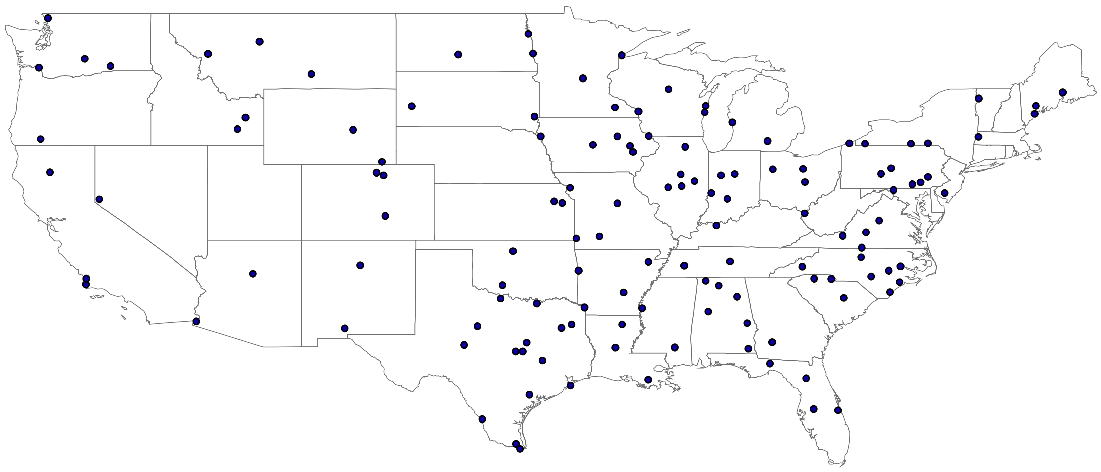
Note:
This table reports the mean value of the inferred bank characteristics based on our estimated demand model. Standard deviations are reported in brackets. The last column calculates the mean difference between multi-branch banks and single-branch banks. Asterisks indicate statistical significance of the difference: * 10%, ** 5%, *** 1%.

Table 4. Market Shares by Bank Types in the Data vs Counterfactual (CF) Experiments

	Bank type	Data	CF1: All WFH	CF2: 67% WFH	CF3: 33% WFH
Panel A: Mean market shares	Whole sample	6.5%	6.5%	6.5%	6.5%
	Work banks	8.1%	7.5%	7.5%	7.8%
	Home banks	4.9%	6.4%	6.4%	6.0%
	Other banks	4.9%	5.4%	5.3%	5.2%
Panel B: Mean changes in market shares (CF - data)	Whole sample		0.0%	0.0%	0.0%
	Work banks		-0.7%	-0.6%	-0.4%
	Home banks		1.5%	1.5%	1.0%
	Other banks		0.5%	0.5%	0.3%

Note:
Panel A reports the mean market shares by bank types in the data and the counterfactual experiments. Panel B calculates the mean changes in market shares under the counterfactual experiments. There are three counterfactual experiments: Counterfactual experiment 1 (CF1: All WFH) assumes all workers work from home; Counterfactual experiment 2 (CF2: 67% WFH) assumes the top 67% workers (i.e., workers with income higher than the 33rd percentile) work from home; Counterfactual experiment 3 (CF3: 33% WFH) assumes the top 33% workers (i.e., workers with income higher than the 67th percentile) work from home.

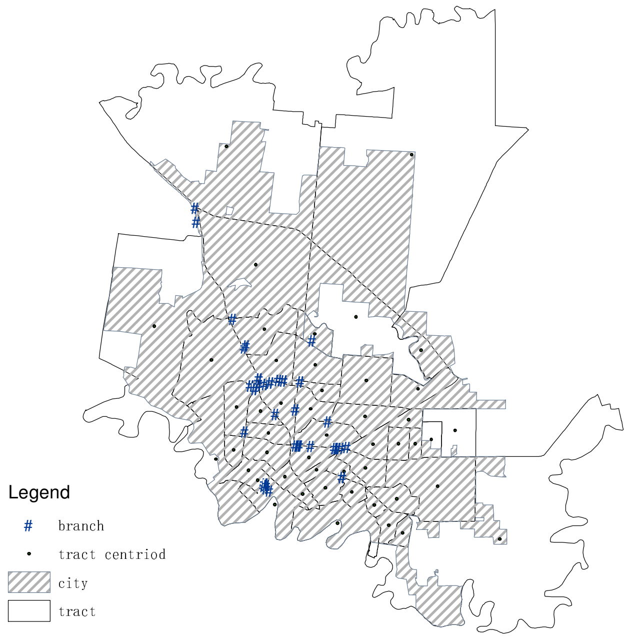
Figure 1: Locations of Markets in Our Sample



Note: Each point in this map represents a city in our sample. They are middle-sized cities with population ranging from 40,000 to 180,000. In addition, they are "isolated" from other cities by satisfying two criteria (Seim, 2006): (i) the largest neighboring city within a 10-mile radius of the focal city's centroid has less than 15,000 population, and (ii) the largest neighboring city within a 20-mile radius of the focal city's centroid has less than 30,000 population.

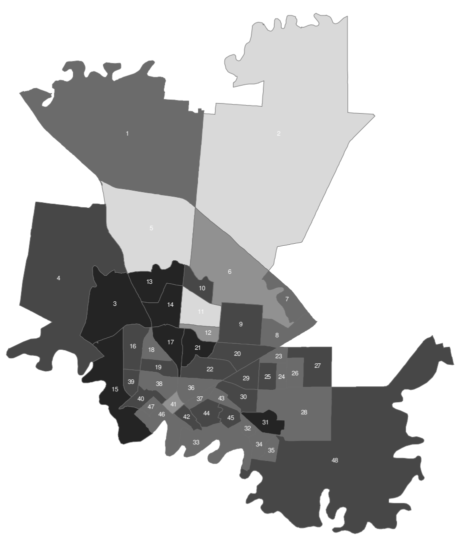
Figure 2: The Map of a Sample Market - Brownsville, Texas

Panel A. Shape of the City, Tracts, and Locations of Bank Branches

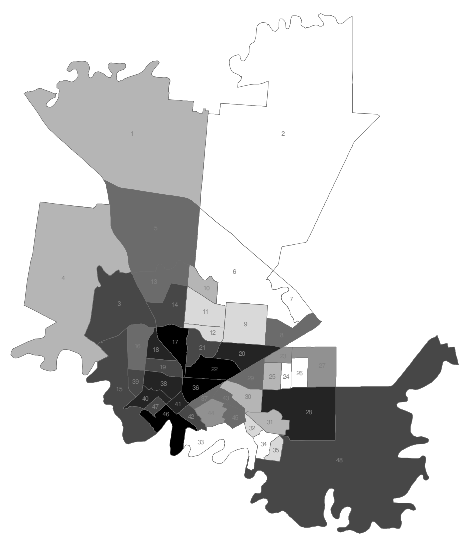


Panel B. Variations in Consumers' Locations

(Darker colors mean higher population shares in the market.)

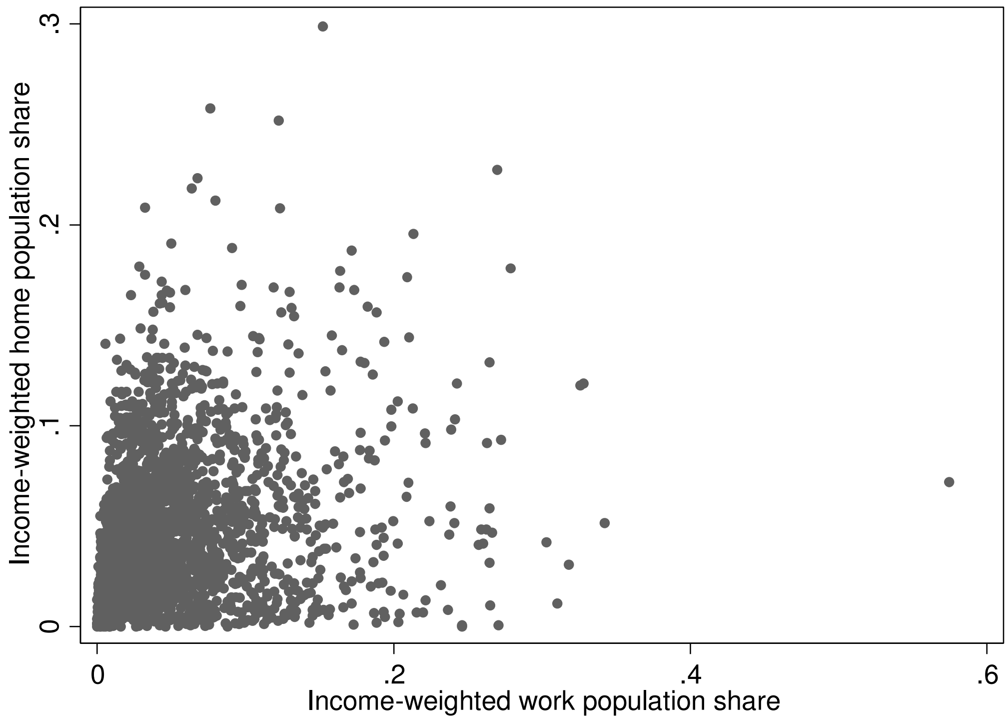


(i) Home Population shares



(ii) Work Population shares

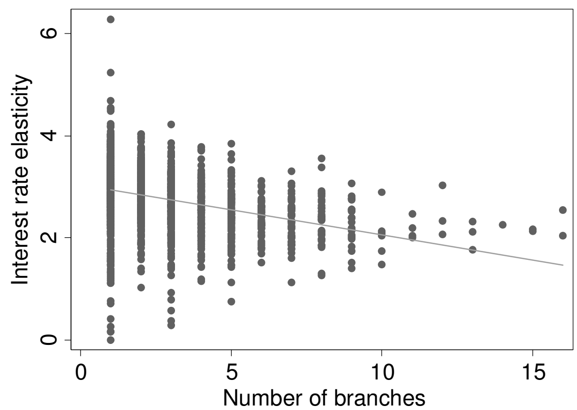
Figure 3. Variations in Tracts' Functionalities as Homes VS. as Workplaces



Note:
Each point in the figure represents a tract. The vertical axis shows tracts' income-weighted population shares in the market when all workers stay at home; the horizontal axis shows tracts' income-weighted population shares in the market when all workers go to work.

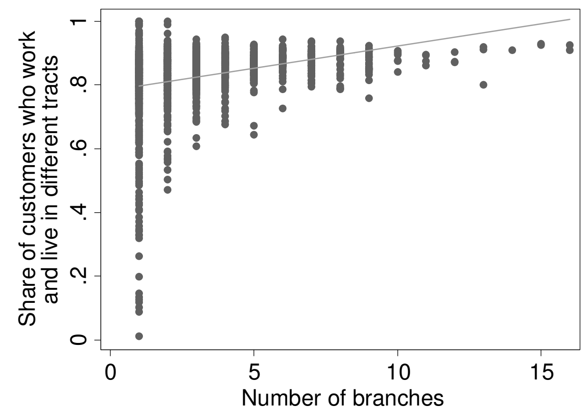
Figure 4. Inferred Characteristics of Banks' Customer Base by Number of Branches

A. Own price (interest rate) elasticities



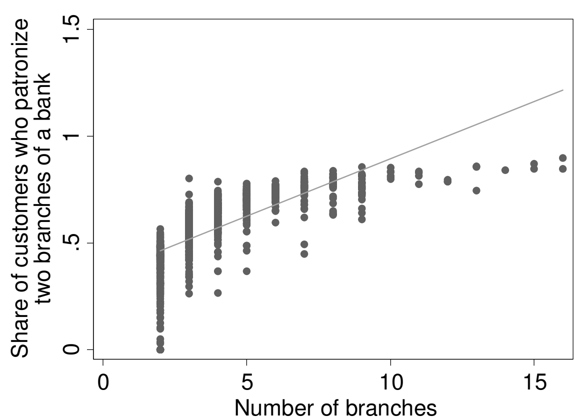
Fitted line: slope = -.099, se = .007

B. Shares of customers who work and live in different tracts



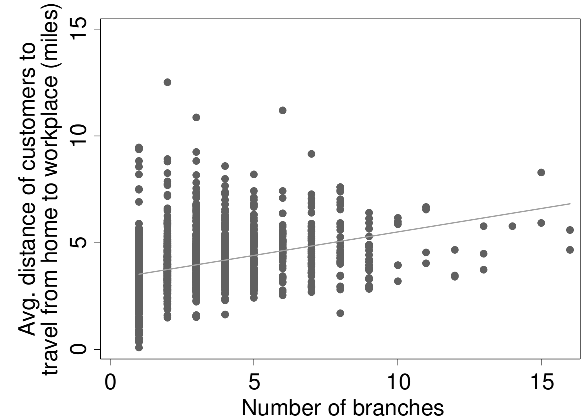
Fitted line: slope = .014, se = .001

C. Shares of customers who patronize two different branches of the bank



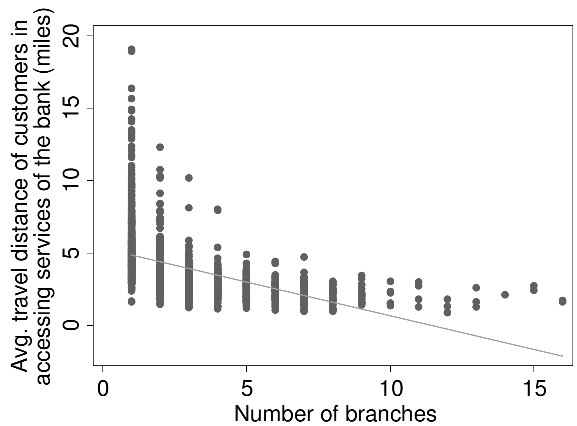
Fitted line: slope = .054, se = .002

D. Average distance of customersfrom home to work



Fitted line: slope = .219, se = .015

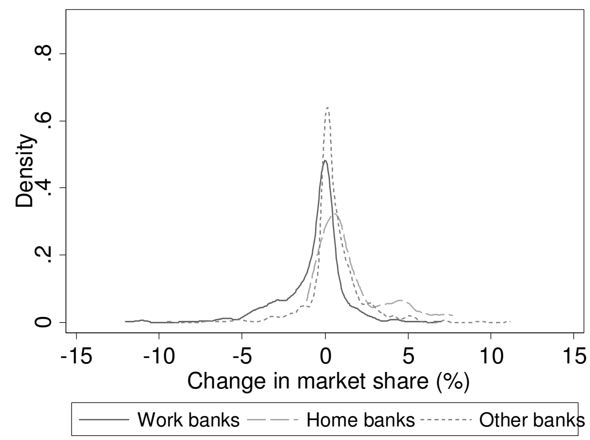
E. Average effective distance of customers to the bank



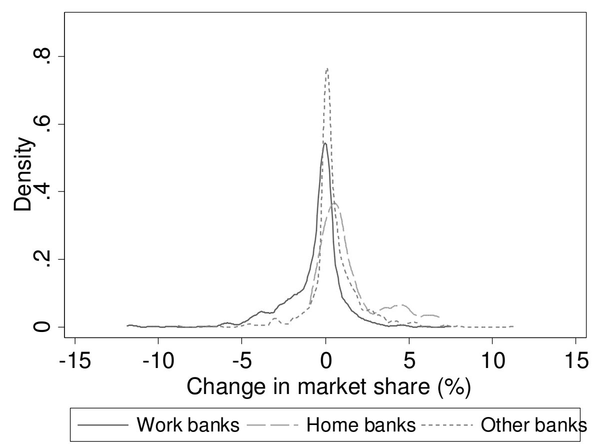
Fitted line: slope = -.465, se = .021

Figure 5. Distribution of Changes in Market Shares (Counterfactual - Data) by Bank Type

Panel A. All workers work from home (All WFH)



Panel B. The top 67% workers work from home (67% WFH)



Panel C. The top 33% workers work from home (33% WFH)

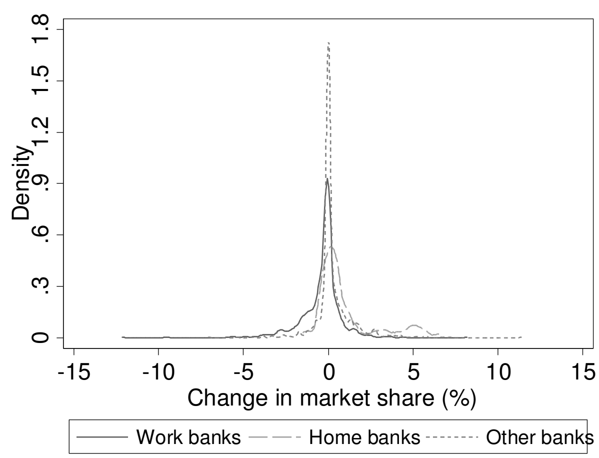
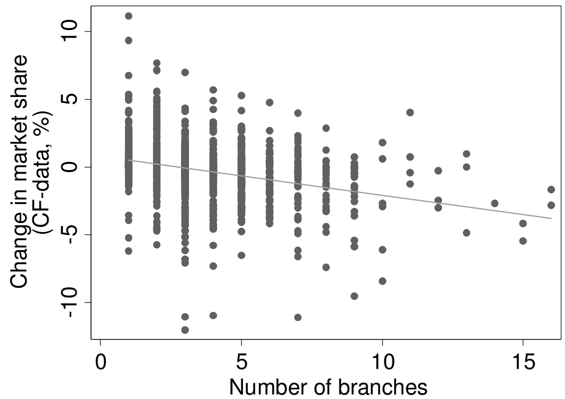


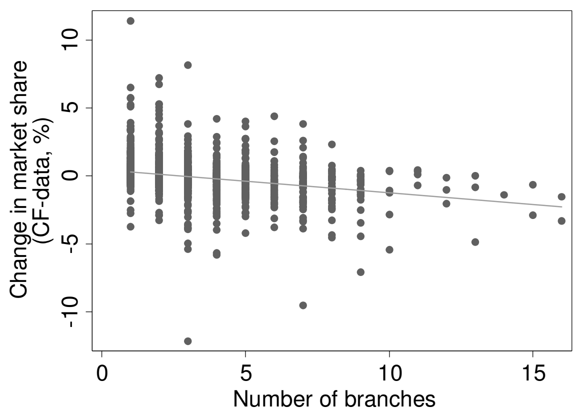
Figure 6. Change in Market Shares (Counterfactual - Data) by Number of Branches

Panel A. All workers work from home (All WFH)



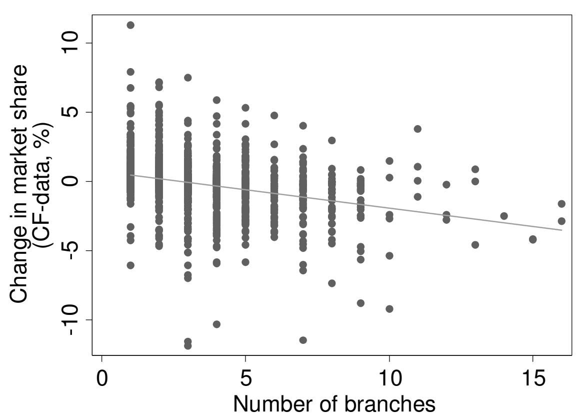
Fitted line: slope = -.29, se = .02

Panel B. The top 67% workers work from home (67% WFH)



Fitted line: slope = -.17, se = .01

Panel C. The top 33% workers work from home (33% WFH)



Fitted line: slope = -.26, se = .02

Note:
Each point in the figure represents a bank in a market. The vertical axis shows the bank's market share changes under our counterfactual experiments; the horizontal axis shows the number of branches of the bank in the market.

Appendices

A. Issues of the Non-workers

As mentioned in the paper, the Journey to Work dataset contains the number of workers who live in tract A (i.e., home location) and work in tract B (i.e., work location) for all markets in our sample. In our model, we assume that this distribution of commuting flow applies to the whole population of each tract.

Obviously, there are non-workers in the population, which include teenagers, spouses who do not work, and retirees. The first two types of people do not generate incomes. We therefore do not include them in our model. The more problematic group is retirees, who used to earn money before retirement and make banking decisions as well. Their deposits are also included in the deposit data used to generate market shares. Because the Census Bureau does not provide the joint distribution of age and work history of residents at the tract level, we do not observe the size of retiree population for each tract directly.[20] The following three assumptions ensure that our construction of market share is valid.

Assumption A1: Retirees made their bank choice before their retirement.

Assumption A2: The preferences for retirees and current workers are the same.

Assumption A3: The distribution of retirees' home and work locations before their retirement is the same as those of current workers' reported in Journey to Work.

We provide an illustrative example below to explain why these conditions can make ignorance of the non-workers innocuous in our estimation.

Consider a market with two tracts: A and B. The numbers of current workers who live in tracts A and B are 100 and 300, respectively. We further assume that, for current workers living in tract A, 20% of them work in tract A and 80% work in tract B; for current workers living in tract B, 80% work in tract A and 20% work in tract B.

We use $P_j^w(AB)$ to denote the choice probability of choosing bank j for current workers (w) who lives in tract A and work in tract B. This can be calculated with our demand model. The conditional market share of bank j due to current workers who live in tract A is:

$$P_j^w(A) = 0.2 \times P_j^w(AA) + 0.8 \times P_j^w(AB).$$

[20] According to the 2000 Census, 12.4% of the U.S. population aged above 65, which is the average retirement age. The labor force participation rate is about 67%. These imply that the retirees account for about 8% of the total population.

Similarly, the conditional market share of bank j due to current workers who live in tract B is:

$$P_j^w(B) = 0.8 \times P_j^w(BA) + 0.2 \times P_j^w(BB).$$

The market share of bank j due to all current workers is:

$$s_j^w = 0.25 \times P_j^w(A) + 0.75 \times P_j^w(B).$$

We use $P_j^r(AB)$ to denote the choice probability of bank j due to retirees (r) who live in tract A and used to work in tract B. Similarly, with Assumptions A1 and A2,

$$P_j^r(XY) = P_j^w(XY), \forall X, Y \in \{A, B\}.$$

Furthermore, with Assumptions A1 and A3,

$$s_j^r = s_j^w.$$

It follows that $s_j = s_j^r = s_j^w$.

This shows that with Assumptions A1-A3, we can replace s_j with s_j^w in our model.

B. Coordinates and Distance Measures

In this study, consumers are assumed to live or work at the centroid of each Census tract. The coordinates of these centroids are available from the Census Bureau's geographic correspondence engine MABLE/Geocorr (https://mcdc.missouri.edu/geography/). The population weighted coordinates are used to take into account the within tract population density variations.

FDIC's Summary of Deposit (SOD) data report the address for each branch. Their addresses are geocoded into coordinates in two steps. First, information on the street number, city, state, and ZIP code is processed by the street locator in the software ArcMap, which generates geographic coordinates with precision at the address level for about 78 percent of all the branches. Second, the unmatched records from the first step are processed by Google Maps API, which further increases the geocoding precision at the address level to 96 percent. The remaining 4 percent of the branches are assigned to the centroids of their zip code.

In our sample, there are 312,253 tract-branch pairs. With the coordinates of tract centroids and branches, two types of distance measures are constructed. The first measure is the spherical distance calculated by the Haversine formula (Sinnott, 1984). The second measure is the driving distance which is obtained using the Google Maps Application Programmable Interface (API). The map data (including the road network structure) used by Google is provided by TeleAtlas, Inc. When calculating the driving time for a given route, Google takes many surface factors into account, including the road networks, the speed limits, road signs (such as U turn), and public traffic records. When multiple routes are available in connecting two locations, we select the one with the fastest speed.

Table B1 compares these two distance measures. Panel A shows that the spherical distance underestimates the actual driving distance. The mean difference is more than 1.5 miles, which is quite large considering the median distance consumers commute between their homes and branches is about 3 miles (e.g., Amel and Starr, 2002). In Panel B of Table B1, it is clear that the discrepancies between these two distance measures varies across different distance ranges. In particular, when the destinations are further away, the difference between the two measures increases, while the relative deviation, defined as the ratio between the spherical difference and the driving distance, decreases.

Table B1. Spherical Distance vs Driving Distance

Panel A: Summary Statistics

	Mean	Median	Std.	Min	Max
Spherical distance (S)	4.10	3.45	2.91	0.00	36.27
Driving distance (D)	5.64	4.72	3.98	0.00	47.38

Panel B: Comparisons across Different Distance Ranges

		Mean(S)	Mean(D)	Mean(D-S)	Mean [(D-S)/D]	N
Overall		4.10	5.64	1.54	0.26	312253
By spherical distance	[0,2)	1.24	1.82	0.57	0.29	76567
	[2,3.4)	2.68	3.77	1.08	0.27	77158
	[3.4,5.4)	4.30	5.94	1.65	0.26	79513
	[5.4,)	8.07	10.86	2.80	0.24	79015

Note:

This table reports summary statistics of the distance between the 312,253 tract-branch pairs in our sample. The spherical distance is calculated using the Haversine formula:

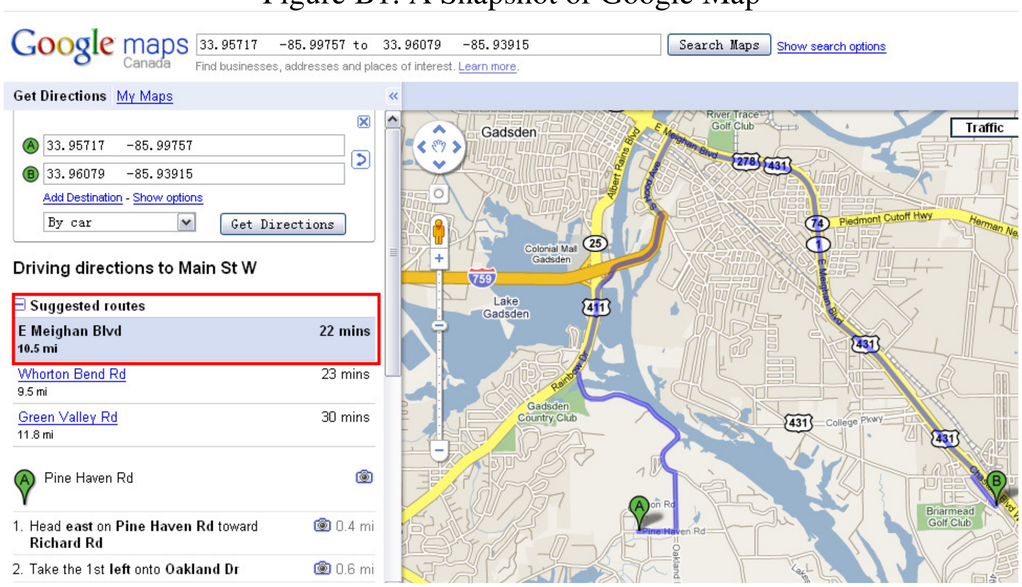
$$d = 2R \cdot arcsin\left(\sqrt{\sin^2\left(\frac{lat_2 - lat_1}{2}\right) + \cos(lat_1)\cos(lat_2)\sin^2\left(\frac{long_2 - long_1}{2}\right)}\right)$$

where R is the Earth radius, and lat and $long$ are the latitudes and longitudes of points.

The driving distance is obtained from the Google Maps Application Programmable Interface.

The results suggest that there are non-negligible differences between spherical distance and driving distance measures. The spherical distance is usually a good proxy when we measure the distance between two locations that are very far away from each other. However, when considering the distance between two locations within a city, the spherical distance could significantly underestimate the actual distance that consumers travel, because it ignores the road network and other geographic conditions, such as rivers and mountains. Figure B1 illustrates such an example, where A is a Census tract and B is a branch location. If consumers can go directly from A to B, the distance is only 3.4 miles, but it is infeasible for consumers to do so, because they have to find a bridge nearby and drive across a river, which increases the actual distance to 10.5 miles.

Figure B1: A Snapshot of Google Map



C. Classification of Residential/Business/Other Banks

In this section, we explain how we classify banks specialized in serving home locations (*home banks*), work locations (*work banks*), or somewhere in between (*other banks*). Towards that end, we need to classify the main functionalities of the tracts where the bank branches are located, which in turn requires the calculation of the population shares of the tracts when people are at home and at work. The procedure is done in three steps.

Step 1. Calculating Income-weighted Home Population Shares and Work Population Shares for Tracts

The Journey to Work data provide the number of people commuting from tract t to t', for all t' and t'. The Census data provide the average income of people living in tract t, for all t. We assume that the average incomes for people living in the same tract but working in different tracts are the same. i.e.,

$$y(t,t') = y(t,t'') = y(t), \text{for all } t, t', \text{and } t'' \qquad \text{(C1)}$$

where $y(t,t')$ means the average income among those who live in tract t and work in tract t', and $y(t)$ means the average income among those who live in tract t.

Based on the data, we can construct the tract-level income-weighted population distribution when people are at home (*income-weighted home population share*) and the tract-level income-weighted population distribution when people are at work (*income-weighted work population share*). These will be used in Step 2 to determine the tracts' main functionalities.

The income-weighted home population share of tract t, $Inc\_PopShare^H(t)$, is hence calculated as

$$\begin{aligned} Inc\_PopShare^H(t) &= \frac{\sum_{t'} y(t,t') \cdot Pop\,(t,t')}{\sum_{t''}\sum_{t'} y(t'',t') \cdot Pop\,(t'',t')} \\ &= \frac{\sum_{t'} y(t) \cdot Pop\,(t,t')}{\sum_{t''}\sum_{t'} y(t'') \cdot Pop\,(t'',t')}, \end{aligned} \qquad \text{(C2)}$$

where $Pop(t,t')$ is the number of people who live in tract t and work in tract t'. The second equality is derived from the assumption that $y(t,t') = y(t)$ for all t, and t'. [21]

The income-weighted work population share of tract t, $Inc\_PopShare^W(t)$, is calculated in a similar way as

[21] The numerator is also the total annual income earned by those whose home tract is t, and the denominator is the total annual income earned by all workers in the market.

$$Inc\_PopShare^W(t) = \frac{\sum_{t'} y(t',t) \cdot Pop\,(t',t)}{\sum_{t''}\sum_{t'} y(t',t'') \cdot Pop\,(t',t'')}$$
$$= \frac{\sum_{t'} y(t') \cdot Pop\,(t',t)}{\sum_{t''}\sum_{t'} y(t') \cdot Pop\,(t',t'')}. \tag{C3}$$

Note that the income-weighted home (work) population shares of all tracts in the city sum up to 1. At the end of Appendix C, we provide a simple numerical example to aid understanding of the two definitions.

Step 2. Classification of Home Tracts, Work Tracts, and Other Tracts

In order to classify how banks position themselves, we would like to identify whether they are primarily for home locations, work locations, or somewhere in between. In addition, we also want to focus on tracts with sufficient income to attract banks' attention. To that end, a tract t is defined as a *home tract* if it satisfies the following two conditions:

(H1) $\frac{Inc\_PopShare^H(t)}{Inc\_PopShare^W(t)} >$ the 75th percentile of the $\frac{Inc\_PopShare^H(\cdot)}{Inc\_PopShare^W(\cdot)}$ distribution (which is 2.59 in the sample);

(H2) $Inc\_PopShare^H(t) >$ the median of the $Inc\_PopShare^H(\cdot)$ distribution (which is 0.0342 in the sample).

Similarly, a tract t is defined as a *work tract* if it satisfies the following two conditions:

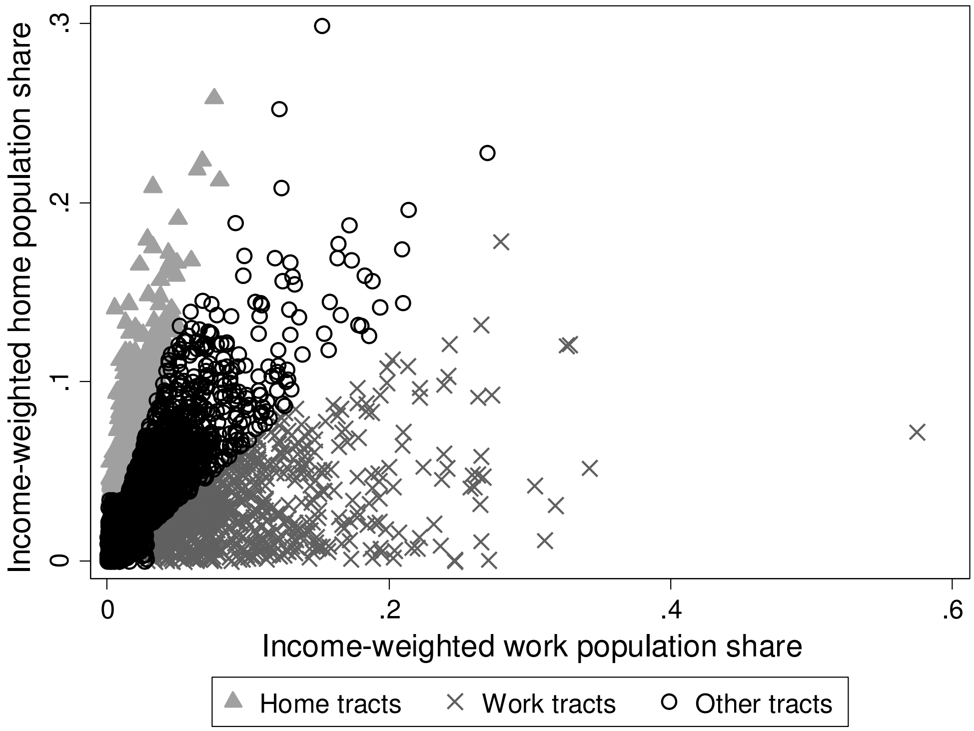
(W1) $\frac{Inc\_PopShare^W(t)}{Inc\_PopShare^H(t)} >$ the 75th percentile of the $\frac{Inc\_PopShare^W(\cdot)}{Inc\_PopShare^H(\cdot)}$ distribution (which is 1.50 in the sample);

(W2) $Inc\_PopShare^W(t) >$ the median of the $Inc\_PopShare^W(\cdot)$ distribution (which is 0.0278 in the sample).

The rest of the tracts (i.e., tracts that are classified as neither home tracts nor work tracts) are classified as *other tracts*.

Figure C1 shows the scatter plot of home tracts (474 tracts, shown in "Δ"), work tracts (676 tracts, shown in "×") and other tracts (1929 tracts, shown in "o").

Figure C1: Classification of Tracts by Their Main Purpose



Note:
Each point in the figure is a tract. The horizontal axis shows a tract's income-weighted population distribution when people are at home and the vertical axis shows a tract's income-weighted population distribution when people are at work (refer to the definitions in Eq(C2) and Eq(C3)).

Step 3. Classification of Home Banks, Work Banks, and Other Banks

After defining a tract's main functionality, we are ready to classify how banks position themselves. To that end, we want to see whether a bank has a branch close to some home (work) tract or not. Specifically, *a bank serves a tract* if the bank-tract distance (i.e. the minimum distance between the tract centroid and the bank's branches) is below the 10th percentile of that distribution (which is 0.97 mile in the sample), i.e., the bank has a branch that is within 0.97 mile of the tract centroid.

Table C1 shows where banks place branches in our sample. Notice that 743 (=609+126+8) banks serve work tracts but not home tracts, while only 42 banks serve home tracts but not business work. It is evident that banks tend to set up branches in work locations. That is not surprising because the opening hours of branches coincide with the normal work hours. Being close to workplaces provides more convenience to banks' customers. However, it is worth highlighting that the previous literature has only used consumers' home locations. Hence, the demand estimation results obtained could be misleading.

Table C1. Distribution of Banks by Numbers of Home/Work Tracts Served

		# work tracts served			
		0	1-2	3-4	5-7
# home tracts served	0	559	609	126	8
	1-2	42	99	36	6
	3-4	0	1	1	1

Now we can define home (work) banks. A bank is a *home bank* if it serves at least one home tract, but no work tracts. A work bank is defined in a similar way: A bank is a *work bank* if it serves at least one work tract, but no home tracts. This definition captures the idea that home (work) banks specialize in serving home (work) locations. We classify the rest of the banks as *other banks*[22]. meaning they serve both home areas and work areas. Note that this definition is bank-market specific. To put it another way, our definition allows for the possibility that a bank which operates in multiple cities is a home bank in one city, but a work bank in another. Empirically, we identify 743 work banks, 42 home banks and 703 other banks in our sample.

An Illustrative Example for Step 1

Suppose a city has two tracts: t_1 and t_2. There are four home-work tract pairs: (t_1, t_1), (t_1, t_2), (t_2, t_1) and (t_2, t_2), where the first element is the home location and the second one is the work location. Suppose the populations of each tract pair are 200, 400, 50 and 150 respectively (so most people live in tract t_1 and work in tract t_2), and the average annual incomes are \$10,000 for people living in tract t_1, and \$20,000 for people living in tract t_2. Then, the income-weighted home population share of tract t_1 is

$$Inc\_PopShare^H(t_1) = \frac{10000\times200+10000\times400}{10000\times200+10000\times400+2000050+20000\times150} = 0.6;$$

[22] Two types of banks might be classified as other banks: (i) banks that serve both *work* tracts and *home* tracts (there are 144 of them in the sample); and (ii) banks that serve no *work* tracts and no *home* tracts (there are 559 of them in the sample). The first case includes big banks that have a lot of branches everywhere. Many banks fall into the second case because we are conservative when classifying tracts (Step 2), and as a result, leave a lot of tracts as *other* tracts. A clarification is that it is not right to say that these banks are located far away from where people are; instead, the tract in which they are located might have a mixed functionality (as home and as workplace).

The home population share of tract t_2 is

$$Inc\_PopShare^{H}(t_2)=\frac{20000\times50+20000\times150}{10000\times200+10000\times400+20000\times50+20000\times150}=0.4;$$

The income-weighted work population share of tract t_1 is

$$Inc\_PopShare^{W}(t_1)=\frac{10000\times200+20000\times50}{10000\times200+10000\times400+20000\times50+20000\times150}=0.3;$$

And the income-weighted work population share of tract t_2 is

$$Inc\_PopShare^{W}(t_2)=\frac{10000\times400+20000\times150}{10000\times200+10000\times400+20000\times50+20000\times150}=0.7.$$

D. Calculation of the Predicted Market Shares

This section details the calculation of market shares using Eq(9). Two assumptions simplify the integration in Eq(9) over consumers. First, we assume that consumers living in the same tract have the same level of annual income[23]. Second, as discussed in Section 3, we assume that all consumers live or work at the centroids of tracts. For these reasons, if there are M tracts in a given market m, there will be $T = M^2$ types of consumers characterized by their home location, work location, and annual income (LC_i^H, LC_i^W, y_i). That is, if consumer i lives in tract t and works in tract t' (denoted by type (t,t')), then LC_i^H is the coordinates of the centroid of tract t, LC_i^W is the coordinates of the centroid of tract t', and y_i is the average annual income among people who live in tract t, $y(t)$. Hence, the disutility from travel parameter is the same across type (t,t') consumers, i.e., $\gamma(t,t') = \gamma_0 + \gamma_1 y(t)$, and the choice probabilities are also the same across type (t,t') consumers, i.e., $\{P_j(t,t')\}_{j=0}^{J}$ as derived in Eq(7). Denote the number of type (t,t') consumers by $POP(t,t')$.

Because there are a finite number of tracts, the integration in Eq(9) can be replaced by the summation as

$$\begin{aligned} s_j &= \frac{\sum_{i=1}^{I} P_{ij} \cdot y_i}{\sum_{i=1}^{I} y_i} \\ &= \frac{\sum_t \sum_{t'} P_j(t,t') \cdot y(t) \cdot POP(t,t')}{\sum_t \sum_{t'} y(t) \cdot POP(t,t')}, \end{aligned} \tag{D1}$$

[23] A discussion on this assumption can be found in Step 1 of Appendix C.

where $\frac{y(t)\cdot POP(t,t')}{\sum_t \sum_{t'} y(t)\cdot POP(t,t')}$ is the income weighted population share of type (t, t') consumers, and hence the market share of bank j in market m, s_j, can be calculated as the average choice probabilities of consumers of different types, weighted by the income weighted population shares.