Bank Competition for Neighborhood Deposits*

Mark J. Kutzbach
FDIC

Gary A. Wagner
University of Louisiana, Lafayette

C. Luke Watson
FDIC

February 2026

Abstract

Spatial markets are typically measured using administrative boundaries chosen for convenience rather than grounded in economic behavior. We estimate tract-level banking concentration using a structural model of household deposit allocation, yielding heterogeneous and overlapping catchment areas. This approach reveals that 89% of variation in local competitive conditions occurs across neighborhoods within MSAs rather than between MSAs. Monte Carlo merger simulations reveal that coarse market definitions can mask localized competitive harm: 2.6% of simulated mergers pass MSA-level screening thresholds while substantially increasing concentration in ten or more neighborhoods. These false negatives arise systematically when merging firms serve similar customers but operate spatially segmented networks. For understanding competition in any spatially differentiated market, knowing where a household lives within a market matters more than knowing which market it lives in.

*We thank Troy Cowell for assistance with CASSIDI data. Views and opinions expressed in this paper reflect only those of the authors and do not reflect those of the FDIC nor the United States. All mistakes are our own.

1 Introduction

Evaluating competition in retail banking requires defining the relevant geographic market, yet there is no consensus on the appropriate definition. The literature variously uses counties (Drechsler et al., 2017; Wang et al., 2022), MSAs (Adams et al., 2007; Dick, 2008; Ho and Ishii, 2011), census tracts (Nguyen, 2019), and zip codes (Becker, 2007). The Federal Reserve provides geographic market definitions for regulatory analysis, but these are heterogeneously determined across district banks, do not cover the entire country, and are based on physical branch locations rather than the customers they serve (DiSalvo, 1999). All of these definitions implicitly assume that consumers have equal access to all branches within broad geographic areas. However, a large literature documents that retail depositors are highly sensitive to distance when choosing a bank branch. If consumers are local but markets are defined broadly, standard concentration measures may mask substantial heterogeneity in competitive conditions across neighborhoods within the same regulatory market.

These definitional choices have real consequences. If markets are defined too broadly, regulators may approve mergers that substantially reduce competition in specific neighborhoods while appearing benign in the aggregate. If markets are defined too narrowly, regulators may block efficiency-enhancing combinations. Getting market definition right matters for both competition policy and household welfare.

This paper develops a tract-level measure of banking competition and uses it to characterize the spatial distribution of competitive conditions across tracts and within traditional bank market definitions. We address three questions. First, what does the competitive landscape for retail deposits look like at fine geographic scales? Second, do concentration levels vary systematically with neighborhood demographics? Third, what do we miss by using coarse market definitions for merger analysis?

We approach these questions using a structural model of household demand for banking services in the spirit of Ellickson et al. (2020). We model a branch choice problems for households, where each representative depositor from each Census tract chooses among all bank branches within an endogenous search distance and an outside option. Utility depends on distance between the depositor and each branch, characteristics of the bank and branch, and tract demographics. We estimate the model by first constructing tract-level liquid savings using the Survey of Consumer Finances and American Community Survey, then matching model-implied deposits at each branch to empirical deposits reported in the FDIC Summary of Deposits.

The estimated model allows us to compute predicted deposit flows from each tract to each branch, and from these flows we construct a tract-level Herfindahl-Hirschman Index (HHI) that characterizes the competitive conditions facing households in each neighborhood. We term this measure Tract-HHI, or THHI.

Our main findings are as follows. First, we document substantial heterogeneity in THHI within traditional market definitions. The average THHI in our sample is 36.5 with considerable dispersion. Using a simple variance decomposition for tracts within MSAs, we find that 89% of variation in THHI is within MSA and only 11% is across MSAs. This implies that knowing *where* a household lives within an MSA matters more for understanding the competitive conditions they face than knowing in *which* MSA the household lives. Second, we show that THHI varies systematically with tract demographics: lower-income and less-educated neighborhoods face higher concentration. Third, we simulate all possible bank mergers independently across MSAs to demonstrate that merger analysis based on coarse market definitions can substantially understate competitive effects in specific neighborhoods. We find that 2.6% of simulated mergers pass MSA-level screening thresholds yet generate $\Delta THHI > 200$ in ten or more tracts. These false negatives arise systematically when merging banks serve similar customers but have spatially segmented branch networks—precisely the cases where customer-level substitution is high but geographic averaging masks local impacts. In addition, we find these effects are large in smaller geographies, implying greater importance for areas outside of MSAs.

To preview the central empirical fact motivating our analysis, Figure 1 maps tract-level banking concentration (THHI) within the Washington, DC metropolitan area. While the DC MSA is typically treated as a single banking market in regulatory analysis, the figure reveals substantial heterogeneity in competitive conditions across neighborhoods. Tracts near the urban core exhibit relatively low concentration, while nearby suburban and exurban tracts face markedly higher concentration. Importantly, these differences occur at spatial scales far smaller than the market boundaries used in merger review.

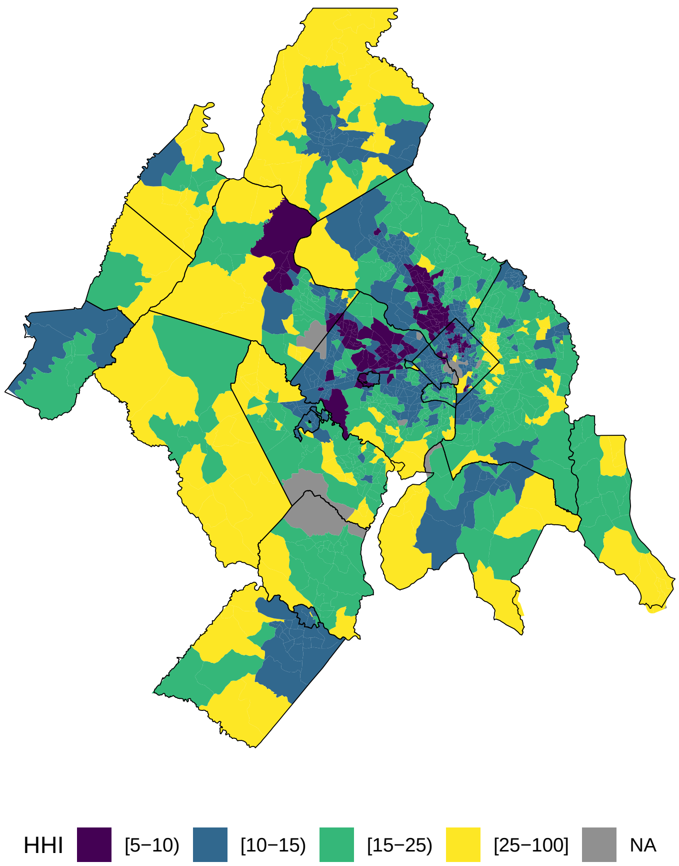
This within-market variation is the focus of the paper. Our goal is not to redefine markets *ex ante*, but to measure competition where households live and to assess how much information is lost when competition is summarized at coarse geographic levels.

We make four contributions. First, methodologically, we provide a framework for measuring banking competition at granular geographic scales that can be applied to any market with branch location data. Second, descriptively, we provide the first systematic documentation of within-MSA heterogeneity in retail deposit competition. Third, we contribute distributional evidence on which communities face more or less competitive banking markets. Fourth, we demonstrate that traditional merger review based on coarse market definitions may miss important local effects, with implications for antitrust policy.

Beyond banking policy, our analysis speaks to a broader set of questions in urban and regional economics concerning how competitive conditions vary across neighborhoods within the same metropolitan area. Many urban services—such as retail, healthcare, education, and financial access—are organized through spatially differentiated networks that are formally regulated or analyzed at coarse geographic scales. When households' effective choice sets are local, but markets are defined broadly, spatial averaging can mask meaningful heterogeneity in access and competition across neighborhoods. Our tract-level approach provides a way to measure competitive conditions where households live, rather than where administrative boundaries are drawn.

Our work relates to several literatures. Most directly, we build on structural approaches to geographic market definition. Ellickson et al. (2020) develop a demand-based framework for measuring competition in grocery retailing; we adapt their approach to banking, where

Figure 1 – Tract HHI for DC Metro



Note: This figure displays tract HHI for the Washington, CD metro area. We find that the average THHI for the metro is 18, which is significantly higher than the Fed-HHI value of 11. Lower HHI implies greater competition; higher HHI implies greater concentration. Generally, we find greater competition in the MSA center and conversely greater concentration further out.

the “expenditure” analog is liquid savings allocated across depository institutions.

Within banking, prior work has either taken geographic market definitions as given and analyzed product differentiation within them (Adams et al., 2007; Ho and Ishii, 2011; Dick, 2008; Cohen and Mazzeo, 2007), or used deposit pricing and branch transactions to assess market boundaries (Radecki, 1998; Heitfield and Prager, 2004). Our contribution is to let consumer demand reveal the relevant geographic scope of competition at the neighborhood level.

We also contribute to understanding spatial variation in access to financial services. The literature on “bank deserts” has documented areas with limited branch presence (Van Leuven et al., 2024; Hegerty, 2016), but branch counts alone do not capture competitive conditions—a tract can have many branches from the same bank (low access, high concentration) or few branches from different banks (low access, low concentration). Our framework distinguishes between access and competition, allowing us to characterize both dimensions of the retail banking landscape.

A key modeling assumption, following Ellickson et al. (2020), is that banks use approximately uniform pricing across branches within a market, allowing us to control for unobserved prices with bank fixed effects. This assumption is supported by Begenau and Stafford (2023) and Granja and Paixao (2023), who show that bank-quarter fixed effects absorb nearly all pricing variation in branch-level rate data from RateWatch.

Conceptually, our approach proceeds in three steps. First, we estimate how households in each census tract allocate deposits across nearby bank branches, allowing preferences to depend on distance, neighborhood characteristics, and bank attributes. Second, we use the resulting tract-by-bank deposit shares to construct a residence-based measure of market concentration that reflects the competitive conditions faced by households in each neighborhood. Third, we use these tract-level measures to assess how competitive conditions vary within traditional banking markets and to evaluate how merger screens based on coarse geographic definitions can miss localized competitive effects. This structure allows us to study spatial competition without imposing sharp market boundaries ex ante.

While our application focuses on retail banking, the framework is more general. The same logic applies to any spatially differentiated service where households choose among multiple locations and where competition is typically assessed using coarse geographic markets. By anchoring competition measures to the household’s place of residence, rather than to administrative market definitions, the approach can be adapted to study competition and access in a wide range of urban settings.

The rest of the paper proceeds as follows. Section 2 presents the household branch choice model. Section 3 describes our two-stage estimation approach. Section 4 presents model coefficient estimates and fit statistics. Section 5 analyzes model output by characterizing the distribution of THHI. Section 6 describes our simulated merger framework and Section 7 describes the results. Section 8 concludes.

2 Household Branch Choice Model

In this section, we describe our model of household branch choice and derived structural model outputs, such as distance elasticities and HHI measures. Afterwards, we describe the data and specific estimation details.

In short, our model supposes a representative depositor in each census tract who allocates their liquid savings across bank branch deposits and other savings opportunities. We nest the choices according to local banks, regional banks, national banks, and other savings opportunities as an outside good.

2.1 Model

Let $l \in L$ index locations. We suppose that each location has a representative depositor, and so we can refer to depositors and locations interchangeably. Let $b \in \mathcal{B}$ index bank branches, with the understanding that every branch is also affiliated with a bank, indexed by $j \in \mathcal{J}$. Finally, let $m \in \mathcal{M}$ index four market-types for firms: (1) local, (2) regional, and (3) national banks, and (4) outside savings options. Each depositor has an allocation of liquid savings, $\mathcal{S}_l$, which the depositor wishes to deposit in banks or other opportunities.

Depositors in each location face a common choice set of branches, denoted:

$$\mathcal{B}_l = \{b \mid x_{lb} < X_l\} \cup \{0\}, \tag{1}$$

where x_{lb} is the distance between location l and branch b and X_l is the location-specific search radius.

For a representative depositor of type $t \in \{H, F\}$ (households and firms) in location l, the utility for saving at branch $b \in \mathcal{B}_l$ is:

$$\begin{aligned} u^t_{lb} &= v^t_{lb} + \epsilon^t_{lb} &(2)\\ &= \mu^t_{lb}(\theta_t) + \xi_{j_b} + \epsilon^t_{lb}, &(3) \end{aligned}$$

where $\mu^t_{lb}(\theta_t)$ collects all observed determinants of utility (parameterized separately for each depositor type below), ξ_{j_b} is a bank fixed effect common to both depositor types, and ϵ^t_{lb} is a type-specific idiosyncratic taste shifter. We assume ϵ^H_{lb} follows a Generalized Extreme Value distribution, yielding a nested logit model for households, while ϵ^F_{lb} follows a Type I Extreme Value distribution, yielding a multinomial logit model for firms.

We collect all model parameters as $\Theta = (\theta_H, \theta_F, \rho, \xi)$, where ρ contains the nesting parameters for households and $\xi = (\xi_j)$ is the vector of bank fixed effects.

For households, we specify the observed component of utility as:

$$\mu^H_{lb}(\theta_H) = \beta^H_x x_{lb} + Y_b\beta^H_Y + Z_l\beta^H_Z + (x_{lb} \cdot Z_l)\beta^H_{xZ} + (Y_b \cdot Z_l)\beta^H_{YZ}, \tag{4}$$

where Y_b includes branch characteristics and Z_l includes tract characteristics. This specification is linear in parameters but includes several interactions with location and branch characteristics for added flexibility.

To allow for endogenous preference for outside savings opportunities, we model the outside good utility as:

$$v_{l0}^{H} = W_l \Pi_W^H + (W_l \cdot Z_l)\Pi_{WZ}^H, \tag{5}$$

where W_l includes geographic characteristics of the tract. We normalize the outside option's fixed effect to zero.

We assume a nested structure to household preferences for types of banks and the outside option, where the nesting is based on the "localness" of a bank. For the nesting structure, we divide branches into mutually exclusive groups: $\mathcal{B}_l = \{M_l^m, \ldots\}_m$, where M_l^m is a collection of branches of type $m \in \mathcal{M}$, with $\mathcal{M}$ indexing local, regional, and national banks. The probability that a representative household in location l allocates savings to branch b of type m is:

$$d_{lb}^{H} = \Pr(\iota_{lb} = 1) = \underbrace{\Pr(\iota_{lb} = 1 \mid b \in M_l^{m_b})}_{\substack{\text{Prob of choosing } b \\ \text{given choice of type } m}} \cdot \underbrace{\Pr(M_l^{m_b} \subseteq \mathcal{B}_l)}_{\text{Prob of choosing type } m}. \tag{6}$$

This assumption allows us to express the probability in two multiplicative parts: the conditional probability of choosing a branch of a specific type and the probability of choosing a type of bank, which we can denote as $d_{lb|m}^{H}$ and $d_{lm'}^{H}$, respectively.

The GEV distributional assumption on ϵ_{lb}^{H} yields the nested logit choice probabilities:

$$d_{lb}^{H}(\theta_H, \rho, \xi) = P_l^H\big(b \mid m(b)\big)\, P_l^H\big(m(b)\big), \tag{7}$$

with

$$P_l^H(b \mid m) = \frac{\exp\big(v_{lb}^H/\rho_m\big)}{\sum_{k \in M_{lm}} \exp\Big(v_{lk}^H/\rho_m\Big)}, \tag{8}$$

$$P_l^H(m) = \frac{\big[\sum_{k \in M_{lm}} \exp\big(v_{lk}^H/\rho_m\big)\big]^{\rho_m}}{\exp(v_{l0}^H) + \sum_{m' \in \mathcal{M}} \Big[\sum_{k \in M_{lm'}} \exp\Big(v_{lk}^H/\rho_{m'}\Big)\Big]^{\rho_{m'}}}, \tag{9}$$

where $\rho = (\rho_m)_{m \in \mathcal{M}}$ contains the nest-specific correlation parameters.

2.2 Model Deposits

Given the above, the model yields location-branch deposits as:

$$D_{lb}^{H}(\theta, \rho; \mathcal{S}_l) = d_{lb}^{H}(\theta, \rho) \cdot \mathcal{S}_l, \tag{10}$$

where $\mathcal{S}_l$ is the pool of liquid savings available in the location. If we let $\mathcal{L}_b$ be the set of tracts that place deposit into branch b, then we can aggregate location-branch deposits to the branch level:

$$D_b^H(\theta, \rho) = \sum_{l \in \mathcal{L}_b} D_{lb}^H(\theta, \rho; \mathcal{S}_l). \tag{11}$$

2.3 Market Concentration

To calculate a residence-based measure of market concentration, we scale the deposit shares by the sum of the location's "inside" shares to account for the outside good. Let $d_{lj}^t \equiv \sum_{b \in j} d_{lb}^t$ denote the share of type-t deposits from location l going to bank j, for $t \in \{H, F\}$. Let $\mathcal{J}_l$ be the set of banks accessible to depositors in location l.

We can compute depositor-type-specific measures of local concentration. The tract-level HHI for household deposits is:

$$THHI_l^H = \sum_{j \in \mathcal{J}_l} \left(100 \cdot \frac{d_{lj}^H}{1 - d_{l0}^H}\right)^2. \tag{12}$$

These measures differ from most existing measures of concentration in several respects. First, the market area for tract l may include branches that are outside of tract l. Second, these measures do not depend on pre-defined geographic boundaries and report localized concentration faced by neighborhoods. Third, unlike the Fed-HHI, our measures of local concentration give equal weight to commercial banks and thrift institutions.

3 Data and Estimation

Here, we describe the data we use and how we estimate the deposit choice model. We use five sources of data: Survey of Consumer Finance (SCF), American Community Survey (ACS), FDIC Summary of Deposits (SOD), FDIC Reports of Conditions and Income ("Call Reports"), and the YourEconomy Time Series (YTS), a database of establishments collected by Data Axle.

The estimation has two stages. First, we estimate the pools of liquid savings available in each location—separately for households and firms—using the Survey of Consumer Finances, American Community Survey, and YTS establishment data. Second, we fit the model's predictions to observed branch-level deposits (see Equation (12)) from the SOD using characteristics of tracts, banks, and branches from the ACS, Call Reports, SOD, and YTS.

All data are for 2010 unless otherwise noted. 2010 is close to the high point of the number of branches in the United States, it is a year for the Decennial Census and SCF, and it mostly predates the rise of online-only banking. We define banks using the highest holder as defined by the National Information Center, using either the RSSD identifier of the highest Bank Holding Company or the FDIC certification number, or Cert, of the standalone bank.

3.1 Stage 1: Liquid Savings

The model supposes a pool of liquid savings that a household wishes to allocate across savings and investment opportunities. Unfortunately, no such measure currently exists. We estimate this pool using the following procedure: (1) use the SCF to estimate a household model of liquid savings, (2) apply this model to ACS households in the Public Use Microdata Sample (PUMS), (3) aggregate the prediction to the public-use microdata area (PUMA), and (4) allocate the PUMA-level savings to census tracts based on the tracts' income shares within the PUMA.

Using the SCF, we estimate the household liquid savings model by summing up various sources of liquid savings within the household. The SCF is the most appropriate data source for this because of the detailed and high-touch surveying of financial information for household finances.[1] We use the following variables to create our liquid savings measure: checking accounts, savings accounts, money market accounts, certificates of deposit, and brokerage call accounts. The SCF does not report holdings of physical cash.

Wealth and savings are even more concentrated than income (Bricker et al., 2016), so any model explaining savings must account for skewness. Following Salem and Mount (1974), we assume the gamma family distribution for savings and and estimate a gamma regression in Stata using a Generalized Linear Model and a log link function, as follows:

$$\mathcal{S}_h = X_h \alpha + e_h, \tag{13}$$

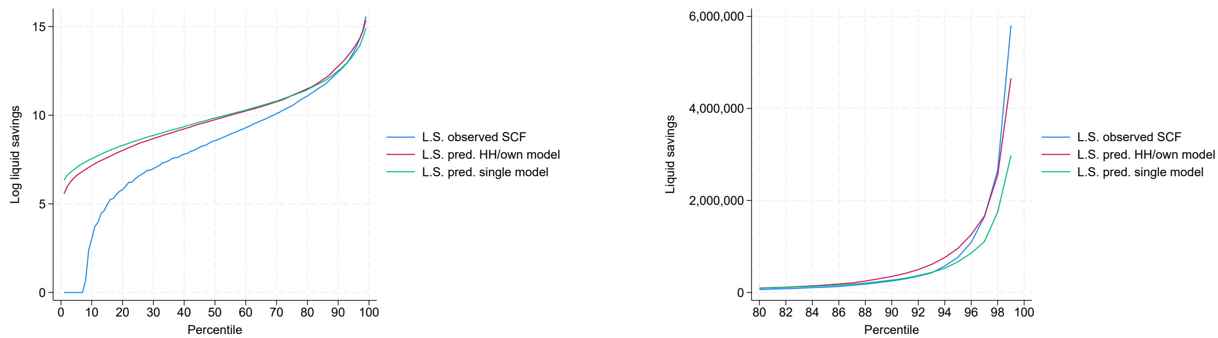
where X includes the following variables: age, race, education, work status, home-ownership, income, household structure, and either home value or monthly rent (depending on housing tenure status). We estimate separate models for four different types of household: couples who own their home, couples who do not own their home, single householders who own their home, and single householders who do not own their home.[2] This yields the object of interest: $\hat{\alpha}$ (see Tables A1 and A2).

We expect there to be substantial, idiosyncratic variation in liquid savings beyond what can be explained with this model. So how well does the model do in prediction? First, in aggregate, the model predicts total liquid savings of \$4.7 trillion, compared to \$4.4 trillion reported in the SCF. Figure 2 shows how the distribution of predicted liquid savings compares with the distribution of reported liquid savings, by percentile. Overall, predicted liquid savings (Model 2) has a flatter distribution than reported liquid savings. Also displayed are predictions from an approach (Model 1) that does not separate out couples and homeowners; predictions from this alternative model are even flatter and do not capture the high end of liquid savings.

[1] Other data sources may ask about financial information (e.g., ACS or Panel Survey of Income Dynamics), but are not as complete in the number of financial products or are focused on low income individuals. The SCF sampling design captures high net-worth households that may be missed in other surveys and are an important segment to capture for bank deposits.

[2] The SCF uses the concept of 'primary economic unit' within a household (e.g., two unmarried adults, a married couple), and where available we include characteristics of additional members of the unit. We use indicator variables for all characteristic except income, for which we use log income measures for the household.

Figure 2 – Liquid savings, observed and predicted distributions



(**a**) Liquid savings logs

(**b**) Liquid savings levels

Note: Authors calculations from the SCF. Panel A displays log values across all percentiles. Panel B displays level values for the 80th percentile and above. Liquid savings observed from the SCF is reported in blue. The HH/own model (in red) predictions are based on separate specifications by household type and home ownership status. The figure also displays a single specification model (in green) for all households.

Next, we use ACS to fit the household liquid savings model and to produce local area estimates. The SCF is a national survey, so it lacks geographic specificity, while the ACS has fine geography, but lacks savings information. As such, we turn to the ACS which is a 1:100 sample annual survey of Americans. Luckily, the ACS and SCF overlap in relevant demographic variables and the primary economic unit concept in SCF aligns with the family unit concept in ACS, so we can directly use the coefficients from the previous step. We again use the gamma family distribution to calculate $\hat{\mathcal{S}}_h = X_h\hat{\alpha}$.

Third, we aggregate the household liquid savings estimates to the PUMA level, which is the smallest geographic area in the ACS. PUMAs must contain a minimum of 100,000 individuals to avoid disclosure of personally identifying information.

Finally, because PUMAs are collections of census tracts, we can reallocate the predicted liquid savings to census tracts. We assume that liquid savings are proportional to the tract share of PUMA personal income. That is, if the total personal income in a tract is 1% then we assume that the tract also has 1% of the PUMA level of liquid savings:

$$\mathcal{S}_l = \mathcal{S}_{PUMA} \cdot \frac{\text{PInc}_l}{\text{PInc}_{PUMA}}. \tag{14}$$

Note, another way of looking at our measure is that it is proportional to total personal income in the census tract with a PUMA level scaling factor. We find that the national aggregate of our tract-level liquid savings estimate is approximately equal to the nationally representative estimate of total liquid savings calculated from the SCF.

3.1.1 Liquid Savings Validation

To validate our tract level liquid savings measure, we compare it with Statistics on Income (SOI) produced from tax data by the IRS. The interest rate implied by the national totals of liquid savings and interest income is 2.89%. This rate is higher than deposit interest rates at the time, but interest income also includes income from bonds, such as U.S. savings bonds.

To investigate geographic correspondence of liquid savings and interest income, we regress tract-level log liquid savings (see Equation (14)) on log reported interest income and log reported wage and salary income, both reported in the IRS Statistics on Income (SOI).[3] We find that the R^2 of the univariate regression of log liquid savings on log reported interest income is 0.60, and when we add log reported wage income the R^2 increases to 0.74 and the effect of interest income remains positive and significant.

3.2 Stage 2: Deposit Model

To estimate our deposit model, we iterate the parameter vector until we minimize the sum of squared residuals. The residual that we minimize is calculated in two steps. First, given a set of parameters, we use the model to predict branch level deposits and calculate the model deviations. Second, we regress these deviations on measures of local business activity, to account for unmodeled business deposits, and calculate the sum of squares of this residual.

3.2.1 Branch-Tract Pairs

Using the SOD and YTS we geolocate all bank branches in the United States. We calculate the distance between the population centroid of each census tract to all branches, x_{lb}, within endogenous search distance, X_l, which we estimate using commuting patterns. These variables form our depositor choice sets, $\mathcal{B}_l$. Thus, each depositor has a unique set of branches that are considered and each branch/bank has a unique set of potential customers.

3.2.2 Utility

We model the utility of a branch using variables from the SOD, Call Reports, and YTS.

Branch characteristics, Y_b, include number of branch employees, whether the branch is a main branch, whether the branch is full service, and a fixed effect for banks. Because there are over seven thousand banks in 2010 and many have fewer than five branches, we do not include a fixed effect for all banks. We describe the full set of rules for bank fixed effects in Appendix A, but in short we include a fixed effect for large banks by asset size, large banks by number of branches, and the largest five banks by state; all others are pooled together.

Tract characteristics, Z_l and W_l, include log median household income, percent of households with cars, percent of households employed, log tract population density, and the square of log tract population density.

[3]The SOI is reported at the zip code level, and so we use a Census Bureau's cross walk (population weighted) from zip code tabulation areas (ZCTAs) to tracts.

3.2.3 Nesting Structure

The nesting structure is based on being a local bank, regional bank, or nationwide bank. If a bank is only located with a single state, then we call it a local bank. If a bank is only within a subset of contiguous states, then we call it a regional bank. All other banks are nationwide banks.[4]

3.2.4 Sum of Squared Errors

To fit the model, we use a non-linear sum of squares routine. We first find the log difference between the model prediction and the SOD observed deposits:

$$\nu_b = \ln[D_b^{\text{Data}}] - \ln[D_b(\theta, \rho)]. \tag{15}$$

Because we specifically model household deposits, we know we are missing firm deposits at bank branches. We regress the model deviation, ν_b, on measures of local business activity, A_b and calculate the residual, υ_b:

$$\nu_b = A_b \Gamma_A + \upsilon_b. \tag{16}$$

We use the YTS data to calculate the number of establishments and employees for retail and non-retail firms within a fourth of a mile of each branch. Using this residual, we calculate the sum of squared residuals:

$$SSR(\theta, \rho) = \sum_b \upsilon_b(\theta, \rho). \tag{17}$$

Finally, our non-linear least squares minimization routine is:

$$\min_{\theta, \rho} \left\{ SSR(\theta, \rho) \right\}. \tag{18}$$

We calculate clustered standard errors using the score of the SSR at the optimal parameters.

4 Model Estimation Results

This section presents parameter estimates, assesses model fit, and derives economic quantities of interest from the estimated model.

4.1 Parameter Estimates

In Table 1 we present a subset of parameter results for the household deposit choice model, specifically the utility parameters related to the main characteristics of tracts and branches.

[4]Note, this nesting structure does not correspond to how banks are typically classified: community banks, regional (and super-regional) banks, and large financial institutions. In future work, we will assess the differences in nesting structure.

All remaining parameter estimates, primarily interaction effects, are displayed in appendix Table A3.

Examining the nesting parameter estimates, we see that they are decreasing in 'bank scope' where the local banks have a nesting parameter of 0.89, regional have a parameter of 0.81, and the nationwide banks have a parameter of 0.75. Thus, the estimates report that the local banks have *less* correlation within nests than regional banks which have *less* correlation than nationwide banks. In economic terms, this implies that the nationwide banks are seen as more substituable relative to regional or local banks. While there is still some within nest correlation for local banks, we can interpret this as local banks are filling different niches within this market segment.

Moving towards the utility parameters, we see that the main effect of distance is negative while most of the distance iteration effects are positive. For example, while distance to the branch decreases deposits, comparing two equidistant tracts, the tract with a higher percent of households with a car will have greater deposits.

Finally, looking at the branch characteristics, we see that HQ branches, branch age, and branch staffing are all positively correlated with increased branch deposits.

4.2 Implied Market Footprints

In Table 2, we present summary information on tract and branch footprints. For tracts, we present all tracts and tracts by above and below median population by above and below median tract density. For branches, we present all branches and branches by the nest of the bank to which it belongs. For both tracts and branches, we calculate the minimum distance such that either 50 or 90 percent of deposits are within that radius of the tract or branch. For each measure, the table presents the 25th, 50th (Median), and 75th percentiles and the mean for each group.

We find that 50% of tracts have 50% of their deposits within 1.7 miles, and that 75% of tracts have 90% of their deposits within 5.9 miles. Across population and density groups, 75% of tracts have 90% of deposits within 11 (Low-Low), 8 (High-Low), and 3 (Low-High, High-High) miles, respectively. We see that all above median density tracts have similar footprints, but for below median density tracts there is a difference between above and below median population.

We find that 50% of branches have 50% of their deposits within 1.9 miles, and that 75% of branches have 90% of their deposits within 9.8 miles. For local, regional, and nationwide, 75% of branches have 90% of deposits within 12.6, 9.7, and 7.2 miles, respectively.

While perhaps not immediately obvious, local banks, which tend to operate few branches, tend to have larger footprints than regional or nationwide banks. Because nationwide banks tend to have many branches, we believe their individual branch footprints are smaller. It could also be that local banks tend to operate in more lower density areas, where distances traveled tends to be greater.

For both tracts and branches, we see that it takes greater distances to account for the

Table 1 – Parameter Estimates

Variable	Estimate
ρ: Local banks	0.89
	(0.13)
ρ: Regional banks	0.81
	(0.09)
ρ: Nationwide banks	0.75
	(0.07)
β_x: Distance	-0.52
	(0.06)
β_{xZ}: Distance X log(Med HH Inc)	0.12
	(0.06)
β_{xZ}: Distance X Emp-Pop Ratio	0.26
	(0.20)
β_{xZ}: Distance X Pct of HH w/ Car	1.68
	(0.58)
π_W: log(Pop Density)	0.62
	(0.08)
β_Y: Distance X log(Pop Density)	-0.13
	(0.03)
β_Y: HQ Branch	0.78
	(0.21)
β_Y: log(Branch Age)	0.30
	(0.04)
β_Y: log(Branch Staff)	0.33
	(0.06)
Other interactions	Y
Tract Characteristics in Outside Option	Y
Bank FEs	Y
R^2	0.41

Note: This table presents a subset of utility parameters from our model of household branch choice. We include the nesting parameters, the coefficients on distance and its interaction with tract characteristics, the main effect coefficients on branch characteristics. Robust standard errors are in parentheses.

same volume of deposits, consistent with out negative distance parameter. To quantify this, we regress the tract-branch log deposits on the distance in miles between the branch and tract; however, we use a tract fixed effect to get the tract distance decay semi-elasticity, and we use a branch fixed effect to get the branch semi-elasticity. For the former, the tract fixed effect ensures we use variation within a given tract to estimate the effect of distance, and likewise the branch fixed effect only uses variation within a branch.

For tracts, we find a distance semi-elasticity of -12.2, which implies that for every additional mile deposits from the tract decay about 12 percent. For example, if two otherwise equal branches are a mile apart from the tract, then the further away branch gets 12 percent less in deposits. For branches, we estimate a semi-elasticity of -8.7. That is, for two otherwise equal customers where one is a mile further away, the branch can expect about 9 percent fewer deposits from the further away customer.

Table 2 – Market Footprints in Miles

Panel A: Tract				
50% of Tract Deposits	Q_{25}	Q_{50}	Q_{75}	Mean
All banks	1.2	1.7	3.2	3.0
Low Pop, Low Density	1.4	3.0	7.2	4.9
High Pop, Low Density	1.4	2.6	5.3	3.8
Low Pop, High Density	1.0	1.4	1.8	1.6
High Pop, High Density	1.0	1.4	1.9	1.7
90% of Tract Deposits	Q_{25}	Q_{50}	Q_{75}	Mean
All banks	1.8	2.5	5.9	4.7
Low Pop, Low Density	2.5	5.9	10.9	7.6
High Pop, Low Density	2.2	4.4	8.9	6.0
Low Pop, High Density	1.7	2.0	2.7	2.5
High Pop, High Density	1.7	2.0	2.7	2.6
Semi-Elasticity, Tract Level	$\mathrm{d}\ln(D)/\mathrm{d}x = -12.2$ (0.07)			
Panel B: Branch				
50% of Branch Deposits	Q_{25}	Q_{50}	Q_{75}	Mean
All banks	1.4	1.9	3.5	2.9
Local banks	1.6	2.9	5.7	4.0
Regional banks	1.4	2.0	3.5	2.8
Nationwide banks	1.3	1.6	2.6	2.6
90% of Branch Deposits	Q_{25}	Q_{50}	Q_{75}	Mean
All banks	2.2	4.9	9.8	6.5
Local banks	4.4	9.3	12.6	9.1
Regional banks	2.2	5.1	9.7	6.5
Nationwide banks	2.0	2.9	7.2	4.9
Semi-Elasticity, Branch Level	$\mathrm{d}\ln(D)/\mathrm{d}x = -8.7$ (0.04)			

Note: This table presents summary statistics on branch catchment areas for all branches and by nest. For each branch, we calculate the minimum distance accounting for 50 or 90 percent of branch deposits, using the model results. The table presents summary statistics for these two variables. Nests are Local, Regional, and Nationwide banks.

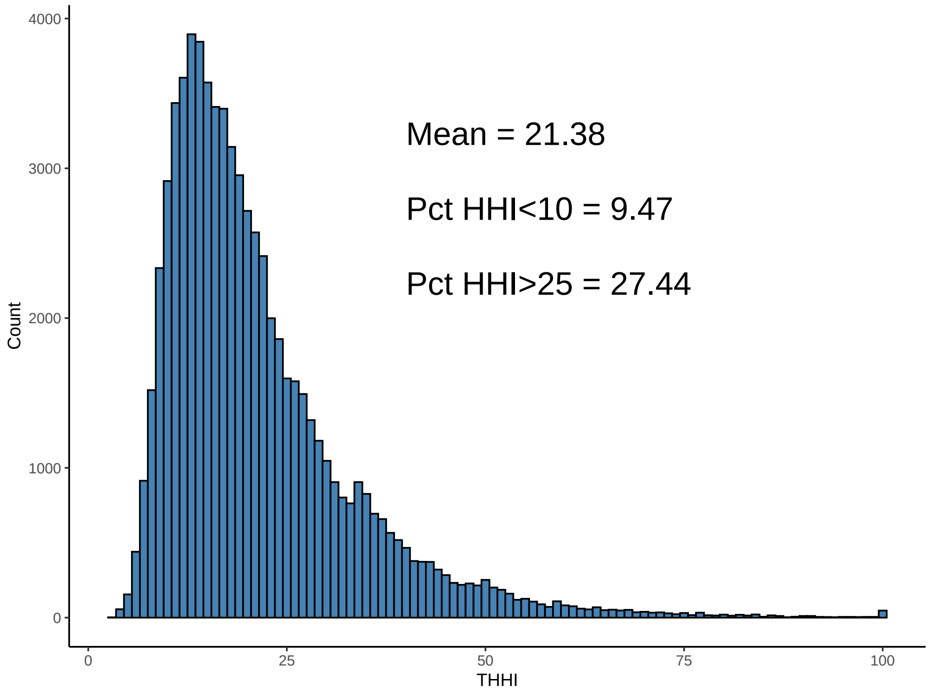
5 Local Banking Competition

This section presents our core empirical findings on the spatial structure of retail deposit competition.

5.1 Distribution of Tract-Level HHI

In Figure 3, we plot a histogram of tract HHI, labeled with the mean and the percent of THHI above and below two traditional thresholds for merger analysis. We find that the distribution is right skewed, with a mean of 21.4 and a median of 18.2. In addition, we calculate that 9.5 percent of tracts are below a THHI level of 10, which is traditionally labeled as competitive, while 27.4 percent are above 25, which is labeled as highly concentrated.

Figure 3 – Tract HHI Distribution



Note: This figure displays a histogram of tract HHI for all tracts. Lower HHI implies greater competition; higher HHI implies greater concentration. Generally, we see most tracts are reasonably competitive; however, there is a long tail of concentrated tracts.

Figure 4 complements Figure 1 by zooming out. We now plot the tract HHI for the multi-state area of DC, MD, NC, SC, VA, and WV (roughly the states of the 5th Federal Reserve district). Outside of major MSAs, tracts become significantly more concentrated. We also see state level differences; the most notable of which is that there is not a single highly competitive census tract in all of West Virginia.

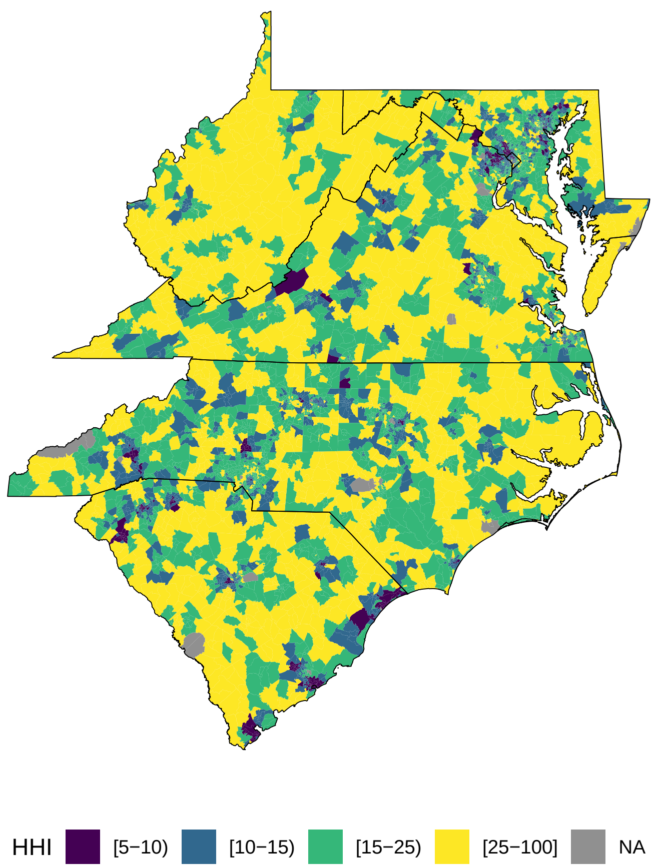
5.2 Variance Decomposition

Current merger policy considers MSAs to be a single banking market. However, given the distribution results above, using very geographically coarse measures of banking markets may obscure heterogeneity in local banking conditions.

For both MSAs and counties, we use a variance decomposition approach to assess how much variation in THHI is within or across these geographies. This is accomplished by using a simple regression of THHI on the coarse-geography fixed effects, and then saving the R^2 value that measures the share of variation between the geographies. The share of variation that is unexplained in the regression, $1 - R^2$, is then the share of variation in THHI that is within the geography. Table 4 displays the results.

For Census tracts that are in MSAs, we find that 88 percent of THHI variation is *within* MSA and only 12 percent is *across* MSA. For all Census tracts, we find that 72 percent of THHI variation is *within* county while 27 percent is *across* counties.

Figure 4 – Tract HHI for DC, MD, NC, SC, VA, WV



Note: This figure displays tract HHI for the District of Columbia, Maryland, North Carolina, Sourth Carolina, Virginia, and West Virginia. Lower HHI implies greater competition; higher HHI implies greater concentration. Generally, we find greater competition in the MSA center and conversely greater concentration further out.

Table 3 – Correlates with Tract HHI

Dependent Variable:	Log Tract HHI				
Model:	(1)	(2)	(3)	(4)	(5)
log(Pop)	-0.03		-0.01	-0.09	-0.01
	(0.01)		(0.01)	(0.01)	(0.01)
log(Med HH Inc)		-0.17	-0.17	0.13	0.08
		(0.01)	(0.01)	(0.02)	(0.02)
log(Pct w/ Car)				-0.43	-0.67
				(0.06)	(0.05)
log(Liquid Savings)					-0.04
					(0.01)
log(Branches)					-0.40
					(0.00)
Other Tract Vars	N	N	N	Y	Y
R^2	< 1%	2.7%	2.7%	26.7%	64.9%

Note: This table displays regression results of log tract HHI on different tract characteristics. Lower HHI implies greater competition; higher HHI implies greater concentration. Robust standard errors are in parentheses.

For accounting for the banking conditions of a depositor, it is by far more important to know where within an MSA or county the customer is than knowing which MSA or county the depositor lives in.

Table 4 – Variance decomposition of tract-level HHI

	Share of THHI Variation	
Geographic Unit	Within	Across
MSA	88%	12%
County	72%	28%

Note: Variance decomposition based on regressions of tract-level HHI on geographic fixed effects. "Within" is the share of variation unexplained by the fixed effects $(1 - R^2)$; "Across" is the share explained by the fixed effects (R^2). The MSA row restricts to tracts within MSAs; the county row includes all tracts.

6 Monte Carlo Design and Testing Framework

This section describes the simulation procedure and the empirical tests used to evaluate how tract-level measures of competition compare with traditional market-level concentration analysis. The goal of the Monte Carlo is not to forecast specific mergers, but rather to systematically map how merger effects vary with customer overlap, geographic overlap, and bank size within a single metropolitan banking market.

6.1 Conceptual Setup

We take as given the estimated deposit allocation model described in Sections 2–3, which produces predicted deposit shares from each census tract to each bank. These shares allow us to compute tract-level concentration measures that reflect competition faced by households at their place of residence, rather than competition averaged over a large geographic market.

The Monte Carlo exercise builds on this structure by asking a simple question: if two banks were to merge, how would local competitive conditions change, and how would those changes be reflected by standard market-level screens? By simulating all potential mergers within a fixed metropolitan area, we can study the distribution of merger effects and the conditions under which coarse market definitions mask local impacts.

Importantly, even if the same two banks exist in multiple MSAs, we only consider the within-MSA effects of this merger to focus on the effect of market definition.

6.2 Simulated Merger Approach

Within an MSA, we consider all unordered pairs of banks that operate at least one branch. Each bank pair constitutes a potential merger candidate.

For each simulated merger, we hold household preferences, branch locations, and deposit totals fixed. The only change induced by the merger is mechanical: deposits previously attributed to the two merging banks are reassigned to a single combined entity. No re-optimization of branch locations, prices, or household behavior is allowed, which allows us to work directly we the estimated tract-bank shares from the choice model. This isolates the first-order competitive effect of consolidation, holding all other features constant. This approach is consistent with typical anti-trust merger screens.[5]

Let s_{jl} denote the share of deposits from tract l allocated to bank j in the pre-merger equilibrium. Consider a hypothetical merger where bank A acquires bank B. By holding consumer choices fixed, the new combined bank-A share becomes $s'_{Al} = s_{Al} + s_{Bl}$. Continently, the change in tract HHI is given by the analytic expression:

$$\Delta THHI_l = 2 \cdot s_{Al} \cdot s_{Bl} \cdot 10,000, \tag{19}$$

where the $10,000$ puts the change in 'HHI points,' as is standard. This expression highlights that local competitive effects arise only where both banks draw deposits from the same tract. Tracts in which one or both banks have negligible presence experience little or no change in concentration, even if the banks are geographically proximate elsewhere in the MSA.

For each merger, we summarize tract-level impacts using several statistics, including the mean change in $THHI$ across tracts, upper-tail measures (such as the 90th percentile), and counts of tracts exceeding conventional screening thresholds.

[5] Allowing re-optimization would not eliminate localized concentration increases when customer overlap is high, but would instead layer behavioral responses on top of the first-order effects we study here.

MSA-level Merger Effects To mirror standard regulatory practice, we also compute the change in MSA-level concentration for each simulated merger. This measure aggregates deposits across the entire MSA and applies the same HHI formula at the market level. Because the MSA-level HHI averages across neighborhoods, it reflects the overall size and market shares of the merging banks, but not how their competitive footprints overlap spatially. Comparing tract-level and MSA-level measures allows us to assess when market-wide screening tools provide an accurate summary of local impacts, and when they do not.

6.3 Merger Pair Variables

We construct four measures of similarity between banks j and k:

Customer Distance. For each bank j, we compute a deposit-weighted average of tract characteristics:

$$\bar{X}_j = \sum_l \omega_{jl} \cdot Z_l, \tag{20}$$

where Z_l is a vector of standardized tract characteristics (median income, percent college-educated, percent with less than high school education, percent homeowners, percent age 65+, and percent employed), and $\omega_{jl} = d_{jl}/\sum_{l'} d_{jl'}$ is the share of bank j's deposits originating from tract l.

Customer Distance between banks j and k is the Euclidean distance between their customer profiles:

$$\text{CustomerDistance}_{jk} = \|\bar{X}_j - \bar{X}_k\|_2. \tag{21}$$

Lower values indicate that the two banks serve demographically similar customers.

Total Variation Distance. Using each bank's deposit distribution across tracts, $\omega_j = (\omega_{j1}, \ldots, \omega_{jL})$, where $\omega_{jl} = d_{jl}/\sum_{l'} d_{jl'}$, we measure Total Variation (TV) Distance as the overlap in the deposit distributions:

$$\text{TV Distance}_{jk} = \frac{1}{2} \sum_l |\omega_{jl} - \omega_{kl}|. \tag{22}$$

This measure ranges from 0 (identical deposit distributions) to 1 (no overlap).

Deposit Share Correlation. This measure captures whether banks tend to have high shares in the same tracts:

$$\text{ShareCorrelation}_{jk} = \text{Corr}(s_{j\cdot}, s_{k\cdot}), \tag{23}$$

where the correlation is computed across all tracts $l = 1, \ldots, L$. This measure is implicitly normalized for bank size: two banks with identical geographic patterns but different total deposits would still have correlation 1.

Branch Distance. For each bank j, we compute a deposit-weighted geographic centroid:

$$(\bar{\text{Lat}}_j, \bar{\text{Lon}}_j) = \sum_{b \in j} \frac{D_b}{D_j} \cdot (\text{Lat}_b, \text{Lon}_b), \tag{24}$$

where D_b is deposits at branch b, $D_j = \sum_{b \in j} D_b$ is total deposits for bank j, and the sum runs over all branches belonging to bank j.

Branch Distance is the Haversine (great-circle) distance in kilometers between centroids:

$$\text{BranchDistance}_{jk} = \text{Haversine}((\bar{\text{Lat}}_j, \bar{\text{Lon}}_j), (\bar{\text{Lat}}_k, \bar{\text{Lon}}_k)). \tag{25}$$

6.4 Testing Framework

The Monte Carlo supports four related empirical tests. Additional details are presented in Appendix D.

First, we test whether mergers between banks with more similar customers generate larger tract-level concentration effects. To do so, we regress MSA averaged $\Delta THHI$ on measures of customer similarity, geographic proximity, and bank size.[6] These regressions assess whether customer overlap predicts competitive harm beyond what is captured by branch distance alone.

Second, we test for disagreement between tract-level and MSA-level screening rules. Each simulated merger is classified according to whether it exceeds a standard threshold of 200 HHI points under each measure. This classification yields four cases: mergers flagged by both measures, mergers flagged by neither, and two forms of disagreement. Of particular interest are "false negatives," in which a merger passes the MSA-level screen but generates substantial tract-level concentration increases.

Third, we test whether "false negatives" are systematic. We examine how the probability of disagreement varies with customer similarity, combined market share, and geographic overlap. This allows us to distinguish between random measurement error and predictable failure modes of coarse market definitions.

Finally, we examine heterogeneity in merger effects across neighborhoods. We limit this analysis to the Washington, DC metro area given the large data requirements.[7] Using tract-by-merger observations, we relate $\Delta THHI_l$ to tract characteristics while including merger fixed effects. This specification asks, within a given merger, which types of neighborhoods experience larger concentration increases. In addition, we compare results specifically for high and low customer similarity mergers, and assess whether distributional exposure is driven primarily by the nature of customer overlap or by aggregate merger size.

Together, this framework provides a structured way to interpret the Monte Carlo results: it clarifies how merger effects arise mechanically, how they are summarized under

[6]While we include MSA level fixed effects, we also standardize variables within-MSA to make comparisons between variables easier.

[7]Additional individual metro areas are available upon request.

alternative screening tools, and which economic features predict discrepancies between local and market-wide measures of competition.

7 Monte Carlo Results

This section summarizes the Monte Carlo results from the simulated mergers across all 340 MSAs in our sample, simulating over 315,000 potential mergers and examining how the relationship between customer similarity and competitive harm varies with market structure. The main message is that tract-level concentration effects can be substantial even when MSA-level screens remain small, and that the discrepancy is systematic: it is largest for mergers between banks with similar customers whose branch networks are not especially close in space.

7.1 Sample and Simulation Design

We apply the same simulation procedure described in Section 6 to each MSA independently. For every MSA, we consider all pairwise combinations of banks operating at least one branch, compute tract-level $\Delta THHI$ for each hypothetical merger, and compare these effects to MSA-level concentration changes.

Table 5 summarizes the scope of the analysis. The 340 MSAs vary considerably in size: the median MSA contains 24 banks and 48 tracts, while the largest (New York) contains 233 banks and 4,610 tracts. This heterogeneity provides leverage to examine how market structure moderates the relationship between customer similarity and merger effects.

Table 5 – Simulation Sample Summary

	Min	Median	Mean	Max
Banks per MSA	7	24	32	253
Tracts per MSA	12	48	148	4,610
Mergers per MSA	21	276	927	31,878

Note: Summary statistics across 340 MSAs. Total simulated mergers: 315,153. Mergers are all pairs of banks in a given MSA where the two banks operate one branch; MSAs treated independently.

7.2 Customer Similarity and Merger Effects

Table 6 presents pooled regressions relating MSA mean $\Delta THHI$ to bank-pair characteristics, with MSA fixed effects and standard errors clustered at the MSA level. All regressors are standardized within MSA to ensure comparability across markets of different sizes.

The results confirm that customer similarity strongly predicts tract-level merger effects. A one standard deviation decrease in customer distance (indicating more similar customers) is associated with a 3.43 point increase in mean $\Delta THHI$ (Column 2). Branch distance also matters: closer branch networks predict larger effects, with a coefficient of −2.28. Combined market share and relative size are strong predictors, with coefficients of 15.79 and −5.66

respectively. When TV distance replaces customer distance (Column 2), it emerges as the strongest similarity predictor, with a coefficient of −7.93, and the within-MSA R^2 rises to 0.19. Share correlation (Column 3) shows that banks with more correlated tract-level deposit shares generate larger merger effects.

Table 6 – Determinants of Mean $\Delta THHI$

	(1)	(2)	(3)
Customer Distance	−3.43***		
	(0.50)		
TV Distance		−7.93***	
		(1.17)	
Share Correlation			3.24***
			(0.67)
Branch Distance	−2.28***	0.22	−2.38***
	(0.24)	(0.48)	(0.24)
Combined Share	15.79***	15.28***	16.92***
	(2.52)	(2.42)	(2.75)
Size Ratio	−5.66***	−5.45***	−6.06***
	(0.95)	(0.89)	(1.03)
MSA Fixed Effects	Yes	Yes	Yes
Observations	315,153	315,153	315,153
Within R^2	0.17	0.19	0.16

Note: Dependent variable is mean $\Delta THHI$ across tracts for each simulated merger. Customer distance, TV distance, and share correlation all measure the degree of customer overlap between the merging banks. Branch distance uses the distance between the branch deposit weighted bank centroids. Combined share and size ratio measure the importance of the banks for the MSA and relative to each other. All regressors standardized within MSA. Standard errors clustered by MSA. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

7.2.1 Heterogeneity by MSA Size

A key advantage of the multi-MSA framework is the ability to examine how market structure moderates the customer-similarity effect. Table 7 splits the sample by MSA size and re-estimates the main specification.

The results reveal striking heterogeneity. In small MSAs ($\leq$ 100 tracts), a one standard deviation decrease in customer distance is associated with a 10.54 point increase in mean $\Delta THHI$, and bank-pair characteristics explain 37% of within-MSA variation. In large MSAs ($>$ 500 tracts), the same coefficient is only 0.41, and within-MSA R^2 falls to 11%. The combined share effect is similarly attenuated: a one standard deviation increase in combined market share raises mean $\Delta THHI$ by 45.97 points in small MSAs but only 3.25 points in large MSAs.

This pattern has a natural interpretation. In small markets, banks' customer bases are more likely to overlap substantially, so mergers between similar banks generate concentrated local effects. In large markets, even banks with similar average customer profiles may draw from different neighborhoods, diluting the tract-level impact of any single merger.

Surprising, we find that branch distance has a much smaller effect once we split MSAs by size. We believe this could indicate that branch distance begins to fail as a diagnostic measure conditional on MSA size.

Table 7 – Heterogeneity by MSA size

	Small (≤ 100 tracts)	Medium (101–500 tracts)	Large (> 500 tracts)
Customer Distance	-10.54^{***}	-3.47^{***}	-0.41^{***}
	(0.86)	(0.35)	(0.10)
Branch Distance	0.22	-0.39^{*}	-0.19^{***}
	(0.59)	(0.21)	(0.06)
Combined Share	45.97^{***}	18.80^{***}	3.25^{***}
	(2.70)	(1.71)	(0.63)
Size Ratio	-16.39^{***}	-6.29^{***}	-0.81^{***}
	(1.03)	(0.65)	(0.19)
MSA Fixed Effects	Yes	Yes	Yes
Observations	69,104	70,932	175,117
Within R^2	0.37	0.26	0.11

Note: Dependent variable is mean $\Delta THHI$. Customer distance, TV distance, and share correlation all measure the degree of customer overlap between the merging banks. Branch distance uses the distance between the branch deposit weighted bank centroids. Combined share and size ratio measure the importance of the banks for the MSA and relative to each other. All regressors standardized within MSA. Standard errors clustered by MSA. Small: MSAs with ≤ 100 tracts. Medium: 101–500 tracts. Large: > 500 tracts. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

7.3 False Negatives Across Markets

We find that 8,204 mergers (2.60%) pass the MSA-level screen ($\Delta HHI_{MSA} \leq 200$) while flagging at least 10 tracts with $\Delta THHI > 200$.

Table 8 shows how false-negative rates vary across the distribution of customer similarity. The pattern mirrors the single-MSA results: false negatives are concentrated among high-similarity mergers. In the most similar quartile, the false-negative rate exceeds 10%, compared to less than 0.3% in the least similar quartile. The mean $\Delta THHI$ for the most similar quartile (22.66) is nearly 20 times larger than for the least similar quartile (1.15).

We note that these results are somewhat mechanical. If two banks have very different customers, then they likely operate in different locations, and as such the first order effect of the merger will be close to zero. Thus, the share correlation of two banks is likely a strong proxy for their customer profile distance.

7.4 Distributional patterns across tracts

Finally, Table 9 studies heterogeneity in tract-level $\Delta THHI$ within a merger for the Washington DC metro area.[8] With merger fixed effects, the coefficients describe which tracts are hit

[8]Additional specific MSAs available upon request.

Table 8 – False negative rates by customer similarity quartile

	Q1 (Most Similar)	Q2	Q3	Q4 (Least Similar)
Mean $\Delta THHI$	22.66	7.42	3.02	1.15
Median $\Delta THHI$	0.21	0.00	0.00	0.00
P90 $\Delta THHI$	52.99	11.88	3.22	0.49
False Negative Rate (%)	10.01	2.87	1.00	0.27
N Mergers	78,920	78,828	78,747	78,658

Note: Quartiles defined by customer distance within each MSA. Q1 contains mergers between the most similar banks. False negative defined as passing MSA screen while flagging $\geq$ 10 tracts with $\Delta THHI > 200$.

hardest within the same merger. The results show systematic within-merger sorting: tract characteristics are associated with larger local concentration increases, and these patterns are substantially stronger when attention is restricted to high customer-similarity mergers.

Table 9 – Which tracts within DC MSA are most affected?

	All Mergers (1)	High Cust. Sim. (2)	Low Cust. Sim. (3)
Median Income	-0.03**	-0.15*	-0.01
	(0.01)	(0.07)	(0.01)
% College	0.08***	0.53***	0.01
	(0.01)	(0.07)	(0.01)
% Less than HS	0.03***	0.21***	0.00
	(0.01)	(0.04)	(0.01)
% Homeowner	-0.04***	-0.18***	-0.01
	(0.01)	(0.05)	(0.01)
% Age 65+	0.03***	-0.05	0.02**
	(0.01)	(0.04)	(0.01)
% Employed	0.04***	0.12**	-0.01
	(0.01)	(0.05)	(0.01)
Merger FE	Yes	Yes	Yes
Observations	6,128,640	1,532,160	1,532,160
R^2	0.18	0.21	0.01

Note: Unit of observation is tract × merger. The dependent variable is tract-level $\Delta THHI$. Specifications include merger fixed effects so coefficients reflect within-merger differences across tracts. "High/Low Cust. Sim." split uses customer similarity quartiles.

7.5 Monte Carlo Summary

Across simulated mergers, MSA-level and tract-level concentration measures often agree for the largest, most obvious cases, but tract-level analysis adds information precisely where customer substitution is strong and branch overlap is spatially segmented. In those settings, MSA averages can mask concentrated impacts in a subset of tracts, generating a meaningful false-negative region under operational screening rules.

Customer similarity consistently predicts tract-level concentration effects and false negatives. This is especially true in smaller metro areas, where bank-pair characteristics explain up to 37% of within-MSA variation in merger effects. False negatives are heavily concentrated among mergers between similar banks: the 10% false-negative rate in the most-similar quartile is 37 times higher than the 0.27% rate in the least-similar quartile.

These findings reinforce the value of tract-level analysis as a complement to traditional market-wide concentration measures, particularly in smaller and mid-sized banking markets where customer overlap is most likely to generate spatially concentrated merger effects.

8 Conclusion

The definition of spatial banking markets is a key ingredient in both empirical banking research as well as policy to ensure competition and access to credit. However, there is no consensus on market definition in the banking literature and Federal Research banking markets do not cover the entire country nor show clear discontinuities between markets.

This paper develops a tract-level measure of banking competition based on a structural model of household deposit allocation and uses it to characterize the spatial distribution of competitive conditions within and across traditional market boundaries. Three findings emerge.

First, competitive conditions vary far more within markets than across them. Our variance decomposition reveals that 89% of variation in tract-level concentration occurs within MSAs rather than between them.

Second, tract-level concentration varies systematically with neighborhood demographics, with lower-income and less-educated areas facing systematically higher concentration. This distributional pattern suggests that coarse market definitions may obscure disparities in competitive conditions across different types of communities.

Third, these measurement issues have practical consequences for merger policy. Our Monte Carlo analysis of all potential mergers within MSAs finds 2.6% of mergers would be approved under current MSA-level screens despite generating $\Delta THHI > 200$ in ten or more neighborhoods. These false negatives arise systematically in mergers between banks serving demographically similar customers whose branch networks exhibit limited geographic overlap—precisely the settings where customer substitution is high but spatial averaging conceals localized competitive harm.

Our framework makes four contributions to banking research and policy. Methodologically, we provide a tractable approach for measuring banking competition at granular geographic scales that can be applied to any market with branch location data and adapted to other retail industries where distance matters. Descriptively, we provide the first systematic documentation of within-market heterogeneity in retail deposit competition and show that this heterogeneity dominates cross-market variation. Distributionally, we document which types of communities face more or less competitive banking markets. For policy, we demonstrate that traditional merger review based on coarse market definitions may systematically miss important local effects, with implications for antitrust enforcement and financial access.

The core message is simple: where you live within a market matters more than which market you live in. For households seeking banking services, for researchers studying banking competition, and for policymakers evaluating mergers, attention to fine-grained geographic variation is not a technical detail; it is central to understanding who faces competitive markets and who does not.

References

Adams, Robert M., Kenneth P. Brevoort, and Elizabeth K. Kiser, 2007, Who competes with whom? the case of depository institutions, *Journal of Industrial Economics* 55, 141—-167.

Becker, Bo, 2007, Geographical segmentation of us capital markets, *Journal of Financial Economics* 85, 151–178.

Begenau, Juliane, and Erik Stafford, 2023, Uniform rate setting and the deposit channel, Technical report, Manuscript.

Bricker, Jesse, Alice Henriques, Jacob Krimmel, and John Sabelhaus, 2016, Estimating top income and wealth shares: Sensitivity to data and methods, *American Economic Review* 106, 641—-645.

Cohen, Andrew M, and Michael J Mazzeo, 2007, Market structure and competition among retail depository institutions, *The Review of Economics and Statistics* 89, 60–74.

Dick, Astrid A, 2008, Demand estimation and consumer welfare in the banking industry, *Journal of Banking & Finance* 32, 1661–1676.

DiSalvo, James V, 1999, Federal reserve geographic banking market definitions, Technical report, Federal Reserve Bank of Philadelphia.

Drechsler, Itamar, Alexi Savov, and Philipp Schnabl, 2017, The deposits channel of monetary policy, *The Quarterly Journal of Economics* 134, 1819–1876.

Ellickson, Paul B, Paul LE Grieco, and Oleksii Khvastunov, 2020, Measuring competition in spatial retail, *The RAND Journal of Economics* 51, 189–232.

Granja, Joao, and Nuno Paixao, 2023, Bank consolidation and uniform pricing, Technical report, Manuscript.

Hegerty, Scott W, 2016, Commercial bank locations and "banking deserts": A statistical analysis of milwaukee and buffalo, *The Annals of Regional Science* 56, 253–271.

Heitfield, Erik, and Robin A Prager, 2004, The geographic scope of retail deposit markets, *Journal of Financial Services Research* 25, 37–55.

Ho, Katherine, and Joy Ishii, 2011, Location and competition in retail banking, *International Journal of Industrial Organization* 29, 537–546.

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

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

Salem, A. B. Z., and T. D. Mount, 1974, A convenient descriptive model of income distribution: The gamma density, *Econometrica* 42, 1115—-1127.

Van Leuven, Andrew J, Dayton Lambert, Tessa Conroy, and Kelsey L Thomas, 2024, Do "banking deserts" even exist? examining access to brick-and-mortar financial institutions in the continental united states, *Applied Geography* 165, 103201.

Wang, Yifei, Toni M. Whited, Yufeng Wu, and Kairong Xiao, 2022, Bank market power and monetary policy transmission: Evidence from a structural estimation, *Journal of Finance* 77, 2093–2141.

Online Appendix of
Bank Competition for Neighborhood Deposits

Mark Kutzbach, Gary Wagner, C. Luke Watson

A Additional Model Details

Here we describe additional model details.

A.1 Bank Fixed Effect Decision Rules

There were over 7,500 banks in 2010 with roughly 20% being unit banks. While our modeling approach can theoretically incorporate bank fixed effects for branching banks due to the heterogeneous choice sets of deposits, we believe we would lose substantial power to estimate all these parameters. As such, we establish four groups of banks: small banks, mid-size banks, locally large banks, and nationally large banks. We only include bank fixed effects for locally large and nationally large banks, and assign an aggregate indicator for small and mid-size banks.

First, we define nationally large banks as all banks above the 95^{th}-percentile nationally with more than five branches. Second, after excluding nationally large banks, we define locally large banks as the top five banks in each state that (a) have more than five branches and (b) have assets above the 85^{th}-percentile nationally. Third, small banks with less than five branches or only exist in a single county. Fourth, all remaining banks are listed as mid-size banks.

A.2 Bank Nesting Structure

We assign banks to nests according to the following:

- Nest 1: banks with branches in a single state,
- Nest 2: banks with branches in contiguous states,
- Nest 3: banks with branches across the many states.

B Liquid Savings Model Results

Here, we present the household regressions to predict liquidation.

C Additional Estimation Results

In Table A3 we display additional parameter estimated from our model. These coefficients are the interaction effects of branch and tract characteristics as well as tract characteristics interacted with density, the latter of which are used for modeling the demand for the outside good.

Table A1 – SCF estimates, householder variables (1 of 2)

VARIABLES	(1) Couple own	(2) Couple rent	(3) Single own	(4) Single rent
ln(Home value)	0.489***		0.112***	
	(0.000372)		(0.000318)	
ln(Monthly rent)		-0.0676***		0.0796***
		(0.000287)		(0.000312)
ln(Income)	0.595***	1.296***	0.658***	0.629***
	(0.000453)	(0.000953)	(0.000678)	(0.00102)
Missing income	7.217***	12.24***	7.210***	6.120***
	(0.00785)	(0.0160)	(0.00900)	(0.0147)
Age 35-44	0.125***	0.182***	0.00947***	-0.465***
	(0.00185)	(0.00220)	(0.00227)	(0.00265)
Age 45-54	0.517***	0.310***	0.738***	-0.396***
	(0.00227)	(0.00292)	(0.00211)	(0.00261)
Age 55-64	0.509***	0.848***	1.227***	-0.0227***
	(0.00262)	(0.00422)	(0.00211)	(0.00282)
Age 65-74	1.145***	0.408***	0.687***	0.561***
	(0.00497)	(0.0121)	(0.00445)	(0.00949)
Age 75-up	0.996***	0.843***	1.103***	1.795***
	(0.00530)	(0.0133)	(0.00438)	(0.00963)
Black non-Hispanic	-0.777***	0.0411***	-0.388***	-0.927***
	(0.00156)	(0.00222)	(0.00180)	(0.00222)
Hispanic	-0.679***	0.0190***	-0.528***	-0.625***
	(0.00134)	(0.00199)	(0.00211)	(0.00247)
Other non-Hispanic	-0.309***	0.989***	-0.433***	-0.240***
	(0.00192)	(0.00324)	(0.00315)	(0.00408)
Less than HS	-0.115***	-0.195***	-0.423***	-0.548***
	(0.00223)	(0.00302)	(0.00334)	(0.00354)
Some college	0.343***	0.394***	0.478***	0.682***
	(0.00107)	(0.00191)	(0.00153)	(0.00220)
College degree	0.238***	0.719***	0.631***	1.177***
	(0.00114)	(0.00275)	(0.00170)	(0.00268)
Grad degree	0.461***	0.932***	1.105***	1.756***
	(0.00141)	(0.00360)	(0.00202)	(0.00358)
Unemployed	-0.161***	-0.941***	-0.792***	0.556***
	(0.00220)	(0.00270)	(0.00310)	(0.00327)
Not workforce	0.430***	-0.282***	0.894***	0.246***
	(0.00135)	(0.00229)	(0.00160)	(0.00245)
PEU persons	-0.0817***	-0.0202***	-0.285***	-0.0651***
	(0.000831)	(0.00179)	(0.00135)	(0.00214)
PEU children	-0.0511***	-0.0709***	0.0743***	-0.225***
	(0.000889)	(0.00179)	(0.00168)	(0.00230)
PEU elders	-0.409***	0.211***	0.253***	-0.490***
	(0.00375)	(0.0111)	(0.00383)	(0.00887)

Note: SCF estimates Robust standard errors are in parentheses.

D Monte Carlo Methodology Details

This appendix provides detailed definitions of the similarity measures, concentration metrics, and screening rules used in the Monte Carlo merger simulation described in Section 6.

D.1 Data and Sample

The simulation uses the Washington–Arlington–Alexandria, DC–VA–MD–WV MSA, which contains 1,344 census tracts and 96 banks operating at least one branch within the MSA boundary. This yields $96 \times 95/2 = 4{,}560$ unique bank pairs, each representing a potential

Table A2 – SCF estimates continued, spouse variables (2 of 2)

VARIABLES	(1) Couple own	(2) Couple rent	(3) Single own	(4) Single rent
Sp: Age 35-44	0.0789***	-0.633***		
	(0.00178)	(0.00242)		
Sp: Age 45-54	0.330***	-0.539***		
	(0.00224)	(0.00355)		
Sp: Age 55-64	0.457***	0.287***		
	(0.00263)	(0.00437)		
Sp: Age 65-74	0.906***	1.083***		
	(0.00489)	(0.0126)		
Sp: Age 75-up	1.534***	-0.110***		
	(0.00533)	(0.0139)		
Sp: Less than HS	-0.413***	0.401***		
	(0.00241)	(0.00304)		
Sp: Some college	0.168***	0.107***		
	(0.00110)	(0.00179)		
Sp: College degree	0.494***	0.379***		
	(0.00119)	(0.00255)		
Sp: Grad degree	0.642***	0.346***		
	(0.00147)	(0.00370)		
Sp: Unemployed	-0.114***	-0.675***		
	(0.00186)	(0.00257)		
Sp: Not workforce	0.250***	-0.308***		
	(0.000944)	(0.00155)		
Female			-0.156***	-0.773***
			(0.00121)	(0.00189)
Constant	-3.434***	-5.071***	0.714***	1.833***
	(0.00610)	(0.0110)	(0.00750)	(0.00997)
Observations	14,960	5,140	5,420	6,890
U.S. Households	49,108,152	18,297,810	24,248,461	25,954,798

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: SCF estimates Robust standard errors are in parentheses.

merger.

For each tract l and bank j, we observe s_{jl}, the share of tract l's deposits allocated to bank j, as predicted by the deposit allocation model described in Section 3. These shares satisfy $\sum_j s_{jl} = 1$ for each tract (excluding the outside option) and form the basis for all subsequent calculations.

D.2 Tract-Level Concentration Measures

Tract HHI. For each tract l, define the Herfindahl–Hirschman Index as:

$$THHI_l = 10{,}000 \times \sum_j s_{jl}^2, \tag{26}$$

where the sum runs over all banks and the scaling ensures $THHI \in [0, 10{,}000]$.

Change in Tract HHI from Merger. Consider a merger between banks A and B. Post-merger, the combined entity has share $s'_{Al} = s_{Al} + s_{Bl}$ in each tract. The change in tract-level HHI is:

$$\Delta THHI_l = 10{,}000 \times \left[(s_{Al} + s_{Bl})^2 - s_{Al}^2 - s_{Bl}^2\right] = 20{,}000 \times s_{Al} \times s_{Bl}. \tag{27}$$

Table A3 – Additional Parameter Estimates

Variable	Estimate
β_{YZ} : log(Med HH Med HH Inc) X HQ Branch	0.218
	(0.799)
β_{YZ} : log(Med HH Inc) X log(Branch Age)	3.697
	(3.584)
β_{YZ} : log(Med HH Inc) X log(Branch Staff)	-0.085
	(0.115)
β_{YZ} : Emp-Pop Ratio X HQ Branch	0.169
	(0.102)
β_{YZ} : Emp-Pop Ratio X log(Branch Age)	0.635
	(0.444)
β_{YZ} : Emp-Pop Ratio X log(Branch Staff)	0.049
	(0.015)
β_{YZ} : log(Pop Density) X HQ Branch	0.366
	(0.237)
β_{YZ} : log(Pop Density) X log(Branch Age)	-0.048
	(0.886)
β_{YZ} : log(Pop Density) X log(Branch Staff)	-0.008
	(0.024)
π_W : log(Pop Density) X log(Pop Density)	0.009
	(0.01)
π_{WZ} : log(Pop Density) X log(Med HH Inc)	-0.028
	(0.083)
π_{WZ} : log(Pop Density) X Emp-Pop Ratio	0.277
	(0.222)
π_{WZ} : log(Pop Density) X Pct of HH w/ Car	0.022
	(0.736)
π_{WZ} : log(Med HH Inc) X Emp-Pop Ratio	-5.48
	(2.503)

Note: This table presents a subset of utility parameters from our model of household branch choice. These are the interaction effects between tract and branch characteristics as well as the interaction of tract characteristics and density. Robust standard errors are in parentheses.

This formula shows that concentration increases only in tracts where both banks have positive presence. Tracts where one or both banks have negligible share experience $\Delta THHI \approx 0$.

MSA-Level HHI. Define bank j's MSA-level market share as $S_j = D_j / \sum_{j'} D_{j'}$, where D_j is total deposits for bank j within the MSA. The MSA-level HHI is:

$$HHI_{\text{MSA}} = 10{,}000 \times \sum_j S_j^2. \tag{28}$$

The change from a merger between banks A and B is:

$$\Delta HHI_{\text{MSA}} = 20{,}000 \times S_A \times S_B. \tag{29}$$

D.3 Merger-Level Summary Statistics

For each simulated merger (j, k), we compute:

- **Mean** $\Delta THHI$: The average of $\Delta THHI_l$ across all tracts in the MSA.
- **P90** $\Delta THHI$: The 90th percentile of $\Delta THHI_l$ across tracts.
- **Max** $\Delta THHI$: The maximum tract-level change.
- **Number of tracts flagged**: Count of tracts where $\Delta THHI_l > 200$.
- **Share of tracts flagged**: Fraction of MSA tracts where $\Delta THHI_l > 200$.

D.4 Screening Rules and Classification

We classify each merger according to whether it exceeds standard regulatory thresholds:

MSA-level screen. A merger is "MSA-flagged" if $\Delta HHI_{\text{MSA}} > 200$. This corresponds to the threshold used in DOJ/FTC Horizontal Merger Guidelines to identify mergers warranting further scrutiny in moderately concentrated markets.

Tract-level screen. We consider two definitions:

1. **Mean criterion**: A merger is "tract-flagged" if mean $\Delta THHI > 200$.
2. **Breadth criterion**: A merger is "tract-flagged" if at least 10 tracts have $\Delta THHI_l > 200$.

Classification. Combining these screens yields four categories:

- **Both flag**: MSA-flagged and tract-flagged. The merger raises concerns under either measure.
- **Neither flags**: Passes both screens. Low competitive concern.
- **MSA only**: MSA-flagged but not tract-flagged. MSA measure may overstate local harm.
- **Tract only ("false negative")**: Tract-flagged but passes MSA screen. The MSA measure fails to capture localized concentration increases.

D.5 Regression Specifications

Merger-level regressions. To assess which merger characteristics predict large concentration effects, we estimate:

$$\overline{\Delta THHI}_{jk} = \beta_0 + \beta_1 \cdot \text{CustomerDistance}_{jk} + \beta_2 \cdot \text{BranchDistance}_{jk} + \beta_3 \cdot X_{jk} + \varepsilon_{jk}, \quad (30)$$

where $\overline{\Delta THHI}_{jk}$ is the mean tract-level change for merger (j, k), and X_{jk} includes additional controls such as share correlation, size ratio (larger bank deposits / smaller bank deposits), and combined market share. All regressors are standardized to facilitate comparison of effect sizes.

False negative prediction. To characterize which mergers fall into the false-negative region, we estimate:

$$\Pr(\text{FalseNegative}_{jk} = 1) = F(\gamma_0 + \gamma_1 \cdot \text{CustomerDistance}_{jk} + \gamma_2 \cdot \text{BranchDistance}_{jk} + \gamma_3 \cdot X_{jk}), \tag{31}$$

where $\text{FalseNegative}_{jk}$ indicates that merger (j, k) passes the MSA screen but is tract-flagged under the breadth criterion. We report both linear probability and probit specifications.

Tract-level regressions. To examine which tracts are most affected within a given merger, we estimate:

$$\Delta THHI_{ljk} = \alpha_0 + \alpha' Z_l + \delta_{jk} + \varepsilon_{ljk}, \tag{32}$$

where $\Delta THHI_{ljk}$ is the concentration change in tract l from merger (j, k), Z_l is a vector of tract characteristics, and δ_{jk} are merger fixed effects. The fixed effects absorb all merger-level variation, so coefficients α describe within-merger differences across tracts. Standard errors are clustered at the tract level.

We also estimate this specification separately for high customer-similarity mergers (top quartile of customer distance) and low customer-similarity mergers (bottom quartile) to assess whether distributional patterns differ by merger type.