Staff Working Paper/Document de travail du personnel —
2020-4

Last updated: February 6, 2020

A Spatial Model of Bank Branches in Canada

by Heng Chen and Matthew Strathearn

Currency Department
Bank of Canada, Ottawa, Ontario, Canada K1A 0G9

hchen@bankofcanada.ca, MStrathearn@bank-banque-canada.ca

ISSN 1701-9397

Acknowledgements

We thank Kim Huynh, Artūras Juodis and Marcel Cristian Voia for their helpful comments. We also thank participants of the Advances in Econometrics Conference: The Econometrics of Networks at Sibiu, Romania, for their helpful comments and suggestions. The views expressed in this paper are solely those of the authors and may differ from official Bank of Canada views. No responsibility for them should be attributed to the Bank of Canada. All errors are our own.

Abstract

This research aims to empirically analyze the spatial distribution of bank-branch networks in Canada. We study the market structure (both industrial and geographic concentrations) within the networks' own or adjacent postal areas. Our empirical framework considers branch density (the ratio of the total number of branches to the area size) by employing a spatial two-way fixed-effects model. Our main finding is that there are no effects associated with market structure; however, there are strong spatial socioeconomic effects from the networks' own and nearby areas. In addition, we also study the effect of spatial competition from rival banks: we find that large banks and small banks tend to avoid markets dominated by their competitors.

Bank topics: Firm dynamics; Market structure and pricing
JEL codes: L1, R3

1. Introduction

For many households and small businesses, physical bank branches are the primary vehicle for accessing financial services. Since most services are supplied on a geographical basis, the "nearness" of a branch has always been considered a source of utility for customers (at least for households and small firms). This is one of the reasons why commercial banks[1] build branch networks with the aim of extending their operating territories. Some contend that changes to the bank-branch network affect access to credit and other important services (see Agarwal and Hauswald (2010) for the effect of branch distance on lenders and lending conditions). However, instead of studying the effect of branch patterns on the provision of financial services, we use a spatial two-way model with fixed effects to identify and analyze driving factors behind changes in branch networks.

Using a novel dataset of bank-branch locations across Canada, we compute an innovative measure of geographic concentration to capture the spatial distribution of branches at the postal-code level. Using this measure, we set out to study spatial distribution impact on bank-branch density: that is, the ratio of the total number of branches to area size. In addition, we also investigate the following explanatory factors: (i) industrial concentration; (ii) socioeconomic variables; and (iii) competition from rival banks.

It is crucial to include geographic concentration in the model, because it is fundamentally different from industrial concentration. The addition of geographic concentration, as measured by the average distance to the closest bank branch, allows us to consider the degree of spatial clustering among bank branches in a given postal area. For example, given two postal codes with the same size branch network, a smaller geographic concentration is associated with a more even distribution of bank branches and, thus, would imply a reduction in travel costs; see Figure 2 for an illustration. By controlling for both industrial and geographic

[1] In this article, the terms "bank," "banking institution," and "banking office" pertain to commercial banks and credit unions (savings and loan associations and savings banks).

concentrations, we can capture not only the degree of competitiveness within a given area but also how well the area is serviced in terms of travel distance to the nearest branch.

Our main findings are as follows: Neither geographic nor industrial concentration significantly impacts bank-branch density at the postal-code level. However, we do find that certain aspects of the demographic profile are significantly correlated with bank-branch density. In terms of competition effects, we find evidence that both big and small entrants prefer to avoid markets saturated with their competitors. The following sections classify our contributions into three distinct categories.

1.1. Disaggregated Area Analysis

While many papers have studied the effects of large aggregate shocks to the national banking system (Kim and Vale (2001)), much less attention has been paid to shocks at the local level. Each year, numerous branches are opened and closed in Canada as banks adjust their physical footprints in response to changes in local market conditions and in firm objectives. While these adjustments occur below the surface of the aggregate banking system, they may have substantial effects at the local level. This paper sets out to study the variation of branch densities across Canadian postal-code areas. For this study we use the first three digits of these postal codes, also known as "Forward Sortation Areas" (FSAs), to be our unit of measurement. This is because the FSA is the finest disaggregated level for which we could obtain precise socioeconomic variables.

1.2. Spatial Dependence

FSAs are not independent submarkets but are very much interdependent within the web of FSAs. Thus we discuss three primary reasons for introducing spatial dependency between two nearby postal-code areas. First, expected relationships between spillovers and branching patterns may be important when the branch networks overlap geographically. Avery et al.

(1999) find that there is evidence of spatial dependence in their "within-market-but-not-within-ZIP" regression. Tranfaglia (2018) also provides a cross-sectional analysis based on Baltimore, Chicago and Philadelphia.

The second reason for introducing spatial dependency to our study is because the FSA level might not be the same as the branching decision level; that is, the specific FSA could be either small or large when the banking institution is making decisions about whether to open or close a branch there. This is because FSA codes were set up for the convenience of the postal service and its ground transportation system, but not for banking managers. De Juan (2003) discusses how to identify the boundary of the local branching market.

The third reason is related to the empirical analysis of entry games. A firm chooses to enter a market if it can obtain a non-negative profit in equilibrium. For example, Bresnahan and Reiss (1990), Bresnahan and Reiss (1991) and Carbo-Valverde and Perez-Saiz (2018) empirically analyze entry decisions in small isolated towns. By restricting their investigation to these isolated markets they can simplify the entry game and ignore competition across different geographic areas. However, these small isolated towns make up only a small portion of the economy and provide limited insight when trying to understand the economy as a whole. Therefore, to analyze the general economy, we have to account for spatial competition across different geographic areas. For example, Nishida (2014) estimates a structural model to investigate the expansion of store networks in the retail industry. His paper investigates the spatial competition within the same chain and across rival chains of convenience stores in Okinawa, Japan.

1.3. Geographical Concentration

Our final contribution is the construction of a new measure of geographic concentration based on the point-pattern process. This new measure has the following characteristics: (i) it accounts for the clustering of branches; (ii) it is amenable to measuring the geographic concentration of disaggregated geographic units; and (iii) it can be easily scaled up for aggregated geographic units and weighted in a variety of ways (i.e. it allows for the inclusion of population weights). First, given that we measure the distance between grid points and the nearest associated branches, our distance measure is able to differentiate various degrees of clustering. Take an example of two FSAs with the same number of branches, where branches in the first one are evenly distributed, while in the second FSA the branches are more clustered at the centroid. Our concentration measure would assign the former FSA with a lower geographic concentration than the latter (refer to Figure 2). Second, the measure used by Duranton and Overman (2005) is only available for the large/aggregated geographic unit (e.g., regions and provinces), whereas our approach can be computed for the small/disaggregated geographic units. This is because Duranton and Overman (2005) compute the bilateral distance between branches: in order for the law of large numbers (LLN) to apply, the number of branches in each area need to be large. However, our approach is based on the distance between the grid point and the nearest branch, where the number of grid points can be as large as possible so that the LLN naturally holds. In the end, since our measure of geographic concentration is built from the finest level, it can be easily aggregated by applying weights for either population or commercial density.

The remainder of the paper is organized as follows. In section 2, we discuss the data used in the empirical analysis. In section 3, we present our empirical framework and provide the results. Section 4 looks at the robustness of our results, and section 5 concludes the paper.

2. Data

Our analysis relies on a novel dataset that combines information on the physical location of bank branches and the demographic characteristics of the FSA. We compile our data from three major sources. First, we geocode the bank branches and compute both the geographic and industrial concentrations by using the Financial Institution File (FIF)[2] provided by Payments Canada. Second, the socioeconomic variables come from the Canadian census.[3] Third, we use the 2011 cartographic FSA boundary file[4] provided by Statistics Canada to map branches into FSAs, and we generate the adjacency matrix (refer to Figure 3 for an example) for the spatial analysis.

To avoid other confounding factors, such as regulation changes, we focus on the period from 2008 to 2018. During this period, there were few regulatory changes that directly impacted bank-branching decisions (expansions or contractions). Moreover, there were relatively few mergers, acquisitions, and consolidations: although there were sizeable changes in the number of branches, the number of the banking institutions remained stable. This unique period differs from those studied by Damar (2007), Allen et al. (2014), and Nguyen (2019), who examine the changes in the number of branches resulting from mergers.

2.1. Payments Canada's Financial Institution File

We used the FIF database from 2008 to 2018 to obtain the physical addresses of all of the bank branches associated with the specific routing numbers in our study. Once we extracted the address information, we classified each observation by latitude and longitude. Next, since each observation was represented by a routing number, we made every effort to exclude non-retail branches and duplicate physical locations.[5] Our exact methodology for compiling the FIF files can be found in Appendix F.

[2]The up-to-date FIF (.pdf) can be found here. We used Payments Canada's annual data for 2008 to 2018, inclusive.

[3]The Census data can be found here

[4]We opted for the cartographic boundary files as they exclude coastal water areas: Statistics Canada (2011).

[5]A duplicate physical location may arise if multiple routing numbers are associated with a single location.

2.1.1. Geographic Concentration

We find that bank branches often cluster. For example, see Figure 1. In order to distinguish between the clustered or dispersed branches in the FSAs, we applied the average empty-space distance function between the grid points and the nearest branches. Our algorithm includes three main steps (see Figure 2 for an example):

1. Create an evenly spaced fine grid of points (red dots) for each FSA;
2. Calculate the Haversine distance between each grid point and the nearest branch (black triangle);[6]
3. Compute the average distance across all grid points within that FSA.

This measure represents, on average, how far an agent would need to travel to reach the closest bank branch. This implies that if branches tend to cluster (disperse), we would observe an increase (decrease) in the average distance traveled (see Figure 2 for an example).

Since our measure is distance-based, we avoid several drawbacks of the existing non-distance-based geographic-concentration measures; see Ellison and Glaeser (1997) and Devereux et al. (2004). For example, most non-distance-based measures allocate points ex-ante (e.g., branches) to spatial units (e.g., the FSA area) at a given level of aggregation. Aggregating data in this way has the obvious advantage of simplifying computations but it throws away a large amount of information (e.g., the exact location of branches) and leads to a range of problems. Furthermore, and importantly, after aggregation has taken place, points in the spatial unit are treated as being evenly distributed, creating a *downward* bias when dealing with clustered branches. However, our distance-based approach avoids the issues associated with aggregation.

Moreover, compared with the existing distance-based approach proposed by Duranton and Overman (2005), whose distance is the bilateral distances between branches, our method is

[6] We allow for the closest branch for a given grid point to fall outside the FSA.

to compute the distance between the grid points and the nearest branches. Here the precision of our approach only requires the number of grid points in a specific FSA area to be large, but it imposes no restrictions on the number of branches in that particular area. Thus, we are able to compute the geographic concentration for each individual FSA area. But, due to the nature of the bilateral distance, Duranton and Overman (2005) require the number of branches to be large. Therefore, if applying their approach, we have to aggregate several FSA areas in order to ensure the number of branches is sufficiently large. Of course, this is at the cost of only having the geographic concentration for the aggregated area rather than for each individual FSA within the area.

2.1.2. Industrial Concentration

It is widely recognized that any informative measure of localization must control for industrial concentration. For instance, it is possible for two FSAs to have similar geographic concentrations, but one could have an equal market share for each branch (low industrial concentration) while the other has a branch with monopoly power (high industrial concentration). Thus our paper adopts the idea of De Juan (2003) by computing the concentration ratio of the big five banks per each FSA: that is, the proportion of the total number of big five bank branches over the total branches in each FSA. These big five banks are the Royal Bank of Canada (RBC), the Bank of Nova Scotia (Scotiabank), Toronto Dominion (TD), the Canadian Imperial Bank of Commerce (CIBC), and the Bank of Montreal (BMO).[7]

2.2. Canadian Census

The demographic variables were drawn from the Canadian censuses (2006, 2011, and 2016). Specifically, we chose the following variables from the demographic profile: median income of individuals aged 15 and older, population density, unemployment rate, percentage of

[7] This ratio is also computed for the European Central Bank (ECB) Data Warehouse, since the size (e.g., the lending/borrowing market share) of each branch is unobserved. Our measurement also differs from that of Carbo-Valverde and Perez-Saiz (2018), who use a structural-game theoretic model.

individuals employed in the financial industry, and median age. For this paper, we benchmarked our time periods at 2009, 2013 and 2017, which correspond to Bank of Canada's Methods-of-Payment (MOP) survey. However, since the years of the MOP do not coincide with the census, we used a simple linear interpolation for the 2009 and 2013 demographic variables. In addition, we extrapolated the census demographic variables for the year 2017. Since interpolation/extrapolation requires multiple observations, any gaps after the interpolations/extrapolation were filled using the demographic values from the nearest FSA (based on the closest centroid).

2.3. Statistics Canada Catographic Boundary File

We use the 2011 cartographic FSA boundary file provided by Statistics Canada to construct the FSA adjacency matrix. We compute the matrix for $1,357$ contiguous urban FSAs. The adjacency matrix has $6,020$ nonzero links, and the average number of links for each FSA is 4.43. The primary reason we consider the urban subsample is that the geographic size of the urban FSAs are relatively small, so the population can be thought as being evenly distributed and our evenly spaced fine grids are a reasonable approximation of this pattern. However, large rural FSAs pose difficulties when small communities are scattered.

3. Spatial Panel Model

One of the lesser explored issues in the literature is the degree of spatial dependency of branch densities among postal areas. To address this issue, we proceed with a spatial panel analysis that allows us to control for unobserved FSA fixed effects and account for technology-driven time-fixed effects. Our analysis considers various spatial specifications that account for spatial dependence among the dependent variable, independent variables, and error term.

A first generation of spatial models had been specified for cross-sectional data (Elhorst (2014)); however, many applications in spatial econometrics are currently based on panel data. As with cross-sectional models, spatial autocorrelation can be taken into account in multiple ways — by either endogenous or exogenous variables, or errors. These three potential spatial terms with two-way fixed effects (α_i, γ_t) can be described as

$$
\begin{aligned}
y_{it} &= \rho \sum_{i \neq j} w_{ij} y_{jt} + x'_{it}\beta + \sum_{i \neq j} w_{ij} x'_{jt}\theta + \alpha_i + \gamma_t + u_{it}, \qquad (1) \\
u_{it} &= \lambda \sum_{i \neq j} w_{ij} u_{jt} + \varepsilon_{it},
\end{aligned}
$$

where the variable $\sum_{i \neq j} w_{ij} y_{jt}$ refers to the spatially lagged endogenous variables. $\sum_{i \neq j} w_{ij} y_{jt}$ is equal to the average value of the dependent variable taken by neighbors of observation i (within the context of the adjacency matrix). Parameter ρ captures the endogenous interaction effect. Next, a contextual effect (or exogenous interaction) is captured by vector θ. Lastly, spatial interaction is also taken into account by specifying $\sum_{i \neq j} w_{ij} u_{jt}$, according to which the unobservable shocks affecting observation i interact with the shocks affecting i's neighbors. Parameter λ captures this correlated effect of the unobservables.

When $\rho \neq 0, \theta = \lambda = 0$, this is referred to a spatial autoregressive model (SAR). When $\theta \neq 0, \rho = \lambda = 0$, we have the spatial lag of X (SLX). If $\lambda \neq 0, \rho = \theta = 0$, then this gives a spatial error model (SEM). When $\rho \neq 0, \theta \neq 0, \lambda = 0$, this is called the Durbin spatial model (DSM) (see, Elhorst (2014)). One motivation of having DSM is to mitigate omitted variable bias in spatial regressions; see Equations (7), (8) and (9) in Fingleton and Le Gallo (2010). Alternatively, the DSM can be motivated by simply rearranging the SEM model in a "spatial" Cochrane–Orcutt transformation; see Equations (2), (3) and (4) in Gibbons and Overman (2012). In addition to these standard spatial econometrics models, our empirical applications also consider the following two variations: (1) a combination of the SLX and the SEM, which can be viewed as a reduced form of the SAR model with only first-order spatial lag as in Gibbons and Overman (2012); and (2) the augmented DSM model by having

$\rho \neq 0, \theta \neq 0, \lambda \neq 0$. The latter augmented DSM model is proposed by Fingleton and Le Gallo (2010) (therein Equations (10) and (11)), where they argue that such a model could further alleviate the problem of the omitted variable and could be estimated by using the 2SLS of Kelejian and Prucha (2007).[8]

There are two kinds of specifications for the spatial fixed effects. The current fixed effects α follows Baltagi et al. (2003), which is not spatially autocorrelated, while the idiosyncratic errors u are spatially autocorrelated. The other option can be followed from Kapoor et al. (2007), where both the fixed effects and the idiosyncratic errors are spatially autocorrelated. Although the two data-generating processes look similar, they imply different spatial spillover mechanisms governed by a different structure of the implied variance-covariance matrix: in Baltagi et al. (2003), only the component that varies over time diffuses spatially; while in Kapoor et al. (2007) both fixed effects and idiosyncratic errors diffuse spatially over time. However, since the within transformation eliminates the individual effects, from an empirical point of view, these two specifications (individuals effects are/are not spatially autocorrelated) are indistinguishable.

Readers familiar with the "neighborhood effects" literature will see immediate parallels between the spatial econometrics models (SAR, etc.) and the "linear-in-means" neighborhood-effects models (or "peer effects" models). The parallels between these fields have already been highlighted by Lee (2004), Lee (2007) and others. The generic neighborhood-effects model described by Manski (1993) and Manski (2000) takes the following form:

$$
\begin{aligned}
y_{it} &= \rho E\left[y_{it}|\Pi\right] + x_{it}\beta + E\left[x_{it}|\Pi\right]\theta + \alpha_i + \gamma_t + u_{it}, \\
u_{it} &= \lambda E\left[u_{it}|\Pi\right] + \varepsilon_{it},
\end{aligned}
$$

where variable Π indexes locations, which are usually nonoverlapping neighborhoods. In

[8]There is another variant of spatial models that includes a spatially lagged dependent variable and a spatially autocorrelated error term, where $\rho \neq 0, \lambda \neq 0, \theta = 0$; see Millo and Piras (2012).

practice, the literature on neighborhood effects uses an empirical version of our equation (1). One important difference is that, in spatial econometrics, equations like (1) are treated as the population-data-generating process, rather than an empirical analogue. Parameter ρ is taken to be the effect of the observed sample mean neighborhood outcome, rather than the effect of an unobserved population mean that is estimated from the data. The crucial difference is that spatial econometrics assumes that w_{ij} is known and represents real-world linkages. Researchers of neighborhood effects argue that the true w_{ij} is almost never known, and is, at best, a means of estimating $E[\cdot|\Pi]$. In other words, the assumption of knowledge about w_{ij} is critical.

Remark: Our current spatial panel model only considers the cross-sectional dependence that is induced by the spatial dependence but not by the common factor model. Sarafidis and Wansbeek (2012) point out that common factors and spatial panels make it possible to capture interactions between individuals. However, they adopt different strategies for this purpose. The spatial econometric models are based on a given structure of interactions between individuals in a panel. This structure is generally constructed from a geographical metric (distance between individuals). However, in common factor panels, the structure of interactions is not constrained a priori. Moreover, the common factor model is "global," while the spatial method is "local," in the sense that the degree of dependence decreases sufficiently quickly as the distance between units increases. Initially, spatial panels were used for panels that comprised a large number of individuals (relative to the temporal dimension), while common factor models were preferred when the temporal dimension was large enough to adequately build common factors. Compared with Holly et al. (2010), where they apply the common correlated effects (CCE) estimator to study U.S. housing prices that have a large time period T, we opt out of their CCE estimator and choose the two-way fixed-effects spatial model because of small T.[9]

Recently, a series of studies highlighted, through applications, the synergies between the two

[9] Westerlund et al. (2019) study a CCE method for fixed-T panels. We will leave that for future research.

approaches (Bhattacharjee and Holly (2011) and Ertur and Musolesi (2017)) and proposed methods that combine spatial effects and common factors (Pesaran and Tosetti (2011)). A recent application is proposed by Vega and Elhorst (2016), who study the development of unemployment disparities between Dutch regions using a model that takes into account not only spatial and temporal dependencies but also the presence of common factors. Their study emphasizes the importance of simultaneously considering these three dimensions. Their results suggest that spatial dependence remains an important factor in understanding the dispersion of regional unemployment rates, even when time dependency and the presence of common factors are taken into account.

3.1. Empirical Results: Baseline Models

We first present the results of the baseline model in which we do not control for the spatial effects across FSAs. The temporal dimension of our model is based on the Bank of Canada MOP survey years, namely, 2009, 2013, and 2017. Below, we discuss the results from the baseline two-way fixed-effects panel model.

3.1.1. Results: All Banks

The regression results with the branch density of all banks as the dependent variable can be found in Appendix A, Column (1) of Table 1. We find that income is positively and significantly correlated with bank-branch density. Banks prefer to locate physical branches in areas where there are higher incomes. One explanation for this is that banks expand their networks into areas with higher value deposits. In addition, higher income individuals may be more likely to make use of the various investment and depository services banks provide. In respect to competitive effects, we find that neither geographic concentration nor industrial concentration significantly influences bank-branch density. This is different from Cohen and Mazzeo (2010), who study the effect of industrial concentration on bank-branching decisions and find that networks are larger in more competitive markets.

3.1.2. Results: Small Banks

The next two regressions split the density between two major classes of banks. The first group consists of Canada's big five banks, and the second group is composed of smaller banks and credit unions. Splitting the two groups allows us to study how branching decisions change with respect to the size of banks. The regression results with small-bank density as the dependent variable can be found in Table 1, Column (2). We find that geographic concentration does not significantly influence small-bank-branch density. However, we do find that the density of large banks is negative and significantly correlated with the density of small banks. This suggests that smaller banks have a tendency to move away from areas where large banks have a foothold. The likely reason for this is that new entrants are unable to capture a large enough share of demand to trigger profitable entry.

As for demographics, both income and age are positive and significant predictors of small-bank density. In terms of age, small banks likely choose to locate in areas where there is higher demand for physical branches. In addition, since older individuals are likely to have higher levels of wealth relative to younger individuals, it may be the case that these areas exhibit higher demand for investment and other depository services. However, we find a negative and significant correlation between population size and small-bank density. This result is counter intuitive but likely arises from biased inference precipitated by excluding spatial effects across FSAs. We will revisit this discussion in Section 3.2.1.

3.1.3. Results: Big Banks

The regression results with the density of big five banks as the dependent variable can be found in Table 1, Column (3). Similar to the results in Section 3.1.2, we find that geographic concentration does not significantly influence big-bank-branch density. However, the density of small banks is negative and significantly correlated with the density of large banks. This suggests that large banks may have difficulty stealing market share in regions typically dominated by smaller banks and credit unions (these smaller institutions tend to target specific geographies and demographics, e.g., labor unions, ethnic groups, or geographic affiliations).

As for demographics, population is a positive and significant predictor of big-bank density. This result is not surprising and indicates that big banks seek out areas with higher expected demand. We find that age is negatively and significantly correlated with big-bank density. This suggests that big banks prefer to locate in areas with younger populations. This is in stark contrast to what we saw with small banks.

Our baseline analysis yields some interesting results; however, since we are dealing with small geographic units, meaningful spillovers are likely occurring between adjacent areas. By ignoring the underlying spatial structure, we risk reporting inconsistent estimates and, at the same time, provide misleading statistical inference. To correct these issues, we will include one of the following spatial elements in the model: either (1) a spatial autoregressive dependent variable; (2) spatial lagged independent variables; or (3) a spatial autoregressive error term. We begin our analysis with a discussion of model specification.

3.2. Main Results: All Banks

The main results of our analysis can be found in Appendix B, Tables 2, 3, and 4. First, we run regressions based on Equation (1) to determine the spatial coefficients of ρ, θ, and λ under the two-way fixed-effects model. This is the most general model to start with and nesting the special cases of the SAR, SLX, or SEM. From Column (4) of Table 2, we can see that the coefficients of ρ and λ are not significantly different from zero, while the coefficients on the spatial lag variables (independent variables) are significantly different from zero. This leads us to rule out the SAR model specification and impose either the SLX or SEM models.

Second, we test whether there is any spatial error dependence conditioning on the SLX by applying Millo (2017). This test is based on the values taken by the CD(p) test (Pesaran

(2004), Section 7) under the permutations of the neighborhood matrix. Our results show that the null hypothesis of independence is consistently rejected under either two-way fixed-effects models or random-effects models. Hence we conclude that both the SLX and SEM are features of the model.[10] We could further justify the inclusion of a spatial error term by comparing the results across the SLX and SLX/SEM models (Table 3, Columns (2) and (3)). We find that the two models are almost identical in terms of direction, magnitude, and significance. It does appear that the inclusion of the spatial error term has a significant effect on the standard errors of the coefficients. This is evidenced by the change in the standard errors across the two models and the significance of the spatial error parameter λ. As such, we argue that it is appropriate to include the spatial error. Furthermore, by computing the Moran I statistics, SLX+SEM (Column 3) reduces the magnitude of the statistics of Column (2) from -0.1363 to -0.0796.

Third, we present the Hausman specification test which makes it possible to arbitrate between a model where the fixed effects are not correlated with the explanatory variables and a model where such a correlation exists. This test determines which estimation method to use. Comparing Columns (1) and (2), we can see that the coefficient of income is switching signs. In addition, the result of the Hausman test (Mutl and Pfaffermayr (2011)) rejects the null hypothesis that the random-effects estimator is consistent. Based on the above discussion, Column (3) of Table 2 should represent the best approximation to the underlying DGP with the SLX and SEM plus the two-way fixed effects.

3.2.1. Results

We present a detailed discussion of the two-way fixed-effects panel model when we account for the spatial lag of the independent variable and the spatial error term (Column 3 of Table

[10]However, we should be cautious about the test result from Millo (2017) in the presence of time fixed effects. Juodis and Sarafidis (2018) point out that such test statistics suffer from severe size distortion when period-specific parameters are present.

2). Specifically, we present the maximum-likelihood estimates.[11]

- **Spatial Socioeconomics**

  We find that there is a positive relationship between income and bank-branch density. To explain this finding, we argue that banks are attracted to wealthier areas because they can guarantee larger deposits. In addition, banks may also benefit from a higher demand for their investment services and other depository services. These results are significant at the 1% level.

  The density of branches should increase with the population density of a market. As a market increases in population, more branches are established. Of course, it would be possible to accommodate an increase in population with the same number of branches by each branch handling more depositors. But an increase in population is likely to entail a disproportionate increase in depositor locations that are less fully served by the existing branch network and, at the margin, the establishment of new branches might be less costly and/or more effective than raising deposits from those existing locations. As expected, we find a positive and highly significant coefficient associated with population density within its own FSA. However, the spatial lag associated with population density is negative and statistically significant. This suggests that the decision to open a new branch within a specific FSA also accounts for the weighted average population density of the adjacent FSAs. In this case, a bank would reduce the number of branches in an area that has a higher population in neighboring areas, because it might be better to simply relocate the branch in this densely populated neighboring area.

  Unemployment is a proxy for the demand for credit. The demand for credit emerges when unemployed individuals are forced to smooth their consumption. Based on this assumption, higher levels of unemployment might be associated with a higher demand

[11] We also apply the generalized method of moments (GMM) and the results are very similar.

for credit and we could expect banks to build their networks in areas with higher credit demand. However, since unemployment is usually negatively correlated with income, the direct effect from unemployment is not statistically significant.

By introducing branches in a certain geographical area, banks can better obtain and process borrower-specific local information and thus can maintain the quality of their loan portfolio. In fact, as has been recently documented by Jayaratne and Strahan (1996), the relaxation of US branching regulations has been an important source of an increase in the rate of real per capita growth in income and output. This growth is shown to have emanated from improved loan monitoring and screening, which was the result of the proliferation of branch networks. Following this argument, the number of branches increases with respect to the increasing demand for financial services. On the one hand, however, we find that the coefficient associated with the size of the finance industry is negative and not significant. On the other hand, the spatial lag associated with this industry is negative and significant. This suggests that banks may avoid areas surrounded by large financial sectors, instead, they rather locate branches in these surrounding areas.

We introduce age to measure the adaptation of new and innovative forms of banking, such as e-banking and mobile banking. We argue that for older individuals, both the fixed and variable costs of adopting innovative forms of banking exceed the gains from these services. Therefore, the demand for physical branch locations will be higher for older individuals than for younger individuals. We would except the number of bank branches to be increasing in the median age of the population. Although the coefficient is positive, it is not statistically significant.

To conclude, we find that both income and population have a positive and significant effect on bank-branch density. In addition, we find that unemployment, the size of the financial sector, and age are poor predictors of bank-branch density. When we study the spatial lag of demographic variables, we find that the spatial lag of population and the spatial lag of the size of the financial sector are both negative and significant predictors.

- **Spatial market structure**

  We argue that markets dominated by Canada's five major banks create barriers to entry for smaller players. Specifically, a smaller entrant may not be able to capture a large enough market share to profitably operate in the long-run. Although concentration impacts competition between the major five banks and the smaller players, it also likely reduces the incentives for competition among the five major banks. As such, a negative coefficient is consistent with the theory. The finding of a negative relationship between market concentration and branching has been reported recently by Rice and Davis (2007). However, we find that the coefficient associated with industrial concentration is negative but not significant. In addition, we do not find any significance for the industrial concentration from nearby areas.

  Surprisingly, the spatial geographical concentration does not play an important role in determining the branch density at the FSA level. This may be the case because within the same postal area there is apparent clustering; the physical distance or localization is small due to the small area size of the usual postal area. In many ways, geographical concentration, from the perspective of the consumer, captures the costs associated with traveling to a physical bank branch. However, from the perspective of the bank, geographical concentration is not a measure of profitability and, as such, should not be a significant factor in the branching decision. We do want to point out that, from the perspective of a central planner, geographical concentration could be used to locate branches in a manner that increases consumer surplus (through a reduction in travel costs).

3.3. Empirical Results: Big Five and Small Banks

In addition to our regression defined in Section 3.2, we also consider a regression of branch density for either the big five banks or other, smaller banks. In addition to the regressors used in Section 3.1, we remove industrial concentration to avoid reverse causality from big-bank densities and industrial concentration having the same numerator: the total number of big five banks. Instead, we replace the industrial concentration with the spatial competition of rival banks, which is measured by the counterpart of the either small- or big-five-branch density. Our regression results are presented below and can be found in Appendix B, Tables 3 and 4.

As for the potential spatial competition within the same banks, our model SAR+SLX+SEM specification does not find strong evidence (Column 4 in Tables 3 and 4): there are no nearby competition effects for the branch density. This suggests that there are few economies of density to be gained by using the same marketing strategies for all branches in the collection of postal areas, or in saving management costs if several branches were close to each other.[12] However, there does appear to be the presence of a spatial error term and this is evidenced by the significance of the λ coefficient in Column (3). Therefore, we direct our focus to the results of Column (3). Since we thoroughly discussed the demographic variables in the previous section, we omit a detailed discussion here and instead focus on the coefficients associated with geographic concentration and competition of rival banks.

3.3.1. Results

- **Spatial geographic concentration**

  For both the big-bank case and the small-bank case, we find that geographic concentration is negative but not statistically significant.

[12] In this paper, we do not observe branch-level profits or revenues. Therefore, we cannot identify the spatial competition within the same bank for branches located in the same FSA. However, using the location data, we can estimate this effect for branches located in nearby FSAs.

- **Spatial competition of rival banks**

  The usual business-stealing effect would imply a negative effect of the number of rival branches on a bank's desire to open a branch (Berry (1992)). Increasing the number of rival branches would reduce the residual demand and hence the profits earned from opening a branch would decline. For both big- and small-bank cases we find that the coefficients associated with the density of competing banks are negative. This suggests big and small banks prefer markets that are not saturated by their competitors. Strengthening this finding, we also observe negative and significant coefficients associated with the spatial lag of the rival's density measure. This suggests that not only are banks averse to dense markets, but they are also averse to markets that are surrounded by dense markets (at the FSA level).

4. Robustness

4.1. Ten Years of Panel Data without Socioeconomic Variables

As a robustness check, we use 10 years worth of data to test the effect of geographical and industrial concentrations on bank-branch density. In this model, we exclude the census data as there is on average a five-year gap between census years. For the same reasons we discussed earlier, we focus our attention on Column (3) in Appendix C. We find that in the all branch model, Appendix C, Table 6, both industrial concentration and geographic concentration are negatively and significantly correlated with bank-branch density. Nevertheless, different from the three-years model in Table 2, the results in Table 6 show the potential bias of omitting the socioeconomic variables when analyzing the branching decision.

Looking at Column (3) of Appendix C, Tables 7 and 8, we observe that the bank-branch density of competitors, the spatial lag of competitor's branch density, and the geographic concentration are all negative and statistically significant. The signs of the coefficients are largely similar in direction to what we found in the three-year regression models. However, again it is shown that omitting the socioeconomic variables would bias the effects of the market structure, and overestimate the importance of the geographic concentration. If we were to run the same model with individual fixed effects only (Appendix C, Table 9), the results would be almost identical in terms of both significance and direction. Comparing the significance of the spatial error term (λ) across both specifications yields similar results.

4.2. Census Metropolitan Area (CMA) Results

In order to check whether the previous results are driven by the FSA level, we also carry out the same two-way fixed-effects panel analysis based on the CMA level. A CMA is defined as having a total population of at least 100,000, of which 50,000 or more live in the core. Thus the CMA area could contain several FSAs and is aggregated from the current FSA units. In addition, each CMA is a large, isolated area so we do not need to apply the spatial analysis as for the FSA. From Appendix D, Table 10, the results are comforting and align well with the FSA-level analysis: the population density is a main driver for the branch density, while the coefficients of the geographic concentration are statistically insignificant across total, big and small branch densities, respectively.

5. Conclusion

This analysis has emphasized the relationship between market characteristics and bank-branching patterns. We find that the branching decision in a specific postal-code area is mainly driven by its within and nearby socioeconomic variables, such as income, population, and the size of the financial sector. Interestingly, the geographic concentration does not play an important role, which implies that the opening/closing decision does not account for branch accessibility in terms of the travel distance between the agents and the nearest branches. Thus, financial institutions may continue to close more branches in remote areas, only based on local economic conditions. This decreases accessibility by reducing the availability of banking services (e.g., cash withdrawals). In order to maintain and improve branch accessibility, a central planner might want to impose some restrictions on the minimum level of geographic concentration within a given area. For example, the Financial Consumer Agency of Canada requires financial institutions to give four months' notice when planning to close a branch or end certain activities. In the European Economic Area, the Swedish parliament proposes to take action and ensure that 99% of Swedes have a maximum travel distance of 16 miles to the nearest cashpoint.

We have not directly measured the delivery of services provided by the bank branch. Thus, an important question, and one that is not answerable with the current data, is whether a change in the number of branches has influenced the supply or pricing of banking services. In addition, we do not explicitly account for every restriction imposed by the regulations. For instance, a bank cannot always freely adjust branch locations. There is also an upper bound on the number of new branches that can be opened in a year. To fully analyze these regulations and evaluate the policy implications, we will develop a structural model to explicitly take care of these restrictions. This is left for future research.

References

Agarwal, Sumit and Robert Hauswald (2010), "Distance and Private Information in Lending." *The Review of Financial Studies*, 23, 2757–2788, URL https://dx.doi.org/10.1093/rfs/hhq001. 1

Allen, Jason, Robert Clark, and Jean-François Houde (2014), "The effect of mergers in search markets: Evidence from the Canadian mortgage industry." *American Economic Review*, 104, 3365–3396, URL http://www.jstor.org/stable/43495323. 5

Amel, Dean F and Kenneth P Brevoort (2005), "The perceived size of small business banking markets." *Journal of Competition Law and Economics*, 1, 771–784.

Angrist, Joshua D (2014), "The perils of peer effects." *Labour Economics*, 30, 98–108.

Avery, Robert B, Raphael W Bostic, Paul S Calem, and Glenn B Canner (1999), "Consolidation and bank branching patterns." *Journal of Banking & Finance*, 23, 497–532. 2

Baltagi, Badi H, Seuck Heun Song, and Won Koh (2003), "Testing panel data regression models with spatial error correlation." *Journal of Econometrics*, 117, 123–150. 10

Berry, Steven T (1992), "Estimation of a model of entry in the airline industry." *Econometrica: Journal of the Econometric Society*, 889–917. 20

Bhattacharjee, Arnab and Sean Holly (2011), "Structural interactions in spatial panels." *Empirical Economics*, 40, 69–94. 12

Bivand, Roger, Jan Hauke, and Tomasz Kossowski (2013a), "Computing the Jacobian in Gaussian spatial autoregressive models: An illustrated comparison of available methods." *Geographical Analysis*, 45, 150–179, URL https://doi.org/10.1111/gean.12008.

Bivand, Roger and Gianfranco Piras (2015), "Comparing implementations of estimation methods for spatial econometrics." *Journal of Statistical Software*, 63, 1–36, URL https://www.jstatsoft.org/v63/i18/.

Bivand, Roger and David W. S. Wong (2018), "Comparing implementations of global and local indicators of spatial association." *TEST*, 27, 716–748, URL https://doi.org/10.1007/s11749-018-0599-x.

Bivand, Roger S., Edzer Pebesma, and Virgilio Gomez-Rubio (2013b), *Applied Spatial Data Analysis with R, Second edition*. Springer, NY, URL http://www.asdar-book.org/.

Bresnahan, Timothy F and Peter C Reiss (1990), "Entry in monopoly market." *The Review of Economic Studies*, 57, 531–553. 3

Bresnahan, Timothy F and Peter C Reiss (1991), "Entry and competition in concentrated markets." *Journal of Political Economy*, 99, 977–1009. 3

Brevoort, Kenneth P, John D Wolken, and John A Holmes (2010), "Distance still matters: the information revolution in small business lending and the persistent role of location, 1993-2003." Technical report, FEDS Working Paper.

Carbo-Valverde, Santiago and Hector Perez-Saiz (2018), "Competition, geographic proximity and pricing in the retail banking industry." Technical report, Working Paper. 3, 7

Chang, Angela, Shubham Chaudhuri, and Jith Jayaratne (1997), "Rational herding and the spatial clustering of bank branches: an empirical analysis." Research Paper 9724, Federal Reserve Bank of New York, URL https://EconPapers.repec.org/RePEc:fip:fednrp:9724.

Cohen, Andrew and Michael J Mazzeo (2010), "Investment strategies and market structure: An empirical analysis of bank branching decisions." *Journal of Financial Services Research*, 38, 1–21. 12

Damar, Evren (2007), "Does post-crisis restructuring decrease the availability of banking services? The case of Turkey." *Journal of Banking & Finance*, 31, 2886–2905. 5

De Juan, Rebeca (2003), "The independent submarkets model: An application to the Spanish retail banking market." *International Journal of Industrial Organization*, 21, 1461–1487. 3, 7

Devereux, Michael P, Rachel Griffith, and Helen Simpson (2004), "The geographic distribution of production activity in the UK." *Regional Science and Urban Economics*, 34, 533 – 564, URL http://www.sciencedirect.com/science/article/pii/S0166046203000735. 6

Duranton, Gilles and Henry G Overman (2005), "Testing for localization using micro-geographic data." *The Review of Economic Studies*, 72, 1077–1106. 4, 6, 7

Elhorst, J Paul (2014), *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*. Springer. 8, 9

Ellison, Glenn and Edward L Glaeser (1997), "Geographic concentration in US manufacturing industries: A dartboard approach." *Journal of Political Economy*, 105, 889–927. 6

Ertur, Cem and Antonio Musolesi (2017), "Weak and strong cross-sectional dependence: A panel data analysis of international technology diffusion." *Journal of Applied Econometrics*, 32, 477–503. 12

Fingleton, Bernard and Julie Le Gallo (2010), "Endogeneity in a spatial context: Properties of estimators." *Progress in Spatial Analysis*, 59–73. 9, 10

Gibbons, Stephen and Henry G Overman (2012), "Mostly pointless spatial econometrics?" *Journal of Regional Science*, 52, 172–191. 9

Holly, Sean, M Hashem Pesaran, and Takashi Yamagata (2010), "A spatio-temporal model of house prices in the USA." *Journal of Econometrics*, 158, 160–173. 11

Hoxby, Caroline M (2000), "Does competition among public schools benefit students and taxpayers?" *American Economic Review*, 90, 1209–1238.

Jayaratne, Jith and Philip E Strahan (1996), "The finance-growth nexus: Evidence from bank branch deregulation." *The Quarterly Journal of Economics*, 111, 639–670. 17

Juodis, Artūras and Simon Reese (2018), "The incidental parameters problem in testing for remaining cross-section correlation." *arXiv preprint arXiv:1810.03715*.

Juodis, Artūras and Vasilis Sarafidis (2018), "Fixed T dynamic panel data estimators with multifactor errors." *Econometric Reviews*, 37, 893–929. 15

Kapoor, Mudit, Harry H Kelejian, and Ingmar R Prucha (2007), "Panel data models with spatially correlated error components." *Journal of Econometrics*, 140, 97–130. 10

Kelejian, Harry H and Ingmar R Prucha (2007), "HAC estimation in a spatial framework." *Journal of Econometrics*, 140, 131–154. 10

Kim, Moshe and Bent Vale (2001), "Non-price strategic behavior: The case of bank branches." *International Journal of Industrial Organization*, 19, 1583–1602. 2

Lee, Lung-Fei (2004), "Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models." *Econometrica*, 72, 1899–1925. 10

Lee, Lung-fei (2007), "Identification and estimation of econometric models with group interactions, contextual factors and fixed effects." *Journal of Econometrics*, 140, 333–374. 10

Manski, Charles F (1993), "Identification of endogenous social effects: The reflection problem." *The Review of Economic Studies*, 60, 531–542. 10

Manski, Charles F (2000), "Economic analysis of social interactions." *Journal of Economic Perspectives*, 14, 115–136. 10

Mazzeo, Michael J (2003), "Competition and service quality in the US airline industry." *Review of Industrial Organization*, 22, 275–296.

Millo, Giovanni (2017), "A simple randomization test for spatial correlation in the presence of common factors and serial correlation." *Regional Science and Urban Economics*, 66, 28–38. 14, 15

Millo, Giovanni and Gianfranco Piras (2012), "splm: Spatial panel data models in R." *Journal of Statistical Software*, 47, 1–38, URL http://www.jstatsoft.org/v47/i01/. 10

Mutl, Jan and Michael Pfaffermayr (2011), "The Hausman test in a Cliff and Ord panel model." *The Econometrics Journal*, 14, 48–76. 15

Nguyen, Hoai-Luu Q. (2019), "Are credit markets still local? Evidence from bank branch closings." *American Economic Journal: Applied Economics*, 11, 1–32, URL http://www.aeaweb.org/articles?id=10.1257/app.20170543. 5

Nishida, Mitsukuni (2014), "Estimating a model of strategic network choice: The convenience-store industry in Okinawa." *Marketing Science*, 34, 20–38. 3

Pesaran, M. Hashem (2004), "General diagnostic tests for cross section dependence in panels." Cambridge Working Papers in Economics 0435, Faculty of Economics, University of Cambridge. 14

Pesaran, M Hashem and Elisa Tosetti (2011), "Large panels with common factors and spatial correlation." *Journal of Econometrics*, 161, 182–202. 12

Rice, Tara and Erin Davis (2007), "The branch banking boom in Illinois: A byproduct of restrictive branching laws." *Chicago Fed Letter*, 238. 18

Sarafidis, Vasilis and Tom Wansbeek (2012), "Cross-sectional dependence in panel data analysis." *Econometric Reviews*, 31, 483–531. 11

Statistics Canada (2011), "Forward Sortation Area Boundary File, 2011 Census." *Statistics Canada Catalogue no. 92-179-X.* 5

Townsend, Robert M and Victor V Zhorin (2014), "Spatial competition among financial service providers and optimal contract design." Technical report, Working paper. http://www. robertmtownsend. net/working_papers.

Tranfaglia, Anna (2018), "Shrinking networks: A spatial analysis of bank branch closures." Technical report, FRB of Philadelphia Working Paper No. 18-12. 3

Vega, Solmaria Halleck and J Paul Elhorst (2016), "A regional unemployment model simultaneously accounting for serial dynamics, spatial dependence and common factors." *Regional Science and Urban Economics*, 60, 85–95. 12

Westerlund, Joakim, Yana Petrova, and Milda Norkute (2019), "CCE in fixed-T panels." *Journal of Applied Econometrics*, 34. 11

A. Regression Results (Baseline Models)

Table 1: Two-way Fixed-effects Panel Results (Urban Sample)

	Dependent variables:		
	Bank Density	Small-Bank Density	Big-Bank Density
	(1)	(2)	(3)
Income ($ 1000's)	0.008***	0.009***	0.0001
	(0.003)	(0.002)	(0.002)
Population (1000's per km^2)	0.018	−0.054***	0.079***
	(0.017)	(0.015)	(0.011)
Unemployment Rate (%)	0.009*	0.005	0.005*
	(0.005)	(0.004)	(0.003)
NAICS Finance (% Employed)	−0.011	−0.012*	0.001
	(0.009)	(0.007)	(0.006)
Age (Median)	0.006	0.014***	−0.009***
	(0.004)	(0.004)	(0.003)
Industrial Concentration (Big 5/All)	−0.011		
	(0.058)		
Big-Bank Density (per km^2)		−0.154***	
		(0.025)	
Small-Bank Density (per km^2)			−0.090***
			(0.015)
Geographic Concentration (km)	−0.024	−0.017	−0.009
	(0.018)	(0.014)	(0.011)
Observations	4,071	4,071	4,071
R^2	0.007	0.035	0.044
Adjusted R^2	−0.495	−0.451	−0.439
F Statistic (df = 7; 2705)	2.578**	14.207***	17.724***

Note: *p<0.1; **p<0.05; ***p<0.01

B. Regression Results (3 Years)

Table 2: Spatial Panel Results (Urban Sample)

	Dependent variable:			
	Branch Density (per km^2)			
	SLX/RE	SLX Two-way FEs	SLX+SEM Two-way FEs	SAR+SLX+SEM Two-way FEs
	(1)	(2)	(3)	(4)
Income (\$ 1000's)	-0.0005	0.007^{***}	0.007^{***}	0.008^{***}
	(0.002)	(0.003)	(0.002)	(0.002)
Population (1000's per km^2)	0.069^{***}	0.070^{***}	0.070^{***}	0.064^{***}
	(0.019)	(0.019)	(0.016)	(0.019)
SL Population	-0.041	-0.175^{***}	-0.177^{***}	-0.180^{***}
	(0.029)	(0.029)	(0.023)	(0.023)
Unemployment Rate (%)	0.006	0.006	0.005	0.006
	(0.005)	(0.005)	(0.004)	(0.004)
NAICS Finance (% Employed)	0.007	-0.002	-0.001	-0.002
	(0.009)	(0.009)	(0.007)	(0.007)
SL NAICS Finance	-0.008	-0.030	-0.030^{**}	-0.031^{**}
	(0.020)	(0.019)	(0.015)	(0.016)
Age (Median)	-0.004	0.001	0.003	0.003
	(0.004)	(0.004)	(0.003)	(0.003)
Industrial Concentration (Big 5/All)	-0.025	-0.019	-0.018	-0.020
	(0.060)	(0.058)	(0.047)	(0.047)
SL Industrial Concentration	-0.121	-0.103	-0.101	-0.106
	(0.105)	(0.102)	(0.080)	(0.081)
Geographic Concentration (km)	-0.023^{*}	-0.024	-0.024	-0.024
	(0.013)	(0.018)	(0.015)	(0.015)
SL Geographic Concentration	-0.030^{*}	-0.010	-0.010	-0.013
	(0.016)	(0.032)	(0.025)	(0.026)
ρ				-0.1104
ρ P-value				0.583
λ			-0.1876	-0.1078
λ P-value			0.0000	0.5938
Moran I	0.0208	-0.1363	-0.0796	-0.0473
Observations	4,071	4,071	4,071	4,071

Note: *p<0.1; **p<0.05; ***p<0.01

Table 3: Spatial Panel Results (Urban Sample)

	Dependent variable:			
	Small-Bank Density (per km^2)			
	SLX/RE	SLX Two-way FEs	SLX+SEM Two-way FEs	SAR+SLX+SEM Two-way FEs
	(1)	(2)	(3)	(4)
Income ($ 1000's)	−0.002	0.008***	0.008***	0.008***
	(0.001)	(0.002)	(0.002)	(0.002)
Population (1000's per km^2)	−0.097***	−0.009	−0.010	−0.012
	(0.015)	(0.016)	(0.013)	(0.015)
SL Population	0.017	−0.135***	−0.130***	−0.134***
	(0.021)	(0.024)	(0.020)	(0.024)
Unemployment Rate (%)	−0.006	0.003	0.002	0.003
	(0.004)	(0.004)	(0.003)	(0.003)
NAICS Finance (% Employed)	0.002	−0.004	−0.004	−0.004
	(0.009)	(0.008)	(0.006)	(0.006)
SL NAICS Finance	−0.056***	−0.032*	−0.032**	−0.033**
	(0.016)	(0.016)	(0.013)	(0.014)
Age (Median)	0.011***	0.010***	0.011***	0.011***
	(0.004)	(0.004)	(0.003)	(0.003)
Big-Bank Density (per km^2)	0.820***	−0.179***	−0.183***	−0.184***
	(0.014)	(0.026)	(0.021)	(0.022)
SL Big-Bank Density	0.325***	−0.126**	−0.155***	−0.163***
	(0.026)	(0.051)	(0.040)	(0.060)
Geographic Concentration (km)	−0.004	−0.015	−0.015	−0.016
	(0.004)	(0.015)	(0.012)	(0.012)
SL Geographic Concentration	−0.003	−0.011	−0.011	−0.012
	(0.005)	(0.026)	(0.021)	(0.021)
ρ				−0.0389
ρ P-value				0.8299
λ			−0.0806	−0.0507
λ P-value			0.0000	0.7813
Moran I	0.0168	−0.0923	−0.0361	−0.023
Observations	4,071	4,071	4,071	4,071

Note: *p<0.1; **p<0.05; ***p<0.01

Table 4: Spatial Panel Results (Urban Sample)

	Dependent variable:			
	Big-Bank Density (per km^2)			
	SLX/RE	SLX Two-way FEs	SLX+SEM Two-way FEs	SAR+SLX+SEM Two-way FEs
	(1)	(2)	(3)	(4)
Income ($ 1000's)	0.002*	0.00001	0.0001	0.0003
	(0.001)	(0.002)	(0.001)	(0.001)
Population (1000's per km^2)	0.025**	0.094***	0.100***	0.095***
	(0.012)	(0.012)	(0.010)	(0.012)
SL Population	0.083***	−0.056***	−0.087***	−0.076***
	(0.017)	(0.019)	(0.014)	(0.020)
Unemployment Rate (%)	0.008**	0.005	0.004*	0.004*
	(0.003)	(0.003)	(0.002)	(0.002)
NAICS Finance (% Employed)	0.016**	0.003	0.003	0.003
	(0.007)	(0.006)	(0.005)	(0.005)
SL NAICS Finance	0.048***	0.002	0.005	0.006
	(0.013)	(0.013)	(0.009)	(0.010)
Age (Median)	−0.014***	−0.010***	−0.007***	−0.008***
	(0.003)	(0.003)	(0.002)	(0.002)
Small-Bank Density (per km^2)	0.521***	−0.095***	−0.093***	−0.097***
	(0.009)	(0.015)	(0.012)	(0.012)
SL Small-Bank Density	0.246***	0.014	−0.083***	−0.097**
	(0.017)	(0.033)	(0.023)	(0.040)
Geographic Concentration (km)	0.001	−0.010	−0.011	−0.011
	(0.004)	(0.012)	(0.010)	(0.009)
SL Geographic Concentration	0.006	0.005	0.005	0.003
	(0.004)	(0.020)	(0.015)	(0.016)
ρ				−0.1402
ρ P-value				0.5468
λ			−0.3281	−0.2032
λ P-value			0.0000	0.3829
Moran I	0.0285	−0.1358	−0.1589	−0.1057
Observations	4,071	4,071	4,071	4,071

Note: *p<0.1; **p<0.05; ***p<0.01

Table 5: Spatial Panel Results: Individual Effects (Urban Sample)

	Dependent variable:			
	Branch Density (per km^2)			
	SLX/RE	SLX/FE	SLX+SEM/FE	SAR+SLX+SEM/FE
	(1)	(2)	(3)	(4)
Income ($ 1000's)	−0.0005	0.002	0.002	0.002*
	(0.002)	(0.002)	(0.001)	(0.001)
Population (1000's per km^2)	0.069***	0.070***	0.071***	0.065***
	(0.019)	(0.019)	(0.016)	(0.020)
SL Population	−0.041	−0.186***	−0.188***	−0.192***
	(0.029)	(0.029)	(0.023)	(0.023)
Unemployment Rate (%)	0.006	0.004	0.003	0.004
	(0.005)	(0.005)	(0.003)	(0.004)
NAICS Finance (% Employed)	0.007	−0.002	−0.001	−0.002
	(0.009)	(0.009)	(0.007)	(0.007)
SL NAICS Finance	−0.008	−0.035*	−0.035**	−0.036**
	(0.020)	(0.019)	(0.015)	(0.016)
Age (Median)	−0.004	−0.003	−0.001	−0.002
	(0.004)	(0.004)	(0.003)	(0.003)
Industrial Concentration (Big 5/All)	−0.025	−0.026	−0.024	−0.027
	(0.060)	(0.058)	(0.047)	(0.047)
SL Industrial Concentration	−0.121	−0.115	−0.112	−0.117
	(0.105)	(0.102)	(0.080)	(0.082)
Geographic Concentration (km)	−0.023*	−0.024	−0.025	−0.025*
	(0.013)	(0.018)	(0.015)	(0.015)
SL Geographic Concentration	−0.030*	−0.018	−0.018	−0.021
	(0.016)	(0.032)	(0.025)	(0.026)
ρ				−0.1078
ρ P-value				0.605
λ			−0.1841	−0.106
λ P-value			0	0.6126
Moran I	0.0208	−0.1362	−0.0781	−0.0465
Observations	4,071	4,071	4,071	4,071

Note: *p<0.1; **p<0.05; ***p<0.01

C. Regression Results (10 Years)

Table 6: Spatial Panel Results (Urban Sample: 10 Years)

	Dependent variable:			
	Bank Density (per km^2)			
	SLX/RE	SLX Two-way FEs	SLX+SEM Two-way FEs	SAR+SLX+SEM Two-way FEs
	(1)	(2)	(3)	(4)
Industrial Concentration (Big 5/All)	-0.070^{*}	-0.065^{*}	-0.065^{*}	-0.065
	(0.038)	(0.038)	(0.037)	(0.050)
SL Industrial Concentration	-0.076	-0.051	-0.050	-0.055
	(0.066)	(0.068)	(0.063)	(0.227)
Geographic Concentration (Km)	-0.030^{***}	-0.032^{**}	-0.032^{***}	-0.032^{***}
	(0.010)	(0.013)	(0.012)	(0.012)
SL Geographic Concentration	-0.022	-0.002	-0.001	-0.003
	(0.014)	(0.022)	(0.020)	(0.091)
ρ				-0.0727
ρ P-value				0.9792
λ			-0.1282	-0.0727
λ P-value			0.0000	0.9792
Moran I	0.1015	0.0349	-0.0522	-0.0303
Observations	14,927	14,927	14,927	14,927

Note: *p<0.1; **p<0.05; ***p<0.01

Table 7: Spatial Panel Results (Urban Sample: 10 Years)

	Dependent variable:			
	Small-Bank Density (per km^2)			
	SLX/RE	SLX Two-way FEs	SLX+SEM Two-way FEs	SAR+SLX+SEM Two-way FEs
	(1)	(2)	(3)	(4)
Big-Bank Density (per km^2)	0.596^{***}	-0.525^{***}	-0.526^{***}	-0.530^{***}
	(0.012)	(0.019)	(0.018)	(0.030)
SL Big-Bank Density	0.344^{***}	-0.414^{***}	-0.433^{***}	-0.463^{**}
	(0.021)	(0.038)	(0.035)	(0.181)
Geographic Concentration (km)	-0.001	-0.019	-0.019^{*}	-0.019^{*}
	(0.004)	(0.012)	(0.011)	(0.011)
SL Geographic Concentration	0.0002	-0.003	-0.003	-0.004
	(0.004)	(0.021)	(0.019)	(0.020)
ρ				-0.045
ρ P-value				0.8573
λ			-0.1016	-0.0667
λ P-value			0.0000	0.7906
Moran I	0.1238	0.0398	-0.0419	-0.0279
Observations	14,927	14,927	14,927	14,927

Note: *p<0.1; **p<0.05; ***p<0.01

Table 8: Spatial Panel Results (Urban Sample: 10 Years)

	Dependent variable:			
	Big-Bank Density (per km^2)			
	SLX/RE	SLX Two-way FEs	SLX+SEM Two-way FEs	SAR+SLX+SEM Two-way FEs
	(1)	(2)	(3)	(4)
Small-Bank Density (per km^2)	0.068^{***}	-0.102^{***}	-0.102^{***}	-0.103^{***}
	(0.004)	(0.004)	(0.004)	(0.005)
SL Small-Bank Density	0.238^{***}	-0.062^{***}	-0.088^{***}	-0.097^{***}
	(0.010)	(0.009)	(0.008)	(0.029)
Geographic Concentration (km)	-0.011^{***}	-0.018^{***}	-0.018^{***}	-0.018^{***}
	(0.003)	(0.005)	(0.005)	(0.005)
SL Geographic Concentration	-0.011^{***}	-0.003	-0.002	-0.004
	(0.004)	(0.009)	(0.009)	(0.009)
ρ				-0.081
ρ P-value				0.6965
λ			-0.199	-0.1282
λ P-value			0.0000	0.5388
Moran I	0.4582	0.2613	-0.0933	-0.0625
Observations	14,927	14,927	14,927	14,927

Note: *p<0.1; **p<0.05; ***p<0.01

Table 9: Spatial Panel Results: Individul Effects (Urban Sample: 10 Years)

	Dependent variable:			
	Bank Density (per km^2)			
	SLX/RE	SLX/FE	SLX+SEM/FE	SAR+SLX+SEM/FE
	(1)	(2)	(3)	(4)
Industrial Concentration (Big 5/All)	-0.070^{*}	-0.071^{*}	-0.071^{*}	-0.072
	(0.038)	(0.038)	(0.036)	(0.050)
SL Industrial Concentration	-0.076	-0.073	-0.073	-0.079
	(0.066)	(0.067)	(0.062)	(0.184)
Geographic Concentration (Km)	-0.030^{***}	-0.032^{**}	-0.032^{***}	-0.032^{***}
	(0.010)	(0.013)	(0.012)	(0.012)
SL Geographic Concentration	-0.022	-0.005	-0.004	-0.007
	(0.014)	(0.022)	(0.020)	(0.066)
ρ				-0.0715
ρ P-value				0.9698
λ			-0.1264	-0.0716
λ P-value			0	0.9697
Moran I	0.1015	0.0348	-0.0515	-0.0299
Observations	14,927	14,927	14,927	14,927

Note: *p<0.1; **p<0.05; ***p<0.01

D. Regression Results (CMA level)

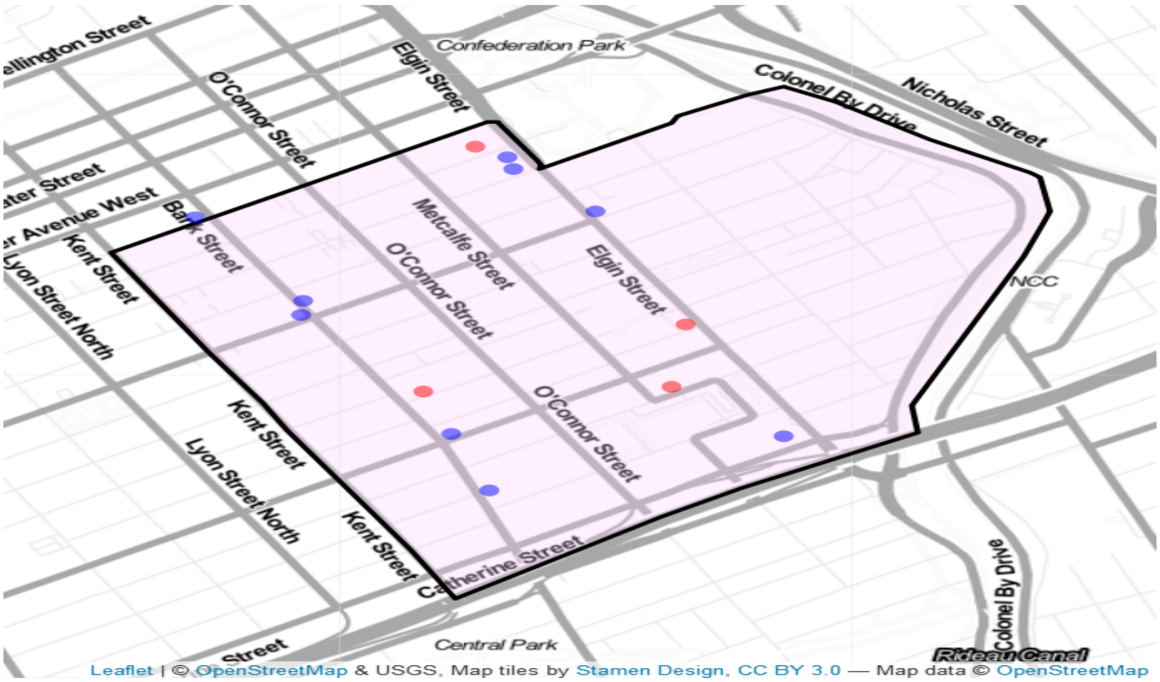
Table 10: Two-way Fixed-effects Panel Results (CMA)

	Dependent variables:		
	Bank Density	Big-Bank Density	Small-Bank Density
	(1)	(2)	(3)
Income (\$ 1000's)	0.0000	0.0001	0.0001
	(0.0002)	(0.0001)	(0.0002)
Population (1000's per km^2)	0.0876***	0.1181***	−0.0436
	(0.0216)	(0.0148)	(0.0317)
NAICS Finance (% Employed)	−0.0034**	−0.0004	−0.0033**
	(0.0017)	(0.0010)	(0.0014)
Age (Median)	0.0061	−0.0002	0.0011
	(0.0011)	(0.0006)	(0.0001)
Industrial Concentration (Big 5/All)	−0.0611***		
	(0.0217)		
Small-Bank Density (per km^2)		0.0741	
		(0.0857)	
Big-Bank Density (per km^2)			0.1689
			(0.1953)
Geographic Concentration (km)	−0.0009	0.0000	−0.0009
	(0.0007)	(0.0003)	(0.006)
Observations	102	102	102

Note: *p<0.1; **p<0.05; ***p<0.01

E. Figures

Figure 1: FSA K2P postal area with physical bank branches



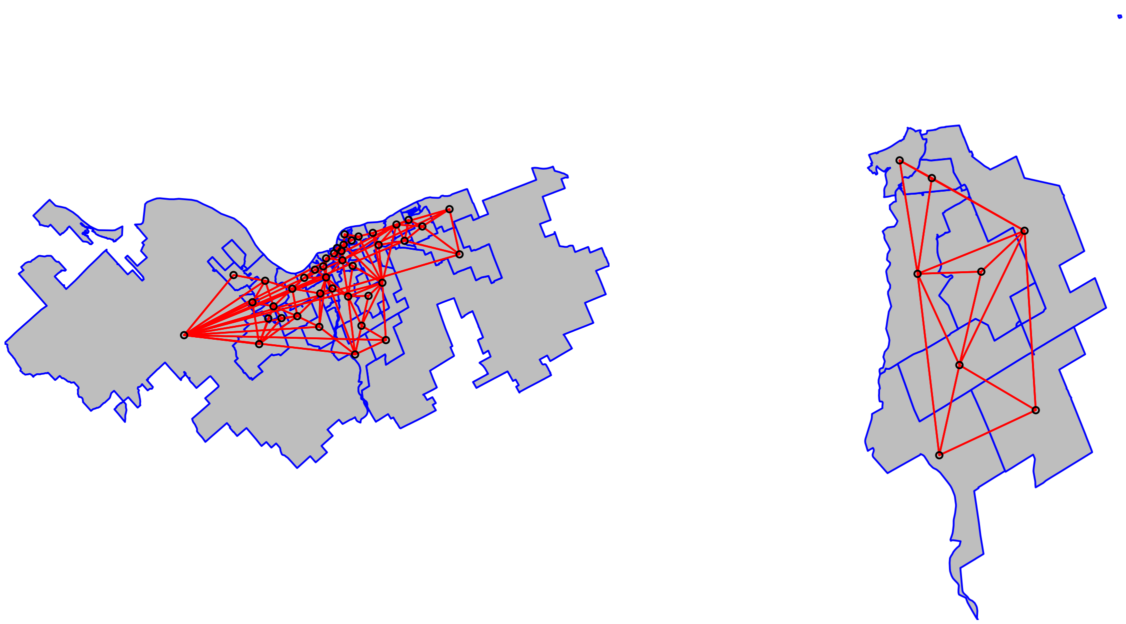
Notes: K2P incurred a reduction in branches from 2009 to 2013 to 2017 (13, 11, 9). Blue dots represent branches in operation in 2017 and red dots represent branches that closed since 2009.

Figure 2: Geographic Concentration (GC): Random vs. Clustered

Random Locations: GC = 1.54 KM

Clustered Locations: GC = 3.01 KM

Figure 3: Graphical Representation of Adjacency Matrix



FSA Group: Ottawa

FSA Group: South Ottawa

F. FIF Cleaning

F.1. Processing FIF Files

The first step is to process the FIF database text files. Each entry has a fixed string length of 400 characters. Each entry has the following variables:

- **Year**: Based on the year in each file name.
- **FileName**: The name of the file the data was extracted from.
- **FIF**: A string that precedes each row (characters 1 through 9).
- **Code_1**: A string on each row (characters 10 through 33). This variable can be broken down even further: 10-11 is the institution type, 12-21 is the identification number, 22 is the status indicator, 23-32 is the cross-reference number, and 33 is the language preference code.
- **Name_E**: Institution name in English (characters 34 to 69).
- **Name_F**: Institution name in French (characters 70 to 105).
- **Branch_Name**: Branch Domicile (characters 106 to 141).
- **Add_1**: Civic address (characters 142 to 177).
- **Add_2**: Postal address (characters 178 to 213).
- **City**: City the branch is located in (characters 214 to 249)
- **Province**: Province the branch is located in (characters 250 to 253)
- **Postal**: Postal code the branch is located in (characters 254 to 283)
- **Code_2**: Institution identification number for routing indirect clearer credits (characters 284 to 305).

- **Code_3**: Indirect clearer MICR encoding data (characters 306 to 327).
- **Code_4**: A string for each row (characters 328 to 344). 328 is the serviceability code, 329 to 334 is the exchange point, and 335 to 344 is the foreign bank transit number
- **Code_5**: A string for each row (characters 345 to 400). 345 to 350 is the effective date, and 351 to 400 is a Filler.

For the first geocoding run, since there are inconsistencies among the address fields, the string is based on the following rules:

1. If Add_1 = Add_2, then set the search string to Add_1.
2. If Add_1 is blank, then set the search string to Add_2.
3. If Add_2 is blank, then set the search string to Add_1.
4. Otherwise, create a string with both addresses separated by a comma.
5. If the "searchString" is blank, we set it equal to "SKIP".

F.2. Geocoding the FIF Data

Based on the search string in Stage 1, we will run the unique branch locations through Google maps locations API.

F.3. Visual Inspection

The third step is to visually inspect the geocoded entries. We go through each line and create a variable called "manualCheck." If the geocoded address returned an error or does not match what was listed in the original address fields, we set this variable to 1. Otherwise we leave it as 0.

F.4. Manual Geocode

We then use Google maps to manually geocode any entry where "manualCheck" is equal to 1. The manually geocoding rule is as follows:

- If you find a match in Google, copy and paste the formatted address into the "manualSearch" field.
- If no match is found and the area is urban, then use the postal code in the "manualSearch" field.
- If no match is found and the area is rural, use the city name in the "manualSearch" field.

F.5. Stage 2 Geocode

We run a second-pass geocoding with the Google-locations API based on the updated manual address entries. This will yield a complete list of branches. Once this is done, we remove any duplicate entries.