GIS in Banking: 
Evaluation of Canadian Bank Mergers

Eliot Hugh MacDonald 
Department of Geography 
University of Toronto 
Toronto, ON M5S 3G3

Introduction

Banks in Canada use geographic information systems (GIS) technology for a wide range of applications. These include, but are not limited to, selecting new branch sites, identifying risks associated with closing a branch, executing targeted marketing campaigns, navigating customers towards their locations, gaining better insights into markets served and expected financial performance given those markets, and adding clarity and impact to reporting through maps. But one such area in which GIS is applied, namely the evaluation of bank mergers, is of particular interest not only to regional scientists but the public in general. In this article, the use of GIS and related methodologies by banks in evaluating mergers and preparing for government review of mergers is profiled.

In 1998, Canadian business news was dominated by merger proposals from some of the country’s largest banking institutions. The Royal Bank of Canada, Canada’s largest bank, and the Bank of Montreal, Canada’s first bank, were first to announce these intentions. This was followed soon after by a merger proposal between the Toronto-Dominion Bank and the Canadian Imperial Bank of Commerce. The Competition Bureau, an agency of the Federal Government, was faced with reviewing these proposals within the context of its general mandate - ensuring that Canada maintains a competitive marketplace which allows the public to enjoy the benefits of competition, including competitive prices, product choice and quality service (Department of Industry 1991, 1997). The outcome of the review was to prevent the mergers in the 1998 proposals.

In making details of the merger review process public, the government re-
vealed that the process lends itself very well to incorporation of GIS methodologies. To put it very simply, the process followed by the government involves trying to identify where a merger will result in market share levels for the merging parties in excess of specified levels. Thus, a merger evaluation of this type requires an understanding of what defines a relevant market, both in terms of what products define a market and the geographical component. It also demands some type of automated processes for handling large amounts of geographic data for all regions of an entire country. The government has established a process which involves defining markets geographically and collecting industry data to gauge market share levels for these markets. The process used is one which is extremely difficult to reproduce by those banks wishing to evaluate mergers by performing their own “self-review” of a merger in relation to competition law compliance, and has been under debate in terms of whether or not appropriate geographic markets result from such a process. This paper focuses on the ways in which GIS-based methodologies can be applied to this merger evaluation process, particularly to derive market boundaries around which the government is expected to structure its merger review process and estimate market share for those regions.

In trying to gauge regional competition issues involving mergers, defining what geographic markets are appropriate has been an issue challenging both government and the corporations. In general, it is agreed that market boundaries should reflect both the locations of service outlets themselves, as well as where the masses of consumers are located in order to predict market share with as much accuracy as possible. For example, defining markets such that boundaries surround only the immediate regions of certain branches can lead to greatly overstated market share estimates (understating the impact of competitors just outside these boundaries). Even if accuracy in market share estimation is achieved, the choice of overall market sizes can present further problems. If the defined markets are too large, regions may exist within these large markets where merging parties dominate the area and can make pricing changes for customers who are still too far away from alternative banking services that they would consider reasonably accessible. Adding to the complexity is the fact that the appropriate geographic market definitions will change as different product markets are considered, such as commercial loans versus consumer deposits.

*Market share* represents the key statistic on which the government relies to gauge the possibility that merging corporations may enjoy too much market power within a particular geographic region. For the banks themselves, the data on competitive business volumes for specific, relatively small, geographic regions needed to measure market share is simply not available. However, using GIS, a complex “spatial interaction model” for predicting market share based on what types of markets (in other words, the settlement pattern of customers) are located near branches of different organisations can be developed (see below). The model can be applied to virtually any defined geographic region, e.g., “non-standard” regions defined by government for a merger review, such that banks need not rely on merely assigned city names in order to deal with the geographical measurement. A methodology for first defining the markets themselves, such that they match as
well as possible the geographic regions expected to be formed by the government for a given merger review, is proposed as well. For simplicity, one specific product line, personal deposits, is used for demonstration purposes.

The paper is presented in five sections. First, a brief background on each of five main concepts involved in this merger review process is presented, viz. how the government agency approaches a merger review, the concept of “trade areas” and defining appropriate geographic markets, measuring market share and an overview of the type of model used. Second, methodologies for defining geographic markets as dictated by government guidelines are reviewed. Third, the key model used to estimate market share (in this case, share of the “personal deposits” market for banks) for these geographic markets is outlined. Fourth, the methodologies are executed using selected case studies in order to test them. Two Canadian banking institutions, Toronto-Dominion Bank and Canada Trust, are used as a target for testing the market definition methodologies. These organisations were chosen because a merger between them took place shortly before this paper was written, and a portion of the results of the merger review process have been published by the government. Finally, a brief set of conclusions are drawn. For some sections of the paper, the added value of having actual data for a selected institution can be offered, and processing of this data is used to support concepts. To protect the privacy of this institution, it will be referred to merely as the F.I.R.M. (the Financial Institution Reviewing a Merger).

Related Research and Concepts

The Competition Bureau and Merger Enforcement Guidelines

The letters, studies, and guidelines developed by or submitted to the Canadian government’s Competition Bureau as a result of the major bank merger proposals in 1998 provide a major source of information on merger procedures and enforcement (e.g., McFetridge 1998; Task Force 1998; C.D. Howe 1998; Neave and Milne 1998; Horstman et al 1996). The Competition Bureau has the mandate of ensuring that Canada maintains a competitive marketplace and that the Canadian public continues to enjoy the benefits of competitive prices, product choice and quality service (Neave 1996; Cardwell 1997). A merger can result in increased market power within certain markets, giving for instance an ability to increase prices when the reduced number of competitors in a market leaves fewer alternatives. The Bureau must perform related market analysis in order to evaluate the presence of such dangers.

A major component of The Bureau’s review is the definition of relevant markets, which has both a product and geographic dimension. The intent of the Bureau in defining the appropriate geographic market is to include all areas in which there are suppliers to which customers would likely turn in response to an attempt by the merging firms to exercise market power. Thus, the size of the
market will vary with different product dimensions. For example, customers looking for a major loan, or commercial clients requiring special services, may not be as sensitive to distance as individual consumers for whom accessibility considerations are more relevant (CCF 1998). As a consequence, “personal deposits” is the product group chosen for analysis in the paper.

Although the Bureau’s specific methodology used to define geographic markets remains confidential, and perhaps even variable or ad hoc depending on the circumstances, a sufficient amount of information on the process has been made available to reach some useful conclusions. For urban markets, it is clear that the 25 markets defined for the latest Census by Statistics Canada as “Census Metropolitan Areas (CMAs)”, and the 112 markets defined as “Census Agglomerations (CAs)” are used (Statistics Canada 1997). These regions are based on urban cores of 100,000 or more population, and 10,000 or more population respectively (Statistics Canada 1993). Their geographic territories are well defined and readily available in electronic format for incorporation into GIS software.

For rural markets, the process is less clear, but the following conclusions can be drawn from published data:

- Initially, pairs of branches of the merging institutions that are within 20 km of one another, and in which both branches are in “rural” areas (outside of CMA or CA territories), are identified.
- Markets are made up of municipalities within the regions surrounding the pairs of branches defined above. The “municipalities” are represented by Statistics Canada’s “Census Subdivisions (CSDs)”, for which specific definitions and electronic representations are also available.
- Markets do not overlap. That is, each CSD is assigned to one and only one market.
- Only “rural” CSDs (CSDs not part of a CMA or CA) are assigned to rural markets.
- A maximum of one rural market per pair of branches defined above is identified, and only for regions surrounding a “line” joining these branches (believed to be based on 20 km around this “line”). This leaves many rural “unassigned” regions throughout the country, for which competitive analysis is deemed unnecessary.

This presents a tremendous stage for showcasing GIS. As will be demonstrated in the next major section, GIS can help the government and merging parties avoid months of tedious effort trying to identify such markets manually. Electronic street address and postal data, combined with GIS software tools, permit us to first assign geographic locations to lists of bank branches involved which the GIS software can further manage. Following this, the task of determining pairs of branches within close proximity and selecting municipalities which surround these pairs can also be automated via GIS.

The Bureau also relies heavily on available data published by organisations such as the Canadian Bankers Association in order to calculate market shares, and
uses these data as an initial screening test. Regarding the last point above, the Bureau states that the markets which are “relevant” are those defined through use of the “hypothetical monopolist” test. Under this test, a relevant market is the smallest geographic area such that a sole supplier of these banking products could profitably maintain a non-transitory price increase versus what could be charged without the merger taking place. In general, if the post-merger market share of the merging parties would be less than 35%, it is assumed that there is NOT a threat of new market power being generated. Otherwise, further in depth review of the market is required.

**Trade Area Definitions**

A trade area can be defined as the geographic region from which a particular sales outlet draws most of its customers (Lea 1998a). There are three important issues. The first relates to the task of defining the appropriate “size” of a trade area for a given store outlet -- how far a customer will go to use a particular store outlet versus another. This varies widely depending on both the product, or service involved, and the attributes of the store outlets themselves. Attempting to define store trade areas using simplified methods such as drawing circles around outlets or splitting up a region into trade areas based merely on proximity is shown to produce significantly inaccurate trade areas in many cases. The distances traveled by customers for a particular banking product needs to be understood, even when branch-level markets are grouped into larger markets.

A second issue concerns the use of polygon based trade area definitions for the purpose of aggregating statistics for markets served by store outlets. For these types of trade area definitions, data are either completely “in” or completely “out”, depending on whether or not its locations are inside or outside of the boundary defined. However, the typical pattern of consumer purchases at retail outlets is one that varies with distance and the attractiveness of the outlets involved. Within a particular neighbourhood, the majority of households may be customers of the nearest store outlet, but a significant percentage may be customers of the next nearest. Thus, trying to estimate market share based on splitting a region into trade area polygons for the individual store outlets can lead to significantly inaccurate results. Therefore, the concept of the “Huff probabilistic trade area”, where a trade area for a given outlet is represented by a set of probabilities for each neighbour- hood of households using the outlet as its prime source, is adopted as a more suitable alternative (Lea 1998b).

A third issue concerns a debate as to whether customers come to a store as a result of its creation, or are served by a store which is built because the customers were already there. There is a two-way relationship between retail outlet growth and population growth. Thus, the “system” of urban growth simultaneously generates markets due to the existence of stores and stores to serve existing markets (Jones and Simmons 1990). This concept underlies some of the research in this paper because, when branch level market share estimation models are created,
branch size is effectively eliminated from consideration as it is assumed to be determined exogenously. The issue of where customers base their branch location choice, such as its proximity to their home or work locations, also plays a key role in determining the appropriate market boundary.

Market Share: Issues and Alternatives

There are a number of varying components of market share, and there are different ways it can be defined (Lea 1999; Simmons et al. 1998). The metrics used can include total customers served, total sales revenue, total units sold, “balances” held by customers and number of stores or companies operating in the industry. The market (geography) defined can vary from immediate (prime) “trade areas” within close proximity to store outlets, to entire major regions within which the institutions have presence. The product dimension can also vary significantly in terms of what set to include in the definitions. This can vary further when an additional “customer” dimension is tied into it, where even one product is further split based on the fact that different customer segments (e.g., major corporate clients versus individual households) are being served by different sets of companies. Furthermore, market share can be calculated on the basis of where customers live (“demand based” approach), or based on where the stores themselves are located (“supply based” approach).

In this analysis, much of this complexity has been removed by the limitation imposed by the data available for modeling. The market data made widely available to banking institutions involves preset definitions for most of these components. Only certain product groupings, sets of competitors, geographic regions and methods of calculation (in terms of branch versus customer based) have been used (provided). Spatial pattern (and thus, GIS) is also central to the estimation of market share, particularly when it comes to estimating the total market (data are usually readily available to give accurate measures of one’s own sales). But the issue of spatial distribution of customers and store outlets, and the associated issues of trade area definitions clearly impact the complexity of the task. Some of the complexity is removed by the fact that one need only deal with relatively large regions for the purpose of merger review, and the issue of market definition is driven by how the government defines the markets. However, it will be necessary to consider submarket issues in cases where the goal is to estimate market share of a specific competitor, and thus “numerator” data are not even directly available for the market share estimates. This is certainly the case for merger planning, particularly when planning and evaluations are needed before the stage when information is being shared between institutions.
The Spatial Interaction Model

A “spatial interaction model (SIM)” is one that is generally used to predict (or explain) the pattern of “flows” (interactions) given the geographic accessibility of destinations in relation to origins, the attractiveness of the destinations and the level of demand at origins, or “customer locations” (Bailey and Gatrell 1995; Fotheringham and O’Kelly 1989). SIMs are particularly well suited for applications such as retail site selection, predicting migration and commuting patterns, locating emergency services such as health care facilities and transportation planning (Lea 1989; Lea and Menger 1990, 1991). A SIM has the following basic form:

$$p_{ij} = \frac{A^o_j / d^{\alpha}_{ij}}{\sum_j (A^o_j / d^{\alpha}_{ij})}$$

where

- $p_{ij} =$ probability of household $i$ patronising site $j$;
- $A^o_j =$ attractiveness of site $j$;
- $d_{ij} =$ distance (separation) from household $i$ to site $j$;
- $\alpha =$ calibration parameter for attractiveness; and,
- $\beta =$ calibration parameter for distance.

The calibration parameters reflect the sensitivity to changes in outlet attractiveness and distances for the specific type of behaviour (or retail activity) being modeled. In the particular case of modeling expected business volumes for different branches, the branches are the sites, and entire neighborhoods (e.g., enumeration areas) are the “origins” being used in place of individual households.

There are four different types, or “families”, of SIMs. As noted above, the models are based on the combination of site attractiveness, demand (or “propulsiveness”) of originating customer locations (“origins”), and spatial separation between sites and origins (Haynes and Fotheringham 1984). If the actual “flows” (values for interactions, however they are defined) are not directly available for the individual sites themselves nor for the individual origins themselves, then the model is termed an unconstrained model. The implication is that the total “flows” across all stores, or across all origins in aggregate (either “side” automatically implies the same total for the opposite “side”), are known. It is just the individual store-by-store (or origin-by-origin) level of information that is unknown.

When the “inflows” are known (and not the “outflows” from origins), the model is termed attraction constrained (or destination constrained). This implies that “indicators” of interaction levels for the origins must be obtained to serve as surrogates for the missing information. When only origin “outflows” (from each origin) are available, the model is termed production constrained (or origin constrained). Finally, when both “sides” are known (i.e., the volume coming from
each individual origin and feeding into each individual site are known), it is a
doubly constrained (or production-attraction constrained) model. In this case, it
is merely the interaction patterns themselves (e.g., which customers go to which
sites) that are being modeled. Note that the set of “origins” need not refer to
individuals themselves, but rather groups of individuals (e.g., “neighbourhoods”),
where central points (“centroids”) for the groups or the edges of boundaries
surrounding the groups are used for the purpose of calculating distances.

GIS Based Market Definition Methodologies
for Merger Review

The focus of this section is on trying to duplicate the pre-defined process set by the
Canadian government for defining geographic markets for merger review. How-
ever, it is worth first reviewing the basic aim of this process and evaluating the
process in place before presenting the methodology.

The Concept of “Natural Markets”

The appropriate market definitions are believed by many to be ones which fall
“naturally” from consumer patterns. “Natural markets” is an unofficial term used
frequently by both the government and the corporations during merger planning
and review. From the perspective of merger review participants, they can be
defined very loosely as isolated geographic regions within which a high degree of
economic integration exists (e.g. commuting to work takes place). One might also
visualise “natural markets” as the regions which result when one splits up the
country using boundary lines which are generally not crossed with high volumes
for the purposes of attending work, shopping or going to the bank. Clearly, trying
to derive boundaries that strictly meet these general conditions would be virtually
unachievable without making the regions very large. For countless urbanised
regions, the overlap in commuting patterns would require that a boundary be drawn
around an enormous population before there is no significant volume of commuting
taking place across the boundary. In theory, these markets should not change
as one reviews a particular proposed merger. However, the government currently
establishes a process which tends to isolate regions around the merging parties’
branch locations, citing that these are the “relevant” areas in question for the given
proposal.

Other factors surrounding market definition for financial services also complicate
the issue. In direct contrast to the urban commuting scenario outlined above,
there are many isolated villages surrounded by hundreds or thousands of square
kilometres of virtually uninhabited territory. In many of these, the village popula-
tion does not come close to warranting the installation of even an automated
banking machine. Another factor is the banking product in question, and how
markets change when different product lines are being reviewed. And finally, despite perhaps representing the key activity, commuting to work does not represent the only way in which economic integration exists between different regions. Certainly examples such as shopping at a regional mall, or relying on information, health and other services of another region, should in theory be given some consideration.

One general approach for defining “natural” rural markets would be one which closely follows the “functional region” concepts used by Statistics Canada in defining CMAs and CA s (Statistics Canada 1997). That is, use the most populated areas as “cores” (or “hubs”) and assign surrounding areas to these “cores” based on some measure of integration between those regions and the cores (a hub and spoke approach, in some sense). Such a process has already been developed by Statistics Canada for developing such regions in urban areas, but not for other rural areas. Figure 1 presents a specific process in the form of a flow chart which can be used to define such “natural” markets in rural regions and using an automated GIS program. Although difficult to interpret without careful review, this demonstrates both the power of today’s GIS tools in being able to automate such a process as well as the complexities involved in enabling the organisation to take advantage of this power. This process is simplified in that distances alone are used as a measure of integration between regions -- though this could be enhanced with greater data availability. Figure 2 shows the regions which result after a GIS program written in ESRI’s ArcView software (using the “Avenue” programming language) is executed. It performs the process described in Figure 1.

Figure 3 demonstrates how following such a process (a “natural market” approach unaffected by which particular organisations are being reviewed) can differ versus rural markets created by using the basic approach employed by the government for merger review. Note also the significant differences in market share figures which can result, with the government’s process almost always resulting in an inflated market share figure for the institutions proposing a merger. This alternative process is first presented in this section merely to demonstrate the theoretical issues involved and where GIS is applied in carrying out a market definition process. It represents a potential improvement, given that the government wishes to maintain both its current urban definitions (the CMAs and CAs), and its desire to keep the process simple by using non-overlapping polygons as markets. It is not necessarily proposed as ideal. As discussed in the previous section on trade areas, polygon-based market definitions, which split market components into those “completely in” or “completely out” of the market, do not generally represent the true marketplace very well.
FIGURE 1a A Simplified 'Hub & Spoke' Approach to Defining Rural 'Natural' Markets
FIGURE 1b A Simplified 'Hub & Spoke' Approach to Defining Rural 'Natural' Markets (continued)
1. The region shown is a major portion of Southwestern Ontario. The thick black lines represent the resulting rural 'natural' market boundaries. Thinner lines running through them represent the Census Subdivision (CSD) boundaries ('municipalities') making up the markets. Regions belonging to Census Metropolitan Areas (CMAs) and Census Agglomerations (CAs), thus considered 'urban' regions, have a 'dotted' shading.

2. The map below shows how the methodology employed by the Competition Bureau can theoretically result in significantly different market share statistics versus an alternative definition process. The stars are fictional branch locations of merging institutions, while
Methodologies for Duplicating the Canadian Government’s Market Definition Process

The above discussion should be given consideration when a bank is evaluating issues of market dominance for itself, but in practical terms it is advisable that the institution attempt to match the process currently being enforced -- a virtually impossible task without the aid of GIS. Given what is known about how the Competition Bureau defines markets, it is possible to use GIS to actually duplicate this same general process. Thus, through a combination of popular GIS and database software, an extremely tedious month-long process involving inspection of branch locations and municipality locations can be converted into one computer program executed within a couple of minutes. The process is used for rural markets alone (given that CMAs and CAs are used to defined urban markets), and is described below, this time without the flowchart representation.

STEP 1:
Identify set of “rural” branches from “opposite” institutions (pairs of branches) within 20 km of each other (where “rural” is defined as outside of CMA and CA regions).
The above graphic depicts Census Subdivisions (CSDs) for a region surrounding Collingwood, Ontario, Canada. Darker zones represent urban municipalities (within CA or CMA regions), and circular markers represent population-weighted “centroids” of the CSDs. Assume the square markers are branches, with white squares belonging to one institution and grey squares belonging to the other institution with which the first is considering merging. Only two pairs of branches would be selected for step 1, as indicated by the dotted lines joining 2 with 4 and 2 with 5. Branch 1 is more than 20 km from branch 4 and branch 4 represents the same institution as branch 5, so these pairs are not selected. Branches 3 and 6 are both within the “urban” zone representing the Collingwood CA, so they are not paired together or with any other branches.

**STEP 2:**
Extract distinct pairs of branches from set created in step 1 using an iterative process of selecting branches closest to each other in same municipality and eliminating all others also within 20 km of either of those branches.

In the graphic above, the pair made up of branches 2 and 5 are selected first, based on proximity, and the pair comprised of 2 with 4 is eliminated.

**STEP 3:**
Create 20 km buffer zones around the set of branch pairs from Step 2.

The distance of 20 km used for the buffer zone represents the best estimate of what the government uses in its own market definition process, from personal experience and the few comments published by the government.
STEP 4:
Create Census Subdivision (CSD) population-weighted “centroids”, based on Enumeration Area (EA) level populations.

The CSD centroids, shown in the above graphics, can be generated by weighting the geographic locations of the EAs making up each CSD according to population. The “centroids” of the EAs (geographic centre points of their boundaries) and their populations can be obtained directly from Statistics Canada.

STEP 5:
Determine, for each buffer zone, which rural CSDs fall into the zone.

In the above graphic, three rural CSDs would be identified for the one buffer zone. These would be stored in a separate table, along with the buffer zone to which they belong. When the process is performed nationally, this will likely result in a huge list of CSD codes matched to buffer zones, the latter of which can be coded via the largest CSD or assigned some kind of pseudo ID numbers. These coded zones now represent the initial set of government-defined “markets”, and each consist of a set of CSD regions (versus the oval-shaped buffer zones themselves). Overlaps between these initial markets still need to be resolved before they are finalised.

STEP 6:
Merge markets resulting from step 5 above which involve an overlap of at least half the population of one of the markets.

Since there may be more than two markets each overlapping at least one of the others in a regional cluster, this process does not simply involve looking at isolated pairs of overlapping markets. Thus, an iterative process is used to determine the procedure for merging markets. For each iteration, the pair of markets involving the highest percentage overlap of population are merged first. For “ties”, the pair of markets which involves the smallest population is chosen first. These markets are merged together, and treated as one market before the next iteration. The process continues until no markets have more than half of the population overlapped by another market. This, of course, can still leave markets which overlap slightly.

STEP 7:
Allocate each municipality (CSD) within a region shared by different markets to one and only one market, based on shortest distance to the nearest branch from branch pair.

In this step, minor overlaps are resolved by splitting up the overlap regions and allocating subregions among the identified markets. Each market has an original associated pair of branches from the earlier steps of this process. During the merging of markets, the “more overlapped” markets were merged into the other (the latter can be considered the “primary” or “receiving” original market...
region for the new expanded region). It is the pair of branches from that region which is retained for the purpose of step 7 here.

Market Share Estimation through GIS

The Spatial Interaction Model:
Institution by Institution Market Share Estimation

Recall, spatial interaction models like the one demonstrated in this section typically involve prediction of customer interactions based on distances and indicators of store or branch attractiveness. As banking institutions do not have extensive attribute data about competitors, indicators of branch attractiveness (the ability for the branch to draw customers versus other branches) would have to be derived from surrogate information such as the following examples:

- The number of staff, or floor space, of the nearest branch belonging to one’s own institution. This indicator offers potential value only in cases where branches of the same size tend to be located together in space, and there is almost always a branch for one’s own institution not far away from each competitor.
- Retail sales levels existing within close proximity to the branch. Business activity data can be used to estimate retail presence within the immediate vicinity of each branch.
- Hours of service. Using individual web sites of banking institutions, it is possible, though tedious, to extract total weekly branch hours of service for each and every branch.
- Institution “type” indicator (or “ratings”). It is possible to provide subjective ratings for branches according to the type of banking institution they represent. For example, a major bank branch can be given a higher “rating” versus, for example, a local credit union, knowing that the typical bank branch tends to draw significantly more personal deposits than the typical credit union.

The particular model presented here for demonstration purposes uses the same attractiveness score across all branches, although experimentation with other indicators of attractiveness such as those listed above may yield stronger results. The market share estimates are derived from a “production-constrained (Huff based) spatial interaction model” (Lea 1990; see above), in which the objective is to estimate the percent of the market in an area (in this case, an enumeration area or EA) banking at each of a number of financial (FI) institutions. Thus, \( p_{ij} \) = estimated percentage of market in enumeration area (EA) \( i \) banking at financial institution (FI) \( j \), or
where

\[ p_y = \frac{A_j^a / d_y^b}{\sum_j (A_j^a / d_y^b)} \]

\[ A_j = \text{Attractiveness of site } F_I_j; \]
\[ d_{ij} = \text{Distance (separation) from EA centroid to site } F_I_j; \]
\[ \alpha = \text{Calibration constant for attractiveness; and,} \]
\[ \beta = \text{Calibration constant for distance} \]

For the model, straight line distances were used, and calculated as the distances between population-weighted EA centroids and the branch locations. The level of demand (“outflow from origins”) was represented by total aggregate interest income of taxfilers in the neighbourhood, adjusted for differences in daytime population in relation to Census population (interest income reweighted, such that figures will still add up to the correct total across all markets, for non-overlapping markets). This was a choice based on some basic analysis of available industry consumer deposit balance data in relation to demographic data, which suggested that using this statistic gave a reasonable indicator of deposit volumes when dealing with regions for which only the demographic data were available. And again, further analysis of what measure to use to indicate market demand for the selected product is part of the process of developing such a model, and better indicators are certainly possible. The taxfiler data used in this model have been aggregated directly to the EA level, by Statistics Canada, from the original raw data at the postal code level (normally only available to the public at higher levels of geographic aggregation than the EA level). For each of the 101 regions for which actual market share data were available, the data (or “market”) being allocated is restricted to that immediate region.

For each model, the attractiveness parameter (\( \alpha \)) was set to 0.8, while the distance parameter (\( \beta \)) was set to 1.2. While these should be determined from rigorous calibration efforts, in this study they have been estimated -- for example, from tests performed on distance decay for F.I.R.M. customers. For this model, attractiveness scores are eliminated (replaced merely with 1’s in the equation). Thus, when allocating the demand within a particular enumeration area to different branches, a calculation of \( (1^{0.8} / \text{distance}^{1.2}) \) was made for each branch in relation to the EA.

In the following diagram, assume the shaded region is an enumeration area with a total demand of 4,000 “units” and a geographic centroid indicated by the circular marker. Assume also that the squares are branches, and that the numbers indicated on the line segments represent distances to the EA centroid in some kind of units.
For this scenario, the expected proportion of demand “interacting with” branch C, with $d_x$ representing the distance from the EA centroid to branch x, would be equal to

$$\frac{(1^{0.8} / d_C^{1.2})}{(1^{0.8} / d_A^{1.2}) + (1^{0.8} / d_B^{1.2}) + (1^{0.8} / d_C^{1.2}) + (1^{0.8} / d_D^{1.2})} = 8.7\%$$

Similarly, the proportions are 60.1 %, 26.2 % and 5 % for branches A, D and B respectively.

In order to arrive at a predicted market share figure for a particular institution in the market, the proportion of demand from each EA is multiplied by the actual demand and then allocated accordingly. Each branch will then have been allocated a certain volume of demand from various EAs. These branch-level demand figures are then totaled across all of the institution’s branches in the market for the numerator. The denominator is the total for the entire market of the statistic used to indicate demand level. Since indicators of demand are being used in place of true demand figures for each EA, the total demand figures across all branches in the market will represent some ratio of the total true demand in the market. Ordinarily, a spatial interaction model used to predict demand (or business volumes) will be formulated such that total predicted demand across all service locations in the market equates to total actual demand in the market. A multiplier is used in these cases when they are not equal. For the models in this paper, since market share (percentage) is being predicted, and the predicted totals will equate to actual totals (100 %), a multiplier is, in effect, being used.

GIS software is heavily leveraged to perform the millions of required distance calculations between all combinations of enumeration areas and bank branches, and to generate the geographic, population-weighted “centroids” for enumeration areas.
tion areas beforehand. For the purpose of this paper, this model was executed by allowing such a GIS programme to run for an overnight period to obtain model results. Overnight processing is a frequent occurrence for organisations taking full advantage of the powers of GIS.

Testing the Methodologies

Evaluating the Spatial Interaction Model

In order to evaluate the results of the model, it was necessary to use the individual branch level figures generated by the model (the demand “allocated” to each of the branches) to produce market share data for only one institution (the F.I.R.M.) for which true market share figures were available. A number of measures were then used for evaluation. In each case, the figure applies to the set of 101 regions used in the model. Again, these regions represent those for which the author has obtained actual personal deposit market share figures for one institution. They also represent the same typical sizes of markets as would be the case in a real world scenario, with most representing CMA or CA definitions. These selected figures involve comparison of the F.I.R.M. market share predicted by the model versus the actual F.I.R.M. market share, and do not necessarily represent the most ideal measures, but rather a sample of some that have been typically requested within the workplace environment of a major Canadian bank. They are as follows:

- The overall $R^2$. Calculated using the set of predicted market share figures from the spatial interaction model (across all markets) in relation to what is being used as the set of actual market share figures.
- The $R^2$ with the 5 regions involving the most error removed – that is, the $R^2$ with “outliers” removed.
- $R^2$ values, with records for Alberta and Quebec removed. Each of these provinces involve incorporation of “non-banks” in the denominator for the actual industry data used (and there is no way to parse them out). This is not the case for any other province.
- $R^2$ with “irrelevant”, or “non-problem”, markets removed. “Non-problem” markets are those in which no two banks have a combined market share greater than 35%. Since error in a prediction which is too low for a “non-problem” market does not generally cause incorrect assumptions about actions which need to take place in the market (in order to resolve a competition problem), the models were evaluated with these records removed.
- Percentage of regions in the “correct category” (“problem” versus “non-problem” categories). Since a goal is to identify a set of problem markets and determine what might need to be done about it, this is another good indicator.
TABLE 1 Model Results: Selected Figures using Given “Predictor” of Personal Deposits Share

<table>
<thead>
<tr>
<th></th>
<th>CBA-Reported Branch Share</th>
<th>Market Share Predicted by Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>21.3%</td>
<td>49.6%</td>
</tr>
<tr>
<td>$R^2$ excluding ‘worst’ 5</td>
<td>28.7%</td>
<td>49.5%</td>
</tr>
<tr>
<td>$R^2$ excl. Alta. &amp; Que.</td>
<td>53.1%</td>
<td>46.0%</td>
</tr>
<tr>
<td>$R^2$ excl. ‘irrelevant’</td>
<td>82.0%</td>
<td>67.6%</td>
</tr>
<tr>
<td>% in Correct Category</td>
<td>65.3%</td>
<td>81.2%</td>
</tr>
<tr>
<td>National % above Actual</td>
<td>19.7%</td>
<td>2.1%</td>
</tr>
<tr>
<td>% within 3 of Actual %</td>
<td>41%</td>
<td>63%</td>
</tr>
</tbody>
</table>

- Overall error in the national average market share. This indicates, on average, whether predictions were too high or low, and by how much. It is not an indicator of quality, of course, since large positive and negative errors become cancelled out.
- Percentage of markets for which predicted share was within 3 % of true share. This measure is more “forgiving” for a model which is consistently incorrect by at least one or two full percent, but significantly incorrect for very few markets.

The results are summarised in Table 1. $R^2$ values of about 70 % are desirable in models used for predictive purposes. Although these results do not reflect such a level in most cases, it should be kept in mind that the types of model discussed here (spatial interaction models), particularly those involving the allocation of small geographic units for the entire country of Canada, typically involve long development (calibrating) periods by many staff members using considerable processing power. The model here is presented primarily to demonstrate the process. The results should still, however, help in evaluating the development of such a model based on GIS for this particular purpose. For reference purposes, the same figures are calculated using reported share of branch locations (number of branches in the market for the selected institution as the percentage of number of all institutions' branches) as a predictor of personal deposit share.

Evaluating the Market Definition Methodology:
A Merger Between TD Bank and Canada Trust

Recall the aim of the GIS-based procedure for defining markets outlined above. It was to try to match the geographic markets which would be defined by the government (the Competition Bureau) when it completes its lengthy process of defining markets through predetermined guidelines for a given merger review. Thus, to test this methodology, actual results of a merger review for a merger proposal made between the Toronto-Dominion Bank and Canada Trust (von
Finckenstein (2000) are used. A listing of the markets defined by the Bureau for this merger proposal was fortunately made public to facilitate this test. As indicated previously, CMAs and CAs were used for urban markets, and rural markets were defined as in Figure 2. The results conformed closely to the Bureau’s markets. From the Competition Bureau’s letter made public, “in rural areas, the Bureau examined the competitive environment of all Toronto-Dominion and Canada Trust branches within 20 kilometres of each other. Using this criteria, Toronto-Dominion and Canada Trust operations overlap in 11 rural markets” (von Finckenstein 2000). The list of rural markets, and whether or not each was also identified successfully by the proposed GIS programme, is as follows:

<table>
<thead>
<tr>
<th>“Bureau Market” Name</th>
<th>Province</th>
<th>Created via GIS Methodology?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dunnville</td>
<td>Ont.</td>
<td>Yes</td>
</tr>
<tr>
<td>Fergus</td>
<td>Ont.</td>
<td>Yes</td>
</tr>
<tr>
<td>Ingersoll</td>
<td>Ont.</td>
<td>Yes</td>
</tr>
<tr>
<td>Kirkland Lake</td>
<td>Ont.</td>
<td>Yes</td>
</tr>
<tr>
<td>Norwich</td>
<td>Ont.</td>
<td>Yes</td>
</tr>
<tr>
<td>Wilmot</td>
<td>Ont.</td>
<td>Yes</td>
</tr>
<tr>
<td>Amherstburg</td>
<td>Ont.</td>
<td>Yes</td>
</tr>
<tr>
<td>Bridgewater</td>
<td>N.S.</td>
<td>Yes</td>
</tr>
<tr>
<td>Hanover</td>
<td>Ont.</td>
<td>Yes</td>
</tr>
<tr>
<td>Kerns</td>
<td>Ont.</td>
<td>No</td>
</tr>
<tr>
<td>Parksville</td>
<td>B.C.</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Unfortunately, there were no other data available, for this merger proposal, which allowed the testing of the actual definitions of these markets. But based on previous exposure by the author to actual market definitions generated by the Bureau, the definitions themselves appear relatively close. It should be kept in mind that there is also the possibility that discrepancies result simply because either the branch location data collected by the government or that collected by the author is in error.

Conclusions

The results show that the spatial interaction models can produce better estimates of market share than those resulting from use of reported branch locations, but under lengthier development efforts than those involved for this demonstration. For efficiency, the simple spatial interaction model demonstrated here almost completely omitted the use of attractiveness indicators and calibration procedures, two very important components of any of these types of models. Yet, the model
still generated fairly useful results. Thus, even with all the simplifications and assumptions, the approach seems quite promising. The paper has demonstrated, in fairly straightforward fashion, how such spatial modeling techniques can be developed. By using these concepts as a starting point, and applying time and resources toward making improvements, a powerful model for predicting market share would likely result.

The most significant potential improvement to the spatial models would almost certainly rest with the competitor branch attribute information, where there is not enough directly related items of data available for other competitors’ branches. In the case of proposed mergers, which involve huge expenses (legal, and so forth), it would seem to be worth the time and effort for a large banking institution to use its resources stationed throughout the country to survey local competitive branches. Extremely valuable data can be collected in a relatively short time, perhaps even through a quick, two-minute assessment at each location. This can involve simply recording scores based on only the most obvious key attributes, such as overall branch size, overall appearance or environment, approximate number of parking spaces and accessibility. Without additional branch information, the model simply estimates the business volumes which must be generated based on where branches are located in relation to the market. This takes many very key factors in determining market share out of the equation, such as level of marketing activity, quality of staff and prices (rates) offered.

In the case of personal deposits, even using share of branches as a proxy for market share is not altogether inaccurate, and can be used in place of generating such a spatial interaction model as demonstrated above if time is simply unavailable. This would imply that the major Canadian banking institutions tend to offer similarly sized service outlets, on average, throughout different regions of the country (at least in terms of retail deposit-related services), which likely is not altogether wrong. However, there is enough error associated with using these simple branch count statistics as market share estimates (and enough importance associated with the task) to warrant development of more robust methodologies in most cases. This is especially true when one starts to analyse other product markets, where number of branches becomes a less useful indicator of true market share.

Most importantly, the research demonstrates how GIS can be play an integral part in helping a financial institution plan more adequately for competition laws related to merger attempts. It should be kept in mind that the aim for this paper was not to present a working model which can successfully be transferred to the real world, but rather to demonstrate the processes and concepts involved in making the corresponding initiative a success via GIS. As made evident, how-to lessons in performing some of these GIS-oriented steps through various software packages was avoided in order to permit a focus on outlining the application itself in simple terms. Fairly detailed steps have been developed which serve as guidelines for the types of work and costs involved for such complex processes -- a necessary stage before getting to the point where the actual steps are carried out within the software environment. It is worth noting also that valuable knowledge about markets
and customer behaviour can be gained by performing such GIS based analyses, even if successful predictive models are not developed. The speed at which otherwise lengthy, tedious tasks can be performed using these types of methodologies is a tremendous benefit -- the market definition process is an excellent example of this. Furthermore, these methodologies need not be limited to the narrow scope of merger planning. Many of the concepts and methodologies contribute almost equally to the understanding of how GIS can assist other important applications for retailers, such as selecting new store sites or evaluating the impact of changes to the branch system of financial institutions.

References


