Which Birds of a Feather Flock Together?

An Examination of Agglomeration/Avoidance Patterns of Homogeneous Retail Outlets

Robert E. Krider
Marketing Area, Faculty of Business Administration, Simon Fraser University

Daniel S. Putler
Marketing Division, Sauder School of Business, University of British Columbia

February 2, 2010

We wish to acknowledge the support of the Social Sciences and Humanities Research Council of Canada under grant 410-2005-0929. We thank the participants of the 2008 UBC – University of Washington Marketing Camp and session participants at the 2007 Marketing Science Conference in Singapore for their helpful comments and suggestions. We would also like to thank José Mora for data support.
Which Birds of a Feather Flock Together?

An Examination of Agglomeration/Avoidance Patterns of Homogeneous Retail Outlets

Abstract

We often observe some types of retailers, such as new auto dealers and bridal boutiques, locating close to each other in spite of the possibility of increased price competition from nearby competitors. Theoretical research has proposed different explanations for this homogeneous agglomeration, but despite the substantial theoretical literature, there are few empirical studies that examine spatial structure of similar retailers. The studies that have been done tend to examine only one, or a few types of retailers, and the findings are not always consistent.

In this study we examine location patterns for 54 different retail types in two Canadian metropolitan areas. A major difficulty in previous studies investigating whether retailers of a particular type agglomerate or avoid one another is the failure to control for forces that induce agglomeration for all retailers equally (e.g., population density, local area income, zoning, etc.). We use a measure that controls for this problem based on assessing whether specific store types are attracted to or avoid each other relative to the background of retailers of all types. The measure also captures the extent of agglomeration or avoidance as a function of store separation, a richer indication of spatial structure than previous global scalar measures. The measure for each store type can be displayed as an intuitive and easily interpreted plot.

Our results indicate a surprising variety of spatial location patterns, including patterns consistent with both maximal and with minimal spatial differentiation, the two types of patterns that tend to emerge in the theoretical literature. One novel pattern, which we call the “auto mall” pattern, consists of local clusters of retailers, but the clusters themselves avoid each other. We argue that this pattern suggests that there is a link between manufacturers’ distribution strategies and the resulting spatial structure of retail outlets. In addition, we examine two categories that show strong agglomeration in one city, but not the other, and find that increasing concentration in retail outlet ownership may mitigate homogeneous agglomeration in retail categories where it would otherwise exist.

Keywords: Retailing, spatial competition, spatial statistics, geographic information systems
Consider the situation faced by the owners of a small, but successful two store pet and pet supply retail chain that is determining where to locate their third outlet. Their first instinct was to avoid any competing pets and pet supply stores, but they also knew of instances where similar retailers located very close to each other despite increasing competitive intensity. They all knew of, for example, retail areas with high concentrations of automobile dealerships, and shopping malls with high concentrations of similar fashion stores. They looked to the academic literature for further insight, and found a number of interesting theoretical approaches to the problem. For example, apparently slightly differentiated product assortments were a factor in similar stores clustering, as were the increased trade areas due to customers’ ability to easily comparison shop across similar stores. These insights were, however, of a highly stylized nature, and it was difficult to translate them to their situation. For example, pets and pet supply stores generally had fairly similar assortments, but at the same time the owners attributed the success of their existing two stores to carrying several exclusive lines of pet products, to fine tuning their assortment to the local market, and to differentiating their stores on the basis of a warm and comfortable store atmosphere.

In contrast to the theoretical literature, they found that the empirical research studying which types of retailers cluster and which avoid was thin, tended to examine store types that were expected to be extreme examples of either clustering or avoidance (such as antique shops and gasoline stations), and was often inconclusive or even contradictory. Practical difficulties with collecting and analyzing location data, with the types of measures used, and with controlling for the very strong influence of spatial demand density were apparent. It occurred to the owners that it would be very helpful if they could find a comprehensive assessment of what sorts of spatial structures were actually observed for a large number of different store types. If pets and pet supply stores were included, they could immediately see what the existing patterns were, and even if not included, they could use the results, along with their new theoretical knowledge to determine the likely spatial patterns in their category, but they had been unable to find such a study.
The purpose of this paper is to provide a rich, consistent, and easily interpreted measure of homogeneous retail spatial structure, and to use it to create a substantive catalog of the spatial structure of a large number of retail store types. In doing this, we study two different western Canadian metropolitan areas to provide an indication of which patterns are likely to be robust across cities: Vancouver, with over 18,000 retail outlets, and Calgary, with over 8,000 retailers. The value of such “ground truth” is in providing managers with insights for location decisions, in having a base for evaluating the comprehensiveness of current theory, and in providing an empirical basis for future theory development.

We find a variety of spatial location patterns for different retail types, with many showing some degree of homogeneous agglomeration, and a few showing avoidance. In some cases these results are consistent with economic theory: gasoline stations and supermarkets avoid each other (due to the comparative lack of differentiated product assortments and relative lack of uncertainty consumers face when purchasing these products), while furniture stores and antique shops cluster (both of which are categories where there is considerable consumer uncertainty that can be mitigated through comparison shopping). In other cases the results are less predictable. Our methodology also allows us to identify more complex patterns than simple agglomeration or avoidance, some of which are quite surprising and hitherto unrecognized. For example, automobile dealers cluster quite strongly within a local area, but, at a larger scale, the clusters of automobile dealerships avoid each other. We argue that this is likely the result of manufacturer distribution arrangements. Many spatial structures are consistent across the two cities, while the ones that differ again raise interesting questions as to why the differences exist. We demonstrate that one possible reason is differences in outlet ownership concentration between the two cities.

There are at least three difficulties with conducting empirical research on homogeneous agglomeration. First, the inference of attraction or avoidance among similar retailers using simple measures of spatial clustering is invariably contaminated by very uneven spatial demand density. One obvious source of such clumpiness is population density. This prob-
lem has resulted in ambiguous, contradictory, and counter intuitive results, such as gasoline stations and convenience stores sometimes showing clustering and sometimes avoidance behavior. The few existing empirical studies that have attempted to control for spatial demand density have had limited success. In our work, we make use of a previously unused method to correct for these effects, which allows better inference of true attraction or avoidance behavior.

Second, collecting the individual spatial locations of stores has been a difficult and time consuming task, so that at most a handful of store types are evaluated for their attraction-avoidance tendency in any one study. After data collection, the analysis of individual locations is also difficult, so that the most commonly used method involves aggregate counts of stores in cells within a two dimensional grid, which is much less accurate than using measures derived from specific individual locations. The recent availability of extensive address databases and improved, readily available geocoding tools has made data collection and analysis a much more manageable task. In particular, we use a complete census of specific locations of all retail outlets (over 26,000) in two cities in our analysis, and measure spatial attraction-avoidance behavior of 54 different retail trade types based on SIC codes.

Third, the spatial structure of a group of homogeneous retailers is complex, and using a global scalar measure of clustering masks much of the interesting detail. In contrast, we assess clustering or avoidance using a vector measure of density as a function of store separation, and present the results in intuitive, easily-interpreted plots.

Although not the object of our research, we note that empirical studies of heterogeneous agglomeration, the clustering of different retailers, are more common. The agglomeration of different retailers and the structure of shopping districts was first explored as Central Places theory by Christaller (1933) and formalized by Lösch (1938, 1944). Empirical models of multipurpose shopping that relate heterogeneous agglomeration to store choice include Arentze et al., (1993), Ghosh and McLafferty (1984), and Harlam and Lodish (1995). We
also note that our research is confined to the spatial structure of retail outlet locations, and does not incorporate store characteristics such as floor area or sales.

In the next section we review the theoretical reasons for homogeneous agglomeration or avoidance and the limited empirical results to date. We then describe the data and methodology, followed by the results and discussion.

Factors Affecting Homogeneous Agglomeration or Avoidance

In the absence of any locational forces, store locations for a particular class of retailers should not be spatially distributed in a way that differs from the distribution of all retail stores in an area. Forces that lead to similar stores being repelled from, or attracted to, each other will lead to spatial structures that show greater avoidance or greater clustering than following a distribution similar to the distribution of all retail stores. Hotelling (1929) first demonstrated attractive forces and the co-location of identical retailers under simple assumptions. However, Hotelling’s minimum differentiation result is not robust. For example, addition of a third firm, (Lerner and Singer, 1937), or including price as a decision variable (d’Aspremont, Gabszewicz, and Thisse, 1979), both of which increase competition, will cause firms to separate.

Generally, the increasing ability to collect monopoly rents as firms become more separated would seem to be a very powerful reason for similar retailers to avoid each other, and this strategic desire to spatially differentiate to avoid price competition is the primary repelling force identified in the literature. A closely related reason for stores to separate is market coverage, which becomes more important if travel costs are convex (d’Aspremont, et al., 1979) and if demand is elastic (Eaton and Lipsey, 1979).

In contrast to a single dominant repelling force, attraction forces that encourage homogeneous agglomeration are many and varied. Customer-side forces include

- comparison shopping motivated by customer uncertainty
• customer taste heterogeneity

• customer expectations of lower prices

• increased customer awareness

• shopping for entertainment

Firm side forces include:

• efficiencies in firm resource utilization

• localized resources

• reduced location choice risk

• follower’s traffic interceptor strategy

The first three of the customer-side reasons are thoroughly explored in the theoretical literature, and involve product differentiation to mitigate the price competition induced by minimal spatial differentiation. Awareness and entertainment reasons have not been formally modeled. Firm side forces are frequently mentioned, but little detailed investigation has been done. Below, we discuss these factors in greater detail.

*Forces encouraging homogeneous agglomeration that involve customer shopping behavior*

Comparison shopping for a single good encourages customers to visit several stores, and the travel-cost economizing shopper will prefer to go to a spatial cluster of similar stores. This rationale for homogeneous agglomeration was first described by Lösch (1944), and the shopping behavior is embodied in the classic notion of a shopping good (Copeland, 1923). Comparison shopping is motivated by consumer uncertainty of price, quality (Bester, 1998), or other attributes (Eaton and Lipsey, 1979) of the good or store. Customers’ uncertainty
of their own tastes (Konishi, 2005) also makes the greater variety associated with a cluster more attractive to shoppers who need to resolve their taste uncertainty. Heterogeneous tastes across customers is satisfied by a variety of similar but slightly differentiated products, and also encourages clustering (Fischer and Harrington, 1996; De Palma, Ginsburgh, Papageorgiou, and Thissse, 1985). Among the implications of this is that customer search, and hence store clustering, should be greater for products that are not standardized (such as antiques), or where there is an extensive range of different product options (such as shoes). Fischer and Harrington, 1996 explicitly focus on the degree of differentiation of the product or store, and their model shows that greater assortment heterogeneity leads to more search and a stronger tendency to agglomerate.

When price is included as a decision variable in theoretical models, consumers have rational expectations of lower prices, and equilibrium prices should be lower in clusters. Consequently, expectations of lower prices increase the attractiveness of clusters (Eaton and Lipsey, 1979; Konishi, 2005; Miller and Finco, 1995).

While most theoretical research is constrained to one-time purchase occasions, Bester (1998) develops a model for repeat-purchase categories. In contrast to situations where clusters have lower prices, in this model firms signal high-quality with high prices to attract first-time customers who are uncertain of quality. This compensates for increased price competition of agglomerated firms. Consumer quality uncertainty and rational expectations of higher quality has a similar logic to consumer price uncertainty and rational expectations of lower prices.

These theoretical models assume that customers know the locations of stores, and that search in the form of comparison shopping occurs after arriving at a known store location. In contrast, the customer search literature recognizes prior search in the media, through personal contacts, and customer’s memories (e.g., Beatty and Smith, 1987; Moorthy, Ratchford, and Talukdar, 1997; for a review see Guo, 2001). Consumer awareness of clusters of retail outlets should be higher than their awareness of a single outlet, making the cluster easier to
remember, which in turn will increase the effective market area of the agglomerated outlets. Although this simple awareness increase has not been addressed in the theoretical literature, Nelson’s (1958) principle of cumulative attraction suggests that we can expect it to be an important reason for agglomeration. Thus, homogeneous agglomeration should not only be beneficial by facilitating search after arriving at a store, it should also confer an advantage over isolated stores during memory search. While the impact of comparison shopping on clustering tendency will be strongly category dependent, the impact of location awareness should not be.

A second behavior not addressed in the theoretical agglomeration literature is shopping as entertainment. A strong interest in a particular product category will lead customers to wish to spend leisure time in a specialized shopping district even if nothing is purchased. Categories where this may be an important factor are art galleries, books, and clothing.

**Supply Side Forces Encouraging Homogeneous Agglomeration**

Efficiencies in resource utilization (e.g. cooperative advertising for an auto mall), shared infrastructure (e.g., boat retailers on a waterway), and localized resources (e.g., ethnic restaurants in an ethnic community), are supply side reasons for homogeneous agglomeration. Locating a new store near existing successful stores may be seen by managers as reducing the risk in location choice (see, for example, Mulligan, 1984). Stores may also locate near well-known existing competitors as a customer “interceptor strategy” (Nelson, 1958).

**Base Spatial Demand Density**

Variations in base demand levels will cause variations in the intensity of retail activity across the landscape. The theoretical literature on homogeneous agglomeration abstracts from these variations in order to identify forces that drive similar retailers closer together. Inferring clustering and avoidance forces from empirical analysis of spatial structure must similarly
control for this clumpiness of demand density, and that is a difficult task because of the
variety of drivers of base demand. These include

- Static demand origin points (e.g., household locations with varying income levels, work-
  places),

- Mobile drive-by demand, and

- Demand attractors (e.g., recreation centers, planned shopping centers, schools).

As discussed below, most of the empirical work on homogeneous clustering recognizes this
limitation to substantive conclusions, and a few attempt to address it by using (for example)
static origin points, such as population density, but with limited success. The control that
we propose is the overall retail intensity in localized areas. To the extent that spatial demand
density, regardless of the drivers, is reflected by overall retail intensity, a measure of total
retail outlet density provides a proxy for demand density.

Other determinants of retail locations and density include geographic constraints (e.g.,
rivers), municipal boundaries, and zoning regulations. Many of these, geographic constraints,
and zoning in particular, will also be reflected in, and controlled for, by overall retail density.

**Empirical Studies**

A few retail categories have been the subject of empirical investigations of clustering. Studies
most relevant to our research are summarized in Table 1, and differ in the geographic location,
the measures used, the attempts—if any—to control for demand density, and the results. The
earliest work approached the problem using density measures operationalized by counts of
stores in grid cells, and inferring the underlying patterns on the basis of determining which
distributions best fit these measures. Rogers (1965) studied six store types in Stockholm,
and found that antique stores cluster the most, and liquor stores avoid each other. In a
second article, Rogers (1969) studied stores in Ljubljana and San Francisco. In both cases
the most clustered stores were clothing stores, and the least clustered were specialty grocery
stores. He recognized that the uneven distribution of population or purchasing power affects the results, but did not attempt to correct for it. In a departure from highly aggregated grid count methods, Lee (1979) developed a nearest neighbor measure to address interdependence stores of the same and different types. In contrast to Rogers, he found that Western grocery stores and Chinese grocery stores in Hong Kong are clustered, and that they also cluster with each other. He applied the method to convenience stores in Phoenix and Atlanta, and found that all convenience stores and each chain separately are spatially random, and that chains avoid each other. Again, the substantive implications are limited because no effort was made to control for demand density. A similar study of gasoline stations in Hong Kong and Denver (Lee and Schmidt, 1980) found that gas stations were clustered (with the counter intuitive implication that gasoline is a shopping good), but in a study that did not control for demand density.

Rogers and Martin (1971) were the first to attempt to control for population density using several complex models, but their model fits were poor, and their substantive conclusions were very limited. Lee and Koutsopoulos (1976) found that convenience stores in Denver showed clustering, contrary to expectations and to Lee (1979). They suspected that this was a result of demand density rather than an attraction effect, since they also showed that the population was clustered. However, population clustering only explained 25% of the variance of store clustering in a regression analysis. Whether the remaining variance was due to homogeneous store attraction or the other drivers of demand density—such as mobile demand or other attractors—remained indeterminate. Fischer and Harrington (1996), in an introduction to their theoretical analysis, examined the clustering of 9 retail categories in Boston, with antiques the most clustered and supermarkets and theaters the least. They also qualitatively judged that greater product differentiation and greater requirements for search lead to greater clustering. Jensen, Boisson, and Larralde (2005) found motorbike shops the most clustered and and banks the least of seven store types, using the mean number of stores in a contiguous retail area relative to the mean of the same stores in the entire city as a
<table>
<thead>
<tr>
<th>Study</th>
<th>Data Type</th>
<th>Agglomeration Measure</th>
<th>Demand Control</th>
<th>City</th>
<th>Store Types and Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rogers (1965)</td>
<td>Density</td>
<td>Distribution fit; scalar</td>
<td>None</td>
<td>Stockholm</td>
<td>From most to least clustered: antiques, clothing, furniture, grocery, tobacco, liquor</td>
</tr>
<tr>
<td>Rogers (1969)</td>
<td>Density</td>
<td>Distribution fit; scalar</td>
<td>None</td>
<td>Ljubljana, San Francisco</td>
<td>From most to least clustered: clothing, non-food, all stores, food stores, grocery stores.</td>
</tr>
<tr>
<td>Rogers and Martin (1971)</td>
<td>Density</td>
<td>Distribution fit; scalar</td>
<td>Population</td>
<td>Ljubljana</td>
<td>Food stores, ambiguous results</td>
</tr>
<tr>
<td>Lee and Koutsopoulos (1976)</td>
<td>Density</td>
<td>Distribution fit; scalar</td>
<td>Regress results on population density (not point), demographics; weak or no results</td>
<td>Denver</td>
<td>Convenience stores of same chain cluster, different chains avoid each other</td>
</tr>
<tr>
<td>Lee (1979)</td>
<td>Point</td>
<td>Nearest neighbor distribution fit; scalar</td>
<td>None</td>
<td>Hong Kong, Phoenix, Atlanta</td>
<td>Western and Chines grocery stores cluster; convenience stores random overall, chains avoid each other</td>
</tr>
<tr>
<td>Lee and Schmidt (1980)</td>
<td>Point</td>
<td>Nearest neighbor distribution fit; scalar</td>
<td>Regress station density on households, demographics, trips into area densities (not point); weak or no effects</td>
<td>Hong Kong, Denver</td>
<td>Gasoline stations cluster</td>
</tr>
<tr>
<td>Fischer and Harrington (1996)</td>
<td>Point</td>
<td>Nearest neighbor</td>
<td>None</td>
<td>Baltimore</td>
<td>From most to least clustered: shoes, antiques, computers, automobiles, clinics, gasoline stations, video stores, supermarkets, theaters.</td>
</tr>
<tr>
<td>Netz and Taylor (2002)</td>
<td>Point</td>
<td>regression coefficient; relative within store type; scalar</td>
<td>Population demographics</td>
<td>Los Angeles</td>
<td>Gasoline stations avoid each other more as competitive intensity increases (not an absolute measure)</td>
</tr>
<tr>
<td>Jensen, Boisson, and Larralde, 2005</td>
<td>point</td>
<td>average store counts in contiguous sites</td>
<td>none</td>
<td>Lyon</td>
<td>From most to least clustered: motorbikes, banks, groceries, hairdressers, laundries, drugstores, savings banks</td>
</tr>
<tr>
<td>Picone, Ridley, and Zandbergen (2009)</td>
<td>Point</td>
<td>Nearest neighbor scalar and vector; relative within store type;</td>
<td>Population demographics on scalar; assume constant effects on compared stores for vector</td>
<td>Birmingham, Chicago, Minneapolis-St. Paul, Oakland, Tampa</td>
<td>On site alcohol retailers cluster more than off site alcohol retailers</td>
</tr>
</tbody>
</table>
measure of clustering in Lyon, and in a simulation of the evolution of same-store spatial patterns.

Two studies address the clumpy nature of demand density by comparing across stores that are assumed to have a similar relation to spatial demand. Netz and Taylor (2002) model a spatial differentiation measure as a function of competitive intensity and demographics across gasoline stations. Since the product is homogeneous and prices posted, there should be little incentive for comparison shopping, so that avoidance should be observed. The resulting positive relation between spatial differentiation and competitive intensity is interpreted as a strategic attempt to increase spatial differentiation as competitive intensity increases, which is interpreted as being consistent with avoidance forces dominating attraction forces. Picone, Ridley, and Zandbergen (2009) handled demand density by contrasting two similar categories, on site and off site alcohol retailers. They use two measures of spatial structure, a scalar nearest neighbor index that measures the degree of clustering relative to a random distribution in a fixed region, and a vector measure of average density as a function of store separation. Assuming that the on site retailers (e.g. restaurants and bars) have a greater ability to differentiate their product than off site retailers (e.g., liquor stores and grocery stores), theory predicts that the on site group will have less need to spatially differentiate (i.e. cluster), and this comparative result is supported.

Other research invokes homogeneous agglomeration, but in a way that is less relevant to our emphasis on direct measurement of attraction and avoidance. Taking a survey-based approach to the relation between shopping behavior and agglomeration, Popkowski-Leszczyc, Sinha, and Sahgal (2004) modeled store choice using a telephone survey of grocery shopping behavior in New Zealand, and found a positive effect on store patronage of both heterogeneous and homogeneous agglomeration for one consumer segment representing 25% of the sample. A similar approach by Fox, Postrel, and McLaughlin (2007) used panel data to model the relation between household expenditures at six focal broad line retailer chains as a function of travel time and agglomeration with six general broad line retail types. Ag-
Agglomeration was measured by a count of the number of stores which were “close enough” to a focal store to facilitate single trip shopping, with the closeness threshold determined by grid search. No significant homogeneous agglomeration effects were found.

A number of studies address the demographic, environmental, marketing mix, and competitive determinants of retail structure. Retail structure is a broad concept that includes spatial structure as well as other characteristics such as the size of stores, sales per store, and the mix of different types within a geographic area (e.g., Bucklin, 1972; Hirschman, 1978; Ingene, 1983; Ingene and Brown, 1987). Most of this work does not address the more narrow question of whether avoidance or attraction dominates spatial structure, with the exception of Miller, Reardon, and McCorkle (1999) who study competitive affects among stores differing by breadth of assortment within the same product category. Their assortment-based classification scheme includes three levels: limited-line specialists, broad-line specialists, and general merchandisers. Studying sporting goods stores and using a density measure of stores per household in metropolitan census areas, they find that both limited and broad line specialists show strong “Darwinian” or avoidance patterns among each other, but that “symbiotic”, or attraction effects exist across adjacent categories. Karande and Lombard (2005), using Miller et al.’s classification scheme, but with the more detailed Lee and Schmidt nearest neighbor measures, investigate broad line retailers in home repair and office supplies. They find either clustering or avoidance, depending on the characteristics of different markets, but again they do not explicitly control for demand density.

**Expected Structure Of 54 Retail Categories**

In the following empirical section, we explore the spatial structure of 54 different types of retailers. There is general theoretical agreement that convenience stores (which serve small neighborhood trade areas with similar convenience goods, and do not inspire comparison shopping) and gasoline stations (which carry a relatively undifferentiated and clearly priced commodity) should avoid each other as there are no obvious strong agglomeration forces to
offset competition avoidance. We also expect that antique shops and clothing stores inspire both recreational and comparison shopping, and should therefore cluster. For most product categories, however, the variety of reasons for clustering, and the mixed empirical results to date do not inspire confidence in making predictions of spatial structure. Such predictions are beyond the scope of this research; nevertheless, to set some expectations based on the literature, we trained six judges by having them read a summary of the literature similar to that reported above. We then gave them a list of the 54 categories and asked them to indicate if they expected clustering or avoidance, on a bipolar seven-point scale, with the midpoint indicating neither. A do-not-know option was also provided.

We classified a store type as expected to cluster if at least five of the six judges gave a rating of six or seven, or if four judges gave that rating and at least one judge gave midpoint ratings (that is, not one or two). A symmetric rule was applied to the other end of the scale to identify stores expected to avoid each other. The categories that were thus classified as clustering, avoiding, or neither are shown in Table 1.

In addition, when they expected clustering, four of the judges were asked to provide their rationale by selecting one or two reasons from a list. We examined the reasons for those stores classified in Table 1 as clustered. A reason was counted as “real” if at least two judges selected it. Comparison shopping is the most common single “real” reason, and applies to doors, children’s clothing, clothing, kitchen cabinets, wedding supplies, and cosmetics. Interestingly, the next largest group are store categories where there are two real reasons, comparison shopping and entertainment: women’s clothing, shoes, antiques, sporting goods, and art galleries. Categories where both reasons of comparison shopping and of a greater awareness of clustered retailers apply are used cars, men’s clothing, and bridal shops. New cars have these two reasons, plus a third, shopping for entertainment. Expectations of lower prices in a cluster appeared, either alone or with other reasons, for furniture, carpets, and consumer electronics.
Table 1. Expectations of Homogeneous Agglomeration by Trained Judges

<table>
<thead>
<tr>
<th>AVOID</th>
<th>NEITHER</th>
<th>CLUSTER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bakers</td>
<td>Paint</td>
<td>Auto New</td>
</tr>
<tr>
<td>Convenience Stores</td>
<td>Lumber</td>
<td>Doors</td>
</tr>
<tr>
<td>Health &amp; Diet</td>
<td>Glass</td>
<td>Auto Used</td>
</tr>
<tr>
<td>Ice Cream Parlors</td>
<td>Supermarkets</td>
<td>Boat Dealers</td>
</tr>
<tr>
<td>Pizza Parlors</td>
<td>Meat</td>
<td>Men’s Clothing</td>
</tr>
<tr>
<td>Pharmacies</td>
<td>Produce</td>
<td>Women’s Clothing</td>
</tr>
<tr>
<td>Liquor, Beer, Wine</td>
<td>Candy &amp; Confectionery</td>
<td>Bridal Shops</td>
</tr>
<tr>
<td>Florist</td>
<td>Auto Parts</td>
<td>Children’s Clothing</td>
</tr>
<tr>
<td>Tobacco</td>
<td>Tires</td>
<td>Clothing</td>
</tr>
<tr>
<td>Orthopedic Appliance</td>
<td>Gasoline Stations</td>
<td>Shoes</td>
</tr>
<tr>
<td></td>
<td>Tailors</td>
<td>Kitchen Cabinets</td>
</tr>
<tr>
<td></td>
<td>Housewares</td>
<td>Furniture</td>
</tr>
<tr>
<td></td>
<td>Music CDs &amp; Tapes</td>
<td>Carpets</td>
</tr>
<tr>
<td></td>
<td>Coffee Shops</td>
<td>Consumer Electronics</td>
</tr>
<tr>
<td></td>
<td>Pubs</td>
<td>Art Galleries</td>
</tr>
<tr>
<td></td>
<td>Bookstores</td>
<td>Sporting Goods</td>
</tr>
<tr>
<td></td>
<td>Jewelry</td>
<td>Gift Shops</td>
</tr>
<tr>
<td></td>
<td>Toys</td>
<td>Wedding Supplies</td>
</tr>
<tr>
<td></td>
<td>Fabric Shops</td>
<td>Antiques</td>
</tr>
<tr>
<td></td>
<td>Opticians</td>
<td>Cosmetics</td>
</tr>
<tr>
<td></td>
<td>Picture Frames</td>
<td>Draperies</td>
</tr>
<tr>
<td></td>
<td>Pets and Pet Supplies</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hearing Equipment</td>
<td></td>
</tr>
</tbody>
</table>


Data

Address data with latitude and longitude information and SIC codes was purchased from InfoCanada for 18,267 retail outlets in the Greater Vancouver Regional District (which we will refer to as Vancouver), and 8,401 retail outlets in Calgary. These addresses are a census of Vancouver and Calgary retail locations for 2005. The metropolitan Vancouver region is approximately 50 by 50 kilometers (or 31 by 31 miles), and Calgary is approximately 25 by 35 kilometers (or 16 by 22 miles). Approximately 90% of the retail locations in the original data set were geocoded by InfoCanada at the address level. The remainder were less precisely located, using only the postal code, and we re-geocoded these to the address level using the open source software PAGC (www.pagcgeo.org) so that all of the retail outlets have address-level spatial locations. For the 90% that were already address-level geocoded, we did a quality check using postal forward sortation area (FSA) polygon GIS layers, to determine whether the geocoded spatial location actually fell into the FSA of the store’s given postal code in an automated way using a point-in-polygon check.\textsuperscript{1} A common problem with automatic geocoding is the existence of multiple streets in a metropolitan area with the same name (which is a significant problem for Main Street and Marine Drive addresses in metropolitan Vancouver), which can result in stores being geocoded to the wrong street. An additional 5% of addresses in Vancouver were re-geocoded for this reason. The mis-geocoding problem was much less common in Calgary. Figure 2 shows store locations in Vancouver at two scales.

\textsuperscript{1} A forward sortation area is identified by the first three of six alpha-numeric characters in the Canadian postal code system, and in urban areas cover geographic areas that are comparable in size to US Zip Code Tabulation Areas.
Figure 2. Store Locations in Vancouver
We selected a subset of 54 retailer types, based on SIC codes, to study homogeneous agglomeration. The criteria for selection were that there were at least 30 locations in Vancouver for that SIC code. In addition, we wanted to keep product assortments fairly narrow, as a result, no general merchandise or department stores (which have broad assortments and therefore compete with a broad range of other types of stores) would be included. Due to the coarseness of SIC codes in this area, most restaurants (other than coffee shops and pizza parlors, which had separate SIC codes) were not included. Grocery stores also suffered from having coarse SIC codes, and outside information was used to determine which of these stores were supermarkets, convenience stores, produce stores, or other food stores (with the “other food store” category, consisting mostly of ethnic foods stores, omitted from the analysis). Table 2 shows the number of outlets of each of the categories.

**Measuring Homogeneous Agglomeration**

A wide range of measures to detect spatial clustering have been proposed in a number of fields, most notably epidemiology and ecology. These measures vary by the type of data and the objective of cluster detection. For example, two common types of spatial data are counts of cases in subareas of the study area, and geographical point locations of individual cases. Objectives include focused tests to determine if a cluster occurs around a predetermined location, such as increased incidence of disease around an industrial site; cluster detection tests designed to pinpoint the location of any cluster anywhere in the study area; and global tests designed to assess whether a particular process clusters overall in a study region. In many cases, the existence of clusters is determined by comparison against a null distribution. Possible null distributions include a random Poisson point process, or a random process within a varying background population density. Haining (2003, pp 237-270) provides a good overview of cluster measures.

Our objective of determining the overall degree of agglomeration of a particular retail outlet type requires a measure of global clustering applied to point location data. The
Table 2. The Number of Outlets for Each of the 54 Retailer Types in Vancouver and Calgary

<table>
<thead>
<tr>
<th>Type</th>
<th>Vancouver</th>
<th>Calgary</th>
<th>Type</th>
<th>Vancouver</th>
<th>Calgary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doors</td>
<td>58</td>
<td>36</td>
<td>Draperies</td>
<td>31</td>
<td>42</td>
</tr>
<tr>
<td>Lumber</td>
<td>44</td>
<td>16</td>
<td>Housewares</td>
<td>35</td>
<td>8</td>
</tr>
<tr>
<td>Paint</td>
<td>67</td>
<td>32</td>
<td>Consumer Electronics</td>
<td>103</td>
<td>85</td>
</tr>
<tr>
<td>Glass</td>
<td>158</td>
<td>84</td>
<td>Music CDs &amp; Tapes</td>
<td>61</td>
<td>30</td>
</tr>
<tr>
<td>Supermarkets</td>
<td>174</td>
<td>50</td>
<td>Ice Cream Parlors</td>
<td>92</td>
<td>69</td>
</tr>
<tr>
<td>Convenience Stores</td>
<td>427</td>
<td>227</td>
<td>Pizza Parlors</td>
<td>413</td>
<td>238</td>
</tr>
<tr>
<td>Meat/Butcher</td>
<td>131</td>
<td>45</td>
<td>Pubs</td>
<td>142</td>
<td>151</td>
</tr>
<tr>
<td>Produce</td>
<td>204</td>
<td>13</td>
<td>Pharmacies</td>
<td>278</td>
<td>158</td>
</tr>
<tr>
<td>Candy &amp; Confectionery</td>
<td>41</td>
<td>25</td>
<td>Liquor, Wine, Beer</td>
<td>172</td>
<td>217</td>
</tr>
<tr>
<td>Bakers</td>
<td>354</td>
<td>96</td>
<td>Antiques</td>
<td>146</td>
<td>27</td>
</tr>
<tr>
<td>Health &amp; Diet</td>
<td>262</td>
<td>105</td>
<td>Sporting Goods</td>
<td>345</td>
<td>174</td>
</tr>
<tr>
<td>Coffee Shops</td>
<td>486</td>
<td>186</td>
<td>Bookstores</td>
<td>147</td>
<td>56</td>
</tr>
<tr>
<td>Auto New</td>
<td>182</td>
<td>76</td>
<td>Jewelry</td>
<td>348</td>
<td>124</td>
</tr>
<tr>
<td>Auto Used</td>
<td>198</td>
<td>55</td>
<td>Toys</td>
<td>45</td>
<td>29</td>
</tr>
<tr>
<td>Auto Parts</td>
<td>130</td>
<td>86</td>
<td>Gifts</td>
<td>332</td>
<td>199</td>
</tr>
<tr>
<td>Tires</td>
<td>96</td>
<td>45</td>
<td>Fabric</td>
<td>76</td>
<td>22</td>
</tr>
<tr>
<td>Gas Stations</td>
<td>294</td>
<td>183</td>
<td>Florists</td>
<td>253</td>
<td>144</td>
</tr>
<tr>
<td>Boat Dealers</td>
<td>51</td>
<td>13</td>
<td>Smoking Supplies</td>
<td>113</td>
<td>41</td>
</tr>
<tr>
<td>Men’s Clothing</td>
<td>122</td>
<td>58</td>
<td>Opticians</td>
<td>181</td>
<td>39</td>
</tr>
<tr>
<td>Women’s Cloth.</td>
<td>695</td>
<td>208</td>
<td>Picture Frames</td>
<td>80</td>
<td>50</td>
</tr>
<tr>
<td>Bridal Shops</td>
<td>55</td>
<td>26</td>
<td>Pets and Pet Supplies</td>
<td>138</td>
<td>44</td>
</tr>
<tr>
<td>Children’s Cloth.</td>
<td>64</td>
<td>25</td>
<td>Orthopedic Supplies</td>
<td>57</td>
<td>14</td>
</tr>
<tr>
<td>Clothing</td>
<td>83</td>
<td>142</td>
<td>Wedding Supplies</td>
<td>34</td>
<td>12</td>
</tr>
<tr>
<td>Shoes</td>
<td>214</td>
<td>120</td>
<td>Art Galleries</td>
<td>220</td>
<td>82</td>
</tr>
<tr>
<td>Tailors</td>
<td>124</td>
<td>71</td>
<td>Hearing Equipment</td>
<td>48</td>
<td>26</td>
</tr>
<tr>
<td>Kitchen Cabinets</td>
<td>109</td>
<td>32</td>
<td>Cosmetics</td>
<td>88</td>
<td>32</td>
</tr>
<tr>
<td>Furniture</td>
<td>291</td>
<td>149</td>
<td>Carpets</td>
<td>107</td>
<td>98</td>
</tr>
</tbody>
</table>
measure we use is Kulldorff’s $D$, which is based on Ripley’s $K$-function (Ripley, 1981). The measure was proposed by Kulldorff (1998, p. 54) and is a variation of Diggle and Chetwynd’s (1991) $D(d)$ function. The reason we use this measure is that it presents clustering as a function of separation of retail outlets, rather than as a single number that only gives the overall degree of agglomeration of stores.\footnote{An example of a global measure of spatial clustering used in the marketing literature is Garber, Goldenberg, Libai, and Muller (2004), who use a cross-entropy measure of spatial clustering for early prediction of new product success.} Aside from providing richer detail than a single number, the graphical presentation of the K-function allows for an intuitive visual interpretation of the global spatial structure of a particular retail category.

**The K-Function**

Analogous to a probability density, the intensity of a spatial point process $\lambda(x)$ over the space $x$ is given by

$$\lambda(x) = \lim_{|dx| \to 0} \left\{ \frac{E[N(dx)]}{|dx|} \right\},$$

where in our case $x$ is implemented as two dimensional geographic coordinates.\footnote{For our analysis, the original longitude and latitude coordinates were reprojected into the appropriate Universal Transverse Mercator (UTM) coordinate system (UTM Zone 10N for Vancouver and UTM Zone 11N for Calgary) so that simple Euclidean distance measure could be used.} $N(dx)$ is the number of events within a region $dx$ that contains the point $x$. A stationary, or homogeneous, process has constant $\lambda$ across the region. The intensity can be estimated as a constant for any finite region by dividing the number of points in the region by its area.

The K-function captures the expected intensity of a stationary point process as a function of the distance $d$ between the points where events of interest occur. The function $K(d)$ is estimated by first counting the number of events within a distance $d$ of each event. In Figure 3, distances of two, three, and four from the focal store give counts of six, eight, and nine stores. The counts are repeated and averaged over all of the focal stores, normalized by dividing by the estimate of the average $\lambda$, and corrected for edge effects (which are defined...
in the next paragraph), resulting in the expression

$$\tilde{K}(d) = n^{-2}|A| \sum_i^n \sum_{j \neq i} W_{i,j}^{-1} I_d(d_{i,j}).$$

where $A$ is the study region with area $|A|$, and contains a total of $n$ points. The inverse average intensity within the study region is given by $n^{-2}|A|$, which normalizes the measure and allows comparisons across different study areas. The distance between points $i$ and $j$ is $d_{i,j}$, $I_d$ is an indicator function taking the value 1 if $d < d_{i,j}$, and $W_{i,j}$ is an edge effect correction. The edge effect correction is necessary to account for the fact that as the radius $d$ increases, it will asymmetrically encounter the boundaries of the study area beyond which there are no further retail outlets. $W_{i,j}$ is defined as the proportion of the circle, with center at $i$ and passing through $j$, that is contained within $A$ (an illustration of which is shown in Figure 4).
If retail outlets are randomly distributed in the region, for example with a two dimensional Poisson point process, the K-function will increase with the square of distance. If clusters occur, the K-function will be greater than the Poisson rate at small distances, and smaller than the Poisson rate at large distances. If outlets strongly avoid each other, leading to a more regular spacing than a random point process, the K-function will be less than the Poisson rate at smaller distances. The K-function (or L-function) by itself could be used as the criteria for clustering. However, as the long history of empirical research on homogeneous agglomeration testifies, teasing out the attraction/avoidance implications for homogeneous retailers is difficult since spatial demand density and zoning regulations strongly influence clustering, and the use of the K-function (or L-function) alone does not control for these effects.

To remedy the clumpy demand density problem, we take advantage of the fact that our data set is a census of all retail outlets. This unique data allows us to use the spatial intensity of all retailers as a proxy for spatial demand density. This will also provide a control for spatial structure arising from zoning regulations. For example, if we wish to infer whether florists benefit more from proximity to other florists, or more from avoiding other florists,

---

4 Picone, Ridley, and Zandbergen (2009) use an area-normalized version the K-function, the L-function, to correct for this increase in their analysis of liquor outlets.
separately from the base demand density and zoning regulations, we should determine if the florist locations are more or less clustered than retailers of all types around each florist. The underlying assumption is that base demand density affects all retailers equally. This is not a perfect control for all of the factors that contribute to spatial structure, but we believe that it is a substantial improvement over previous methods, and furthermore can be used as a consistent control to infer attraction or avoidance across many different store types. As an indication, consider population density, which is a major component of spatial demand density. Figures 5 and 6 provide maps of Vancouver and Calgary that illustrate the relationship between population density (based on the sum of the residential and workplace population in an area) and retail locations. These figures reveal a strong, albeit not perfect, relationship between population density and retail intensity. The figures also reveal that in less densely populated areas with a substantial number of retailers, the outlets often trace out an arc, which is consistent with being along arterial thoroughfares which would provide a source of mobile demand.
A common method from epidemiology to implement the control (Diggle and Chetwynd, 1991) is to simply take the difference between the K-function for the “cases” (florists) and the K-function for “controls” (all other retail outlets).\(^5\) This works well provided there are few cases that are all spatially embedded within many controls. However, since the two K-functions are calculated independently, if case clusters are spatially isolated from otherwise identical control clusters, the K-functions for cases and controls could be the same. The difference between them will be zero, and the real clustering of the cases will not be detected. Kulldorff (1998) proposed a variation that resolves this problem by taking into account the proximity of cases and controls.\(^6\) Let $K_{c,c}(d)$ be the standard K-function that counts the number of cases within a distance $d$ of each case, and let $K_{c,k}(d)$ be the K-function that counts the number of controls within a distance $d$ of each case, then

\[^5\] We adopt the jargon “cases” and “controls” from the epidemiological problem of isolating disease density from the background population density.

\[^6\] To the best of our knowledge, there has been no previous empirical application of this measure.
Figure 7. An example of strong “Case” (white crosses) agglomeration relative to the “controls” (black crosses).

\[ D_{c,k}(d) = K_{c,c} - K_{c,k} \]  \hspace{1cm} (3)

is a measure of the degree of clustering of the cases (specific stores) relative to the background of controls (all other retail stores). It can be thought of as the difference in the probability of encountering a case and the probability of encountering a control as one moves further out from a particular case. In Figure 7, the stores of interest (cases, represented by open crosses) would be measured as clustered, while in Figure 8 (which has identical case locations) the measure corrects the density of the case stores for the increased density of all other types of retailers (controls, represented by filled crosses), and the cases would not be considered clustered.

The sampling properties of (3) are such that it is not possible to derive formulas to calculate confidence intervals. Instead, significance is assessed by constructing Monte Carlo confidence intervals based on randomly labeling a proportion of all retailers in the study area as cases and recalculating \( D_{c,k}(d) \). For example, if we have 50 bridal shop outlets that we are studying, we randomly select 50 retail locations from the set of all retail locations, and calculate \( D_{c,k}(d) \) for the random set. Repeat this exercise 100 times, and obtain the 90% confidence interval by taking the fifth largest and fifth smallest values as the limits.
Results

A Range of Patterns from Clustering to Avoidance

Figure 9 shows 55 bridal shops in Vancouver, which exhibit some clustering and some isolated shops, and a closeup of one particular cluster of 10 bridal shops.

The $D$ statistic for bridal shops (the dark line in Figure 10) shows a rapid increase in the proximity of bridal shops to each other at short distances. The fan-shaped curves are the second, fifth, 95th, and 99th percentile Monte Carlo limits. If one is at a bridal shop, one is much more likely to encounter another bridal shop within a half kilometer than would be the case if the shops were distributed with the same spatial intensity as all retailers.

If each store is randomly located with a spatial intensity that simply follows the overall retail intensity, the measure will stay close to zero at all distances. Figure 11 show that picture framers follow this pattern.

Past empirical research on spacing of gasoline stations, which theoretical considerations suggest are excellent candidates for avoidance, has been mixed, or required indirect approaches of detection. The $D$-plots cleanly detect the expected avoidance, as shown in Figure 12 for both Calgary and Vancouver.
Figure 9. Locations of Bridal Shops in Vancouver

Figure 10. Kulldorff’s $D$ for Vancouver Bridal Shops
Figure 11. Kulldorff’s $D$ for Picture Framing Shops

Figure 12. Kulldorff’s $D$ for Gasoline Stations
A Typology of Spatial Patterns

To assist with comparing and contrasting spatial patterns across stores and cities, we developed a measure to assist us in categorizing the patterns. The measure \( (M_i) \) is based on normalizing the value of the \( D \) statistic by the relevant 95\textsuperscript{th} percentile confidence limit of the \( D \) statistic if \( D \) is positive, and the fifth percentile limit if \( D \) is negative. Specifically, for \( D \) taken at distance \( i \), the measure \( M_i \) is calculated as

\[
M_i = \begin{cases} 
  D_i / D^U_i, & D_i > 0 \\
  -D_i / D^L_i, & D_i \leq 0 
\end{cases}
\]

where \( D^U_i \) is the 95\textsuperscript{th} percentile confidence level for the \( D \) statistic at distance \( i \), and \( D^L_i \) is the 5\textsuperscript{th} percentile confidence level for the \( D \) statistic at distance \( i \). We calculated this value at 200 meter intervals for each retail category in each city. We next averaged the \( M_i \) values over 2 kilometer quintiles of the 10 kilometer range, and used this measure to classify retail store types into different patterns.

We then looked at those categories that had an average \( M_i > 1 \) for all five quintiles, the first four quintiles, the first three quintiles, and so on. If the average value of \( M_i \) was greater than one in at least three of the quintiles we labeled the spatial structure as “hyper agglomeration”. These strongest clustering stores in both Vancouver and Calgary are furniture, antiques, coffee shops, art galleries, smoking supplies, gift shops, used automobiles, and bookstores (see Table 3). If \( M_i \) was greater than one only for the first one or two quintiles, we call it “local agglomeration”. Many categories that exhibited local agglomeration in one city were classified as hyper agglomeration in the other, reflecting small differences on the continuum of agglomeration intensity. Among those categories that showed local agglomeration many are those that readers will recognize as being common in malls, such as Men’s, Women’s and Children’s Clothing, Shoes, and Jewelry.
Table 3. Agglomeration Patterns Exhibited by Different Retail Types

<table>
<thead>
<tr>
<th>Retail Types with Same Nonrandom Spatial Pattern in Vancouver and Calgary</th>
<th>Hyper Agglomeration</th>
<th>Local Agglomeration</th>
<th>Auto Mall Pattern</th>
<th>Avoidance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Furniture</td>
<td>Carpets</td>
<td>New Autos</td>
<td>Gas Stations</td>
<td></td>
</tr>
<tr>
<td>Antiques</td>
<td>Children’s Clothing</td>
<td>Boat Dealers</td>
<td>Supermarkets</td>
<td></td>
</tr>
<tr>
<td>Coffee Shops</td>
<td>Clothing</td>
<td></td>
<td>Health and Diet</td>
<td></td>
</tr>
<tr>
<td>Arte Galleries</td>
<td></td>
<td></td>
<td>Pizza Parlors</td>
<td></td>
</tr>
<tr>
<td>Smoking Supplies</td>
<td></td>
<td></td>
<td>Liquor, Beer, Wine</td>
<td></td>
</tr>
<tr>
<td>Used Autos</td>
<td></td>
<td></td>
<td>Pets and Supplies</td>
<td></td>
</tr>
<tr>
<td>Gift shops</td>
<td></td>
<td></td>
<td>Paint</td>
<td></td>
</tr>
<tr>
<td>Bookstores</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nonrandom Spatial Patterns Unique to Vancouver</th>
<th>Hyper Agglomeration</th>
<th>Local Agglomeration</th>
<th>Auto Mall Pattern</th>
<th>Avoidance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cosmetics*</td>
<td>Bridal*</td>
<td></td>
<td></td>
<td>Pubs</td>
</tr>
<tr>
<td>Women’s Clothing*</td>
<td>Delicatessen*</td>
<td></td>
<td></td>
<td>Ice Cream Parlors</td>
</tr>
<tr>
<td>Men’s Clothing*</td>
<td>Kitchen. Cabinets*</td>
<td></td>
<td></td>
<td>Lumber</td>
</tr>
<tr>
<td>Shoes*</td>
<td>Hearing Supplies*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jewelry*</td>
<td>Auto Parts*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Music CDs &amp; Tapes</td>
<td>Sporting</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fabric</td>
<td>Bakery</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tailors</td>
<td>Meat</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Opticians</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nonrandom Spatial Patterns Unique to Calgary</th>
<th>Hyper Agglomeration</th>
<th>Local Agglomeration</th>
<th>Auto Mall Pattern</th>
<th>Avoidance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bridal*</td>
<td>Cosmetics*</td>
<td>Housewares</td>
<td>Meat</td>
<td></td>
</tr>
<tr>
<td>Delicatessen*</td>
<td>Women’s Clothing*</td>
<td>Wedding Supplies</td>
<td>Fabric</td>
<td></td>
</tr>
<tr>
<td>Kitchen Cabinets*</td>
<td>Men’s Clothing*</td>
<td>Doors</td>
<td>Pharmacies</td>
<td></td>
</tr>
<tr>
<td>Hearing Supplies*</td>
<td>Shoes*</td>
<td>Lumber</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Auto Parts*</td>
<td>Jewelry*</td>
<td>Pubs</td>
<td>Convenience Stores</td>
<td></td>
</tr>
<tr>
<td>Lumber</td>
<td>Tires</td>
<td></td>
<td>Opticians</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Glass</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Differ between cities by hyper versus local agglomeration only
In the case of avoidance, we classified a category as exhibiting avoidance if \( M_i \) was less than one for at least one quintile, and no values were greater than one. Both Calgary and Vancouver show avoidance for the retail categories of gasoline stations (see Figure 12), supermarkets, health and diet, pizza parlors, liquor-beer-wine, pets and pet supplies, and paint. For most of these stores, the curve is uniformly negative and declining over the first four quintiles.

There are also more complex patterns. An unexpected one is a structure where a local cluster exists, but at larger scales, the clusters themselves show avoidance. This pattern is very strong in Vancouver for new automobile dealerships (hence our moniker the “auto mall” pattern) which is a retail category that has long been presented as anecdotal evidence of retail clustering. Using a criteria that either the first or second quintile had \( M_i > 1 \) (exceeding the Monte Carlo 95th percentile) and at least one of the last three quintiles had \( M_i < 1 \) (below the fifth percentile), new automobile dealers and boat dealers fell into this category in both cities, as did five other retailer types in Calgary. Several other retail types showed similar patterns but did not quite reach the lower significance criteria, including Auto Parts, Kitchen Cabinets, and Men’s Clothing. The observation that the clusters themselves avoid each other is novel and interesting. One partial explanation for boats in Vancouver is that the shops tend to be on the water, and thus there are limited locations available, a possible example of local resource limitations forcing agglomeration. However, this explanation is not available for boat dealers in Calgary, which lacks the waterways. In an attempt to begin to find an explanation for this pattern, in the next subsection we examine the case of Vancouver new automobile dealers in greater detail. Interestingly, \textit{used} auto dealerships do not exhibit this pattern; rather they show a tendency to agglomerate over the full 10 kilometer range investigated. Figure 13 shows the \( D \) statistic plot for Vancouver boat dealers to illustrate the pattern.

Several store types are quite different between the two cities. Retail classes that exhibit avoidance in Vancouver, but not in Calgary, are lumber and pubs, (both of which exhibit an
“auto mall” pattern in Calgary), and ice cream parlors. The categories that show some level of avoidance in Calgary, but not in Vancouver, are fabric stores, pharmacies, convenience stores, meat/butcher shops, and opticians, the latter two of which exhibit local agglomeration in Vancouver. The categories that show hyper agglomeration in Vancouver, but no agglomeration in Calgary are music CDs and tapes, tailors, and fabric. This interesting difference is investigated in the next subsection.

All in all, we show five patterns: (1) Hyper to very strong agglomeration (8 categories exhibit this pattern in both cities); (2) Moderate local agglomeration with no real pattern at larger geographic scales (3 categories in both cities); (3) Local agglomeration with more distant avoidance (2 categories in both cities); (4) Overall avoidance (7 categories in both cities); and finally (5) no tendency towards agglomeration or avoidance, instead simply having a distribution of stores that reflects the background intensity of all retailers.
Further Observations, Implications, and Predictions

The role of retail brand concentration in moderating homogeneous agglomeration

One factor that might help explain some of the differences between spatial patterns between Calgary and Vancouver for some retail types are differences between the two cities in the percentage of outlets under the same retail brand. Specifically, two stores in the same retail category under the same retail brand are likely to have a much higher overlap in their retail assortments than two stores in the same category under different brand names. Moreover, in locating stores, the firm controlling the retail brand (either through direct ownership or franchising) will likely seek to minimize sales cannibalization between the retail brand’s outlets by keeping them well separated. As a result of these two factors, fewer independent stores and more single-brand chain stores within a retail category, which we refer to as a higher retail brand concentration, should lead to less agglomeration in the retail category.

To assess whether increasing retail brand concentration decreases the tendency towards homogeneous agglomeration, we examine two categories (music CDs and tapes, and fabric) that hyper agglomerate only in Vancouver.\(^7\) The Kulldorff’s $D$ statistic analysis for music CDs and tapes is given in Figure 14, while the same analysis for fabric is shown in Figure 15. Both figures clearly show that the two categories hyper agglomerate in Vancouver, while music CDs and tapes exhibits only a weak agglomeration pattern in Calgary, and fabric essentially exhibits a pattern consistent with the density of all stores over most of the range of the statistic, and a weak avoidance pattern at the most distant range of the statistic.

To measure the extent of retail brand concentration, a Herfindahl index (based on each retail brand’s share of outlets in each city) is calculated for both categories, and is presented in Table 4. An examination of Table 4 indicates that the Herfindahl index for both retail categories is between four and five times higher in Calgary than in Vancouver. Since higher levels of the Herfindahl index indicates greater concentration, the numbers indicate that

\(^7\) Tailors also exhibit hyper agglomeration in Vancouver, but we did not include this category in our analysis since many of the tailor shops are located in Vancouver’s historic “China Town” district, suggesting its agglomeration is at least partially due to historical “supply-side” considerations.
Figure 14. Kulldorff’s $D$ for Music CD and Tape Stores

Figure 15. Kulldorff’s $D$ for Fabric Stores
the level of retail brand concentration for both of these categories much higher in Calgary than it is in Vancouver (the Herfindahl index is between four and five times higher for both categories), consistent with the proposition that increasing retail brand concentration mitigates the tendency towards agglomeration for these two retail categories.

A deeper look at the “auto mall” pattern

Given the high risk and the consequent high levels of comparison shopping associated with new car purchases, one would expect to see a tendency towards homogeneous agglomeration on the part of new auto dealers. Our results indicate that while new auto dealers (as well as several other retail categories) do tend to locally agglomerate, on a larger scale, the resulting clusters of outlets avoid one another. A natural question to ask is what drives this pattern of local agglomeration followed by more distant avoidance? To simplify the discussion, we will develop and examine our argument focusing on new automobile dealers. Most new auto dealers carry a particular automobile brand (e.g., a Ford dealer, a Toyota dealer, etc.) under a contract with that automobile manufacturer which places restrictions on the dealer such as limiting the assortment of new automobile brands that the dealership can carry, typically only a single brand for one of the major manufacturers. The resulting narrow assortment (in terms of brands) offered by any new automobile dealership provides some differentiation, and the theoretical incentive for these dealers to agglomerate with others carrying different brands. However, to avoid the intense price competition that would likely occur if dealerships for the same brand were to locate near one another, dealerships for any particular brand have an incentive to avoid one another. The incentive for a particular dealer to locally agglomerate

<table>
<thead>
<tr>
<th>Retail Category</th>
<th>Calgary</th>
<th>Vancouver</th>
</tr>
</thead>
<tbody>
<tr>
<td>Music CDs and Tapes</td>
<td>0.1133</td>
<td>0.0261</td>
</tr>
<tr>
<td>Fabric</td>
<td>0.1116</td>
<td>0.0235</td>
</tr>
</tbody>
</table>
with dealers for different brands, while avoiding dealerships carrying the same brand, could
well result in no duplication of dealers within clusters. At the same time, each cluster has a
very similar overall composition, so that each dealer within any cluster has an incentive to
strongly avoid other clusters. Thus, contractual product assortment limitations could lead
to the conditions that result in the auto mall pattern. On top of this, manufacturers may
provide some spatial exclusivity to dealerships. This would amplify the tendency for dealers
of the same brand to avoid each other, and hence for the clusters to avoid each other.

While it is beyond the scope of this paper to examine whether the auto mall effect would
emerge given these countervailing forces on new automobile dealerships, we can examine
whether the implied detailed structure exists by comparing the agglomeration of all new
car dealers and of specific brands in Vancouver. As indicated above (and as is illustrated
in Figure 16), new automobile dealers (ignoring the dealership brand) do strongly locally
agglomerate within a range of about two kilometers, consistent with our argument that new
automobile dealers have a general incentive to agglomerate. From that distance until around
nine kilometers, dealerships are very sparse, at which point we are just beginning to see the
D-curve turn, indicating dealerships are again being encountered. This pattern is so strong
it can readily be seen in a map of new automobile dealership locations in Vancouver (Figure
17).

What remains to be determined is the extent to which dealers for the same brand spatially
avoid one another. To examine this, we again make use of Kulldorff’s $D$ statistic, however,
we define the set of cases and controls differently from our previous analysis. Specifically,
for this analysis, the relevant control population is the set of new automobile dealers in
Vancouver, with the cases being dealerships for a particular brand. The question this allows
us to address is whether the dealers for a particular brand agglomerate or avoid one another
relative to automobile dealerships in general.

We examine this question for the four brands with the greatest number of dealerships
in Vancouver, Chrysler (15 dealers), Honda (14 dealers), Ford (14 dealers), and Toyota (13
Figure 16. Kulldorff’s $D$ in Vancouver for New Automobile Dealers

![Kulldorff’s D Statistic](image)

Figure 17. The Locations of New Automobile Dealers in Vancouver

![Map of Vancouver](image)
dealers). The Kulldorff’s $D$ analysis for each of the four new automobile brands is presented in Figure 18. Inspection of this figure reveals that dealerships for all four of these brands strongly avoid one another up to about eight kilometers where the D-curve starts to curve upward, consistent with our argument above. The conditions that we propose as being conducive to the emergence of the auto mall pattern appear to be at work in the Vancouver area.

**Weddings, funerals, and the importance of comparison shopping for homogeneous agglomeration**

Theoretical models suggest that comparison shopping is one important criterion for homogeneous agglomeration to be an equilibrium outcome. Many stores which cluster, such as automobiles and furniture stores, have somewhat differentiated products and involve large and infrequent expenditures which inspire comparison shopping. A question which we now pose: Are these product criteria enough to ensure comparison shopping and clustering? Bridal shops, for example, meet these criteria, and, as we have seen, also cluster. Weddings have some similarities to funerals, in that they are expensive, high involvement, and infrequent. Do funeral homes therefore also cluster? From a behavioral standpoint, however, we would predict that the circumstances surrounding the two are such that the ability and willingness to comparison shop funeral homes is substantially lower than the ability and willingness to comparison shop bridal stores, and hence funeral homes should be less willing to agglomerate. While funeral homes are not considered retailers, to test this hypothesis we collected data on funeral home locations and calculated Kulldorff’s $D$, again using all retailers as the controls. The results (Figure 19) support the hypothesis. Funeral homes strongly avoid each other in Vancouver, and at least do not agglomerate in Calgary.\(^8\) While both bridal shop and funeral home customers incur large and infrequent expenditures, and

\(^8\) In Calgary there are supply-side resource constraints that come into play associated with the comparative concentration of cemeteries relative to Vancouver.
Figure 18. Kulldorff's $D$ Statistic for Automobile brands in Vancouver
are likely to involve a great deal of uncertainty about prices and quality, the willingness to comparison shop is lower for funeral homes dominating the category’s spatial structure.

**Conclusions and Future Research**

We conduct a comprehensive examination of the spatial structure of a set of 54 different types of retail stores within a complete census of retailers in two cities. While much theoretical work has been done to determine equilibrium configurations, and to identify the drivers of homogeneous agglomeration, empirical work presents difficulties and has been limited to date. These difficulties include collecting, geocoding, and validating large address data sets, difficulties in separating spatial structure that arises from demand density from structure that arises from attraction or avoidance, and the use of scalar global measures that mask interesting detail. We have taken advantage of new databases and technologies to work with over 26,000 geographically located retail outlets, introduced a means of isolating the spatial structure relevant to homogeneous agglomeration, and introduced an information-rich vector measure of structure which produces visually intuitive plots.
We find that 21 of the 54 retail store types show significant agglomeration tendencies in both cities. Stores in this group which previous researchers have also found to cluster, and which our judges panel identified as likely to cluster based on theory, are antiques, furniture, clothing stores, and carpets. Additional stores that cluster that were also identified by our judges panel as consistent with theory are new autos, used autos, boat dealers, gifts, art galleries, cosmetics, shoes, bridal shops, and kitchen cabinets. Stores that agglomerate that are a bit more puzzling are coffee shops, bookstores, and smoking supplies. The latter contradicts Rogers (1965) study in Stockholm, which found that tobacconists were second only to liquor stores in avoidance.

Gasoline stations and convenience stores are the most frequently studied retail types in the agglomeration literature. Theory predicts that gasoline stations and convenience stores should avoid each other, but numerous studies have found the opposite, most likely from failing to control for demand density. We find that gasoline stations unambiguously avoid each other in both cities, but convenience stores only do so in Calgary. We also confirm Rogers (1965) finding that liquor stores avoid each other. The judges agreed on convenience stores and liquor stores, but not on gasoline stations. They also expected, and we found in both cities avoidance for health and diet shops and pizza parlors. Expected avoidance that we found in one city were for ice cream parlors, convenience stores, and pharmacies. Retail classes where avoidance exists in both cities, but unexpected by the panel, are supermarkets, pets and pet supplies, and paint.

Overall, many of the retail types we investigated show spatial patterns consistent with theoretical expectations. Discrepancies from those expectations raise interesting questions about the forces that drive homogeneous agglomeration or avoidance, and we use our data and methodology to explore three of these. First, cannibalization avoidance is a strong moderator, as increasing concentration of retail brands decreases the tendency to agglomerate. Second, while product characteristics that lead to infrequent, uncertain, and expensive purchases indicate likely comparison shopping and store agglomeration, such product char-
acteristics can be insufficient to create agglomeration. The difference between the spatial structure of funeral homes and bridal shops, which have a very different shopping circumstances, but are similarly expensive, have high levels of uncertainty, and are infrequent product purchases show that the context can dominate the product characteristics. Third, we proposed that the unusual pattern of local clustering and large scale avoidance dubbed the “auto mall” pattern arises from institutional arrangements with manufacturers. The limitation on the number of brands a dealer may carry should lead to a willingness to be close to dealers of different brands, but far away from same-brand dealers, which in turn could plausibly lead to the pattern. In addition, any exclusivity arrangements would amplify the distant avoidance effect. We then show that individual brands do indeed avoid each other strongly up to the scale of the cluster separation.

This research leads to new research questions of both a theoretical and empirical nature. Two theoretical research areas relate to the “auto mall” pattern and the effect of retail brand concentration on the spatial distribution of retailers. In terms of the auto mall pattern, we conjecture that this is a result of the exclusive nature of the distribution arrangements between retailers and manufacturers, however, we do not formally show this. As a result, it would be worthwhile developing a theoretical model that sheds light on whether, and under what conditions, manufacturers’ distribution strategies influence the spatial distribution of retailers. Our empirical results also suggest that retail brand concentration results in less retail agglomeration than would otherwise occur. Again, this is a conjecture on our part, and a theoretical analysis of whether, and under what conditions, retail brand concentration causes this to occur (and what the implications are for both consumers and other retailers) should be explored.

Areas for future empirical research include developing measures of the underlying attraction and avoidance forces and incorporating them into the analysis. These would include not only comparison shopping, but other behaviors such as shopping for entertainment, as well as supply side factors, and integrating them into the location data. The use of a vector rather
than scalar measure of clustering allows uncovering of interesting patterns that suggest that overall attraction and avoidance forces change, sometimes dramatically (as in the auto mall pattern) with store separation. Can we predict more of these variations from institutional considerations?

More detailed examination of the stores themselves (both theoretically and empirically) is obviously warranted. The effect of store size (measured by floor area or revenues) on outlet spatial structure is an area worthy of additional attention. One interesting question is whether large category killers are substitutes for homogeneous agglomerations. On the other hand, casual observation shows that there are instances where larger specialty stores such as outdoor shops appear to create a basin around which other similar stores collect, so that the “category killer” becomes a “category attractor.” What is the difference between the two, and why? The unexpected result that coffee shops consistently hyper agglomerate may be related to this, and begs further investigation.

Finally, the pet store owners we met in the paper’s introduction can see that pets and pet supply stores consistently avoid each other (the possible level of differentiation may not be great enough to warrant the clustering of stores in this category). The owner’s first concern should be whether or not this pattern reflects a true competitive advantage of avoidance over clustering, or is caused by a high concentration of same brand stores. In Vancouver, of the 138 stores in this category, the largest chain has only 18 stores, so cannibalization avoidance is not the likely explanation. We recommend that the owners avoid other pets and pet supply stores in locating their new outlet.

References


