A Dissertation

entitled

The Influence of Competition on Retail Firm Location: Theory and Measurement

by

Isabelle M. Nilsson

Submitted to the Graduate Faculty as partial fulfillment of the requirements for the Doctor of Philosophy Degree in Spatially Integrated Social Science

Dr. Oleg A. Smirnov, Committee Chair

Dr. Kristen Keith, Committee Member

Dr. Peter S. Lindquist, Committee Member

Dr. Neil Reid, Committee Member

Dr. Linda Wårell, Committee Member

Dr. Patricia R. Komuniecki, Dean
College of Graduate Studies

The University of Toledo
May 2015

An Abstract of

The Influence of Competition on Retail Firm Location: Theory and Measurement

by

Isabelle M. Nilsson

Submitted to the Graduate Faculty as partial fulfillment of the requirements for the Doctor of Philosophy Degree in Spatially Integrated Social Science

The University of Toledo
May 2015

Store location is regarded as a primary driver of retail competition and one of the most important factors in a consumer's store choice decision. In order to explain differences in the relative location between competing retailers a game theoretic model is applied to the study of firm location choices under varying degrees of strategic interaction. The theoretical results contribute to the existing literature by explaining why certain types of retailers may want to locate in close proximity to their competition despite increased price competition. While theoretical results are dependent on the assumptions made, realized outcomes of firms' location decisions provide insights regarding underlying location behaviors.

For this purpose a multivariate spatial statistic is developed aimed at identifying different interaction patterns between competing stores. In order to define when two outlets are located relatively close to each other, a topological proximity criterion is derived based on the theoretical framework. Multivariate spatial statistic has received relatively little attention in the literature when compared to univariate methods. Therefore the statistic developed in this thesis not only proposes a new methodology aimed at distinguishing between different firm location behaviors, it also contributes to the existing literature on multivariate point pattern methods.

The proposed statistic has asymptotic properties, and its distribution is approximately normal for larger samples. Tests of finite sample properties and robustness

checks are conducted through simulations by varying assumptions with regards to population size, expected values, as well as interactions between categories of points and underlying spatial processes. The simulation results confirm that the proposed statistic has the ability to capture not only asymmetrical relationship but also distinguish pairwise categorical associations from clustering in the joint population.

To demonstrate the usefulness of the statistic, it is applied to competing stores in two different retail sectors in order to detect any differences in the relative location between competitors. The application of the statistic to real world location data shows that observed patterns are effectively captured by the proposed statistic and easily interpreted using presented theory. This dissertation also examines whether access to transportation infrastructure may induce retailers to locate in close proximity to one another. For this analysis the statistic is applied to measure whether there is a significant difference in interaction patterns between competing outlets located in proximity to important transportation infrastructure and those that are not. The results suggest that transportation infrastructure affects the nature of the location behavior of retail firms with respect to the relative location of their competitors.

The findings from this dissertation suggest that public policy has different welfare effects depending on the demand conditions in different retail sectors. Therefore broader impacts arise from establishing theoretical and empirical evidence that not only improves our understanding of difference in relative location between competitors in different retail sectors but also clarifies the consequences of public policy.

For Jeff.

Acknowledgments

This project was completed with the support and partnership of the National Center for Freight and Infrastructure Research and Education (CFIRE), a Tier 1 University Transportation Center (UTC) funded by the US Department of Transportation (USDOT) Research and Innovative Technology Administration (RITA).

I would like to thank Oleg A. Smirnov for his support and guidance throughout this exciting research. Furthermore, I would like to express my appreciation towards Peter S. Lindquist for providing me with many exciting research opportunities throughout these years. The work and my professional development has benefitted a great deal from many hours of their time. I would also like to thank my other committee members for their thoughtful input and assistance throughout this process.

Contents

List of Tables

List of Figures

Chapter 1

Introduction and Summary

Retailers strive to find sites having great demand potential, which can be characterized by being located in areas with high accessibility, traffic flow and representing a sound investment in terms of real estate value. The focus on prime locations often leads to clustering of retailers. Another factor that influence retailers' location decision is the location of competing outlets (Ghosh and McLafferty, 1987). Store location is regarded as a primary driver of retail competition and one of the most important factors in a consumer's store choice decision (Zhu and Singh, 2009). The theoretical literature offers a variety of explanations to why stores may strive to either locate in close proximity or distance themselves in relation to their competitors (Anderson and de Palma, 1988; d'Aspremont et al., 1979; De Fraja and Norman, 1993; Dudey, 1990; Hotelling, 1929; Konishi, 2005). The effect that competing retailers may have on each others location choices can be summarized via two opposing forces: a market power effect and a market share effect Netz and Taylor (2002). According to the market power effect, retailers selling similar goods strive to locate far away from each other in order to establish local monopoly power. However, firms also have an incentive to move towards each other in order to increase their market share (the market share effect). Most of the theoretical literature on retailer location choice has found evidence that support the "market power effect". In this dissertation a theoretical framework is developed that contributes to the literature by explaining frequently observed attraction patterns among closely competing firms, *i.e.* when the market share effect dominates the market power effect. However, theoretical results are dependent upon the assumptions made and what forces dominate the decision making process may differ between retail sectors.

The empirical literature is small and most studies tend to focus on clustering (or dispersion) of firms within the same retail sector (Fischer and Harrington, 1996; Krider and Putler, 2013; Picone et al., 2009). As demand and/or supply side benefits of certain locations may induce retail firms to cluster within the same area, a more interesting question is related to how competing outlets locate in relation to each other within these high density areas. More specifically, do specific pairs of competing firms (brands) tend to locate in close proximity to each other given the pattern of the joint population? Being able to measure such tendences can help distinguish between different firm location behaviors as they relate to those scenarios outlined in the theoretical literature.

For example, consider the two markets in Figure 1-1, which displays the location of fast food restaurants in Indianapolis, Indiana (Figure 1-1(a)) and "big-box" discount stores in Pittsburgh, Pennsylvania (Figure 1-1(b)). At a smaller geographic scale, both markets appear to be clustered around a single source of demand. Attraction to a common source of demand leads to clustering with the density of competing stores in both markets coinciding in space. On the other hand, at a larger geographic scale, both markets appear dispersed with no visible signs of clustering of competing outlets. However, the relative location between competing stores within each market is different. While the map of fast food restaurants (Figure 1-1(a)) displays tendencies of competing stores located in close proximity to each other in pairs or triples. Such a pattern is not observed in the map of "big-box" discount stores (Figure 1-1(b)) which shows a more evenly distributed pattern with two instances where competing stores are located in close proximity to each other.

This dissertation is motivated by this empirical phenomenon and its purpose is to explain the economic forces behind it and develop a practical method for measuring it. First, a game theoretic model is developed that explains the economic forces behind different relative location patterns of retail firms under varying assumptions of product differentiation and the policy implications associated with such. While clusters of firms arise to some extent because of demand benefits of a central location, strategic competition over consumers also tends to bring firms closer geographically (Dudey, 1990; Hotelling, 1929; Marshall, 1920; Schuetz, 2014). Strategic interactions,[1] or competition over consumers, occur when firms sell similar products. Therefore, to explain the differences between different retail firm location patterns a game theoretic model is applied to study firm location choices under varying cross-price elasticity of demand. Under varying assumptions of product differentiation, this framework enables an in-depth examination of strategic interactions among retail firms in the spatial economy.

The model suggests that firms locate in close proximity to each other for different reasons which can be categorized in two ways. The first arises due to demand benefits of being located in the center of the market and is mostly independent of behavior of competing firms. The second occurs from strategic interactions between firms competing over market share. In the presence of strategic interactions, the equilibrium is characterized by a non-cooperative Nash-equilibrium where by both firms and consumers would be better off if firms were to locate more uniformly throughout space. Theoretically, this suggests that under certain demand conditions public policy can have a welfare-improving effect on firm location choices. In the presence of strategic interactions, regulations (such as zoning laws) could force firms to move away from the

[1]By strategic interactions I mean that each firm's payoff (profit) depends not only on its own actions (choice of location and price) but also on the actions of the other firm.

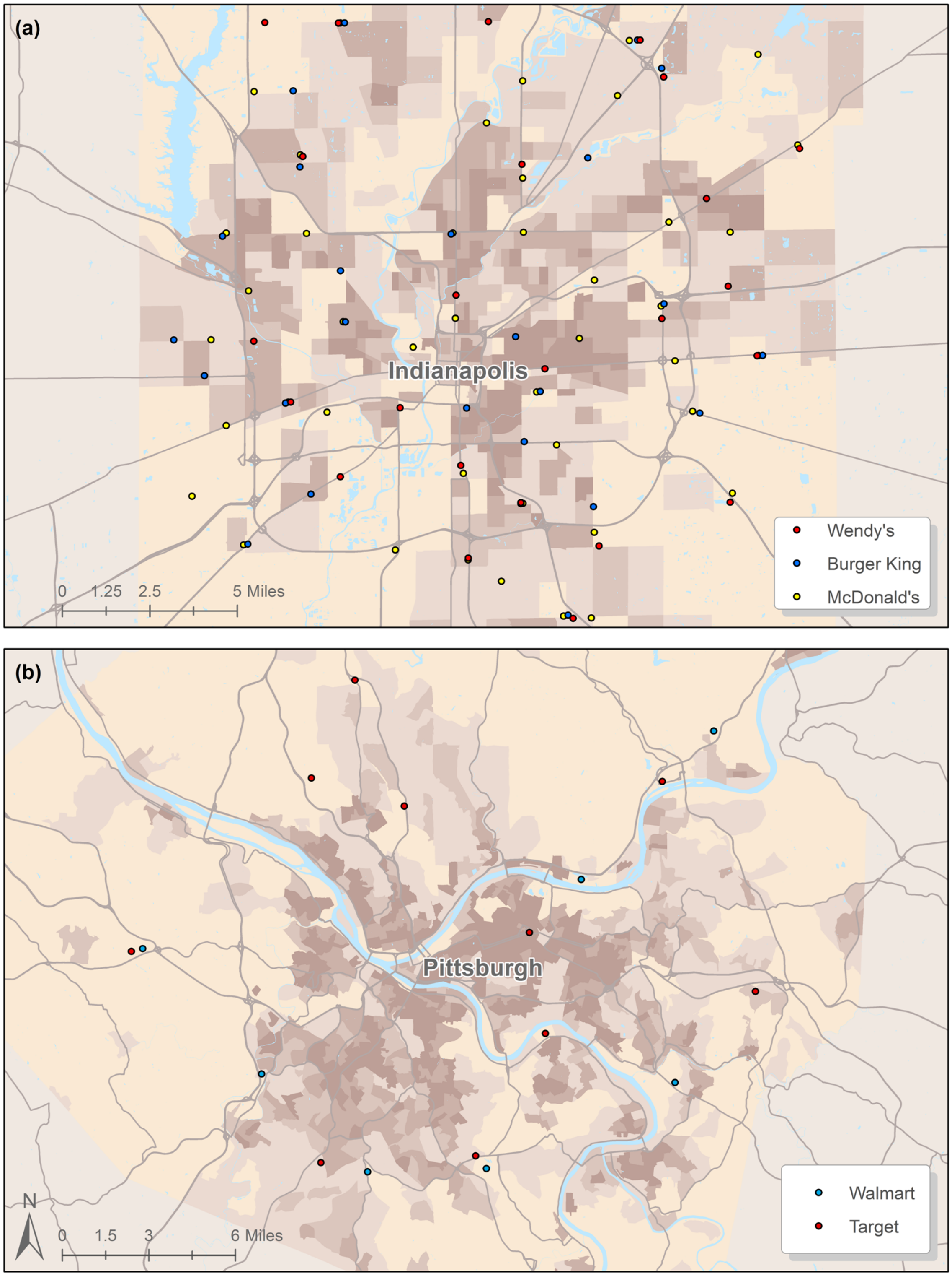


Figure 1-1: Location Patterns of Firms in Selected Markets

Nash-equilibrium location to the socially optimal location where firms and consumers alike are better off. It also has implications for transportation planning. Through the even distribution of such firms, special arrangements and costs associated with congestion could be mitigated and the impact on roads (or the demand for public infrastructure, in general) more evenly distributed.

Second, in order to identify different location patterns which arise from realized location choices, a multivariate spatial statistic is developed. The statistic measures whether specific pairs of categories of points (brands of firms) tend to be located relatively close to each other. While existing measures for detecting multivariate (or categorical) spatial association in point data can detect such tendencies under certain conditions, these measures are not concerned with measuring the same type of location patterns examined in this study. There are two commonly used types of multivariate spatial point pattern methods; the first is concerned with measuring whether the number of points in one category is higher than expected within a given distance, d, of points from another category. The second class of methods are concerned with whether specific pairs of categories tend to be nearest neighbors more often than is to be expected. However, in order to define when two outlets are located relatively close to each other, a different conceptualization of "near" is needed as two stores being nearest neighbors does not necessarily mean that they are located near each other and the same distance between two stores can mean different things in different locations (*e.g.*, urban vs. rural or car centric vs. walkable parts of a city). For this purpose a topological definition of "relatively close" is presented which takes into account the relative spatial separation between outlets and heterogenity in the pattern of the joint population. As such, this research contributes to the existing literature on multivariate spatial association by introducing a new statistic that measures a new phenomenon: positive (attraction) and negative (avoidance) categorical assocation in terms of whether categories of points tend to be located relatively close to each other given the pattern of the joint population. To demonstrate the robustness of the statistic, it is applied to a variety of simulated bivariate point patterns with different interaction structures.

Third, to demonstrate the usefulness of the statistic it is applied to location patterns of competing firms in two different retail sectors. The two sectors are fast food restaurants and "big-box" discount stores. While all three fast food chains included in this study target the same customer segment with similar prices and product offerings, the two "big-box" discount store chains included in the analysis target two differing customer segments. The results suggest that while there are strong attraction tendencies among the fast food chains, the two "'big-box" retailers show no significant tendencies of attraction or avoidance. The results are interpreted in terms of firm location behaviors described in the theoretical literature. The results for these two retail sectors confirm the findings from the theoretical framework developed in this dissertation.

Finally, the effect of transportation infrastructure on retail firm location patterns is analyzed in a state of equilibrium with a focus on the location behavior of competing outlets. The spatial distribution of firms is a primary determinant of the demand for transportation infrastructure. What complicates the analysis is the inherent endogeneity of this relationship where changes in the transportation system also affect the distribution of firm location patterns. Not only does transportation infrastructure attract firms but it also changes the nature of their location behavior with respect to their competition, which in turn has further consequences for travel demand.

There are two aspects to this analysis. First, the game theoretic framework is used to describe the connection between relative firm location and transportation demand. Second, the multivariate spatial statistic is applied to observed, equilibrium firm location patterns to measure the effect transportation infrastructure has on firm location behavior. The results suggest that transportation infrastructure affects the location behavior of retail firms with respect to their location in relation to their competitors. Access to important transportation infrastructure induces competing firms to locate next to each other – a tendency not necessarily observed among firms without such access. This does not only affect economic outcomes, but also travel demand and consumer accessibility.

This dissertation is organized as follows: Chapter 1 provides an introduction and summary of the dissertation. Chapter 2 gives an overview of related literature. Chapter 3 provides the set up of the theoretical framework together with derived theoretical results for retail firm location behavior, prices and welfare implications in a state of equilibrium for various cross-price elasticities of demand. Chapter 4 provides a definition of the proposed multivariate spatial statistic and its properties together with an application to simulated data. In Chapter 5 and 6 the proposed theoretical framework and statistic is applied to analyze competing retail firms location patterns in two retail sectors. Chapter 7 provides a comparison of the relative location between competing firms in the presence and absence of access to major road transportation infrastructure. Chapter 8 summarizes the findings from this dissertation and outlines future research needs.

Chapter 2

Literature Review

2.1 Retail Firm Location and Competition

Retailers strive to find sites having great demand potential, which can be characterized by being located in areas with high accessibility, traffic flow and representing a sound investment in terms of real estate value. Due to the focus on high potential locations, clustering often occurs. Another factor influencing retailers' location decision is the location of competitors (Ghosh and McLafferty, 1987). While clusters of retailers from the same sector arise to some extent because of demand benefits of a central location, strategic competition over consumers also tends to bring firms closer geographically (Ago, 2008; Anderson et al., 2013; Dudey, 1990; Hotelling, 1929; Klier and McMillen, 2008; Marshall, 1920; Schuetz, 2014). Space gives competiton a particular form. Since consumers patronize the firm with the lowest price (including transportation cost) when firms sell similar products, each firm only competes directly with a few neighboring firms regardless of the total number of firms in the industry. Therefore, the nature of spatial competiton is oligopolistic and should be studied within a game theoretic framework (Fujita and Thisse, 1995).

The spatial competition literature begins with Hotelling (1929)'s seminal work in which he studied the strategies of two competitors selling a homogenous product with respect to their location and price in a linear market with a uniformly distributed demand. His main conclusion was that each firm would always have an incentive to move towards the other. This tendency is usually referred to as the Principle of Minimum Differentiation. The general conclusion within the spatial competition literature following Hotelling (1929) is that firms with differentiated products tend to cluster, while firms selling close substitutes tend to disperse (Anderson and de Palma, 1988; d'Aspremont et al., 1979; De Fraja and Norman, 1993; De Palma et al., 1985; Irmen and Thisse, 1998). The rationale behind these findings is that differentiation in some other dimension is sufficient to reduce price competition so that the agglomerating equilibrium can be sustained. Even though these results are economically appealing, forces exist that oppose complete dispersion of firms in the spatial economy. Although firms may strive to spatially differentiate themselves for strategic purposes, this literature does not derive any intuition as to why such strategic incentives dominate those of being close to the demand (Tirole, 1988).

The argument for complete dispersion is that firms could always increase profits by moving away from each other in order to reduce price competition. However, firms also have an incentive to move towards the other in order to increase demand and thereby profits. This is to say there are two opposing forces at play – a market power effect and a market share effect (Netz and Taylor, 2002). The question becomes which of these effects dominate. De Fraja and Norman (1993) provide some insights as to why firms choose to locate in close proximity to each other even though this may result in fierce price competition. De Fraja and Norman (1993) argue that under spatial competiton there are additional forces at work.[1] In contrast to non-spatial oligopolistic competition, given the location of its competitor, when a firm moves away from the market center the total amount it sells at any given price decreases.

[1]For an overview of the differences between spatial and non-spatial competition, and why many of the findings in classical price theory are reversed in a spatial context, the reader is referred to Capozza and Order (1978).

Therefore, the firm has an incentive to lower its price in order to compensate for the reduction in quantity and this spatial influence outweighs the competitive effect of more proximal locations. In an empirical analysis, Schuetz (2014) find that "big-box" stores tend to locate in close proximity to their direct competitors. As one of the reasons for these tendencies, Schuetz (2014) hypothesizes that "big-box" stores may prefer to share market areas with a competitor rather than cede large number of consumers.

Externalities that induce firms to locate in close proximity to one another may also exist, as evidenced by common installations and trade centers, transportation infrastructure, *etc* (Miller et al., 1999; Tirole, 1988). This dissertation has an explicit focus on how transportation infrastructure affects the location behavior of competing retailers. While the location and intensity of economic activity is one primary determinant of the demand for transportation infrastructure, changes in the transportation system also affects the distribution of economic activity due to the benefits associated with being located close to important transportation infrastructure (Elgar et al., 2009; Forkenbrock, 2002; Hicks, 2006; Maoh and Kanaroglou, 2009; Targa et al., 2006). For example, gasoline stations strive to find sites on high-traffic routes that are accessible to a large number of motorists (Ghosh and McLafferty, 1987). While there have been considerable efforts focused on examining the relationship between access to transportation infrastructure and agglomerating economies or economic development (Hicks, 2006; Targa et al., 2006), the effect of transportation infrastructure on the location choices made by competiting firms has not received much attention. Therefore the analysis performed in Chapter 5 is intended to examine the effect that transportation infrastructure has on the location choices of competing retailers in relation to each other.

Consumers' search for products is another type of externality that may provide incentives for firms to cluster. Search-based models that examine demand benefits of agglomerations, tend to support a clustering equilibrium under competition with close substitutes. This literature follows a similar argument to the one made by Marshall (1920): that firms can get additional consumer traffic by locating close to each other, which can outweigh the negative effects associated with more intense competition. Alternatively, citing Nystrom (1930) "Known competition in itself attracts trade..." (p. 138). Wolinsky (1983), Dudey (1990), Bester (1998) and Konishi (2005) suggest that firms selling similar products may choose head-on competition by clustering together in order to attract consumers by facilitating price and quality comparison. Yang (2012) provides other reasons for why fast food chain stores may end up locating in close proximity to each other in terms of crowding (that the presence of rivals may draw larger crowds), learning (due to uncertainty, chains may have an incentive to learn from experienced incumbents) and free-riding off market research.

2.2 Multivariate Spatial Statistics for Point Data

While the theoretical literature associated with retail firm location choice is large, the results are highly dependent upon the assumptions made (Anderson and de Palma, 1988; d'Aspremont et al., 1979; De Fraja and Norman, 1993; Dudey, 1990; Hotelling, 1929; Irmen and Thisse, 1998; Konishi, 2005). Therefore, as mentioned by Krider and Putler (2013), empirical research could help shed light on realized location choices made by retailers in different sectors. The empirical literature, however, is limited and few studies measure whether firms tend to be attracted to each other [or not] (Krider and Putler, 2013; Netz and Taylor, 2002; Pinkse and Slade, 1998). Ripley's K-function (Ripley, 1977) and its normalized version, the L-function, are two measures that have been used for this purpose (Picone et al., 2009). The nearest neighbor statistic has also been used to detect clustering among firms (Fischer and Harrington, 1996; Krider and Putler, 2013; Picone et al., 2009). However, these are both commonly used measures for analyzing univariate spatial point patterns, meaning that multivariate (or categorical) spatial association between two or more populations (or different categories within a single population) cannot be detected using such measures. For the purpose of measuring spatial association between two or more types (or categories) of points, a multivariate point pattern method is needed.

While most research has focused on univariate spatial point patterns, multivariate (or categorical) spatial association has received less attention (Ceyhan, 2010; Cressie, 1991; Goreaud and Pelissier, 2003; Kulldorff, 2006). The most commonly adopted measure for detecting bivariate spatial association in point data is the cross K-function (Lotwick and Silverman, 1982) which is an extension of the univariate K-function for bivariate spatial point patterns. Other multivariate point pattern methods include nearest-neighbor contingency tables (NNCT) (Dixon, 1994; Pielou, 1961), the co-location quotient (CLQ) (Cromley et al., 2014; Leslie and Kronenfeld, 2011), comparison of univariate K-functions and extensions and modifications of the cross K-function (Besag, 1977; Diggle and Chetwynd, 1991; Duranton and Overman, 2005; Marcon and Pueach, 2010; Okabe and Yamada, 2001). However, these measures do not necessarily measure the same type of phenomenon. Therefore, we will distinguish between two types of multivariate spatial point pattern methods. First, there are methods that measure whether the density of points in one category are higher than expected in the vicinity of points from another category (the cross K-function and variations of it). Second, there are methods that aim at measuring whether specific pairs of categories tend to be nearest neighbors (the CLQ and NNCT).[2]

In this dissertation the goal is to measure whether specific pairs of categories tend

[2]NNCT measures similar tendencies as the joint count statistic (Dacey, 1968) but for point data which is spatial autocorrelation. That is whether the nearest neighbors of category A include other category A points more frequently than expected (*i.e.* positive spatial autocorrelation) or if category A tends to have category B points as its nearest neighbor (*i.e.* negative spatial autocorrelation). Similarily the CLQ measures the ratio of observed to expected proportions of one category among another category's nearest neighbors. Although similar to the joint count statistic, the CLQ uses nearest neighbor counts instead of pairwise joins.

to be located *relatively close* each another. Therefore, even if the proposed measure is similar to the second kind of methods, these measures are not able (and it is not necessarily their purpose) to detect the kind of categorical association described here as these measures do not account for the relative spatial separation between points. For example, in Figure 2-1(d), each category has another category as its nearest neighbor but that does not mean that they are attracted to each other. If the nearest neighboring firm of another category is miles away – should the two firms still be considered attracted to each other? In short, the fact that two points are nearest neighbors does not necessarily imply that they are located near each other.

The measure presented in this dissertation can be likened to the cross-K function in that it measures if the share of points in one category is higher than expected in the vicinity of points from another category. However, while the cross-K function measures these tendencies at differing scales (at different fixed distance threshold), this neighbor criteria does not vary throughout the study area (it does not take into account heterogeneity in the spatial structure). As we are concerned with conceptualizing the notion of when two points may be considered located relatively near each other (regardless of variations in the intensity of points throughout the study area) and especially with regard to same category points, a distance threshold will not suffice. The main reason is that the same distance between two firms might mean different things in different locations of the study area. For example, while two firms half-a-mile apart in a rural or suburban area of a city may be considered neighbors or located relatively close, two firms half-a-mile apart in the innercity or a commerical business district would not.[3] In order to define *relatively close* (or near) in the presence of heterogeneity in the pattern of the joint population (such as illustrated in Figure 2-1(e)), a criterion is needed that adjusts itself to variations in the joint

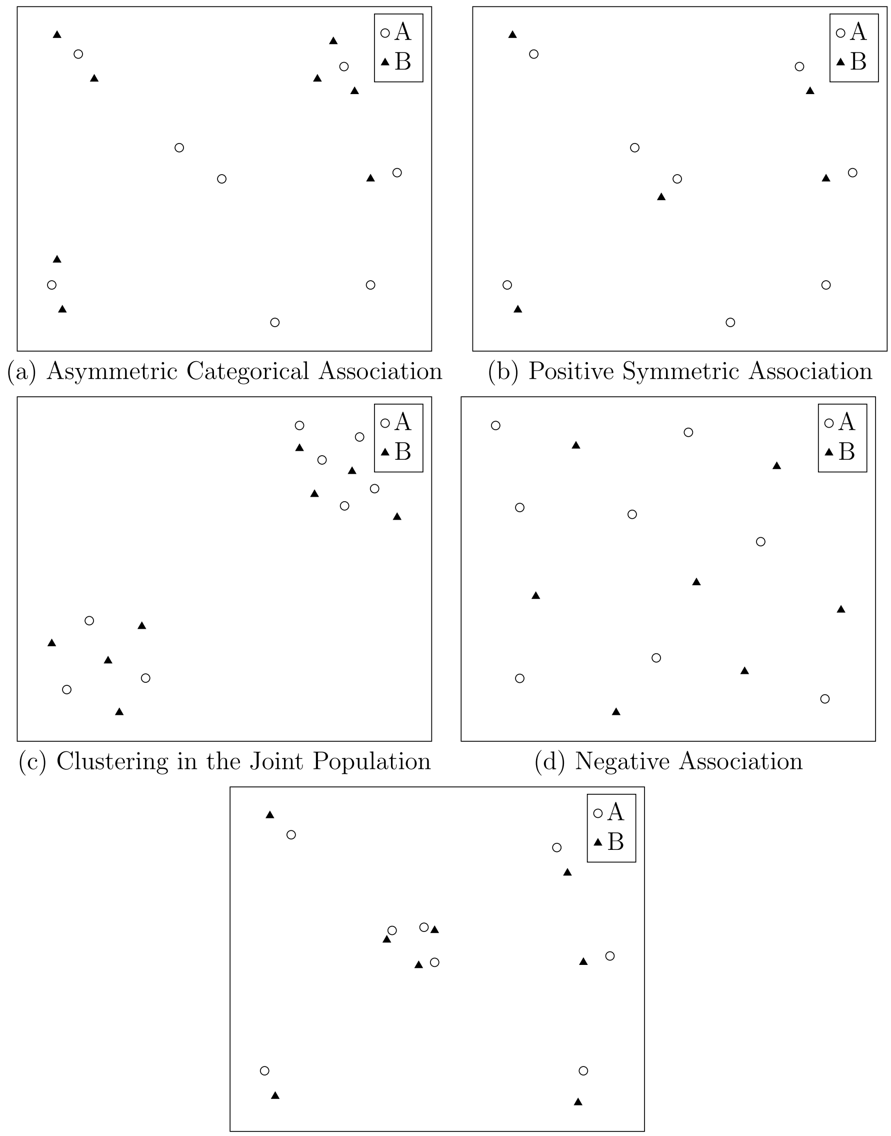
[3]Whether a city or parts of a city is car centric versus walkable may also effect what distance defines if two stores should be considered near each other [or not].

pattern thoughout the study area.[4] Therefore, a new statistic is proposed where relatively close (or near) is defined in terms of topology.

One important feature of the proposed measure is the ability to distinguish between whether positive association (attraction) between two categories is due to clustering in the joint population (*e.g.*, overall concentration of economic activity within a city or business district, such as in Figure 2-1(c)) or if specific pair categories of points tend to locate near each other given clustering in the joint population (*e.g.*, the tendency for two types of firms to be located next to each other within a commerical district, such as in Figure 2-1(b)). Failure to separate pairwise positive association between categories of points from clustering in the joint population can result in spurious findings (Leslie and Kronenfeld, 2011). For example, even though the density of category A and B points coincide at the same place (such as in Figure 2-1(c)), this does not necessarily imply that there is any attraction between specific pairs of these categories within those high density areas. A zoomed-in image on either high density area in Figure 2-1(c) could display a pattern such as in Figure 2-1(d) with (what will here be defined as) negative categorical association (avoidance).

Measures such as the cross K-function have a tendency to find many pairwise positive associations due to clustering in the joint population (Leslie and Kronenfeld, 2011). While the CLQ attempts to distinguish between positive association between pairs of categorical subsets and clustering in the joint population, the alternative hy-

[4]For areal data the solution is to use continuity (topological) neighbor criterion, where neighbor relations are independent of scale and inherent differences between locations. One way of taking advantage of this more robust neighbor criteria is to construct Thiessen (or Voronoi) polygons around each point. However, converting the whole population of points into polygons is equivalent of transforming a multivariate pattern into a univariate pattern where both different and same brand firms will share boarders. Such a simplification implies a loss of information as it will not take into account the relative spatial separation between firms similarly to the nearest neighbor criteria. For areal data, spatial concentration measures such as the Gini coefficient, the location quotient, and the Ellison-Glaeser index (Ellison and Glaeser, 1997) are commonly used. However, these measures does not take into account spatial relations and suffer from other shortcomings (for a more detailed discussion see Duranton and Overman (2005) and Brachert et al. (2011)). Then there are measures of spatial autocorrelation, such as the Moran's I, Geary's C and the joint count statistic (Aldstadt, 2010).



(a) Asymmetric Categorical Association

(b) Positive Symmetric Association

(c) Clustering in the Joint Population

(d) Negative Association

(e) Positive Symmetric Association

Figure 2-1: Examples of Multivariate Spatial Point Patterns.

pothesis of the CLQ and NNCT is different from that of the cross-K and the statistic in this paper. The CLQ and NNCT measures whether category A has more (or less, meaning more A than B) nearest neighbors of category B than what would be expected (given by the share of B points in the joint population), meaning that for all patterns in Figure 2-1 (expect 2-1(a)) this would be the case. While it is not affected by spatial heterogeneity in the joint population, it does not account for the relative distance between points. The cross-K function on the other hand test whether the density of category B points around category A points at a distance d is higher than expected in the case of two independent patterns. First, it may at certain distances indicate positive categorical association being present in Figure 2-1(c). Second, given the fixed distance threshold it may not be able to detect the positive categorical association illustrated in Figure 2-1(e) as it occurs at different distances simultaneously. Another important feature is the ability to detect asymmetrical relationships (*e.g.*, even though B is attracted to A, A is not necessarily attracted to B such as in Figure 2-1(a)) which the cross-K function, NNCT and the CLQ all allows for.

As the theoretical distributions of the most commonly used multivariate measures are unknown, confidence intervals are commonly estimated through Monte Carlo simulations of a specified null hypothesis (Goreaud and Pelissier, 2003). Monte Carlo based hypothesis testing is a widely accepted method in various fields (Ceyhan, 2010; Dixon, 2002c; Kulldorff, 2006; Lotwick and Silverman, 1982; Marcon and Pueach, 2010; Ripley, 1977). With this approach, random data are generated under the null hypothesis conditional on the total number of observed points. After choosing a null hypothesis, the test statistic is calculated for the observed dataset and a large number (R) of randomly generated datasets. The null is rejected at the α significance level if the test statistic calculated from the observed dataset is among the highest $\alpha \times R$ test statistics among all datasets (Kulldorff, 2006). The establishment of study-specific critical values through the use of Monte Carlo simulations has also been mentioned as a way of dealing with edge effects in the estimation of nearest neighbor statistics (Dixon, 2002c).

To interpret the spatial interaction between two categories of points, there are two commonly used null hyotheses: independence or random labeling (RL). As shown by Goreaud and Pelissier (2003) an inappropriate choice of the null hypothesis can lead to misinterpretation of the results as these two null hypothesis correspond to different confidence intervals. For the null case of independence the locations of the two categories of points are *a priori* the result of different processes (*e.g.* individuals of different species or age cohorts). The null case of RL is defined as the result of some processes affecting *a posteriori* the individuals of a single population (*e.g.* living vs. dead or diseased vs. non-diseased individuals of a single species) (Ceyhan, 2010; Goreaud and Pelissier, 2003).[5] For the purpose of this study, which is to detect interactions between different brands/chains of retail firms, we chose to the null of independence as firms may have different business models and location decision making processes. Under both null hypotheses the random redistribution of points or labels may produce patterns where same-category points are located next to each other. This would not be a realistic pattern for same-brand firms. Seldom do we observe two same brand stores locate right next to each other. This is another reason why the null hypothesis of independence is chosen as the proposed measure. In order to take into account such tendencies under the null of independence, simulation experiments are performed in Chapter 4.4.1 where relocated A points cannot be within a certain distance of each other.[6]

Another challenge is the determination of market (or study) area. According to

[5]As biological examples Goreaud and Pelissier (2003) mentions between-species or between-cohorts interaction for independence and a disease attack or accidental disturbances within a population.

[6]A more realistic null hypothesis would be to restrict the random relocation of points to appropriately zoned areas in order to account for the fact that firms are not able to locate anywhere. This is left for future research.

Beckmann (1968), a market area extends as far as demand exists. Theoretical work assumes that the spatial extent of a market is well defined (Netz and Taylor, 2002). Beckmann (1999) defines a single sellers market as "...bounded by the distances at which price plus transportation cost causes demand to reach zero" (p. 13). In reality, determining the geographic boundary of a market area is difficult and becomes an empirical question (Netz and Taylor, 2002). Empirically, there are different ways in which the study area can be defined. The most commonly used approach is the "hypothetical monopolist" test (Davis, 2006). However, this test necessitates data on prices which can be difficult to obtain. In order to define a market area in the absence of available price data, alternative approaches have been developed. For example Netz and Taylor (2002) use different market radii of one-half miles, one mile and two miles to define a single establishment's market area. For the purpose of the statistic developed in this thesis, the aim is to determine the market boundary for a certain retail industry or a collection of certain types of firms (*i.e.* not individual establishments) within a formal political geographic area unit (such as a city or metropolitan area). For this purpose many studies use the political boundaries alone. For example, Seim (2001) uses a city's boundaries to describe the total market area, while Picone et al. (2009), Jia (2008) and Thomadsen (2007) define their study area by using county boundaries. The definition of the market area are probably of less weight when dealing with large data sets. However, for moderate and small data sets the potential influence of the definition of market area is likely to be greater.

Chapter 3

Model of Retail Location Choice

Most of the theoretical literature on retailers location choices has found evidence that support the "market power effect", *i.e.* that firms competing with close substitutes strive to locate away from each other in order to avoid fierce price competition. Although firms may strive to spatially differentiate themselves for strategic purposes, this literature does not derive any intuition as to why such strategic incentives dominate those of being close to the demand (Tirole, 1988). That is, why the "market power effect" would dominate the "market share effect" which states that firms also have an incentive to move towards the other in order to increase demand. In this chapter an alternative game theoretic framework is presented which aims at explaining scenarios where the market share effect dominates the market power effect. The findings from this theoretical analysis adds to the existing literature by explaining why certain types of retailers may want to locate in close proximity to their competition despite increased price competition.

First, the setup and assumptions of the game is described. Next, theoretical results are derived for firm location behavior, prices and welfare implications in a state of equilibrium for various cross-price elasticities of demand. This enables an in-depth examination of strategic interactions among firms in the spatial economy under different assumptions related to product differentiation. The motivating factor for this chapter is that firms cluster for different reasons depending on the degree of interaction. As such, location patterns among competitors exhibit different characteristics depending upon the reasons for which firms engage in clustering activity.

The proposed framework make the case that these characteristics can be categorized in two ways: due to demand benefits of being located in the center of the market or if the clustering emerges from strategic interactions between firms. The theoretical results can be generalized into three possible scenarios: clusters with strategic interactions, clusters without strategic interactions and an intermediate case. In the presence of strategic interactions, the equilibrium is characterized by a non-cooperative Nash-equilibrium where both firms and consumers would be better off if firms were to locate more uniformly throughout space. These results are illustrated using simulated data. All derivations are outlined in the Appendix A.

3.1 Setup

There are two products q_i, i={1,2} each produced by a different firm. Firms are located along a linear market of length L (as in Hotelling's 'Main Street' example) at at distances x_1 and x_2 respectively from the ends of this line ($0 < x_1 < L - x_2; x_1 + x_2 \leq L$). Consumers are uniformly distributed along the market, $s \in [0, L]$. Each consumer pays a delivered price, which is the price of the product plus transportation cost. The transportation cost is assumed to be linear with respect to the distance. Firms pay a constant marginal cost of production c per unit. It is assumed that the degree of differentiation, reflected by the cross-price elasticity of demand, is symmetric and given exogenously. There are no barriers to enter the market, and firms act in a non-cooperative fashion with location and price as strategic variables.

3.1.1 Consumer Behavior

A quadratic utility function is adopted in order to distinguish between absolute advantage in demand and cross-price effects:[1]

$$U(q_1, q_2) = \alpha_1 q_1 + \alpha_2 q_2 - \frac{1}{2}(\beta_1 q_1^2 + 2\gamma q_1 q_2 + \beta_2 q_2^2) + m - \sum_{i=1}^{2} P_i q_i \tag{3.1}$$

Where $\alpha_i > 0$, $\beta_i > 0$, $\gamma \geq 0$ are model parameters and m denotes income. The quadratic formulation of the utility function captures two aspects of product differentiation. First, a higher α_i implies an absolute advantage in demand enjoyed by firm i, i={1,2}.[2] Second, γ is a parameterization of cross-price effects. One important assumption is that $\beta^2 \geq \gamma^2$, impling that own-price effects dominate (or in the case of perfect substitutes, equals) cross-price effects. Consumers maximize utility subject to the budget constraint $P_1 q_1 + P_2 q_2 \leq m$ where P_i is the delivered price for products i={1,2}. From the first order conditions the following demand functions are derived for the two products, each produced by a different firm:

$$\begin{aligned} q_1 &= \frac{(\alpha - P_1)\beta - (\alpha - P_2)\gamma}{\beta^2 - \gamma^2} \\ q_2 &= \frac{(\alpha - P_2)\beta - (\alpha - P_1)\gamma}{\beta^2 - \gamma^2} \end{aligned} \tag{3.2}$$

Where the delivered prices are $P_1 = p_1 + t\,|s - x_1|$ and $P_2 = p_2 + t\,|s - (L - x_2)|$ for firm 1 and 2 respectively, p_i is the price of firm i's product and t is the unit

[1]The quadratic utility model is common in the industrial organization literature (Ago, 2008; Dixit, 1979; Picone et al., 2009; Shy, 1995; Singh and Vives, 1984), in demand analysis (Phlips, 1974), and also in international trade (Andersson et al., 1995; Krugman and Venables, 1990; Ottaviano et al., 2002).

[2]Since there are no barriers to enter the market, α is assumed to be sufficiently high. This ensures a non-negative demand at all locations, and hence deters new entrants during the duration of this game. However, α cannot be too large, so that it prevents effective competition when preferences between firms' products are nearly identical.

transportation cost. Since this framework focuses on changes in demand due to cross-price effects under symmetric condition, it is further assumed that $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$. For simplicity, without loss of generality, further assume $\beta = 1$. The degree of product differentiation can then be measured by $\gamma \in [0, 1]$. Where γ is ranging from zero when goods are independent (i.e. products are highly differentiated when $\gamma \to 0$) to one when goods are perfect substitutes (*i.e.* products are less differentiated when $\gamma \to 1$).

In the special case of completely differentiated products ($\gamma = 0$), the demand takes the following form:

$$q_i = \frac{(\alpha - P_i)}{\beta} \tag{3.3}$$

which implies that each firm i, i={1,2} acts as a monopolist. The other special case is when products are considered perfect substitutes ($\gamma = 1$). In this case, the demand follows the Bertrand game set-up for homogeneous products, which necessitates explicit assumptions about consumers behavior under all possible price configurations: (i) consumers always purchase from the seller with the lowest delivered price (i.e. at locations where $P_1 < P_2$ firm 1 faces the demand given in Equation (3.3) and the demand for q_2 is equal to zero); (ii) if a consumer is facing the same delivered price $P_1 = P_2$, she purchases half from firm 1 and half from firm 2. Formally:

$$q_1 = \begin{cases} 0 & \text{if } P_1 > P_2 \text{ and/or } P_1 > \alpha, \\ \frac{\alpha - P_1}{2\beta} & \text{if } P_1 = P_2 > \alpha, \\ \frac{\alpha - P_1}{\beta} & \text{if } P_1 < P_2 \text{ and } P_1 < \alpha. \end{cases} \tag{3.4}$$

By symmetric conditions, the reverse holds for firm 2. The consumer location where $P_1 = P_2$ is usually referred to as the point of indifference, $\hat{s}$. Given $x_1 < \hat{s} < L - x_2$:

$$\hat{s} = \frac{L - x_2 + x_1}{2} + \frac{p_2 - p_1}{2t} \tag{3.5}$$

3.1.2 Firm Behavior

Firms engage in a two-stage, non-cooperative game. In the first stage, firms simultaneously choose their location given the location of its competitor. The cost of relocating is assumed to be sufficiently high[3] such that location becomes an irreversible decision for the duration of the game. The second stage consists of finding a Bertrand equilibrium in price strategies[4] given the location parameter x_i, $i = \{1,2\}$. Firms maximize the profit:

$$\Pi_i = (p_i - c)Q_i \tag{3.6}$$

with respect to x_i in the first stage and price p_i in the second stage. Note that for completely differentiated products or imperfect substitutes firms have overlapping markets:

$$Q_i = \int_0^L q_i ds$$

while in the case of perfect substitutes they do not:[5]

$$Q_i = \int_0^{\hat{s}} q_i ds$$

3.2 Equilibrium

The firms' price-location optimization problem is dependent on whether consumers view firms' products as independent, imperfect substitutes or perfect substitutes. The resulting equilibrium location and price for each degree of product differentiation is treated successively.

[3]It seems reasonable to assume that once a firm has chosen its location in the first stage, this involves a relatively large amount of fixed costs (expences related to the licensing process, aquiring the property, etc.) that the firm will have an incentive to pay-off in the short-run.

[4]One attractive feature with the Bertrand set-up applied in the current set-up is that firms are able to change prices faster than quantity produced, since there is no need to change capacity (*i.e.* size of establishment).

[5]With the exception of location $\hat{s}$ which is only a point of zero dimension.

3.2.1 Independent Products

The analysis begins with the limited case of completely differentiated products ($\gamma = 0$). Since firms' products are independent in consumption and with demand given by Equation (3.3), the firm does not take into consideration the price or the location of its rival in its profit maximization decision.

Proposition 1. For $\gamma = 0$, there exists an equilibrium in locations which is in the middle of the market.

Proof. Taking the integral of the demand function given by Equation (3.3) over the whole market of length L and substituting it into Equation (3.6) yields the following profit function:

$$\Pi_i = \frac{1}{\beta}(p_i - c)\left[L\left(\alpha - p_i + t(x_i - \frac{1}{2}L)\right) - tx_i^2\right] \tag{3.7}$$

Maximizing Π_i with respect to p_i yields the following expression:

$$p_i = \frac{\alpha + c + t\left[x_i - \frac{1}{2}L\right]}{2} - \frac{tx_i^2}{2L} \tag{3.8}$$

Substituting Equation (3.8) into (3.7) and maximizing Π_i with respect to x_i returns a cubic function. The only root that obeys the model assumptions is:

$$x_i = \frac{L}{2}$$

Taking the first order partial derivatives of (3.7) and (3.8) with respect to x_i confirms that $\frac{L}{2}$ is the equilibrium location for both firms. Therefore, it may be concluded that in the case of completely differentiated products, firms have an incentive to locate at the middle of the market.

3.2.2 Perfect Substitutes

The other special case arises when firms' products are considered to be perfect substitutes ($\gamma = 1$). In this case, firms face the demand structure for homogeneous products as outlined in Equation (3.4).

Proposition 2. For $\gamma = 1$, the two firms will locate next to each other in equilibrium.

Proof. Taking the integral of Equation (3.4) and substituting the resulting aggregate demand function into Equation (3.6) yields the following profit function:

$$\Pi_1 = \frac{1}{\beta}(p_1 - c)\left[\hat{s}\left(\alpha - p_1 + t(x_1 - \frac{1}{2}\hat{s})\right) - tx_1^2\right]$$

$$\Pi_2 = \frac{1}{\beta}(p_2 - c)\left[L\left(\alpha - p_2 + t(x_2 - \frac{1}{2}L)\right) - \hat{s}\left(\alpha - p_2 - t(L - x_2 - \frac{1}{2}\hat{s})\right) - t(L - x_2)^2\right] \tag{3.9}$$

Maximizing Π_i with respect to p_i yields a quadratic function where one of the roots obeys the model assumptions, namely:

$$\begin{aligned} p_1 = \frac{1}{9}&(4\alpha + 2p_2 + 3c + t(6x_1 - 2x_2 + 2L) \\ &- \sqrt{16\alpha^2 - p_2(20\alpha - 13p_2) - c(12\alpha + 6p_2 - 9c + 6Lt) + t[x_1(u_1) + x_2(v_1) - L(z_1)]}\Big) \end{aligned}$$

$$\begin{aligned} p_2 = \frac{1}{9}&(4\alpha + 2p_1 + 3c + t(6x_2 - 2x_1 + 2L)) \\ &- \sqrt{16\alpha^2 - p_1(20\alpha - 13p_1) - c(12\alpha + 6p_1 - 9c + 6Lt) + t[x_2(u_2) + x_1(v_2) - L(z_2)]}\Big) \end{aligned} \tag{3.10}$$

where $u_1 = 12\alpha + 6p_2 - 18c + t(81x_1 - 6x_2 + 6L)$, $v_1 = 20\alpha - 26p_2 + 6c + t(13x_2 - 26L)$ and $z_1 = 20\alpha - 26p_2 - 13Lt$ for firm 1. For firm 2, $u_2 = 12\alpha + 6p_1 - 18c + t(81x_2 + 6L)$, $v_2 = 20\alpha - 26p_1 + 6c + t(13x_1 - 26L - 6x_2)$ and $z_2 = 20\alpha - 26p_1 - 13Lt$. For this case, it becomes infeasible to substitute the price back into the profit function and solve for the optimum location. The location equilibrium is defined as the situation where no firm can increase its profit by relocation when the location of its competitor is given. In this set-up both price and profit are monotonically increasing functions of a firm's own location, which is to say, that the maximum is found where the function intersects with the market boundary.[6] These results are illustrated in Figure 3-1 for $x_1 \in [0, x_2]$ given $x_2 = \frac{L}{2}$ and $x_2 = \frac{L}{4}$. That is, firms that compete with perfect substitutes have an incentive to locate next to their competitor in order to maximize profit. The resulting equilibrium becomes one where both firms are located next to each other.

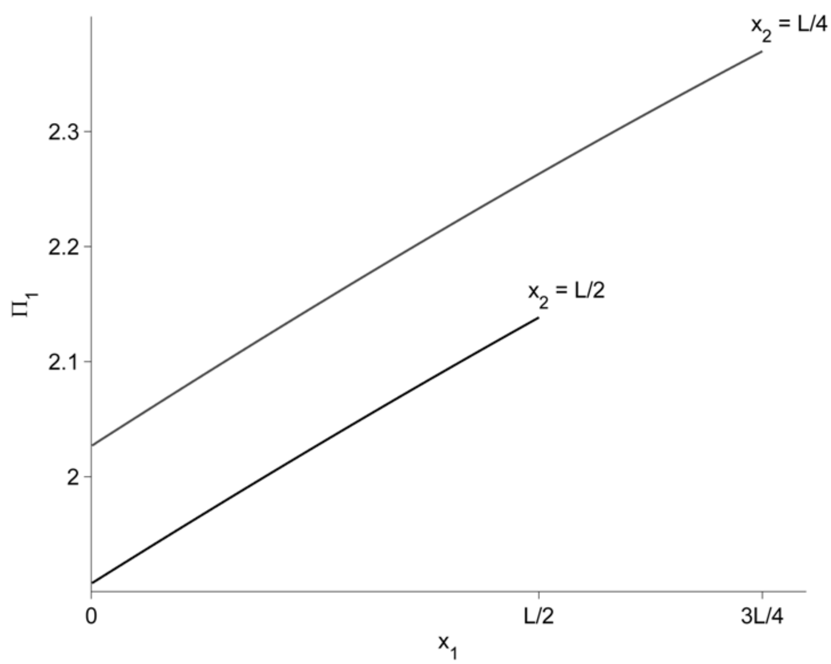


Figure 3-1: Firm 1's Profit as a Function of x_1 given $x_2 = \frac{L}{2}$ and $x_2 = \frac{L}{4}$

The result $\frac{\partial p_i}{\partial x_i} > 0$ might seem counter-intuitive as one would expect the opposite should hold due to more fierce price competition the closer the two firms locate. These results follows the argument put forward by De Fraja and Norman (1993) although

[6]Which is at $\hat{s} = \frac{L}{2}$ for all symmetric cases.

they only consider competition with imperfect substitutes. That is, in spatial competition, when a firm moves away from the market center the total amount it sells at any given price decreases (given the location of its competitor). Thus, the firm has a incentive to lower its price in order to compensate for the reduction in quantity, and this spatial influence outweighs the competitive effect of more proximate locations. That is, the market share effect dominates the market power effect.

3.2.3 Imperfect Substitutes

In the intermediate case, represented by imperfect substitutes with $\gamma \in (0,1)$, firms face the demand given in Equation (3.2). The varying cross-price elasticity allows firms to capture demand over the whole market. Given the results obtained for the extreme cases of independent products and perfect substitutes, one would expect that as $\gamma \to 0$ and $\gamma \to 1$, the resulting demand structure should start approaching that of the two extremes respectively.

Proposition 3. For $\gamma \in (0,1)$, the central agglomeration is an equilibrium.

Proof. Integrating Equation (3.2) over the whole market of length L and substituting the aggregate demand for each firms' product into Equation (3.6) results in the following profit function:

$$\Pi_1 = \frac{1}{\beta^2-\gamma^2}(p_1-c)\left[L\left(\beta(\alpha-p_1+t(x_1-\frac{1}{2}L))-\gamma(\alpha-p_2-t(x_2-\frac{1}{2}L))\right)-t(\beta x_1^2-\gamma x_2^2)\right] \tag{3.11}$$

$$\Pi_2 = \frac{1}{\beta^2-\gamma^2}(p_2-c)\left[L\left(\beta(\alpha-p_2+t(x_2-\frac{1}{2}L))-\gamma(\alpha-p_1-t(x_1-\frac{1}{2}L))\right)-t(\beta x_2^2-\gamma x_1^2)\right]$$

Maximizing Π_i with respect to p_i and solving for price yields the following expression for firm 1 and 2 respectively:

$$p_1 = \frac{\alpha + c + t(x_1 - \frac{1}{2}L)}{2} - \frac{\gamma(\alpha - p_2 + t(x_2 - \frac{1}{2}L))}{2\beta} - \frac{t(\beta x_1^2 - \gamma x_2^2)}{2L\beta}$$
$$p_2 = \frac{\alpha + c + t(x_2 - \frac{1}{2}L)}{2} - \frac{\gamma(\alpha - p_1 + t(x_1 - \frac{1}{2}L))}{2\beta} - \frac{t(\beta x_2^2 - \gamma x_1^2)}{2L\beta} \quad (3.12)$$

Substituting Equation (3.12) into (3.11), differentiating Π_i with respect to x_i, and solving for x_i results in a cubic function. As in the case of completely differentiated products, the only root that obeys the model assumptions is:

$$x_i = \frac{L}{2}$$

The first and second order conditions of (3.11) and (3.12) with respect to x_i, confirms that $x_i = \frac{L}{2}$ is a true maximum. That is, in the case where firms sell imperfect substitutes, the equilibrium location is at the center of the market. For $\gamma \rightarrow 0$ and $\gamma \rightarrow 1$ the previous results indicate that this should be the case. More intermediate values of γ, can be thought of as either each consumer having some preferences for both products, or each consumer location consisting of several consumers where a portion prefers q_1 and another q_2. In either case, since the firm faces a positive demand at all locations and is also competing to a varying degree with the other firm, there are both pure demand side benefits with being located in the center, but also some strategic incentives. As $\gamma \rightarrow 0$ or $\gamma \rightarrow 1$, the incentive structure starts becoming dominated by one or the other.

3.3 Social Optimality of Equilibrium

In the case of competition with perfect substitutes, the equilibrium location is characterized by a non-cooperative Nash equilibrium.[7] That is to say that both firms and consumers would be better off if firms were to locate some distance away from each other. However, firms end up locating next to each other due to strategic competition over customers. This argument can be illustrated in a game theoretic setting, as summarized in Figure 3-2.

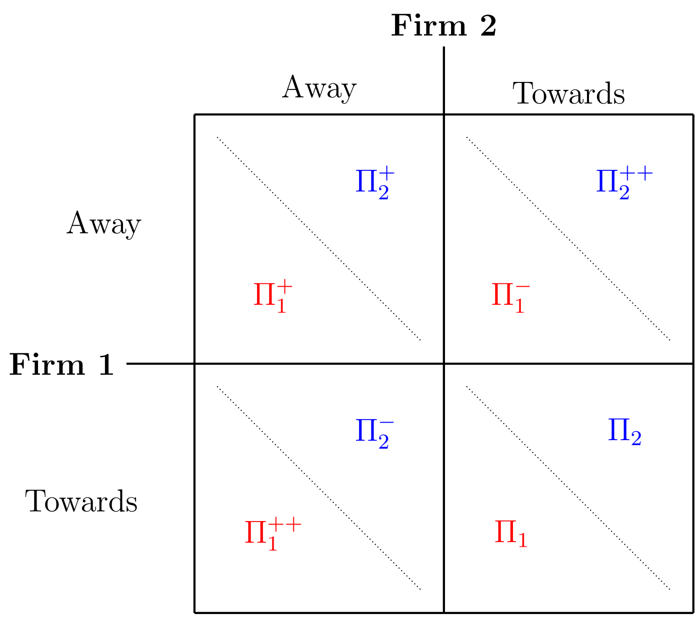


Figure 3-2: Payoff Matrix

First, suppose both firm 1 and 2 decide to locate away from each other. In this situation, by moving towards the other, either firm could increase its market share and hence receive a higher profit (Π_i^{++}), holding the location of its competitor constant (as illustrated in Figure 3-1).[8] That is, the dispersed location pattern, or the strategy

[7] This can also be extended to the intermediate case for values of γ close to 1.

[8] Recall that α is assumed to be high enough to ensure that each consumer location is served by one of the existing firms in the market, no matter the equilibrium location configuration. Meaning that the firm does not risk losing consumers at the end of the market when moving towards the

("Away", "Away"), is not an equilibrium. Since this movement of one firm increases that firm's profits at the expense of its rival, the only way its rival can reclaim its lost market share, and hence its lost profit, is by also moving towards the other. For each firm, the dominant strategy is to move towards the other, regardless of what the other firm does. Locating next to each other is therefore an equilibrium, despite the fact that both firms would be better off if they were to locate at equal intervals away from the market center where profits are higher.

The equilibrium follows that of the classical "Prisoner's Dilemma". If both firms were to initially locate equi-distance away from the center of the market, there is always an incentive for the firm to move towards the center in order to steal market shares from its competitor. No firm can benefit from changing its strategy given that the other player's strategy remains unchanged. Therefore each firm will always have an incentive to locate next to its competitor in order to maximize profits, given the location of its competitor.

The socially optimal solution arises when both firms are located around $\frac{L}{4}$ distance away from the center of the market. This is the location at which the aggregate distance, and therefore the aggregate transportation costs payed by consumers, is minimized.[9] By defining total profits, $\Pi_1 + \Pi_2$, made by firms as producer surplus and total quantity demanded by consumers, $Q_1 + Q_2$, as consumer surplus, it can also be shown that total surplus is higher if firms locate more uniformly throughout the

middle. This is in order to prevent entry during the duration of this game, keeping the game setting to two firms, since there is assumed to be no barriers to enter for new firms.

[9]Total transportation costs under $x_1 = x_2 = \frac{L}{2}$ can be found by evaluating:

$$\int_0^L t\left|\frac{L}{2} - s\right| ds = t(\frac{L}{4})$$

Compare that with evaluting the total transportation cost when $x_1 = x_2 = \frac{L}{4}$:

$$2\int_0^{\frac{L}{2}} t\left|\frac{L}{4} - s\right| ds = t(\frac{L}{8})$$

market.

Proposition 4. The equilibrium location $x_i = \frac{L}{2}$ for i={1,2} is characterized by a non-cooperative Nash equilibrium when the two firms are selling perfect substitutes. Both firms and consumers are better off if firms are located at $x_i = \frac{L}{4}$, $i = \{1,2\}$.

Proof. By applying the symmetric equilibrium conditions $x_1 = x_2$ and $p_1 = p_2$ to Equation (3.9) and (3.10) for $x_1 = x_2 = \frac{L}{2}$ and $x_1 = x_2 = \frac{L}{4}$ it can be shown that $(Q_i^{\frac{L}{4}} - Q_i^{\frac{L}{2}}) > 0$:

$$\frac{L}{2}(p_i^{\frac{L}{2}} - p_i^{\frac{L}{4}}) + \frac{L^2}{16}t > 0 \tag{3.13}$$

where $(p_i^{\frac{L}{2}} - p_i^{\frac{L}{4}}) < 0$:

$$\frac{1}{8}\left(Lt + \sqrt{(16\alpha^2 + Lt(8c - 8\alpha + 57Lt) - 32\alpha c + 16c^2)}\right) - \frac{1}{2}\sqrt{(\alpha^2 + 5L^2t^2 - 2\alpha c + c^2)} < 0 \tag{3.14}$$

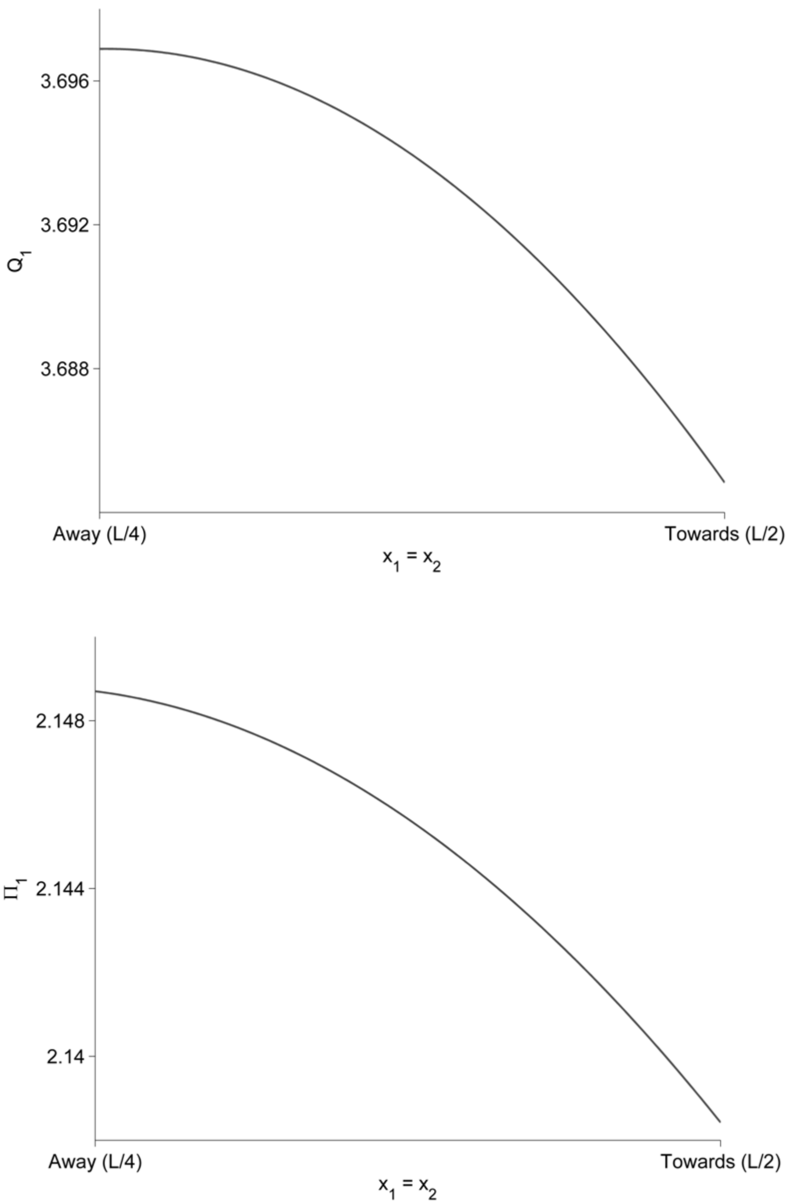
since the closer to its competitor a firm is locating, the lower the price due to more intense competition.[10] The inequality in Equation (3.13) indicates a trade-off for the consumer between lower prices at $x_i = \frac{L}{2}$ and lower transportation costs at $x_i = \frac{L}{4}$. Given the model assumptions, the difference in price is relatively small in comparison to the lower transportation cost experienced by the average consumer. This implies that consumer surplus, Q_1+Q_2, is higher at $x_i = \frac{L}{4}$ when firms sell perfect substitutes.

Since firms face both a higher demand and a higher price around location $x_i = \frac{L}{4}$, producer surplus is also higher if both firms locate more uniformly (*i.e.* if both choose the strategy *"Away"*, relating back to the game matrix in Figure 3-2):

$$(Q_i^{\frac{L}{4}} - Q_i^{\frac{L}{2}}) + (p_i^{\frac{L}{4}} - p_i^{\frac{L}{2}}) > 0$$

[10] In the current setting, price is a positive function of a firm's own location and the price of its competitior, and a negative function of its competitors' location, *i.e.* $\frac{\partial p_1}{\partial x_1} > 0$, $\frac{\partial p_1}{\partial p_2} > 0$ and $\frac{\partial p_1}{\partial x_2} < 0$.

These results are illustrated in Figure 3-3, where the symmetric equilibrium conditions $x_1 = x_2$ and $p_1 = p_2$ have been applied to simulate firms' profits and consumer demand when firms adopt the same strategy, *("Away", "Away")* or *("Towards", "Towards")*. Compare Figure 3-3(b) to Figure 3-1, which shows the pay-off for firm 1 when it chooses to either stay *"Away"* or move *"Toward"* given the location of its competitor.



[a]

[b]

Figure 3-3: Distribution of Benefits when Firms Choose the Same Strategy: (a) shows the distribution of consumer surplus, while (b) shows the distribution of producer surplus.

3.4 Extension to Two-Dimensional Space

This section provides an extension of the model to a two-dimensional space. For this purpose, a circular market is adopted where firms locate along the diameter of length L. The center of the market is given by the coordinates $(x_c,\ y_c) = (\frac{L}{2},\ \frac{L}{2})$. Consumers are uniformly distributed throughout the market. Distance to the closest firm is modeled using Euclidean distance. The market is illustrated in Figure 3-4. As in the one-dimensional case, given symmetric conditions, firms split the market at point $\hat{s}$, which in this case is given by the vertical line going through the center of the market, dividing the market into two semicircles. For simplicity, firms are restricted to only being able to move horizontally, holding $y_1 = y_2 = \frac{L}{2}$.

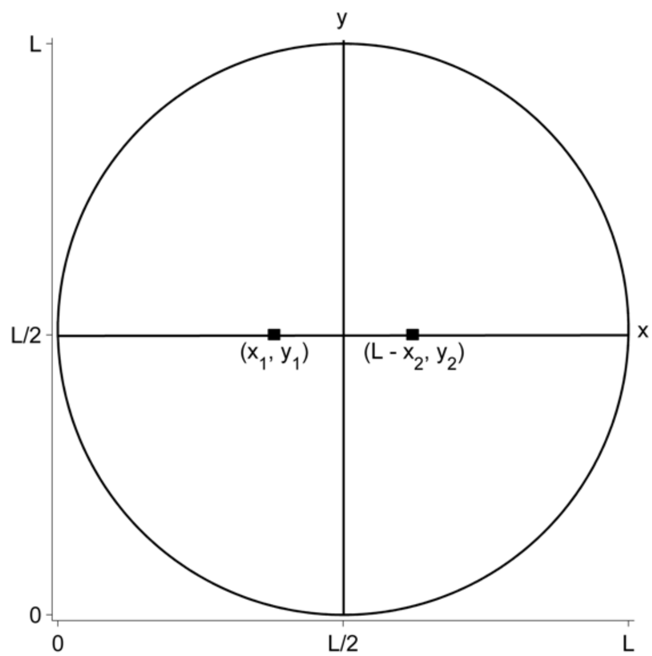


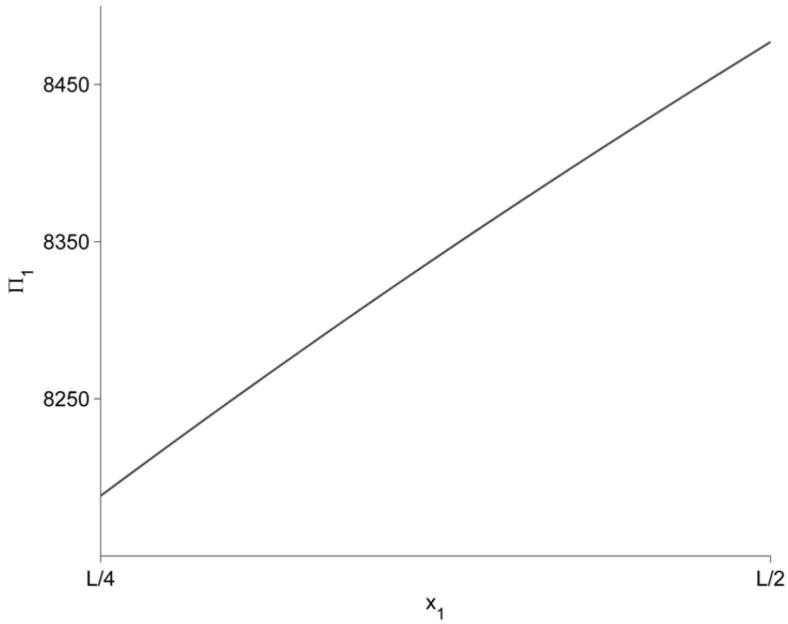
Figure 3-4: The Circular Market.

To show that the results from the model proposed in this paper are generalizable to a two-dimensional space, simulations are presented. Figure 3-5(a) illustrates firm

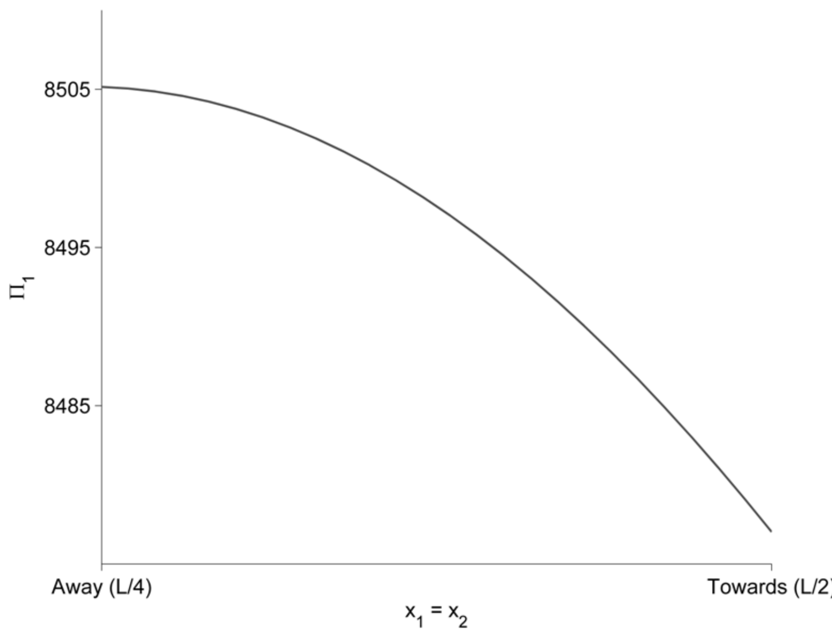
1's profit as a function of its own location, holding the location of its competitor constant at $x_2 = \frac{L}{2}$. This is used in order to illustrate one firm's best response given the location of its competitor. The result mirrors those derived for the one-dimensional case (see Figure 3-1) with the maximum found at the market boundary, $\hat{s} = \frac{L}{2}$. Figure 3-5(b) illustrates firm 1's profit as a function of its own location, assuming its competitor adopts the same location strategy, *i.e.* either *("Away", "Away")* or *("Towards", "Towards")*. In this case, the maximum is found slightly further away from the center where intense price competition is avoided.

As in the one-dimensional case, consumer surplus is maximized around the point where the aggregate transportation costs payed by consumers are minimized.[11] This is illustrated in Figure 3-6.

[11] In the semicircle, aggregate distance traveled by consumers is minimized if firms are located $\frac{4r}{3\pi}$ distance away from the center of the circle (holding $y_i = \frac{L}{2}$).



[a]



[b]

Figure 3-5: Firm 1's Profit given the Location of Firm 2 vs. when Firm 2 Adopts the Same Strategy. (a) Shows firm 1's profit as a function of x_1 (given $x_2 = \frac{L}{2}$), while (b) shows firm 1's profit when both firms adopts the same strategy ($x_1 = x_2$).

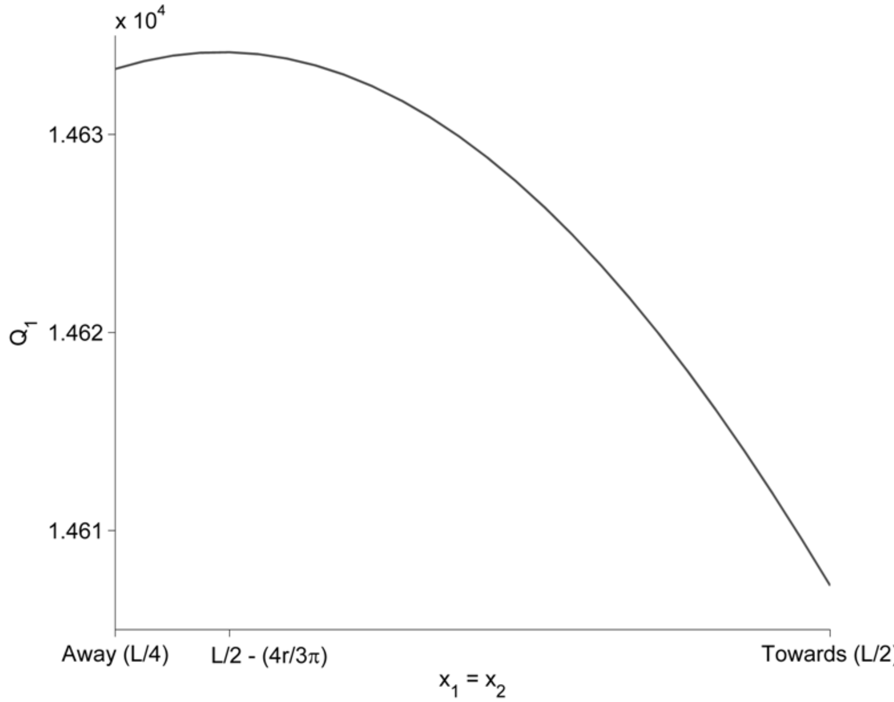


Figure 3-6: Consumer Surplus when Firms choose the Same Strategy

3.5 Concluding Comments

The purpose of this chapter has been to develop a game theoretic framework for the study of retail firm location choices under varying assumptions related to product differentiation. Beginning the analysis with the assumption of a linear market, the analysis is later extended to a two-dimensional (circular) market. One important finding from this model is that properties of the equilibrium differ depending on the degree of product differentiation. When products are highly substitutable, firms end up locating next to each other due to strategic competition over customers. The resulting equilibrium is characterized by a non-cooperative Nash equilibrium. In the case of highly differentiated products, firms locate in the middle of the market due to demand side benefits.

Unlike the case of non-spatial oligopolistic competition, where consumers are better off when firms compete, in spatial competition the clustering of firms through strategic interactions is not socially optimal. From the consumers' standpoint, if both firms sell similar or homogeneous products, consumers do not have any (or small) preference for either firms product and therefore buy from the seller with the lowest delivered price. This implies that consumers would be better off if these two firms were to locate uniformly throughout the market as this arrangement would serve to minimize the total distance traveled by the consumers. Firms would still be able to reach the same amount of consumers, while reducing the intense price competition which is often associated with being located close to rival firms.

If products are viewed as independent in consumption or highly differentiated, there is no meaningful game present as each firm acts as a monopolist. In this case, the clustering equilibrium is socially optimal. Whether consumers buy some quantity of both products, or have strict preferences for one over the other, their accessibility is maximized when both firms are located in the middle of the market.

Instead of having to travel to two different locations to purchase each good, the average consumer can now enjoy the benefits of multi-purpose shopping while at the same time minimizing their total distance traveled. Alternatively, in the case where consumers have strict preferences for one product, the distance traveled by the average consumer is minimized.

Theoretically, these findings indicate that, under certain demand conditions, public policy can have a welfare-improving effect on firm location choices. If firms selling close substitutes were persuaded (for example through zoning laws or through economic incentives) to move away from the Nash equilibrium location to the socially optimum location where firms and consumers alike are better off, such a policy could also have a efficiency-boosting effect for transportation infrastructure investments. Through the even distribution of such firms, special arrangements and costs associated with congestion could be mitigated and the impact on roads (or the demand for transportation infrastructure, in general) more evenly distributed, as one consequence of everyone traveling to the same location is congestion and more intensive use of certain roads.

The model findings can help interpret empirical location patterns of retailers. The findings from this theoretical analysis adds to the existing literature by explaining why certain types of retailers may want to locate in close proximity to their competition despite increased price competition (*i.e.*, situations where the market share effect dominates the market power effect). In order to identify between different location behaviors among firms, a multivariate spatial statistic is develop in Chapter 4.

Chapter 4

An Empirical Approach to Analyzing Firm Location Patterns

Based on the theoretical framework outlined in Chapter 3, a multivariate spatial statistic is developed that can identify between different location behaviors. For this purpose a definition of the proposed statistic is formulated and its properties defined. The proposed statistic has the following desirable features: a theoretical basis from which a robust, topological definition of *relatively close* is derived; ability to distinguish between positive association (attraction) between two categories and clustering in the joint population; detect asymmetrical relationships between different categories, and; it is intuitive in terms of visualization of the results on the map. To demonstrate its utility and robustness, the statistic is applied to a variety of simulated bivariate point patterns.

4.1 Proximity Criteria

In order to define when two points are to be considered located relatively close to each other, the method described in this chapter utilizes relative spatial separation. This relative spatial separation is conceptualized in the following way: given the location of firms of category **B**, are firms of category **A** located in such a way that they avoid firms of category **B** or are they located closer than what would be the case if they were trying to maximize the distance from category **B** firms? That is, in relation to firms of another category how are firms of category **A** located?

Let P denote a joint point population of size N within which there are f different categories. For convenience, assume a bivariate pattern (i.e. $f = 2$) with the two point processes $p_i^A, i = \{1, 2, ..., n_A\}$ and $p_j^B, j = \{1, 2, ..., n_B\}$ both belonging to P, where n_A and n_B denotes the population size of category **A** and **B** respectively.

For illustrative purposes, assume a one-dimensional, linear space (as in Hotelling's 'Main Street' example) with $n_A = 1$ and $n_B = 3$ as shown in Figure 4-1. At the one extreme, if a category **A** firm is trying to avoid firms of category **B**, it would place itself right in between two category **B** firms at $x_{\frac{1}{2}}$. At the other extreme, if firm **A** is very attracted to firms of category **B** it would place itself right next to one of firm **B**'s locations. The intermediate case would be at the locations $x_{\frac{1}{4}}$ or $x_{\frac{3}{4}}$ – right in between the two extremes. The line can therefore be divided into two regions, one where firm **A** is considered located relatively close to one of the category **B** firms which is within the interval (p_i^B, $x_{\frac{1}{4}}$) or ($x_{\frac{3}{4}}$, p_i^B) (the relative attraction area) and one region of relatively avoidance respresented by the interval ($x_{\frac{1}{4}}$, $x_{\frac{3}{4}}$) (the relative avoidance area).

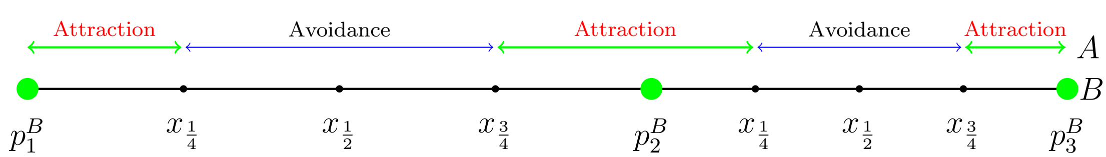


Figure 4-1: Defining Areas of Relative Attraction

In order to extend the criteria to a two dimensional space, Thiessen polygons are derived around the points of the category of interest. In order to test whether category **A** is attracted to category **B**, Thiessen polygons $T_i^B, i = \{1, 2, ..., n_B\}$ are derived around each category **B** point, p_i^B. Within each Thiessen polygon, T_i^B, "relative attraction" areas are derived, $h_i^B, i = \{1, 2, ..., n_B\}$. These areas of relative attraction are constructed by finding the midpoint, $m_{i,e}^B, e = \{1, 2, ..., k\}$, between the point p_i^B and each vertex, $v_{i,e}^B, e = \{1, 2, ..., k\}$, of the Thiessen polygon T_i^B around point p_i^B. The attraction area h_i^B for point p_i^B is then the polygon formed by connecting all midpoints, $m_{i,e}^B$.

Figure 4-2 provides an illustration of the proposed criteria where the Thiessen polygons are restricted to the study area (here defined by the extreme points in each direction in the joint population). A category **A** point is considered to be located relatively close to a category **B** point if it is located within its attraction area. In this small randomly generated example, two category **A** points are considered attracted to a category **B** point (highlighted in red). This definition of attraction is not sensitive to heterogeneity in the density of the joint population as it "adjust itself" while still taking into account the relative spatial separation between points. It is also constructed in such a way that asymmetrical associations can be captured.

4.2 Multivariate Statistic

To determine whether category **A** points are positively (attracted) or negatively (avoiding) associated with points of category **B**, relative attraction areas are constructed around category **B** points, $p_i^B, i = \{1, 2, ..., n_B\}$, and the number of category **A** points located within attraction areas h_i^B are calculated and compared against an expected value. Let $C_{A \to B} = \sum c_j^A$ denote the sum of category **A** points that falls within category **B**'s attraction areas, h_i^B. Formally:

$$c_j^A = \begin{cases} 1 & \mathit{if}\ p_j^A \in h_i^B, \\ 0 & \mathit{if}\ p_j^A \notin h_i^B. \end{cases} \tag{4.1}$$

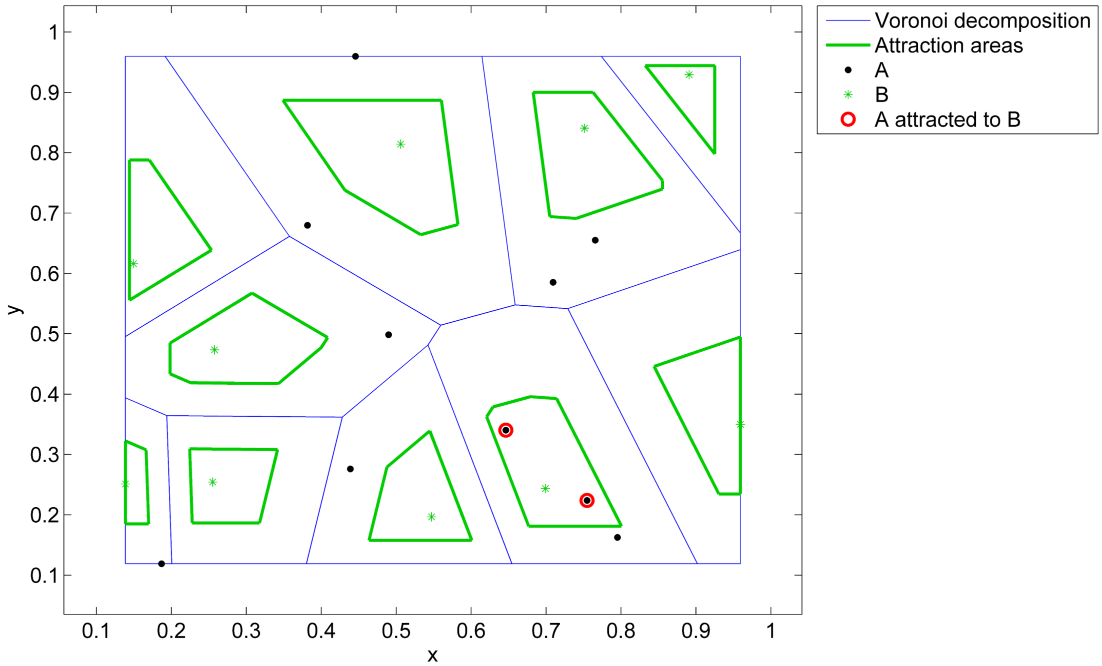


Figure 4-2: Deriving Areas of Relative Attraction from Thiessen Polygons

The statistic is then defined as the observed proportion of category **A** points that are co-located with category **B** points, formally:

$$Q_{A\to B} = \frac{C_{A\to B}}{n_A} \tag{4.2}$$

The $Q_{A\to B}$ statistic follows a Bernoulli distribution with mean $E[Q_{A\to B}]$ and variance $E[Q_{A\to B}](1 - E[Q_{A\to B}])$. The expected value $E[Q_{A\to B}] = \frac{\sum y_i^B}{Y}$ of the statistic is given by the share of attraction area, $\sum y_i^B$, in relation to the total study area, Y. Given the way the relative attraction areas are constructed, the expected value is a constant (independent of the way the total study area is defined), more specifically $E[Q_{A\to B}] = \frac{1}{4}$. That is, the probability of a random category **A** point to be co-located with a category **B** point is 0.25 (or 25 percent).

If $c_j = c$, for all i, by the Gauss-Markov central limit theorem, the statistic is asymptotically normally distributed, *i.e.*, $Q_n \sim N\left(c, \frac{c(1-c)}{n}\right)$. Given its asymptotic properties, the test statistic for $Q_{A\to B}$ for larger samples is:

$$z_{Q_{A\to B}} = \frac{Q_{A\to B} - E[Q_{A\to B}]}{\sqrt{E[Q_{A\to B}](1 - E[Q_{A\to B}])/n_A}} \tag{4.3}$$

The $z_{Q_{A\to B}}$ statistic says that if category **A** is to be considered positively associated with (attracted to) category **B**, a large share of category **A** points should be located within category **B**'s relative attraction areas in relation to the portion of attraction areas in the study region. When interpreting the value of $z_{Q_{A\to B}}$, the following holds:

$$z_{Q_{A\to B}} \begin{cases} < 0 & \textit{if } \boldsymbol{A} \textit{ is negatively associated with } \boldsymbol{B} \textit{ (avoidance),} \\ \approx 0 & \textit{if } \boldsymbol{A} \textit{ is neither positively nor negatively associated with } \boldsymbol{B} \textit{ (independence),} \\ > 0 & \textit{if } \boldsymbol{A} \textit{ is positively associated with } \boldsymbol{B} \textit{ (attraction).} \end{cases} \tag{4.4}$$

That is, the value of $z_{Q_{A\to B}}$ will be positive/negative, if there are category **A** points located in category **B**'s attraction areas more/less frequently than expected. If the observed proportion of category **A** points in the attraction areas is equal to what is expected (which is given by the proportion of attraction area in the total study area), the statistic will take the value of 0.

Since $Q_{A\to B}$ is unidirectional, $Q_{B\to A}$ may take on a different value. Meaning that the interaction between category **A** and **B** can be asymmetric. For example, there may be a larger share of category **B** points co-located with category **A**, while there are few category **A** points located in close proximity to category **B**, *i.e.* $Q_{B\to A} > Q_{A\to B}$, meaning that category **B** is more attracted to category **A** than category **A** is to category **B**.

4.3 Inference and Interpretation

This section considers how the significance of the results can be tested and how to control for the underlying spatial structure. The null hypothesis for the $z_{Q_{A \to B}}$ statistic is *given the locations of category **B**, the locations of category **A** in relation to **B** is no different from the case of two independent patterns.* If the null can be rejected, the statistic will tell whether points are negatively ($z_{Q_{A \to B}} < 0$) or positively associated ($z_{Q_{A \to B}} > 0$).

Since the statistic has asymptotic properties, for smaller samples Monte Carlo simulations should be used to generate an empirical distribution. The empirical distribution is generated by holding the observed locations of the base category (**B**) fixed, while randomly redistributing all category **A** points through a large number of Monte Carlo simulations. In this study, 600 iterations are used, *i.e.* to generate an empirical distribution category **A** is redistributed 600 times. In the redistribution process the number of category **A** points is held fixed while horizontal and vertical coordinates respectively are drawn from the standard uniform distribution on the interval given by the study area. In each iteration the $Q_{A \to B}$ statistic is calculated in order to construct a distribution from which critical values for different confidence levels can be established. The observed $Q_{A \to B}$ is then compared against these critical values in order to determine whether or not the null of independence can be rejected in favor for one of the alternative hypotheses (positive or negative association between categories).

4.3.1 Relating Theoretical and Statistical Results

This section describes how to interpret the statistic by relating the theoretical results obtained in Chapter 3 and existing theoretical literature to each potential outcome of the statistic. The statistic has three possible outcomes: the null hypothesis (no categorial association/independence) and two alternative hypotheses (positive or negative categorical association).

Negative categorical association ($z_{Q_{A\to B}} < 0$) can be interpreted as avoidance if the two retailers in question are competing with homogeneous products (as suggested by for example d'Aspremont et al. (1979)). If the two retailers sell differentiated products it can be interpreted as them locating independently of each other where they target different customer segments located in different parts of the markets. That is, the two retailers are not strategically interacting and they locate in the middle of their respective market as described in Chapter 3.2.1.

The null of independence ($z_{Q_{A\to B}} \approx 0$), neither attraction nor avoidance, can be interpreted as simply two retailers who locate independently of each other because they are not affecting each others decision (*i.e.*, they sell completely differentiated products). Alternatively, it can be the case of retailers selling imperfect substitutes[1] where there is still some degree of strategic interactions present. In some areas[2] firms will locate next to its competitors due to strategic competition over customers, while in other parts of the market they will locate as close as possible to the demand. It may also be an unsaturated market where some stores still enjoy local monopoly power in the middle of some local market areas, but in the future competitors might come in and locate in close proximity if demand is high enough to support two stores selling close substitutes.

The second alternative hypothesis of positive categorical association ($z_{Q_{A\to B}} > 0$), may be interpreted as the case with firms selling close substitutes and therefore end up locating next to each other in competition over market shares (as described in Section 3.2.2). If the two firms' product offerings and prices are very similar, this is likely the case as they will be affecting each others performance and are not likely to end up

[1]As described in Chapter 3.2.3.
[2]Where demand is high enough to support two similar stores.

next to each other at random. The location decision is a long-term fixed investment and have significant impact on market share and profitability (Ghosh and McLafferty, 1987). If such patterns are found among firms with differentiated products, it is likely due to demand side benefits of certain locations where many retailers tend to locate.

Interpretation of results should be accompanied with either estimating the cross-price elasticity between the firms under investigation[3] or by a case study where the business model of each firm and the competitive relationship between the two firms are identified.

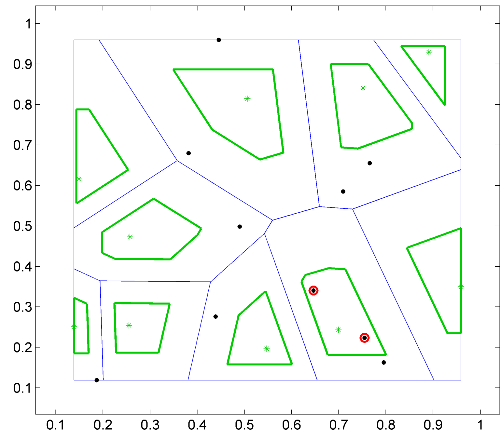
4.4 Application to Simulated Data

In order to test whether the statistic is sensitive to unequal sample sizes of two point patterns, three scenarios are developed. In case 1, two random and independent patterns of equal size ($n_A = n_B$) are generated for a variety of population sizes. In case 2 and 3, two random and independent patterns of unequal size ($n_A > n_B$ and $n_A < n_B$) are generated. For all cases and population sizes, the $Q_{A \to B}$ is calculated and inference is made based on the associated $z_{Q_{A \to B}}$ score and critical values obtained from the empirical distribution. While the statistic is assumed to be approximately normal for larger samples, simulation based inference is recommended for smaller samples as described in Chapter 4.3.

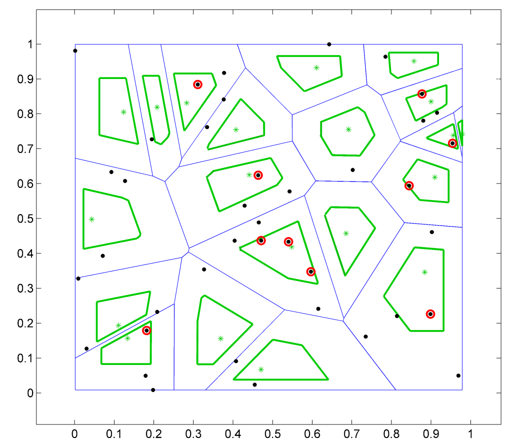
Table 4.1 presents, for each population size (n_i), the $Q_{A \to B}$ and associated $z_{Q_{A \to B}}$ statistic. It also shows the mean, standard devation (Std), and critical values at the 5 percent significance level ($Q^*_{A \to B}$) obtained from the empirical distribution. Inference based on $z_{Q_{A \to B}}$ and $Q^*_{A \to B}$ is also noted for each case. The results in Table 1 are as expected in all three cases. Some of the results are visualized in Figure 4-3 in order to

[3]Such an analysis might be hard as the data required for such an estimation is hard to collect and in many cases proprietary.

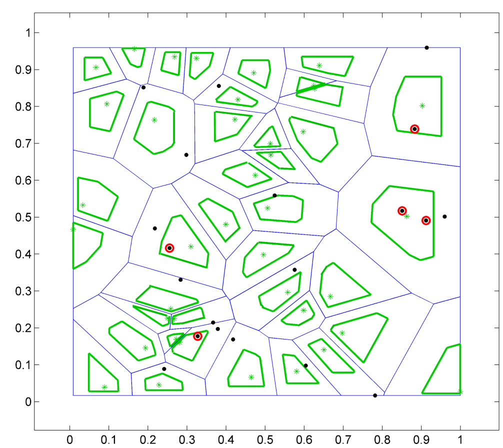
see whether the statistic complies with what might be expected from visual inspection. For two random and independent patterns, the statistic reports independence which is confirmed by both the z-statistic and the critical values obtained from the empirical distribution. The mean from the empirical distribution approaches the expected value of the theoretical distribution ($E[Q_{A \to B}] = 0.25$). The variance of the statistic decreases with the sample size increases together with the non-critical region for the statistic. In addition, the statistic does not appear to be sensitive to relative sample sizes between categories.



[a]



[b]



[c]

Figure 4-3: Two Random and Independent Patterns: (a) Case 1 with $n_A = n_B = 10$; (b) Case 2 with $n_A = 40, n_B = 20$, and; Case 3 with $n_A = 20, n_B = 40$.

Next, an experiment was performed for a larger expected value (*i.e.*, larger attraction areas), $E[Q_{A\to B}] = 0.5$. As expected, this only causes a shift in the values of the statistic and its critical values. The results from this simulation experiment are presented in Table 4.2. This indicates that the stastic is not sensitive to changes in the parameters.

Table 4.2: Two Independent Patterns with Larger Expected Value

	$Q_{A\to B}$	$z^{\mathbf{Inf}}_{Q_{A\to B}}$	Mean (Std)	$Q^*_{A\to B}$ (lower, upper)$^{\mathbf{Inf}*}$
$n_A = n_B$				
5	0.4	$-0.4472^{\mathbf{I}}$	0.4977 (0.2314)	$(0, 1)^{\mathbf{I}}$
10	0.4	$-0.6325^{\mathbf{I}}$	0.5018 (0.1586)	$(0.2, 0.8)^{\mathbf{I}}$
20	0.5	$-9.93E-15^{\mathbf{I}}$	0.4953 (0.1164)	$(0.25, 0.7)^{\mathbf{I}}$
40	0.525	$0.3162^{\mathbf{I}}$	0.4993 (0.0773)	$(0.35, 0.65)^{\mathbf{I}}$
80	0.5444	$0.8433^{\mathbf{I}}$	0.4972 (0.0534)	$(0.3889, 0.6)^{\mathbf{I}}$

Note: The table shows the $Q_{A\to B}$ and $z^{\mathbf{Inf}}_{Q_{A\to B}}$ (where $^{\mathbf{Inf}}$ shows the inference – attraction $^{\mathbf{A}}$, independence $^{\mathbf{I}}$ and repulsion/avoidance $^{\mathbf{R}}$ – at the 5% significance level). It also shows the mean, standard deviation and the 5% critical values with inference $Q^*_{A\to B}(lower, upper)^{\mathbf{Inf*}}$ from the empirical distribution.

To show that the statistic can distinguish between clustering in the joint population and pairwise categorical association tendencies given the pattern of the joint population, two overlaying Matern cluster processes are realized using the same parameters. To describe the Matern cluster process, let G be a point process of "centers" where each $g \in G$ is associated with a point process P^g. For a Matern cluster process (Matern, 1960) each point in the cluster with center g is uniformly distributed within a disc of radius r at g (Kroese and Botev, 2014). For this experiment, the goal is to generate two overlaying but still independent cluster processes, the same cluster centers are used but two separate sets of random points are generated within each disc (of the same radius, $r = 0.1$) around each cluster center. The results are presented in Table 4.3. The results for the case with $n_A = 28$ and $n_B = 30$, is visualized in Figure 4-4. While there is an overall clustering in the population, within each of the three clusters generated (which all contain both categories), the two categories of points are independent of each other.

Table 4.3: Two Realized Matern Cluster Processes

n_A, n_B	$Q_{A\to B}$	$z^{\mathbf{Inf}}_{Q_{A\to B}}$	Mean (Std)	$Q^*_{A\to B}$ (lower, upper)$^{\mathbf{Inf}*}$
$27, 16$	0.2593	$0.1111^{\mathbf{I}}$	0.2556 (0.0849)	$(0.1111, 0.4444)^{\mathbf{I}}$
$28, 30$	0.1552	$-1.6678^{\mathbf{I}}$	0.2492 (0.0572)	$(0.1379, 0.3621)^{\mathbf{I}}$
$49, 46$	0.2857	$0.5774^{\mathbf{I}}$	0.2509 (0.0636)	$(0.1429, 0.3673)^{\mathbf{I}}$

Note: The table shows the $Q_{A\to B}$ and $z^{\mathbf{Inf}}_{Q_{A\to B}}$ (where $^{\mathbf{Inf}}$ shows the inference – attraction $^{\mathbf{A}}$, independence $^{\mathbf{I}}$ and repulsion/avoidance $^{\mathbf{R}}$ – at the 5% significance level). It also shows the mean, standard deviation and the 5% critical values with inference $Q^*_{A\to B}(lower, upper)^{\mathbf{Inf*}}$ from the empirical distribution.

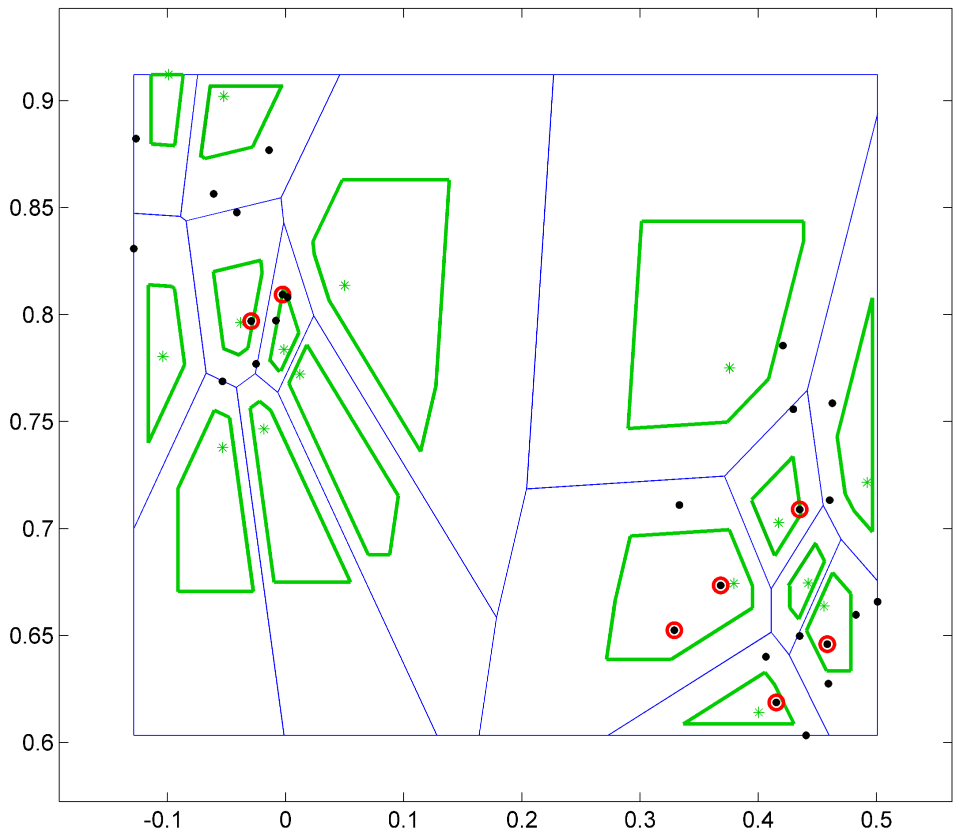


Figure 4-4: Two Realized Matern-Cluster Processes

4.4.1 Robustness to Buffering

Thus far, in all the simulation experiments presented in Tables 4.1-4.3, category **A** points are redistributed based on a random distribution. However, a random generation of points may produce patterns where some points are located in close proximity to each other. This would not be a realistic pattern for same brand firms. Seldom do we observe two same brand stores locate right next to each other. However, the average distance between same brand firms may vary depending on the overall density of stores in certain areas. For example, the average distance between same brand firms may be longer in rural or suburban areas, while the distance between same brand stores within the central business district may be smaller.

In order to take into account such tendencies, simulation experiments are performed where relocated **A** points can not be within a certain distance of each other. This distance is given by $\frac{1}{3}$ and $\frac{1}{4}$ of the average distance between the three nearest neighboring firm **B** points. The purpose of this restriction is twofold: (1) to reflect the fact that same brand firms are unlikely not locate right next to each other; (2) to take into account heterogeneity in the underlying spatial arrangement of economic activity. However, as it turns out, applying such a restriction does not alternate the results significantly which is confirmed by the simulation results presented in Table 4.4. The same holds for case 2 ($n_A > n_B$) and 3 ($n_A < n_B$). The results for these cases are given in Appendix C.

Table 4.4: $\frac{1}{3}$ and $\frac{1}{4}$ Varying Buffer

$n_A = n_B$	$Q_{A\to B}$	$z^{\mathbf{Inf}}_{Q_{A\to B}}$	Mean (Std)	$Q^*_{A\to B}$ (lower, upper)$^{\mathbf{Inf}*}$
Varying buffer ($\frac{1}{3}$)				
5	0.2	$-0.2582^{\mathbf{I}}$	0.235 (0.1924)	$(0, 0.6)^{\mathbf{I}}$
10	0.2	$-0.3651^{\mathbf{I}}$	0.2518 (0.1399)	$(0, 0.5)^{\mathbf{I}}$
20	0.3	$0.5164^{\mathbf{I}}$	0.2478 (0.0943)	$(0.1, 0.45)^{\mathbf{I}}$
40	0.175	$-1.0954^{\mathbf{I}}$	0.2505 (0.066)	$(0.125, 0.375)^{\mathbf{I}}$
80	0.3	$1.0328^{\mathbf{I}}$	0.2515 (0.0456)	$(0.1625, 0.3375)^{\mathbf{I}}$
160	0.2625	$0.3651^{\mathbf{I}}$	0.2501 (0.0324)	$(0.1812, 0.3125)^{\mathbf{I}}$
320	0.2594	$0.3873^{\mathbf{I}}$	0.2491 (0.0245)	$(0.2031, 0.3)^{\mathbf{I}}$
Varying buffer ($\frac{1}{4}$)				
5	0.2	$-0.2582^{\mathbf{I}}$	0.2503 (0.1865)	$(0, 0.6)^{\mathbf{I}}$
10	0.2	$-0.3651^{\mathbf{I}}$	0.2495 (0.1366)	$(0, 0.5)^{\mathbf{I}}$
20	0.3	$0.5164^{\mathbf{I}}$	0.2514 (0.1042)	$(0.05, 0.5)^{\mathbf{I}}$
40	0.175	$-1.0954^{\mathbf{I}}$	0.2466 (0.0694)	$(0.125, 0.4)^{\mathbf{I}}$
80	0.3	$1.0328^{\mathbf{I}}$	0.2486 (0.0486)	$(0.15, 0.35)^{\mathbf{I}}$
160	0.2625	$0.3651^{\mathbf{I}}$	0.2509 (0.0333)	$(0.1812, 0.3125)^{\mathbf{I}}$
320	0.2594	$0.3873^{\mathbf{I}}$	0.25 (0.0246)	$(0.2031, 0.3)^{\mathbf{I}}$

Note: The table shows the $Q_{A\to B}$ and $z^{\mathbf{Inf}}_{Q_{A\to B}}$ (where $^{\mathbf{Inf}}$ shows the inference – attraction $^{\mathbf{A}}$, independence $^{\mathbf{I}}$ and repulsion/avoidance $^{\mathbf{R}}$ – at the 5% significance level). It also shows the mean, standard deviation and the 5% critical values with inference $Q^*_{A\to B}(lower, upper)^{\mathbf{Inf*}}$ from the empirical distribution.

4.4.2 Dependent Bivariate Point Patterns

Table 4.5 shows simulation results for two positively associated (attracted) point patterns. To account for different patterns in the join population, two different spatial point processes are used: a non-homogeneous Poission process and a Matern cluster process.[4] To generate two positively associated patterns, category **A** is generated according to either process and the category **B** is created by shifting each point of category **A** by a small distance in random direction. In addition, a perfectly uniform distribution is generated using a square grid where every other vertex of a square grid represents a point of category **A** while category **B** is created by shifting category **A** by a small distance in random direction. Examples of results for the non-homogeneous Poission process and the uniform distribution are visualized in Figure 4-5.

As expected significant positive categorical association is detected for all realized patterns at the 5 percent significance level according to the z-statistic as well as the critical interval obtained from the empirical distribution. However, given the way in which the data was generated, one would expect $Q_{A \to B}$ to be one for all three underlying distributions and population sizes. The primary reason for the slightly lower $Q_{A \to B}$-values is the way the study area is constructed with the outermost points in each direction defining the boundaries. While these points are located in very close priximity, the statistic will not count them as attracted given the way the relative attraction areas are constructed. As indicated by the results, the effect from boundary points is greater in smaller samples and does not have as much of an influence in larger samples.

One straight forward solution to this problem is to apply a buffer around the

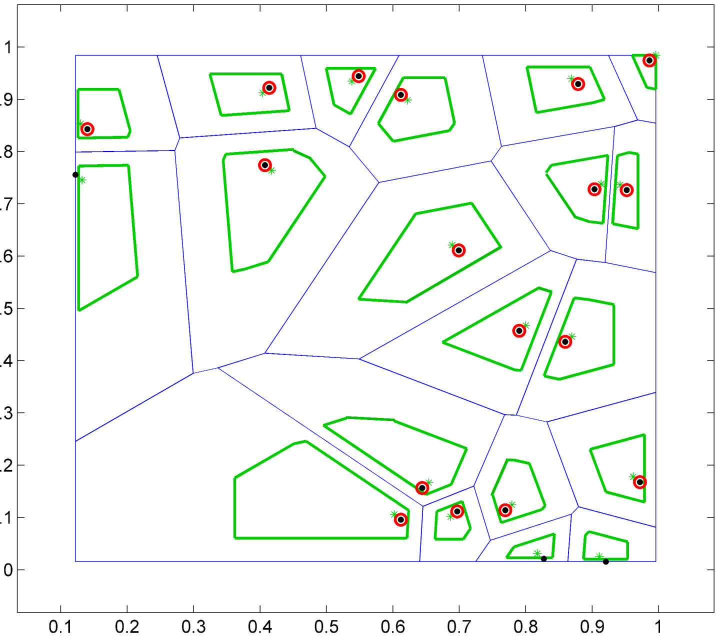
[4]In a Poisson process, point locations are generated under the following conditions: (i) each location in the study area has an equal probability of receiving a point; and (ii) a point location is selected independently of the location of existing points (Aldstadt, 2010). The non-homogeneous Poisson process used in these simulations is generated by thinning out a homogeneous Poission process (Kroese and Botev, 2014). For a more detailed description of both processes see Kroese and Botev (2014).

study area. One way in which this can be done is by taking a fraction of the average distance between all points in the base category (category **B**) and apply it as a buffer around the existing study area. This will not affect the properties of the statistic given the way the expected value is calculated and the attraction areas are constructed. The results of this solution with a buffer of $\frac{1}{4}$ of the average distance between category **B** points around the study area, is given in Table 4.6 for all three underlying distributions in Table 4.5. Figure 4-6 illustrates the results for the non-homogeneous Poission process and Matern cluster process with a buffer around the study area. As expected, boundary points that are in fact located in close proximity to each other are now accounted for and the statistic reaches a $Q_{A\rightarrow B}$ of 1 (*i.e.*, absolute positive categorical association between the two categories) for all three underlying distributions.

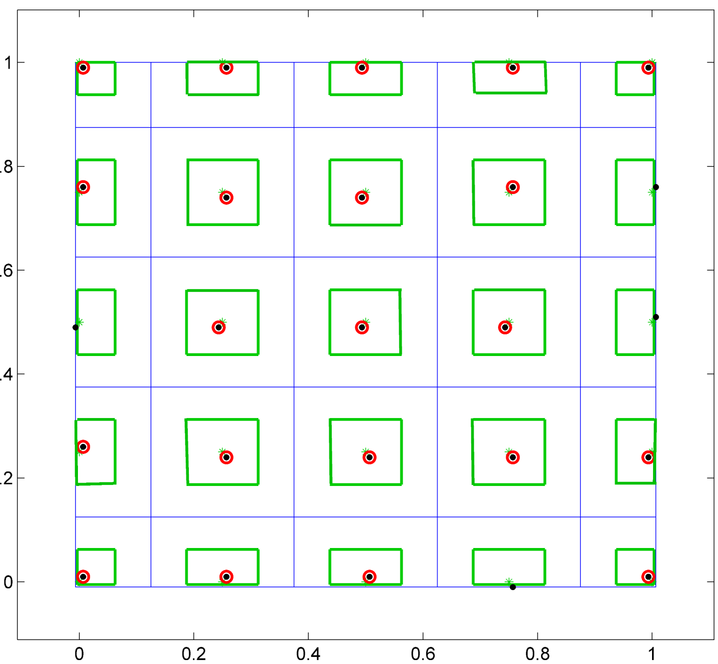
Table 4.5: Positively Associated Patterns

$n_A = n_B$	$Q_{A\to B}$	$z^{\mathbf{Inf}}_{Q_{A\to B}}$	Mean (Std)	$Q^*_{A\to B}$ (lower, upper)$^{\mathbf{Inf}*}$
Non-Homogeneous Possion				
10	0.8	$4.0166^{\mathbf{A}}$	0.2485 (0.1441)	$(0, 0.6)^{\mathbf{A}}$
20	0.85	$6.1968^{\mathbf{A}}$	0.255 (0.097)	$(0.1, 0.45)^{\mathbf{A}}$
40	0.875	$9.1287^{\mathbf{A}}$	0.2432 (0.0694)	$(0.125, 0.375)^{\mathbf{A}}$
Matern Cluster Process				
10	0.8	$4.0166^{\mathbf{A}}$	0.2497 (0.1316)	$(0, 0.5)^{\mathbf{A}}$
20	0.85	$6.1968^{\mathbf{A}}$	0.2506 (0.0921)	$(0.1, 0.45)^{\mathbf{A}}$
40	0.975	$10.5893^{\mathbf{A}}$	0.2507 (0.0667)	$(0.125, 0.375)^{\mathbf{A}}$
Uniform				
16	0.6875	$4.0415^{\mathbf{A}}$	0.2801 (0.1137)	$(0.0625, 0.5)^{\mathbf{A}}$
25	0.84	$6.8127^{\mathbf{A}}$	0.2651 (0.085)	$(0.12, 0.44)^{\mathbf{A}}$
36	0.7778	$7.3131^{\mathbf{A}}$	0.2586 (0.0735)	$(0.1111, 0.3889)^{\mathbf{A}}$

Note: The table shows the $Q_{A\to B}$ and $z^{\mathbf{Inf}}_{Q_{A\to B}}$ (where $^{\mathbf{Inf}}$ shows the inference – attraction $^{\mathbf{A}}$, independence $^{\mathbf{I}}$ and repulsion/avoidance $^{\mathbf{R}}$ – at the 5% significance level). It also shows the mean, standard deviation and the 5% critical values with inference $Q^*_{A\to B}(lower, upper)^{\mathbf{Inf*}}$ from the empirical distribution.



[a]



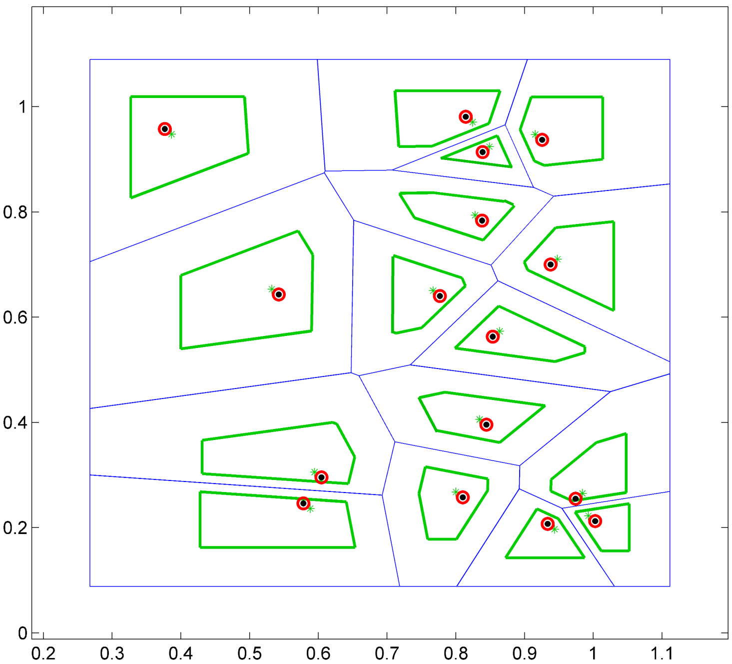
[b]

Figure 4-5: Positively Associated Patterns: (a) Non-Homogeneous Possion Process ($n_A = n_B = 20$) and (b) Uniform Distribution ($n_A = n_B = 25$).

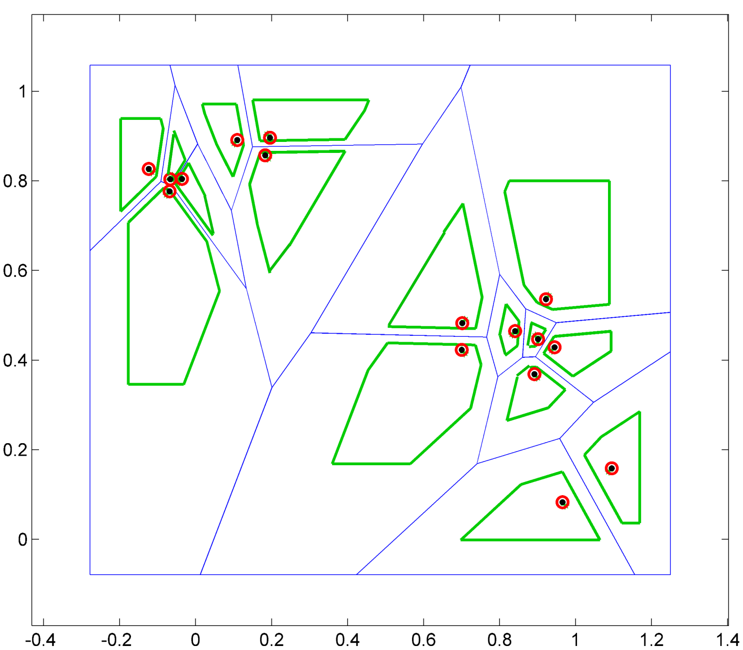
Table 4.6: Positively Associated Patterns with Buffer around Study Area

$n_A = n_B$	$Q_{A\to B}$	$z^{\mathbf{Inf}}_{Q_{A\to B}}$	Mean (Std)	$Q^*_{A\to B}$ (lower, upper)$^{\mathbf{Inf}*}$
Non-Homogeneous Possion				
16	1	$6.9282^{\mathbf{A}}$	0.2467 (0.1109)	$(0.0625, 0.4375)^{\mathbf{A}}$
Matern Cluster Process				
16	1	$6.9282^{\mathbf{A}}$	0.2538 (0.1091)	$(0.0625, 0.5)^{\mathbf{A}}$
Uniform				
16	1	$8.6603^{\mathbf{A}}$	0.2915 (0.0948)	$(0.12, 0.48)^{\mathbf{A}}$

Note: The table shows the $Q_{A\to B}$ and $z^{\mathbf{Inf}}_{Q_{A\to B}}$ (where $^{\mathbf{Inf}}$ shows the inference – attraction $^{\mathbf{A}}$, independence $^{\mathbf{I}}$ and repulsion/avoidance $^{\mathbf{R}}$ – at the 5% significance level). It also shows the mean, standard deviation and the 5% critical values with inference $Q^*_{A\to B}(lower, upper)^{\mathbf{Inf*}}$ from the empirical distribution.



[a]



[b]

Figure 4-6: Positively Associated Patterns with Buffer around Study Area: (a) Non-Homogeneous Possion Process ($n_A = n_B = 16$) and (b) Matern Process ($n_A = n_B = 16$).

The next experiment considers negatively associated patterns (repulsed/avoiding) patterns. To generate two extremely repulsed categories, every other vertex of a square grid is assigned to one category. Table 4.7 shows the results for this pattern. Results for the case of $n_A = n_B = 25$ is visualized in Figure 4-7. As expected, $Q_{A \to B} = 0$ for all three sample sizes.

Table 4.7: Negatively Associated Patterns

$n_A = n_B$	$Q_{A \to B}$	$z^{\mathbf{Inf}}_{Q_{A \to B}}$	Mean(Std)	$Q^*_{A \to B}$ (lower, upper)$^{\mathbf{Inf}*}$
16	0	$-2.3094^{\mathbf{R}}$	0.2919 (0.1117)	$(0.125, 0.5)^{\mathbf{R}}$
25	0	$-2.8868^{\mathbf{R}}$	0.2727 (0.0891)	$(0.12, 0.48)^{\mathbf{R}}$
36	0	$-3.4641^{\mathbf{R}}$	0.2732 (0.0779)	$(0.1389, 0.4167)^{\mathbf{R}}$

Note: The table shows the $Q_{A \to B}$ and $z^{\mathbf{Inf}}_{Q_{A \to B}}$ (where $^{\mathbf{Inf}}$ shows the inference – attraction $^{\mathbf{A}}$, independence $^{\mathbf{I}}$ and repulsion/avoidance $^{\mathbf{R}}$ – at the 5% significance level). It also shows the mean, standard deviation and the 5% critical values with inference $Q^*_{A \to B}(lower, upper)^{\mathbf{Inf*}}$ from the empirical distribution.

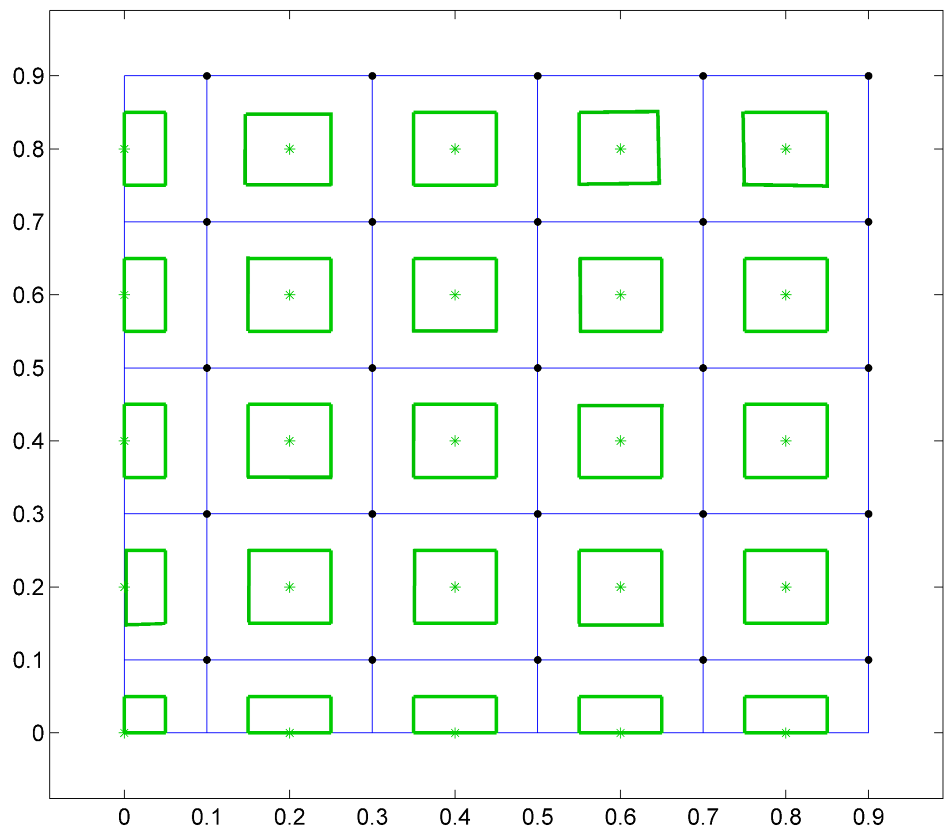


Figure 4-7: Negatively Associated Patterns ($n_A = n_B = 25$)

4.5 Concluding Comments

This chapter presents a new multivariate spatial statistic aimed at identifying different interaction patterns between competing outlets. It is established that the statistic has asymptotic properties, and that its distribution is approximately normal for larger samples. Finite sample properties and robustness are tested through Monte Carlo simulations by varying assumptions regarding population size, definition of study area, expected value of the statistic and other parameters. The simulation results are consistent and approaches those of the theoretical distribution as the population size increases.

The results from the statistic, using various simulated data, suggest that the topological criteria of relatively closeness, derived from the theoretical framework, is independent of heterogeneity in the underlying spatial structure. These results also confirm that the proposed statistic has the ability to capture not only asymmetrical relationship but also the potential to distinguish pairwise categorical associations from clustering of the joint population. The next two chapters illustrate how the statistic can be used to quantify real world location patterns and how the results can be interpreted based on the theoretical literature (including the framework presented in Chapter 3).

Chapter 5

Relative Location Among Retail Chain Stores: The Quick Service Restaurant Sector

To illustrate how the statistic presented in Chapter 4 performs on observed data, it is applied to competing chains within the quick service restaurant (QSR) segment of the fast food industry. The QSR industry is represented by the three largest firms in the industry: McDonald's, Wendy's and Burger King. This allows for an analysis of relative location among competitors selling close substitutes. Among these three chains, each outlet that is constructed by each retail chain is nearly identical in terms of outlet size and product offerings. Second, the prices are similar and the product offerings by the three competing fast food restaurants are reasonably close substitutes (Toivanen and Waterson, 2005) (*i.e.*, they compete more directly with each other (Melaniphy, 2007)).

Since the data necessary to estimate actual cross price elasiticities between the different firms in each market is difficult to obtain, a brief overview of the competitive environment in the quick service restaurant segment is provided in the following section. This information should help connect statistical outcomes to findings from the theoretical literature. Do competitors in these two markets locate next to each other to capture some share of their competitors business, do they seek to establish spatial monopolies by locating away from their competition or do they target different customer segments and locate independently of each other?

5.1 The Quick Service Restaurant Sector

In the United States, the fast food industry is divided into four major categories: quick service restaurants (QSR); takeaways; mobile and street vendors; and leisure locations. The fast food industry is characterized by high levels of industry competition and limited product differentiation with firms offering similar products, services and distribution outlets (Kamal and Wilcox, 2014; Wendy's, 2013). The largest segment of the fast food industry is QSR, representing 73.3 percent of the U.S. fast food market (Kamal and Wilcox, 2014). The QSR segment is skewed towards chains with major restaurant chains representing 64 percent of total industry traffic (Magazine, 2014a).

The QSR segment contains the three largest fast food chains (in terms of annual sales): McDonald's, Wendy's and Burger King (Magazine, 2014b; Thomadsen, 2007). All three chains offer products that are very standardized within each chain, not only in terms of the food but also in terms of menu boards, uniforms and architectural style. While some outlets are operated directly by each respective corporations, most U.S. outlets are operated independently by franchisees within a framework of a national brand (purchasing their inputs from approved suppliers and setting their own prices) (Thomadsen, 2007).

In the fast food industry these chains are each others principal competitors (BurgerKing, 2013; Friedman, 2014; Gibson, 1997). They also compete with regional hamburger restaurant chains (*i.e.*, Carls Jr., Jack in the Box and Sonic) and, to a lesser extent, with national, regional and local quick service restaurants that offer alternative menus, casual and "fast casual" restaurant chains, street stalls or kiosks, and convenience stores and grocery stores (BurgerKing, 2013; Little, 2015; McDonald's, 2013; Wendy's, 2013). Furthermore, the industry is characterized by few barriers to entry, and therefore new competitors may emerge at any time (BurgerKing, 2013). All three chains claim to compete on the basis of price/affordability, convenience, service, product quality and variety. While McDonald's does not mention location as part of competitive factors in its business model, both Burger King and Wendy's do (BurgerKing, 2013; McDonald's, 2013; Wendy's, 2013).

5.2 Description of the Data and the Market

To represent the QSR industry in Indianapolis, Indiana, location data for McDonald's, Wendy's and Burger King are collected. The location data are displayed in Figure 5-1 and 5-2. The data are verified by crosschecking the Yellow Pages, Google Earth and company-specific web pages. The city of Indianapolis is located in Marion County, Indiana. In this application, the county is interpreted as an isolated market with market boundaries being set by the outermost quick service restaurants in the county. Within Marion County, 100 QSR outlets are identified. In line with national data, McDonald's is the leading player in this market with 48 establishments followed by Burger King (27 outlets) and Wendy's (25 outlets).

Marion County has an estimated population of 928,281 people with a median household income of $42,334 (in 2013 dollars) as of year 2013. In 2007, the total accommodation and food services sales reached $2,248,380,000. Out of Marion County's population, approximately 843,393 people live within the city limits of Indianapolis. In the city, the median household income is similar to that of the county ($41,962 in 2013 dollars) (Census, 2013). The distribution of population and median household income within Marion County are mapped in Figure 5-1 and Figure 5-2, respectively. These maps are overlayed with the locations of the three fast food chains representing the QSR segment in this analysis. As expected, most outlets are located in or near areas with higher population density. Given the "carry-out" nature of these establishments it is not surprising to see many of the outlets being located in close proximity to highways which is further addressed in Chapter 7. The map of median household income in Figure 5-2 shows that median income by block group increases as one moves out of the city center. The location of quick service restaurants does not seem to have any particular relationship to income distribution. However, visual inspection reveals that they do not seem to be present inside block groups with higher median income as may be expected by their focus on low prices (affordability).

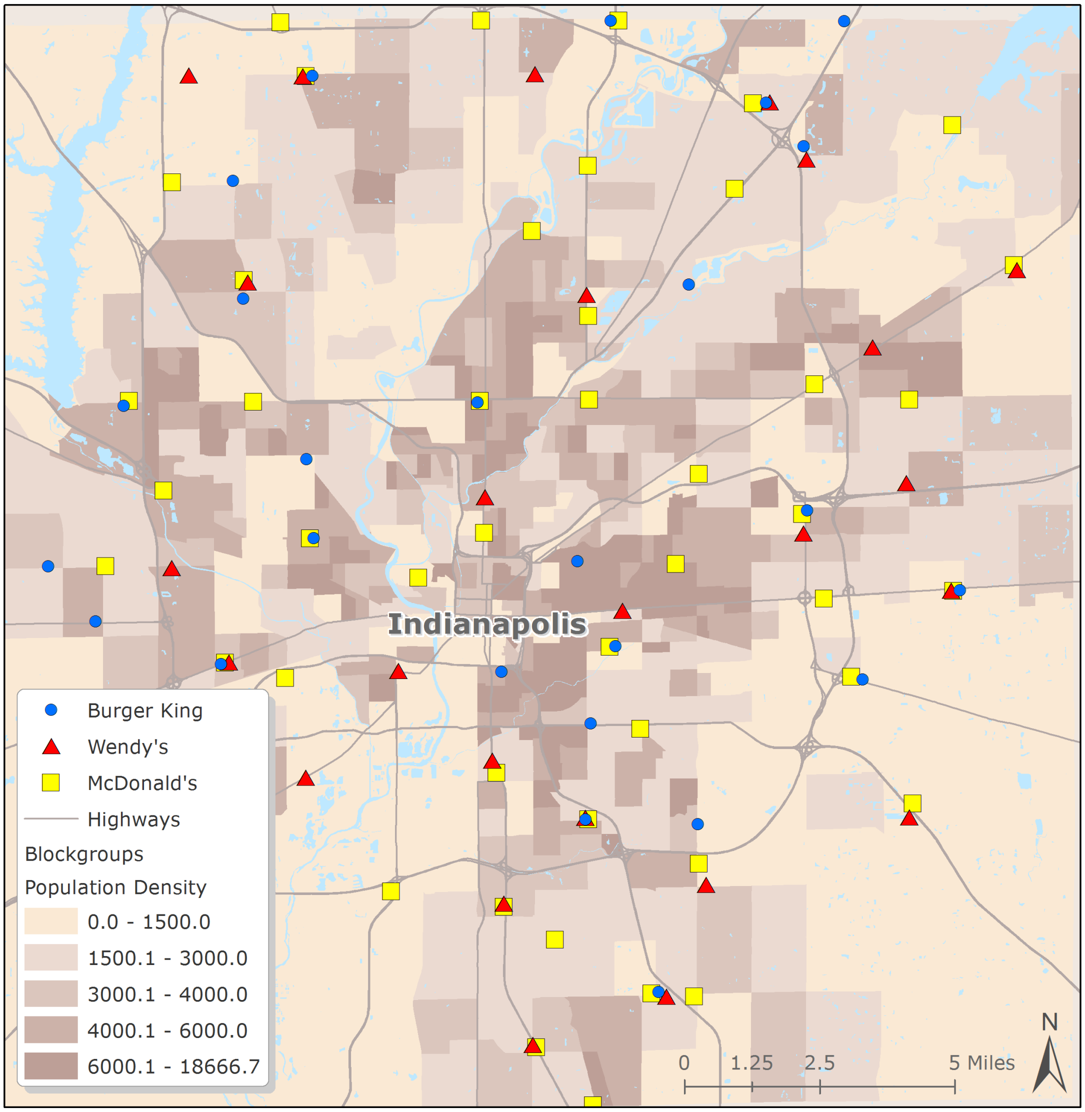


Figure 5-1: Locations of Quick Service Restaurants and Population Density

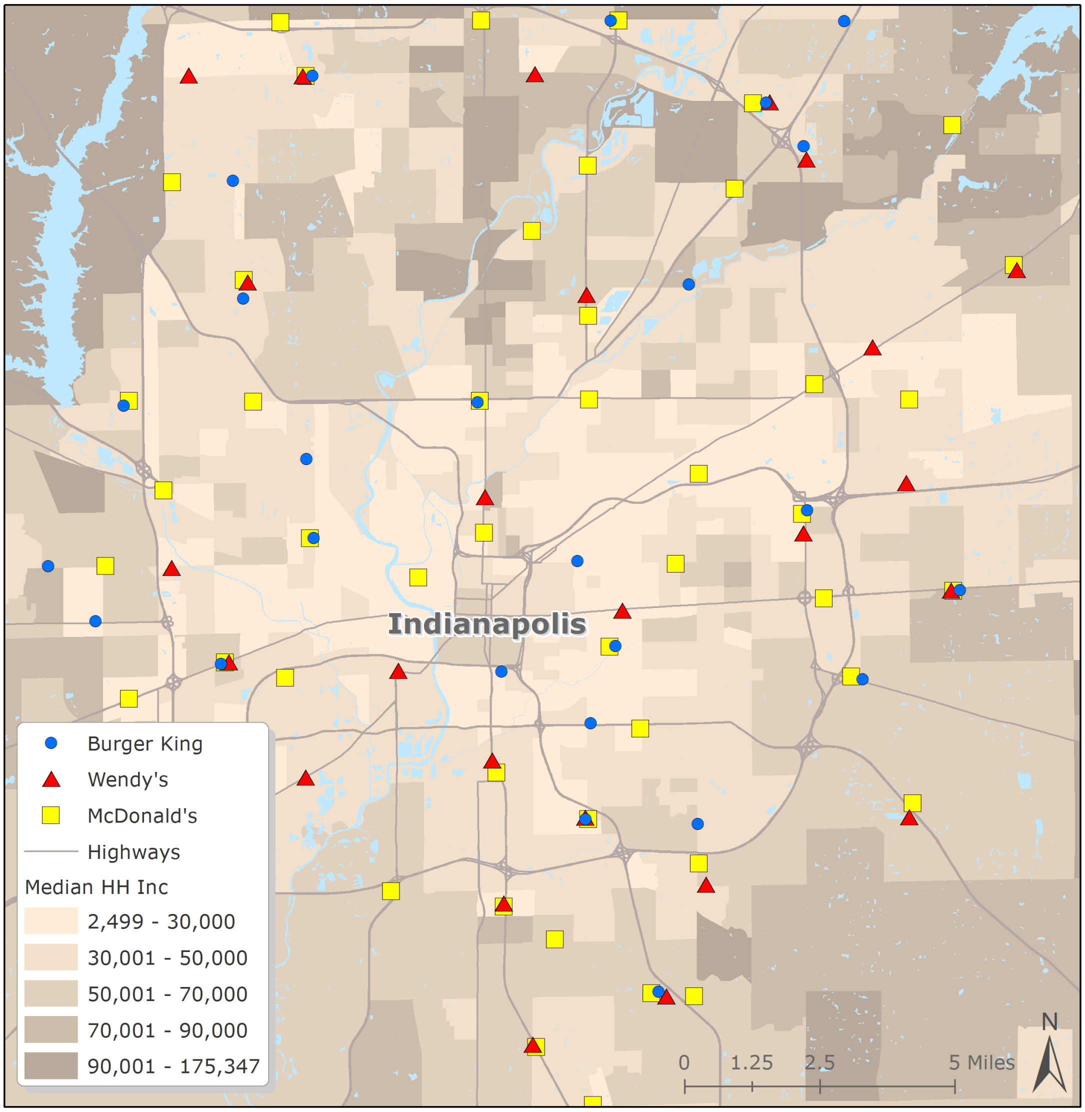


Figure 5-2: Locations of Quick Service Restaurants and Income Distribution

5.3 Results

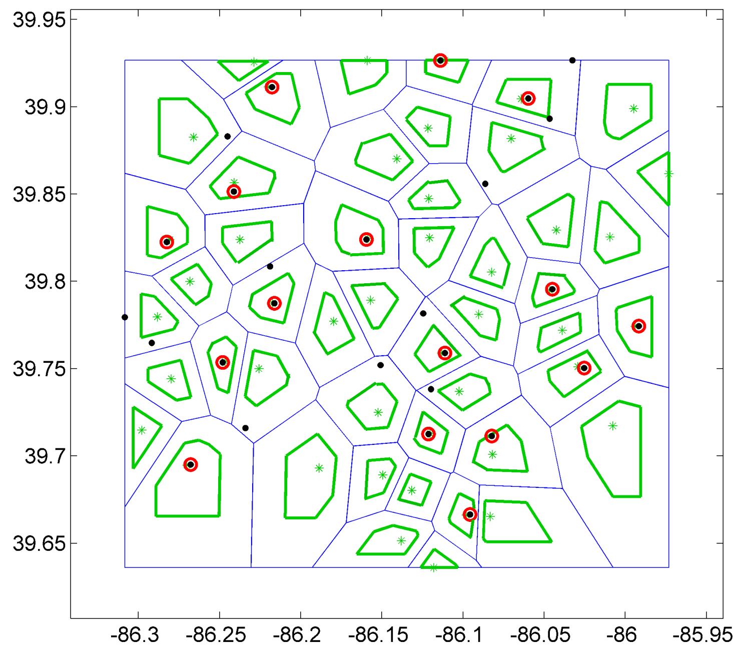
The results for Indianapolis QSR segment are presented in Table 5.1. The results suggest that there is significant positive interaction (attraction) among all three fast food chains in Indianapolis. For all six possible combinations of firms, the null hypothesis of independence can be rejected in favor of the alternative hypothesis of positive association. The results show a slight asymmetry in attraction between the three firms where Burger and Wendy's tend to be slightly more attracted to McDonald's (the largest firm in this market and nationwide), than the reverse. It seems reasonable to assume that these three stores target the same customer segment with similar products and prices. Therefore, given that stores make conscious decisions regarding their location, the results indicate that the market share effect is stronger than the market power effect for this industry. That is, quick service restaurants prefer to share market areas with their competitors, rather than giving up large number of customers as suggested by the theoretical framework in Chapter 3 regarding location of stores competing with close substitutes.

In order to see whether these results correspond to what may be expected by visual inspection, maps of the results are displayed in Figure 5-3, 5-4 and 5-5 for the QSR segment.

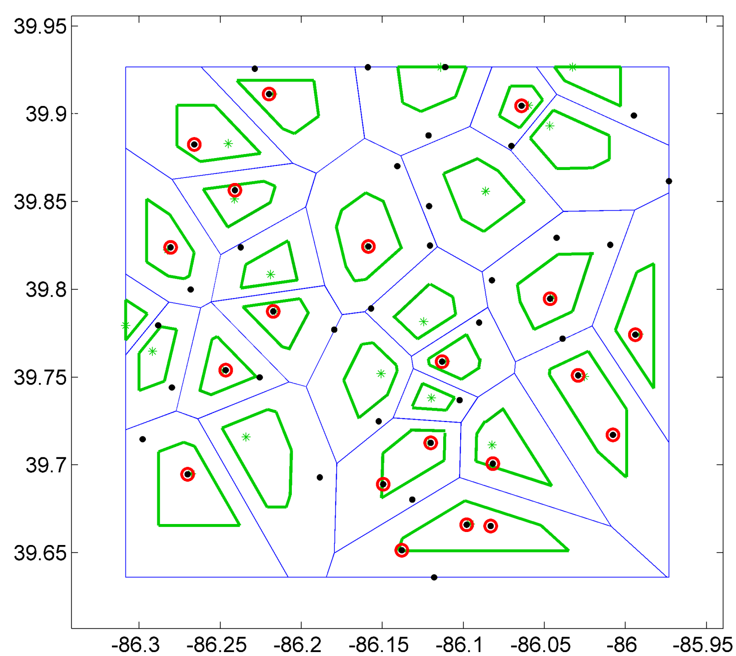
Table 5.1: Test Statistic for Quick Service Restaurants in Indianapolis

A/B	Burger King	McDonald's	Wendy's
Burger King (27)	-	0.5926, $4.1111^{\mathbf{A}}$ (0.1111, 0.4444)	0.4444, $2.3333^{\mathbf{A}}$ (0.0741, 0.4074)
McDonald's (48)	0.4167, $2.6667^{\mathbf{A}}$ (0.125, 0.375)	-	0.4167, $2.6667^{\mathbf{A}}$ (0.1458, 0.375)
Wendy's (25)	0.48, $2.6558^{\mathbf{A}}$ (0.08,0.44)	0.52, $3.1177^{\mathbf{A}}$ (0.08,0.44)	-

Note: The number in parentheses after the company name indicates the number of observations n_A. Each cell provides the $Q_{A\to B}$, $z^{\mathbf{Inf}}_{Q_{A\to B}}$ (where $^{\mathbf{Inf}}$ shows the inference – attraction $^{\mathbf{A}}$, independence $^{\mathbf{I}}$ and repulsion/avoidance $^{\mathbf{R}}$ – at the 5% significance level or $^{Inf^*}$ for 10%), and the 5% critical values from the empirical distribution ($Q^*_{A\to B}$) generated through Monte Carlo simulations.

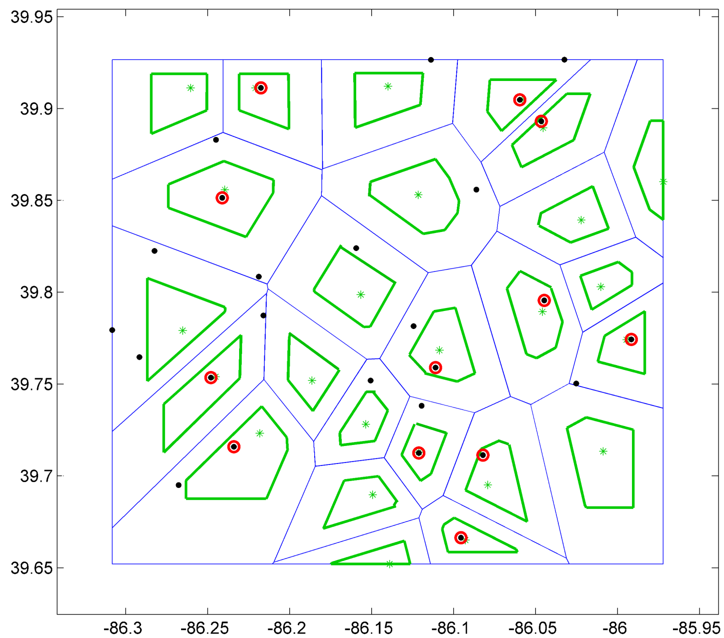


[a]

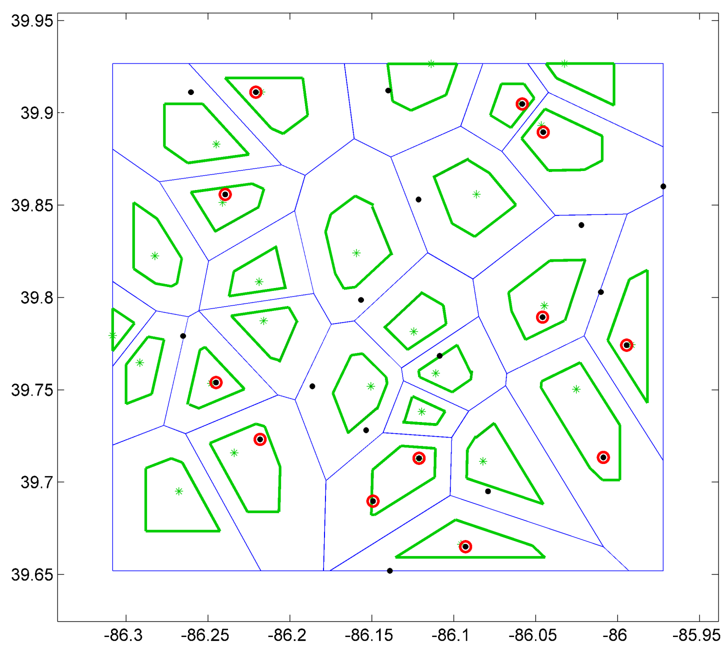


[b]

Figure 5-3: Interactions Between McDonald's and Burger King: (a) shows the location pattern of Burger King in relation to McDonald's and (b) the locations of McDonald's restaurants in relation to Burger King restaurants.

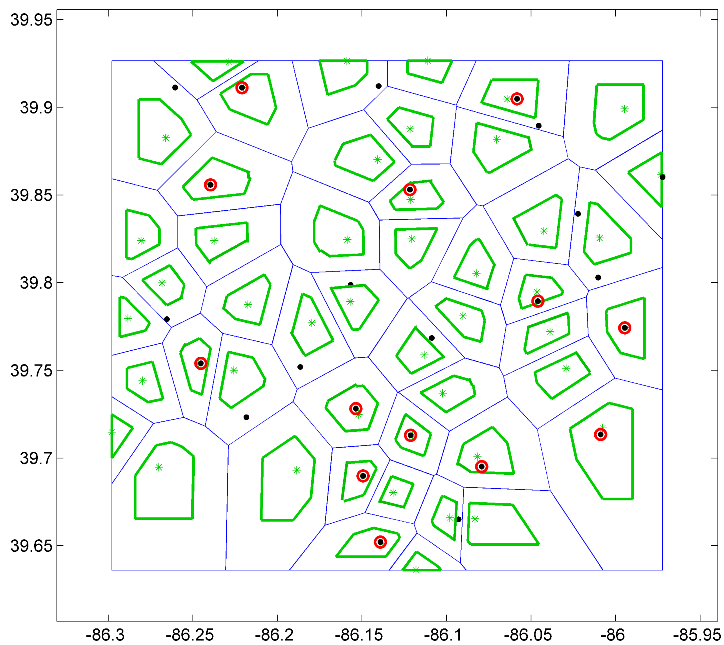


[a]

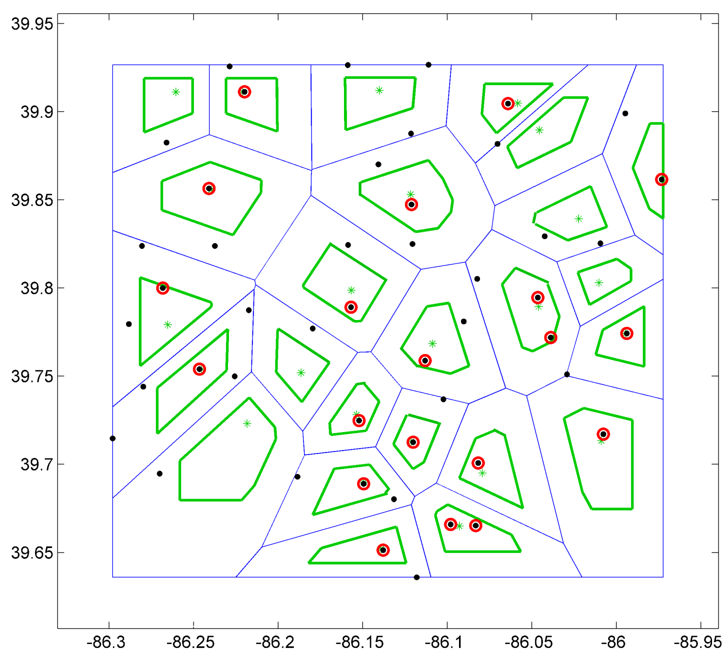


[b]

Figure 5-4: Interactions Between Burger King and Wendy's: (a) shows the location pattern of Burger King in relation to Wendy's and (b) the locations of Wendy's restaurants in relation to Burger King outlets.



[a]



[b]

Figure 5-5: Interactions Between Wendy's and McDonald's: (a) shows the location pattern of Wendy's in relation to McDonald's and (b) the locations of McDonald's restaurants in relation to Wendy's outlets.

5.4 Concluding Comments

In this chapter the statistic from Chapter 4 is applied to competing chain stores in the quick service restaurant segment in the fast food industry. The QSR segment is represented by the three largest firms in the industry: McDonald's, Wendy's and Burger King. The results suggest that there are significant positive interaction (attraction) among all three fast food chains. These results suggest that fast food restaurants may prefer to share market areas with their competitors, rather than give up large number of customers. That is, the market share effect dominates the market power effect. This application shows that observed patterns are effectively captured by the statistic presented in Chapter 4 and that the results can be interpreted based on the theoretical framework in Chapter 3.

Chapter 6

Relative Location Among Retail Chain Stores: The Discount Retailing Sector

To illustrate how the statistic presented in Chapter 4 performs on observed data, it is applied to competing chains within the "big-box" discount retailing sector. The "big-box" discount store segment is represented by Walmart and Target. Together with the analysis performed in Chapter 5 this setting provides a suitable laboratory for studying the relative location of retail chains in two different markets in order to detect potential differences. First, in both industries, each outlet that is constructed by each retail chain is nearly identical in terms of outlet size and product offerings. Second, the prices are similar and the product offerings by the three competing fast food restaurants are reasonably close substitutes (Toivanen and Waterson, 2005) (*i.e.* they compete more directly with each other (Melaniphy, 2007)), while prices and product offerings between the two discount chains differs to a larger extent (*i.e.*, they compete less directly with each other) (Schuetz, 2014; Steverman, 2009).

As mentioned in Chapter 5, since the data necessary to estimate actual cross price elasiticities between the different firms in each market is difficult to obtain, a brief overview of the competitive environment in the discount retailing sector is outlined in the next section. This information should help connect statistical outcomes to findings from the theoretical literature. Do competitors in this market locate next to each other to capture some share of their competitors business, do they seek to establish spatial monopolies by locating away from their competition or do they target different customer segments and locate independently of each other?

6.1 The Discount Retailing Sector

The discount retailing sector in the United States is almost entirely controlled by chains with the top three chains – Walmart, Kmart and Target – accounting for around 73 percent of total sector sales and 54 percent of the discount stores (Jia, 2008). These stores have experimented with different retail formats, most prominently the supercenter which combines a full-service grocery store and a general merchandise outlet. They may also include several ancillary services such as pharmacies, vision centers, *etc.*, to provide consumers with "one-stop" shopping opportunities (Wal-Mart, 2013; Zhu and Singh, 2009). These stores are sometimes referred to as "big-box" stores. "Big-box" stores can be grouped into two broad categories: general merchandise stores (such as Walmart and Target) and specialized stores (such as Lowes, Home Depot, Staples and Office Max) (Schuetz, 2014).

In this application, the focus is directed towards the relative locations of the general merchandice stores Walmart and Target. The business models of Walmart and Target are somewhat different. While Walmart (the leading player) focuses on its "everyday low prices" (Wal-Mart, 2013, p. 4) strategy, Target tries to establish itself as an upscale discount store with focus on the shopping experience, customer service and selling "both everyday essentials and fashionable, differentiated merchandise at discounted prices" (Target, 2013, p. 2), (Schuetz, 2014; Steverman, 2009). That is, while selling similar products, they target two differing customer segments. This is supported by findings in Zhu and Singh (2009) who finds that Target prefers markets with significantly higher income and education while Walmart prefer markets with lower income levels.[1]

6.2 Description of the Data and the Market

To represent "big-box" discount stores in Pittsburgh, Pennsylvania, locations are gathered for Walmart and Target.[2] The location data are displayed in Figure 6-1 and 6-2. The data are verified by crosschecking the Yellow Pages, Google Earth and company-specific web pages. The city of Pittsburgh is located in Allegheny County, Pennsylvania. In this analysis, the county is defined as the market with market boundaries being set by the outermost "big-box" outlets in the county. Within this market area a total of 20 outlets of the two discount retail chains are identified (9 Walmart and 11 Target stores).

Allegheny County has an estimated population of 1,231,527 people with a median household income of $51,366 (in 2013 dollars) as of year 2013. In 2007, the total retail sales reached $20,075,411,000 (or $16,456 per capita). Out of Allegheny County's population, approximately 305,841 people live within Pittsburgh's city limits. In the city, the median household income is lower ($39,195 in 2013 dollars) than to that of the county (Census, 2013).

The distribution of population and median household income within Allegheny county are mapped in Figure 6-1 and Figure 6-2, respectively. These maps are overlayed with the locations of the two "big-box" discount chains considered in this analysis. Most outlets are located in outskirts of the city in areas with relatively lower

[1] Zhu and Singh (2009) also finds that Target stores fare well under competition from other discount stores except when they are located particularly close.

[2] Walmart locations only includes Walmart Supercenters.

population density. This is likely due to the amount of land these big establishments require in comparison to other retailers. Another reason is that consumers are willing to travel a longer distance in order to get to these kind of stores which offers "one-stop" shopping opportunities. These establishments are also often located in close proximity to highways which increases accessibility not only to customers but also to suppliers. It is hard to detect any tendencies between the location of these establishments and particular income areas from the map of median household income in Figure 6-2. However, it does show that median income by block group is increasing as one moves from of the city center.

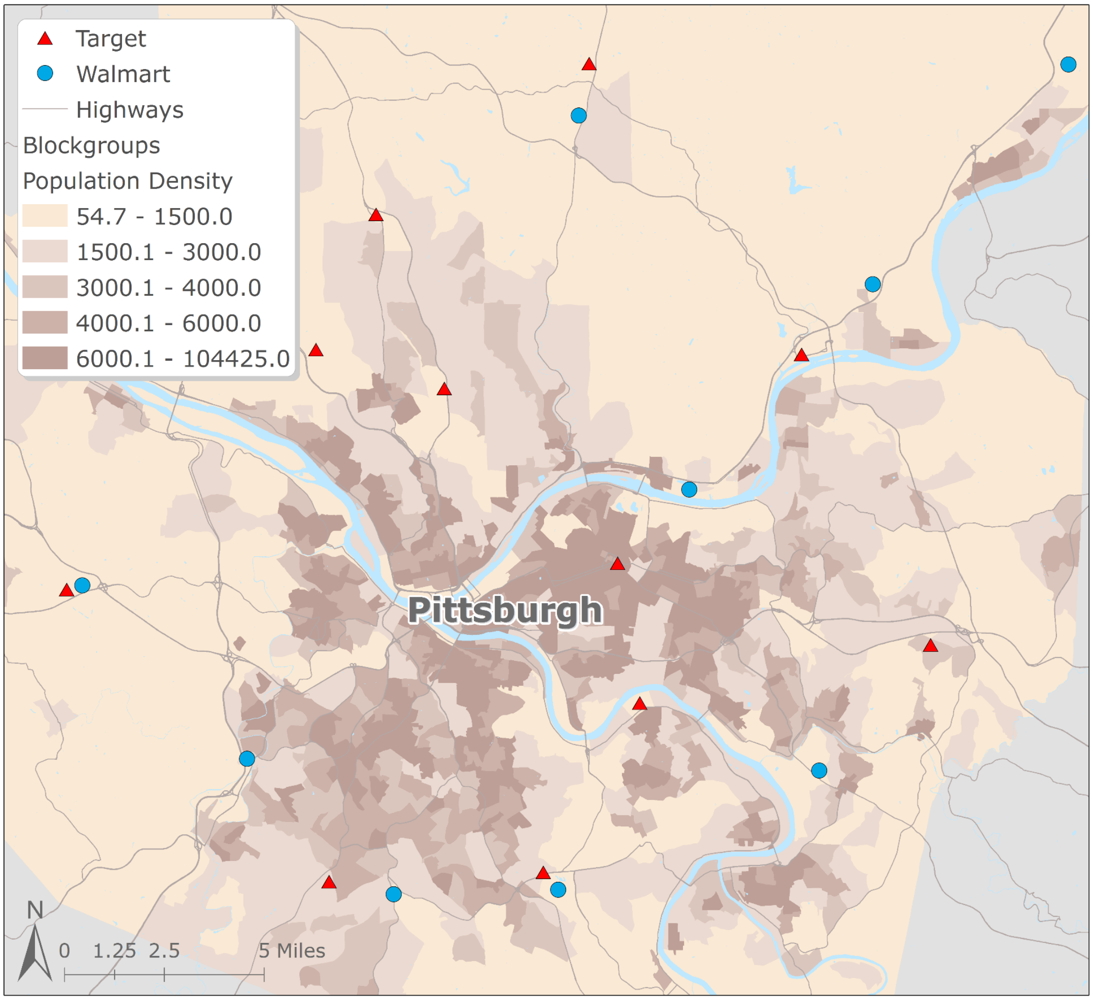


Figure 6-1: Locations of "Big-Box" Retail Stores and Population Density

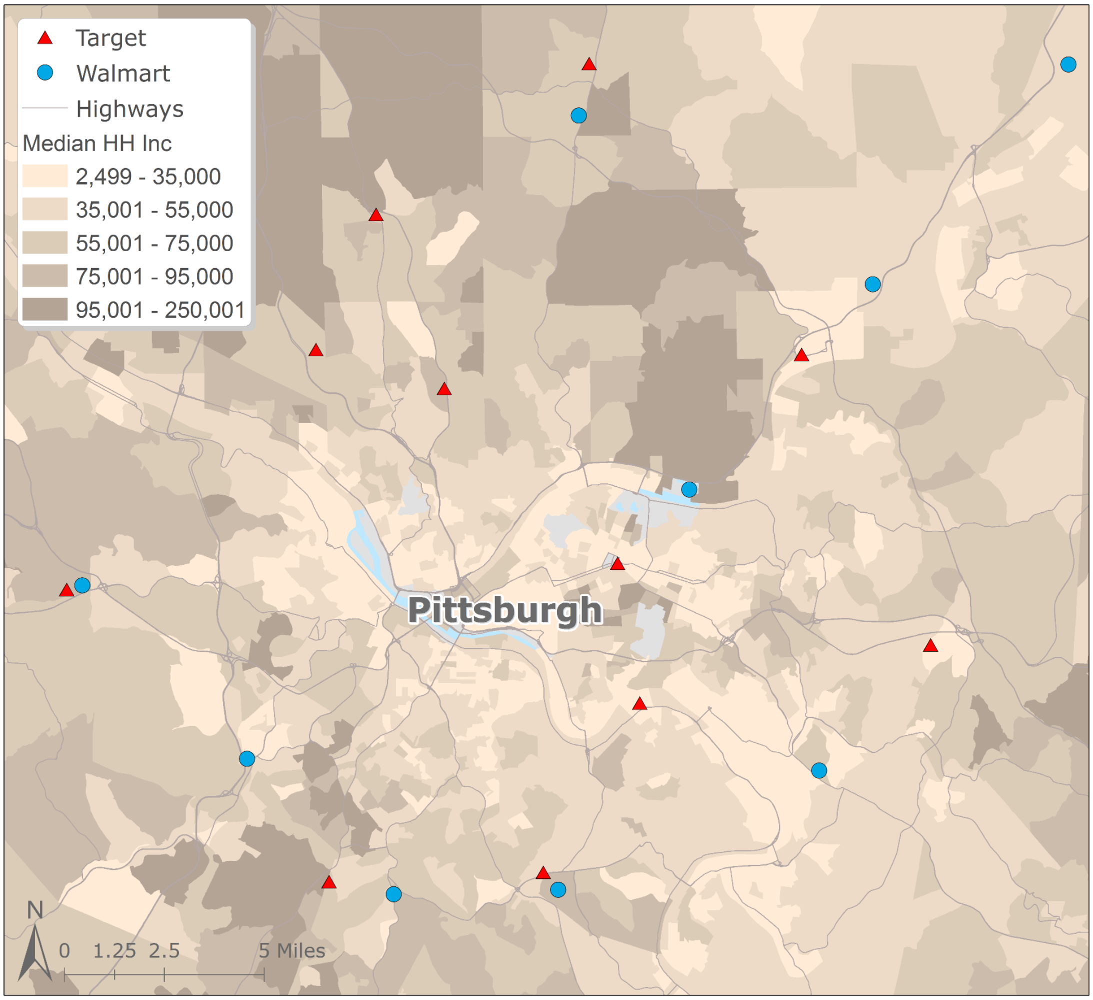


Figure 6-2: Locations of “Big-Box” Retail Stores and Income Distribution

6.3 Results

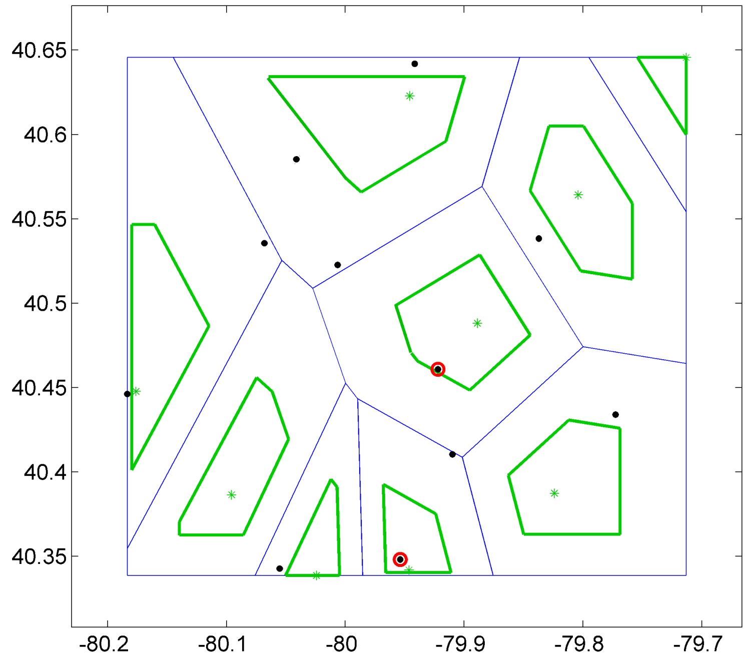
Table 6.1 shows the results for discount "big-box" retailers in Pittsburgh. For this market, the null hypothesis cannot be rejected. This is to say that these two competitors are neither positively or negatively associated, an indication that the two are located independently of each other. One reason for this might be that the two chains are targeting different customer segments, unlike for example the fast food chains described in the previous chapter. While Walmart focuses on its low-price strategy, Target focuses more on the shopping experience and customer service. These results are in-line with previous findings by Zhu and Singh (2009) who find that Target prefers to locate in areas with higher income and education levels while Walmart prefers areas with lower income levels. In light of the theoretical results found in Chapter 3, Target and Walmart may be considered selling imperfect substitutes, where there is still some degree of strategic interactions present. In some areas (where demand is high enough to support two similar stores) firms will locate next to their competitors due to strategic competition over customers, while in other parts of the market they will locate as close as possible to their respective target customer segment.

In order to see whether these results correspond to what may be expected by visual inspection, the results for the "big-box" discount store sector are visualized in Figure 6-3.

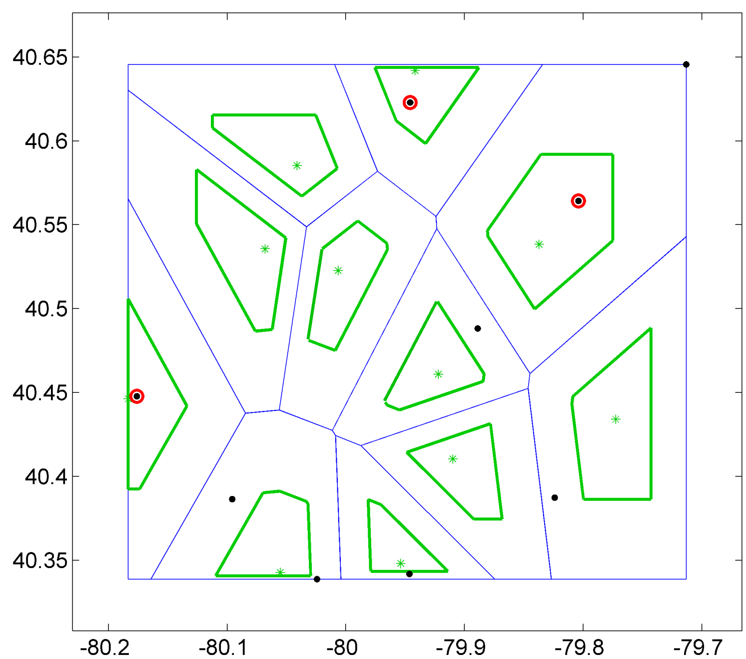
Table 6.1: Test Statistic for "Big-Box" Stores in Pittsburgh

A/B	Walmart	Target
Walmart (9)	-	0.1818, $-0.5224^{\mathbf{I}}$ (0.0909, 0.5455)
Target (11)	0.3333, $0.5774^{\mathbf{I}}$ (0, 0.5556)	-

Note: The number in parentheses after the company name indicates the number of observations n_A. Each cell provides the $Q_{A\to B}$, $z^{\mathbf{Inf}}_{Q_{A\to B}}$ (where $^{\mathbf{Inf}}$ shows the inference – attraction $^{\mathbf{A}}$, independence $^{\mathbf{I}}$ and repulsion/avoidance $^{\mathbf{R}}$ – at the 5% significance level or $^{Inf^*}$ for 10%), and the 5% critical values from the empirical distribution ($Q^*_{A\to B}$) generated through Monte Carlo simulations.



[a]



[b]

Figure 6-3: Results for "Big-Box" Discount Stores: (a) shows the location pattern of Target in relation to Walmart and (b) the locations of Walmart stores in relation to Target stores.

6.4 Concluding Comments

In this chapter the statistic from Chapter 4 is applied to competing chain stores in the "big-box" discount store sector in order to measure the relative location of competitors within this industry. The "big-box" discount store segment is represented by Walmart and Target. Together with the analysis performed in Chapter 5 the relative location of retail chains in two different markets can be compared in order to detect potential differences. While the results suggest that there is significant positive interaction among the competing fast food chains in Chapter 5, the null hypothesis cannot be rejected for the discount "big-box" retailers. One reason for this might be that the two chains are targeting different customer segments, unlike for example fast food chains, making them imperfect substitutes. These results are in-line with previous findings suggesting that Target prefers to locate in areas with higher income and education levels while Walmart prefers areas with lower income levels. This application shows that observed patterns are effectively captured by the statistic and that the results can be interpreted based on the theoretical framework in Chapter 3.

Chapter 7

Effect of Transportation Network on Retail Firm Location

Demand for transportation infrastructure is driven by the spatial separation of consumers and producers. Thus, it can be considered as a derived demand that arises from the need to perform market transactions, which involve consumer travel (Boyer, 1999; McCarthy, 2001; Van Wee, 2002). As such, one primary determinant of the demand for transportation infrastructure is the location and intensity of economic activity that generally forms non-trivial patterns of many establishments. Consequently, predicting demand for transportation infrastructure requires an understanding of firm location behavior and resulting firm location patterns (De Bok and Sanders, 2005; Elgar and Miller, 2006; Kumar and Kockelman, 2008). The inherent endogeneity of the relationship further complicates this relationship as changes in the transportation system also affects the distribution of economic activity due to the benefits associated with being located close to important transportation infrastructure (Elgar et al., 2009; Forkenbrock, 2002; Hicks, 2006; Maoh and Kanaroglou, 2009; Targa et al., 2006).

New transportation infrastructure attracts firms and spurs new development within spatial proximity to it. The same phenomenon makes firms and their consumers more aware of their competition. This gives rise to the strategic interaction (competition) between firms that might not have existed without the transportation infrastructure. The results from the theoretical model in Chapter 3 suggest that strategically interacting firms have an incentive to locate next to each other through competition over market shares. That is, they form nested, positive interaction patterns. Under such circumstances the location patterns deviate from socially-efficient outcomes in terms of congestion, accessibility, and consumer welfare (Nilsson et al., 2014).

The analysis performed in this chapter is intended to examine the effect that transportation infrastructure has on the location choices of competing firms in relation to each other within the overall cluster. For this purpose, the multivariate spatial statistic from Chapter 4 is applied to observed location patterns of competing retail firms to measure interaction patterns among outlets located in proximity of major road transportation infrastructure (versus those that are not) in order to identify any measureable differences. The results provide empirical evidence that proximity to transportation infrastructure changes the location behavior of competing firms with respect to each other.

7.1 Data Description

To illustrate how transportation infrastructure can affect the location behavior of retailers, the statistic presented in Chapter 4 is applied to the quick service restaurant (QSR) industry. For this illustration, location data for the three largest firms in the industry (McDonald's, Wendy's and Burger King) in Indianapolis, Indiana, described in Chapter 5 is used. In addition, location data for these three QSR chains in Memphis, Tennessee, is collected. The two cities chosen for this study, Indianapolis and Memphis, cover a total of 195 outlets in the year 2014. Each city is interpreted as an isolated market with market boundaries being set by the outermost outlets in that market.

To limit the anaylsis, the effect of major road transportation infrastructure on firm location patterns is examined. Major road transportation infrastructure is herein represented by highways. In order to define accessbility, an outlet is considered being located in proximity to the highway if it is within 1/4 mile distance from the highway (with an exit or direct access to it). A distance of 1/2 mile was also tested, but did not change the results significantly. Based on this criterion, two subsets – outlets with and without access – were created for each city. Figure 7-1 provides a visualization of the two markets and a 1/4 mile buffer around highways leading through each respective market. These layers are overlayed on a population density layer showing population per square mile at the block group level to show the distribution of demand.

Overall, a majority of the fast food outlets in Indianapolis and Memphis are concentrated around highways, 124 out of 195 outlets (63.6 percent) are located within $\frac{1}{4}$ mile of a highway. In Indianapolis the corresponding number is 65 percent while Memphis has 62 percent. Among the different chains in both markets, the share of outlets located with access to the highway system is 67, 65 and 58 percent for Wendy's, McDonald's and Burger King, respectively.

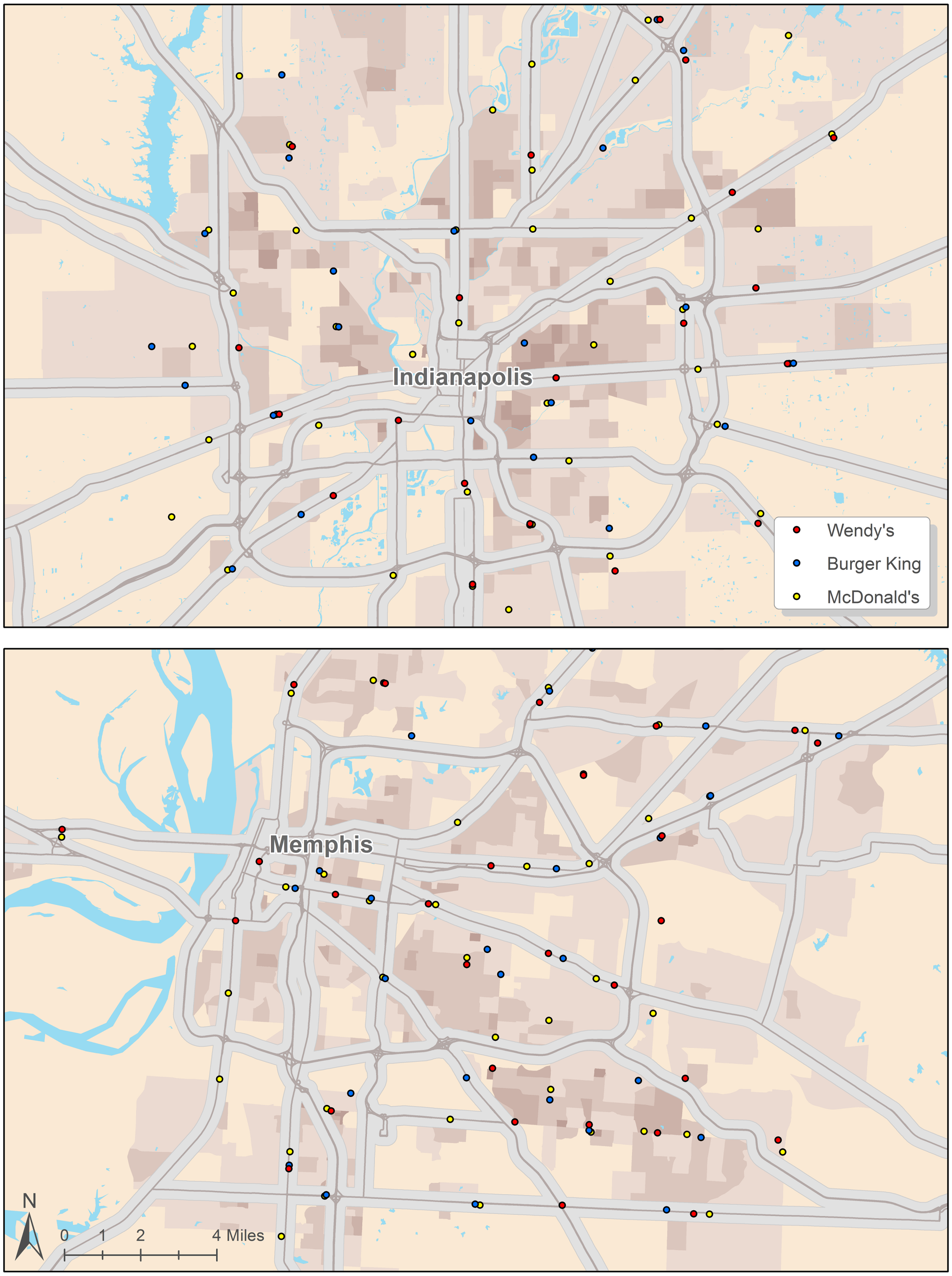


Figure 7-1: Quick Service Restaurants in Indianapolis and Memphis

7.2 Results

Common transportation infrastructure may induce firms to locate in close proximity to one another. To measure whether there is a significant difference in location behavior between competing outlets located in proximity to highways and those that are not, the Q statistic is applied to the full population as well as the two subsets. The results for the Indianapolis and Memphis market are presented in Table 7.1 and 7.2, respectively.

Table 7.1: Test Statistic for Quick Service Restaurants in Indianapolis

A/B	Burger King	McDonald's	Wendy's
Whole population			
Burger King (27)	-	0.5926, 4.1111$^{\mathbf{A}}$ (0.1111, 0.4444)	0.4444, 2.3333$^{\mathbf{A}}$ (0.0741, 0.4074)
McDonald's (48)	0.4167, 2.6667$^{\mathbf{A}}$ (0.125, 0.375)	-	0.4167, 2.6667$^{\mathbf{A}}$ (0.1458, 0.375)
Wendy's (25)	0.48, 2.6558$^{\mathbf{A}}$ (0.08,0.44)	0.52, 3.1177$^{\mathbf{A}}$ (0.08,0.44)	-
With access			
Burger King (15)	-	0.6, 3.1305$^{\mathbf{A}}$ (0.0667, 0.4667)	0.4667, 1.9379$^{\mathbf{A*}}$ (0.0667, 0.4667)
McDonald's (32)	0.4063, 2.0412$^{\mathbf{A}}$ (0.125, 0.4375)	-	0.4375, 2.4495$^{\mathbf{A}}$ (0.125, 0.4063)
Wendy's (18)	0.5, 2.4495$^{\mathbf{A}}$ (0.0556, 0.4444)	0.5556, 2.9938$^{\mathbf{A}}$ (0.0556, 0.4444)	-
Without access			
Burger King (12)	-	0.5, 2.0$^{\mathbf{A}}$ (0, 0.5)	0.3333, 0.6667$^{\mathbf{I}}$ (0.0833, 0.5)
McDonald's (16)	0.4375, 1.7321$^{\mathbf{A*}}$ (0.0625, 0.4375)	-	0.3125, 0.5774$^{\mathbf{I}}$ (0.0625, 0.5)
Wendy's (7)	0.2857, 0.2182$^{\mathbf{I}}$ (0, 0.5714)	0.2857, 0.2182$^{\mathbf{I}}$ (0, 0.5714)	-

Note: The number in parentheses after the company name indicates the number of observations n_A. Each cell provides the $Q_{A\to B}$ and $z^{\mathbf{Inf}}_{Q_{A\to B}}$ (where $^{\mathbf{Inf}}$ shows the inference – attraction $^{\mathbf{A}}$, independence $^{\mathbf{I}}$ and repulsion/avoidance $^{\mathbf{R}}$ – at the 5% significance level). It also shows the 5% critical values from the empirical distribution (lower, upper).

The results from Table 7.1 suggest that when considering the whole population there is significant attraction among all three fast food chains in Indianapolis. However, when applying the statistic on the two subsamples, the results indicate that while there is significant attraction among outlets of the three firms located in proximity to the transportation network, these results do not hold in the subsample that only includes those outlets without such access. In this subsample the positive interaction between Wendy's and the other firms is no longer present. The attraction between McDonald's and Burger King ($Q_{McD \to BK}$) has also weakened, where attraction is only evident at the 10 percent significance level.

In Memphis, the results suggest that overall there is significant attraction among the three chains, except between Wendy's and Burger King. The same results hold for the subsample of outlets with access to the highway system. In the subsample with those outlets not located in proximity to the highway, the attraction between McDonald's and its competitors are no longer present ($Q_{McD \to BK}$ and $Q_{McD \to Wendy's}$ cannot reject the null hypothesis of independence).

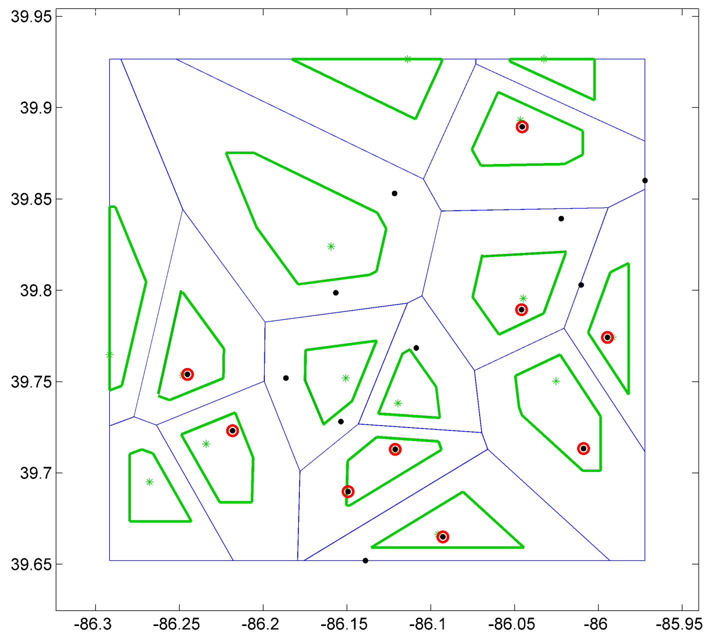
Overall, the results indicate that access to common transportation infrastructure induces firms to locate in close proximity to one another, interactions that may not have been present in the absence of the transportation network. Many of the QSR chains that displayed significant attraction in the joint population and among outlets located in close proximity to important road transportation infrastructure, shows no such tendencies among the outlets with no such access. Two examples of such tendencies are displayed in Figure 7-2, which illustrates the interactions between Wendy's and Burger King in Indianapolis ($Q_{Wendy's \to BK}$) as well as McDonald's and Burger King in Memphis ($Q_{McD \to BK}$). Figure 7-3 shows the lack of interactions among outlets of these two chains when they are not located in proximity to the highway network. These results provide evidence that the presence of major road infrastructure affects the location behavior of competing firms.

Table 7.2: Test Statistic for Quick Service Restaurants in Memphis

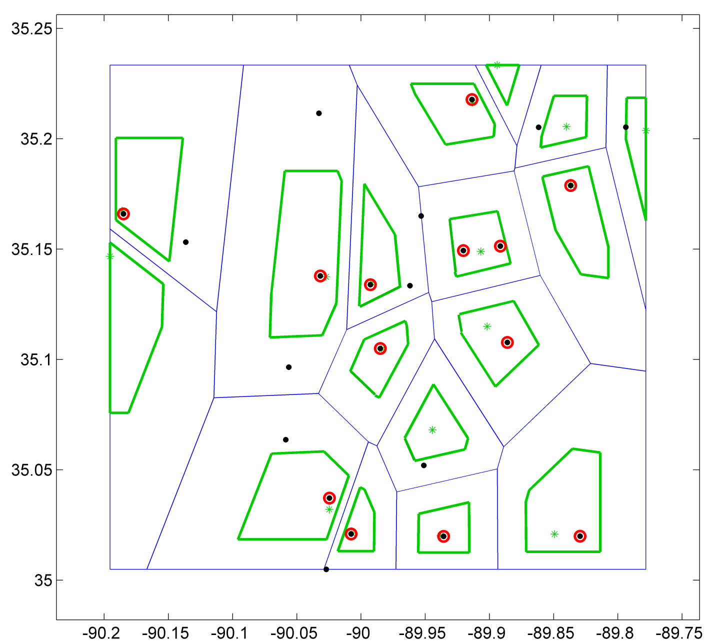
A/B	Burger King	McDonald's	Wendy's
Whole population			
Burger King (28)	-	0.6071, 4.3644$^{\mathbf{A}}$	0.3929, 1.7457$^{\mathbf{I}}$
		(0.1071, 0.3929)	(0.1071, 0.4286)
McDonald's (37)	0.4595, 2.9424$^{\mathbf{A}}$	-	0.4865, 3.3221$^{\mathbf{A}}$
	(0.1081, 0.3784)		(0.1081,0.3784)
Wendy's (30)	0.3667, 1.4757$^{\mathbf{I}}$	0.5667, 4.0056$^{\mathbf{A}}$	-
	(0.1, 0.4)	(0.1, 0.4)	
With access			
Burger King (17)	-	0.6471, 3.7808$^{\mathbf{A}}$	0.2941, 0.4201$^{\mathbf{I}}$
		(0.0588, 0.4706)	(0.0588, 0.4706)
McDonald's (23)	0.5652, 3.4912$^{\mathbf{A}}$	-	0.4348, 2.0466$^{\mathbf{A}}$
	(0.087, 0.4348)		(0.087, 0.4348)
Wendy's (19)	0.3684, 1.1921$^{\mathbf{I}}$	0.5263, 2.7815$^{\mathbf{A}}$	-
	(0.0526, 0.4211)	(0.0526, 0.4737)	
Without access			
Burger King (11)	-	0.7273, 3.6556$^{\mathbf{A}}$	0.3636, 0.9312$^{\mathbf{I}}$
		(0, 0.5455)	(0, 0.5455)
McDonald's (14)	0.4286, 1.543$^{\mathbf{I}}$	-	0.3571, 0.9928$^{\mathbf{I}}$
	(0.0714, 0.5)		(0, 0.4286)
Wendy's (11)	0.2727, 0.1741$^{\mathbf{I}}$	0.6364, 2.9593$^{\mathbf{A}}$	-
	(0, 0.5455)	(0, 0.5455)	

Note: The number in parentheses after the company name indicates the number of observations n_A. Each cell provides the $Q_{A\to B}$ and $z^{\mathbf{Inf}}_{Q_{A\to B}}$ (where $^{\mathbf{Inf}}$ shows the inference – attraction $^{\mathbf{A}}$, independence $^{\mathbf{I}}$ and repulsion/avoidance $^{\mathbf{R}}$ – at the 5% significance level). It also shows the 5% critical values from the empirical distribution (lower, upper).

These results also point to the dynamic and endogeneous interactions between firm location and travel demand. The location of firms are primary determinants of the demand for transport infrastructure, however, the location of transportation infrastructure affects the location behavior of firms. Therefore, these results stress that when predicting changes in travel demand from new transportation infrastructure, there is a need to account for how this affects the (long-term) location behavior of competing firms which in turn also affects the distribution of traffic flows and traffic volumes.

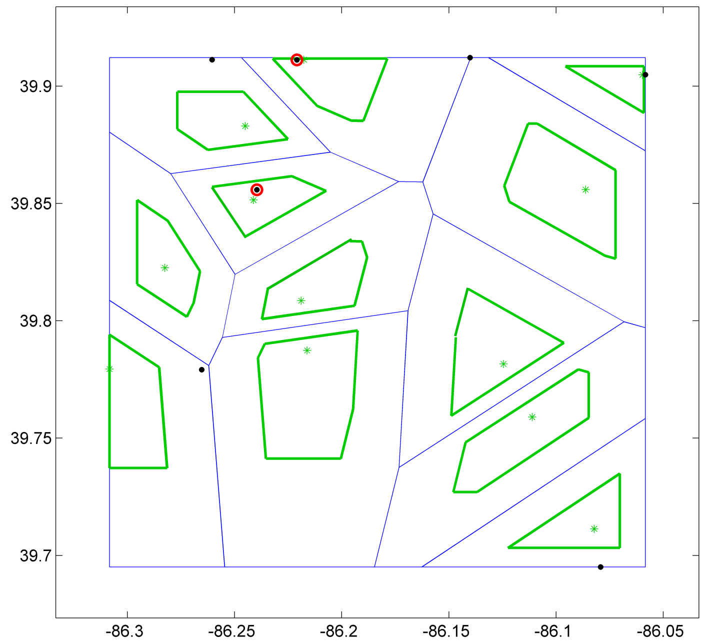


[a]

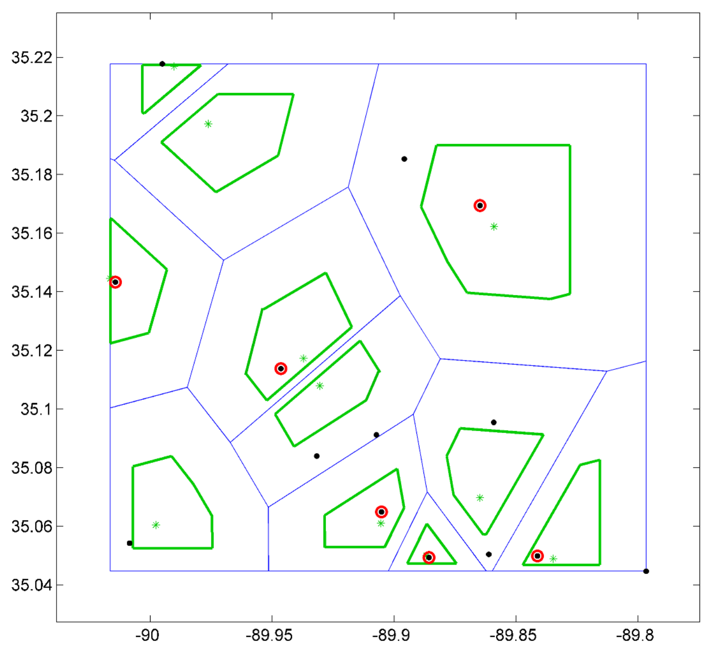


[b]

Figure 7-2: Interactions Among Quick Service Restaurants with Access to the Highway Network: (a) the locations of Wendy's in relation to Burger King, Indianapolis; (b) shows the location pattern of McDonald's in relation to Burger King, Memphis.



[a]



[b]

Figure 7-3: Interactions Among Quick Service Restaurants without Access to the Highway Network: (a) the locations of Wendy's in relation to Burger King, Indianapolis; (b) shows the location pattern of McDonald's in relation to Burger King, Memphis.

7.3 Concluding Comments

This chapter analyzes the effect of transportation infrastructure on retail firm location patterns. Economic activity tends to be clustered around important transportation infrastructure. However, the focus herein is on how transportation infrastructure affects the relative location of competitors. The results suggest that transportation infrastructure affects the nature of the location behavior of firms with respect to their relative location to their competitors which in turn affects the distribution of travel demand and congestion.

The contribution of this application lies in providing empirical evidence that substantiates the notion that transportation infrastructure attracts firms (data shows that a majority of establishments are located in proximity to major transportation infrastructure). It also shows that transportation infrastructure induces firms to locate less uniformly (firms within proximity to highways tend to locate in close proximity to each other). The results show significant attraction among competing firms located with access to major transportation infrastructure. Regarding potential endogeneity, it is unlikely that the kind of retailers studied in this chapter would have an effect on the location of transportation infrastructure as their economic footprint is relatively small.

Chapter 8

Summary of Findings

One of the factors that influence retailers' location decisions is the location of competing outlets. According to the existing theoretical literature there are two opposing forces at play: (i) the market power effect, where firms locate away from each other in order to avoid intense price competition, and (ii) the market share effect, where firms want to locate close to their competitors in competition over market shares. Most of the theoretical literature on retailers location choices has found evidence that support the "market power effect". In this dissertation an alternative game theoretic framework is presented which aims at explaining scenarios where the market share effect dominates the market power effect. The findings from this theoretical analysis adds to the existing literature by explaining why certain types of retailers may want to locate in close proximity to their competition despite increased price competition.

With regards to public policy, the theoretical results suggest that under certain demand conditions public policy can have a welfare-improving effect on firm location choices. In the presence of strategic interactions, regulations (such as zoning laws) could force firms to move away from the Nash-equilibrium location to the socially optimal location where firms and consumers alike are better off. It also has implications for transportation planning. Through the even distribution of such firms, special arrangements and costs associated with congestion could be mitigated and in turn, the impact on roads (or the demand for public infrastructure, in general) be more evenly distributed. The theoretical framework contributes to the literature by explaining frequently observed attraction tendencies among closely competing firms.

While theoretical results are dependent on the assumptions made and the empirical literature is small, realized outcomes of firm location decisions can provide insights into which one of these effects dominate in different retail sectors. For this purpose a multivariate spatial statistic is developed that is aimed at identifying different interaction patterns between competing outlets. In order to define when two outlets are located relatively close to each other, a topological proximity criterion is derived based on the theoretical framework. This criterion takes into account the relative spatial separation between outlets and heterogenity in the pattern of the joint population. The statistic developed has asymptotic properties, and its distribution is approximately normal for larger samples. Finite sample properties and robustness are tested through simulations by varying assumptions regarding population size, expected value of the statistic, interactions between categories of points and underlying spatial processes. The simulation results are consistent and approaches those of the theoretical distribution as the population size increases. The results also confirm that the proposed statistic has the ability to capture not only asymmetrical relationship but also the potential to distinguish pairwise categorical associations from clustering in the joint population.

To demonstrate the usefulness of the statistic, it is applied to competing stores in two different retail sectors in order to detect any differences in the relative location. These two industries are quick service restaurants (QSR) and "big-box" discount chain stores. The results suggest that there are significant positive interaction (attraction) among the competing fast food chains while the null hypothesis cannot be rejected for the discount "big-box" retailers. One reason for this might be that the two "big-box" chains are targeting different customer segments, unlike the fast food chains, thus making them imperfect substitutes. This application shows that observed patterns are effectively captured by the statistic presented in Chapter 4 and that the results can be interpreted based on the theoretical framework in Chapter 3.

This dissertation also takes a closer look at whether common transportation infrastructure may induce retailers to locate in close proximity to one another. More specifically, an analysis is performed to determine how transportation infrastructure affects the relative location of competing firms in the quick service restaurant segment. The contribution of this analysis lies in providing empirical evidence that substantiates the notion that transportation infrastructure attracts firms (the data show that a majority of establishments are located in proximity to major transportation infrastructure). The results also suggest that transportation infrastructure affects the nature of the location behavior of firms with respect to their relative location to their competitors, which in turn affects the distribution of travel demand and congestion. For this analysis the multivariate spatial statistic developed in this thesis is applied to measure whether there is a significant difference in interaction patterns between competing outlets located in proximity to highways and those that are not. The findings from this analysis stress the importance of taking the endogenous relationship between transportation infrastructure and firm location into account, as well as strategic interactions (competition) between firms, in travel demand frameworks.

Future research should investigate the case of firms selling products considered as complements or vertically integrated firms, where [theoretically] co-location should be socially optimal. That is, to extend both the theoretical and the empirical framework to study firms in different product markets and how they interact with products ranging from perfect complements, weak complements to non-related products. Regarding the proposed statistic, future research involves comparing the statistic to commonly used multivariate spatial point patterns methods such as the cross-K function in order to demonstrate differences. As zoning restrictions may sometimes force firms to locate within proximity to each other, the statistic will be extended to take such restrictions into account. This can be done in the simulation of empirical distributions for the statistic by only allowing redistributed points to fall inside areas with proper zoning restrictions. Another area which deserves further investigation is the role of saturation. That is, how much of the attraction tendendencies we observe are due to a saturated market with few available locations left to locate within? For this purpose, a sequential game theoretic framework can be developed. To find empirical evidence, the multivariate statistic can be applied to location patterns over time and its measurement can be used as a dependent variable in a panel data setting.

References

T. Ago. Central agglomeration of monopolistically competitive firms. *Journal of Economic Geography*, 8:811–823, 2008.

J. Aldstadt. Spatial clustering. In M. M. Fischer and A. Getis, editors, *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*, pages 279–300. Springer-Verlag Berlin, 2010.

A. Anderson, D.E. Andersson, Z. Daghbashyan, and B. Harsman. Location and spatial clustering of artists. *Regional Science and Urban Economics*, 47:128–137, 2013.

S. P. Anderson and A. de Palma. Spatial price discrimination with heterogeneous products. *Review of Economic Studies*, 55:573–592, 1988.

S. P. Andersson, N. Schmitt, and J.-F. Thisse. Who benefits from antidumping legislation? *Journal of International Economics*, 38:321–337, 1995.

M. J. Beckmann. *Location Theory.* Random House, Inc., New York, 1968.

M. J. Beckmann. *Lectures on Location Theory.* Springer, New York, 1999.

J. E. Besag. Contribution to the discussion of Dr Ripley's paper. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39:193–195, 1977.

H. Bester. Quality uncertainty mitigates product differentiation. *RAND Journal of Economics*, 29:828–844, 1998.

K. D. Boyer. The accuracy of non-spatial techniques for analyzing transportation demand. *Research in Transportation Economics*, 5:53–74, 1999.

M. Brachert, M. Titze, and A. Kubis. Identifying industrial clusters from a multidimensional perspective: Methodical aspects with an application to Germany. *Papers in Regional Science*, 90:419–440, 2011.

BurgerKing. Annual report. Form 10-k, Burger King Holdings, Inc., 2013.

D. R. Capozza and R. Van Order. A generalized model of spatial competition. *American Economic Review*, 68:896–908, 1978.

Census. Quickfacts beta. www.census.gov/quickfacts Retrieved March 4th, 2015, United States Census Bureau, 2013.

E. Ceyhan. On the use of nearest neighbor contingency tables for testing spatial segregation. *Environmental Ecological Statistics*, 17:247–282, 2010.

N. A. C. Cressie. *Statistics for Spatial Data*. Wiley, New York, 1991.

R. G. Cromley, D. M. Hanink, and G. C. Bentley. Geographically weighted colocation quotients: Specification and application. *The Professional Geographer*, 66:138–148, 2014.

M. F. Dacey. A review of measures of contiquity for two and k-color maps. In B. J. L. Berry and D. F. Marble, editors, *Spatial Analysis: A Reader in Statistical Geography*, pages 479–495. Prentice-Hall, 1968.

C. d'Aspremont, J. Jaskold Gabszewicz, and J.-F. Thisse. On Hotelling's "Stability in competition". *Econometrica*, 47:1145–1150, 1979.

P. Davis. Spatial competition in retail markets: Movie theaters. *RAND Journal of Economics*, 37:964–982, 2006.

M. De Bok and F. Sanders. Firm relocation and accessibility of locations: Empirical results from the Netherlands. *Transportation Research Record: Journal of the Transportation Research Board*, 1902:35–43, 2005.

G. De Fraja and G. Norman. Product differentiation, pricing policy and equilibrium. *Journal of Regional Science*, 33:343–363, 1993.

A. De Palma, V. Ginsburgh, Y. Y. Papageorgiou, and J.-F. Thisse. The principle of minimum differentiation holds under sufficient heterogeneity. *Econometrica*, 53: 767–781, 1985.

P. J. Diggle and A. G. Chetwynd. Second-order analysis of spatial clustering for inhomogeneous populations. *Biometrics*, 47:1155–1163, 1991.

A. Dixit. A model of duopoly suggesting a theory of entry barriers. *Bell Journal of Economics*, 10:20–32, 1979.

P. M. Dixon. Testing spatial segregation using a nearest-neighbor contingency table. *Ecology*, 75:1940–1948, 1994.

P. M. Dixon. Nearest neighbor methods. In A. H. El-Shaarawi and W. W. Piegorsch, editors, *Encyclopedia of Environmetrics*, pages 1370–1383. John Wiley & Sons, Ltd, 2002c.

M. Dudey. Competition by choice: The effect of consumer search on firm location decisions. *American Economic Review*, 80:1092–1104, 1990.

G. Duranton and H. Overman. Testing for localisation using microgeographic data. *Review of Economic Studies*, 72:1077–1106, 2005.

I. Elgar and E. J. Miller. Conceptual model of location of small office firms. *Transportation Research Record: Journal of the Transportation Research Board*, 1977: 190–196, 2006.

I. Elgar, B. Farooq, and E. J. Miller. Modeling location decisions of office firms: Introducing anchor points and constructing choice sets in the model system. *Transportation Research Record: Journal of the Transportation Research Board*, 2133: 56–63, 2009.

G. Ellison and E. Glaeser. Geographic concentration in u.s. manufacturing industries: A dartboard approach. *Journal of Political Economy*, 105:889–927, 1997.

J. H. Fischer and J. E. Harrington. Product variety and firm agglomeration. *RAND Journal of Economics*, 27:281–309, 1996.

D. J. Forkenbrock. Transportation investments and urban form. *Transportation Research Record: Journal of the Transportation Research Board*, 1805:153–160, 2002.

N. Friedman. Burger King hits McDonald's where it hurts. *The Motley Fool*, March 16, 2014.

M. Fujita and J.-F. Thisse. *Economics of Agglomeration*. Cambridge University Press, Cambridge, 1995.

A. Ghosh and S. L. McLafferty. *Location Strategies for Retail and Service Firms*. Lexington Books, Lexington, 1987.

R. Gibson. The burger wars start to sizzle as McDonald's clones Whopper. *The Wall Street Journal*, September 17, 1997.

F. Goreaud and R. Pelissier. Avoiding misinterpretation of biotic interactions with the intertype K_{12}-function: Population independence vs. random labelling hypotheses. *Journal of Vegetation Science*, 14:681–692, 2003.

M. J. Hicks. Transportation infrastructure, retail clustering, and local public finance: Evidence from Wal-Mart's expansion. *Federal Reserve Bank of St. Louis Regional Economic Development*, 2:100–14, 2006.

H. Hotelling. Stability in competition. *Economic Journal*, 39:41–57, 1929.

A. Irmen and J.-F. Thisse. Competition in multi-characteristics spaces: Hotelling was almost right. *Journal of Economic Theory*, 78:76–102, 1998.

P. Jia. What happens when Wal-Mart comes to town: An empirical analysis of the discount retailing industry. *Econometrica*, 76:1263–1316, 2008.

S. Kamal and G.B. Wilcox. Examining the relationship between advertising expenditures and sales of quick-service restaurants in the United States. *Journal of Food Products Marketing*, 20:55–74, 2014.

T. Klier and D. McMillen. Clustering of auto supplier plants in the United States. *Journal of Business Economics and Statistics*, 26:460–471, 2008.

H. Konishi. Concentration of competing retail stores. *Journal of Urban Economics*, 58:488–512, 2005.

R. E. Krider and D. S. Putler. Which birds of feather flock together? Clustering and avoidance patterns of similar retail outlets. *Geographical Analysis*, 45:123–149, 2013.

D. P. Kroese and Z. I. Botev. Spatial process generation. In To appear in: V. Schmidt, editor, *Lectures on Stochastic Geometry, Spatial Statistics and Random Fields, Volume II: Analysis, Modeling and Simulation of Complex Structures*. Springer-Verlag, 2014.

P. Krugman and A. J. Venables. Integration and the competitivness of peripheral industry. In C. Bliss and J. Braga de Macedo, editors, *Unity with Diversity in the European Community*, pages 56–77. CEPR/Cambridge University Press, 1990.

M. Kulldorff. Tests of spatial randomness adjusted for an inhomogeneity: A general framework. *Journal of the American Statistical Association*, 101:1289–1305, 2006.

S. Kumar and K. M. Kockelman. Tracking size, location and interactions of businesses: Microsimulation of firm behavior in Austin, Texas. *Transportation Research Record: Journal of the Transportation Research Board*, 2077:113–121, 2008.

T. F. Leslie and B. J. Kronenfeld. The Colocation Quotient: A new measure of spatial association between categorical subsets of points. *Geographical Analysis*, 43:306–326, 2011.

K. Little. Fast food vs better burger: Inside the burger wars. *CNBC*, January 29, 2015.

H. W. Lotwick and B. W. Silverman. Methods for analysing spatial processes of several types of points. *Journal of the Royal Statistical Society. Series B (Methodological)*, 44:406–413, 1982.

QSR Magazine. NPD: Chains outperforming independent restaurants. http://www.qsrmagazine.com/news/npd-chains-outperforming-independent-restaurants September 2, 2014, Industry News, 2014a.

QSR Magazine. The QSR 50 burger segment. http://www.qsrmagazine.com/reports/qsr50-2014-burger-segment-breakdown August, 2014, Special Report, 2014b.

H. Maoh and P. Kanaroglou. Intrametropolitan location of business establishments:

Microanalytical model for Hamilton, Ontario, Canada. *Transportation Research Record: Journal of the Transportation Research Board*, 2133:33–45, 2009.

E. Marcon and F. Pueach. Measures of the geographic concentration of industries: Improving distance-based methods. *Journal of Economic Geography*, 10:745–762, 2010.

A. Marshall. *Principles of Economics*. MacMillian, London, 1920.

B. Matern. Spatial variation: Stochastic models and their application to some problems in forest surveys and other sampling investigations. Thesis 49:5, Statens Skogsforskningsinstitut, 1960.

P. S. McCarthy. *Transportation Economics: Theory and Practice: A Case Study Approach*. Blackwell Publishers Inc., Malden, 2001.

McDonald's. Annual report. Form 10-k, McDonald's Corportation, 2013.

J.C. Melaniphy. *The Restaurant Location Guidebook*. International Real Estate Location Institute, Inc., Chicago, 2007.

C. E. Miller, J. Reardon, and D. E. McCorkle. The effects of competition on retail structure: An examination of intratype, intertype, and intercategory competition. *Journal of Marketing*, 63:107–120, 1999.

J. S. Netz and B. A. Taylor. Maximum or minimum differentiation? Location patterns of retail outlets. *Review of Economics and Statistics*, 84:162–175, 2002.

I. Nilsson, O. Smirnov, and P. Lindquist. Demand for transportation infrastructure and strategic interactions in firm location choices. *Proceedings of the CLTT Symposium, February 26-27, Gulfport, MS*, 2014.

P. H. Nystrom. *The Economics of Retailing.* Ronald Press Company, New York, 1930.

A. Okabe and I. Yamada. The K-function method on a network and its computational implementation. *Geographical Analysis*, 33:271–290, 2001.

G. Ottaviano, T. Tabuchi, and J.-F. Thisse. Agglomeration and trade revisited. *International Economic Review*, 43:409–435, 2002.

L. Phlips. *Applied Consumption Analysis.* North-Holland, Amsterdam, 1974.

G. A. Picone, D. B. Ridley, and P. A. Zandberg. Distance decreases with differentiation: Strategic agglomeration by retailers. *International Journal of Industrial Organization*, 27:463–473, 2009.

E. C. Pielou. Segregation and symmetry in two-species populations as studied by nearest-neighbor relationships. *Journal of Ecology*, 49:255–269, 1961.

J. Pinkse and M. E. Slade. Contracting in space: An application of spatial statistics to discrete-choice models. *Journal of Econometrics*, 85:125–154, 1998.

B. D. Ripley. Modelling spatial patterns. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39:172–212, 1977.

J. Schuetz. Why are Wal-Mart and Target next-door neighbors? Working paper, Board of Governors of the Federal Reserve System, 2014.

K. Seim. *Spatial Differentiation and Markets Structure: The Video Retail Industry.* Dissertation, Yale University, 2001.

O. Shy. *Industrial Organization.* MIT Press, Cambridge, 1995.

N. Singh and X. Vives. Price and quantity competition in a differentiated duopoly. *RAND Journal of Economics*, 15:546–554, 1984.

B. Steverman. Target vs. Wal-Mart: The next phase. *Bloomberg Business*, August 18, 2009.

F. Targa, K. J. Clifton, and H. S. Mahmassani. Influence of transportation access on individual firm location decisions. *Transportation Research Record: Journal of the Transportation Research Board*, 1977:179–189, 2006.

Target. Annual report. Form 10-k, Target Corporation, 2013.

R. Thomadsen. Product positioning and competition: The role of location in the fast food industry. *Marketing Science*, 26:792–804, 2007.

J. Tirole. *The Theory of Industrial Organization*. MIT Press, Cambridge, 1988.

O. Toivanen and M. Waterson. Market structure and entry: Where's the beef? *RAND Journal of Economics*, 36:680–699, 2005.

B. Van Wee. Land use and transportation: Research and policy challenges. *Journal of Transport Geography*, 10:259–271, 2002.

Wal-Mart. Annual report. Form 10-k, Walmart Stores, Inc., 2013.

Wendy's. Annual report. Form 10-k, The Wendy's Company, 2013.

A. Wolinsky. Retail trade concentration due to consumers' imperfect information. *Bell Journal of Economics*, 14:275–282, 1983.

N. Yang. Burger King and McDonald's: Where's the spillover? *International Journal of the Economics of Business*, 19:255–281, 2012.

T. Zhu and V. Singh. Spatial competition with endogenous location choices: An application to discount retailing. *Quantitative Marketing and Economics*, 7:1–35, 2009.

Appendix A

Mathematical Appendix

Proof of Proposition 1

Substituting p_i in Equation (3.8) into Π_i in (3.7), taking the partial derivative of Π_i with respect to x_i and solving for x_i yields the following three roots:

$$x_i^- = \frac{L}{2} - \frac{\sqrt{t\left(\alpha - c - \frac{tL}{4}\right)}}{t} \tag{A.1}$$

$$x_i = \frac{L}{2} \tag{A.2}$$

$$x_i^+ = \frac{L}{2} + \frac{\sqrt{t\left(\alpha - c - \frac{tL}{4}\right)}}{t} \tag{A.3}$$

Clearly, Equation (A.1) and (A.3) are outside the market. Furthermore, the first order partial derivatives of (3.7) and (3.8) show that they both are zero at $x_i = \frac{L}{2}$:

$$\frac{\partial p_i}{\partial x_i} = \frac{t}{2} - \frac{tx_i}{L}$$

$$\frac{\partial \Pi_i}{\partial x_i} = \frac{1}{\beta}(p_i - c)\,(t(L - 2x_i))$$

Simulating profits as given by substituting (3.8) into (3.7) using parameters that obeys the model assumptions for $x_i \in [0, L]$ confirms that $\frac{L}{2}$ is a maximum. The simulation results are displayed in Figure A-1.

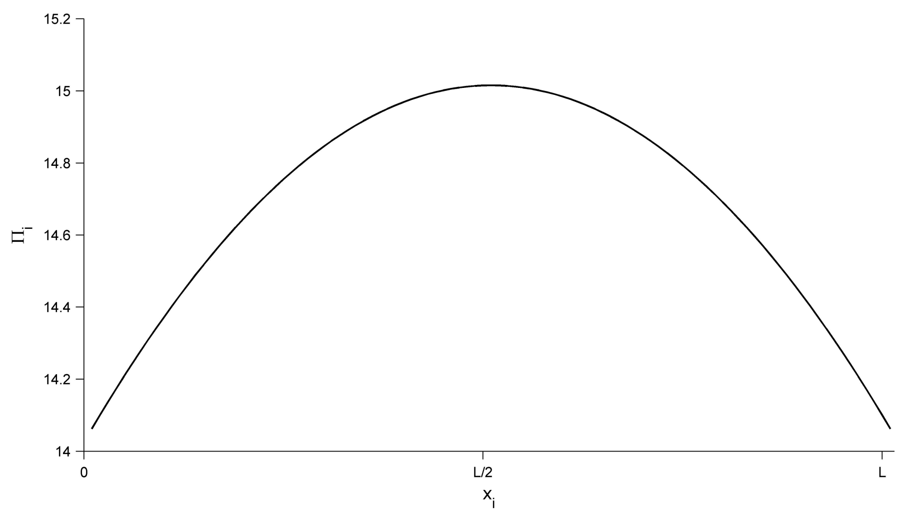


Figure A-1: Firm i's Profit as a Function of x_i

Proof of Proposition 2

Maximing Π_i with respect to p_i yields a quadratic function. When plotted against p_2, p_1^+ is a convex increasing function and is greater than p_2 for all simulated values of p_2 up to α which goes against economic intuition when firms are selling perfect substitutes.[1] The lower root, p_1^-, follows the expected relationship with the other variables.

Both Π_i in Equation (3.9) and p_i in (3.10) are increasing functions of x_i. This can be confirmed by showing that the first order partial derivatives of these functions

[1]Consequently, p_1^+ creates negative profits since firm 1 does not try to undercut or match its rivals price (holding the rivals price p_2 constant).

with respect to x_i are both positive:

$$\frac{\partial p_1}{\partial x_i} = \frac{2}{3}t - \frac{t\left(12\alpha + 6p_2 - 18c + t(162x_1 - 6x_2 + 6L)\right)}{18\sqrt{16\alpha^2 - p_2(A) - c(B) + t(C)}}$$

$$\left(\frac{2}{3}t\right)^2 > \left(\frac{t(12\alpha + 6p_2 - 18c + t(162x_1 - 6x_2 + 6L))}{18\sqrt{16\alpha^2 - p_2(A) - c(B) + t(C)}}\right)^2$$

where $A = 20\alpha - 13p_2$, $B = 12\alpha + 6p_2 - 9c + 6Lt$ and $C = x_1(u_1) + x_2(v_1) - L(z_1)$. For simplicity let $y_1 = 16\alpha^2 - p_2(A) - c(B) + t(C)$.

$$\frac{4}{9}t^2\left(324\left(y_1\right)^2\right) > t(12\alpha + 6p_2 - 18c + t(162x_1 - 6x_2 + 6L))^2$$

This inequality holds for a sufficiently high α, which is already satisfied by the model assumptions aiming at not introducing new entry during the duration of this game. That is, p_i is an increasing function of x_i (confirmed by simulation output as indirectly illustrated in Figure 3-1) and the maximum of this function is found at the (upper) market boundary, $\hat{s} = \frac{L}{2}$. The first order partial derivative of Π_i with respect to x_i yields the follow expression:

$$\frac{\partial \Pi_i}{\partial x_i} = \frac{1}{\beta}(p_i - c)\left[\frac{3}{4}t\hat{s} + \frac{1}{2}(\alpha - p_i + t(x_i - \frac{1}{2}\hat{s})) - 2tx_i\right] \tag{A.4}$$

Given symmetric conditions, in equilibrium $x_1 = x_2 = x$ and hence $p_1 = p_2 = p$ such that $\hat{s} = \frac{L}{2}$, Equation (A.4) can therefore be reduced to:

$$\frac{\partial \Pi_i}{\partial x_i} = \frac{1}{2}\left(\alpha - p + t(x - \frac{3}{2}L)\right)$$

Given that α is sufficiently large, $\frac{\partial \Pi_i}{\partial x_i} > 0$ for $x_i \leq \frac{L}{2}$, and at $x = \frac{L}{2}$:

$$\frac{\partial \Pi_i}{\partial x_i} = \frac{1}{2}(\alpha - p - tL)$$

The above indicates that α needs to be greater than $p + tL$ for this to hold.

Proof of Proposition 3

Substituting p_1 in Equation (3.12) into Π_1 in (3.11), taking the partial derivative of Π_1 with respect to x_1 and solving for x_1 yields the following three roots:

$$x_1^- = \frac{L}{2} - \frac{\sqrt{L^{2\beta t} - 2L^{2\gamma t} - 4x_2^{2\gamma t} - 4\beta t L\left[\gamma(\alpha + tx_2) - \beta(\alpha - c) - p_2\right]}}{2\beta t} \tag{A.5}$$

$$x_1 = \frac{L}{2} \tag{A.6}$$

$$x_1^+ = \frac{L}{2} + \frac{\sqrt{L^{2\beta t} - 2L^{2\gamma t} - 4x_2^{2\gamma t} - 4\beta t L\left[\gamma(\alpha + tx_2) - \beta(\alpha - c) - p_2\right]}}{2\beta t} \tag{A.7}$$

Since Equation (A.5) and (A.7) are outside the market, the optimum location is $x_i = \frac{L}{2}$ for $i = \{1,2\}$. Furthermore, the first order partial derivatives of (3.12) and (3.11) with respect to x_i show that they both are zero at $x_i = \frac{L}{2}$:

$$\frac{\partial p_i}{\partial x_i} = \frac{t}{2} - \frac{tx_i}{L}$$

$$\frac{\partial \Pi_i}{\partial x_i} = \frac{1}{\beta^2 - \gamma^2}(p_i - c)\left(\beta t(L - 2x_i)\right)$$

Simulating profits as given by substituting (3.12) into (3.11) using parameters that obeys the model assumptions confirms that $\frac{L}{2}$ is a maximum. These results are displayed in Figure A-2 for $\gamma = 0.25,\ 0.5,\ 0.75,\ 0.9$.

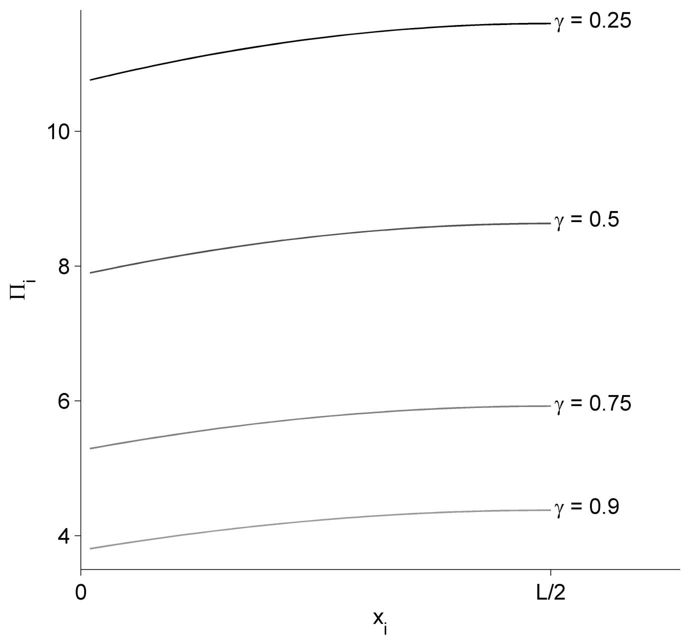


Figure A-2: Firm i's Profit as a Function of x_i under Different Values of γ

Proof of Proposition 4

Defining $\Pi_1 + \Pi_2$ as producer surplus and $Q_1 + Q_2$ as consumer surplus, show that total surplus is higher at $\frac{L}{4}$ than at $\frac{L}{2}$ when $\gamma = 1$. First, recall that given symmetric conditions in equilibrium $x_1 = x_2$ and $p_1 = p_2$. Applying these equilibrium conditions to Equation (3.9) and (3.10) for $x_1 = x_2 = \frac{L}{2}$ and $x_1 = x_2 = \frac{L}{4}$ yields the following:

$$\Pi_i^{\frac{L}{2}} = \left(p_i^{\frac{L}{2}} - c\right) Q_i^{\frac{L}{2}}$$

$$\Pi_i^{\frac{L}{4}} = \left(p_i^{\frac{L}{4}} - c\right) Q_i^{\frac{L}{4}}$$

$$Q_i^{\frac{L}{2}} = \frac{1}{\beta}\left[\frac{L}{2}\left(\alpha - p_i^{\frac{L}{2}} + \frac{L}{4}t\right) - \left(\frac{L}{2}\right)^2 t\right]$$

$$Q_i^{\frac{L}{4}} = \frac{1}{\beta}\left[\frac{L}{2}\left(\alpha - p_i^{\frac{L}{4}}\right) - \left(\frac{L}{4}\right)^2 t\right]$$

$$p_i^{\frac{L}{2}} = \frac{1}{2}(\alpha + c) + Lt - \frac{1}{2}\sqrt{(\alpha^2 + 5L^2t^2 - 2\alpha c + c^2)}$$

$$p_i^{\frac{L}{4}} = \frac{1}{2}(\alpha + c) + \frac{7}{8}Lt - \frac{1}{8}\sqrt{(16\alpha^2 + Lt(8c - 8\alpha + 57Lt) - 32\alpha c + 16c^2)}$$

Show that $Q_i^{\frac{L}{4}} > Q_i^{\frac{L}{2}}$:

$$\frac{L}{2}(p_i^{\frac{L}{2}} - p_i^{\frac{L}{4}}) + \frac{L^2}{16}t > 0$$

where $(p_i^{\frac{L}{2}} - p_i^{\frac{L}{4}}) < 0$:

$$\frac{1}{8}\left(Lt + \sqrt{(16\alpha^2 + Lt(8c - 8\alpha + 57Lt) - 32\alpha c + 16c^2)}\right)$$
$$-\frac{1}{2}\sqrt{(\alpha^2 + 5L^2t^2 - 2\alpha c + c^2)} < 0$$

Meaning that for consumers it becomes a trade-off between a lower price at $x = \frac{L}{2}$ and lower transportation costs at $x_i = \frac{L}{4}$. Given the model assumptions and the relatively small difference in prices, the following inequality holds:

$$\frac{L}{2}(p_i^{\frac{L}{2}} - p_i^{\frac{L}{4}}) < t\frac{L^2}{16}$$

Meaning that for consumers the benefits from lower transportation costs outweighs those of a lower price. Recall that, at $x_i = \frac{L}{4}$ firms still compete about the indifferent consumer. Although the price competition is not as fierce as when they are located next to each other, it is still present, such that the difference in prices between the two locations is small. Since firms are facing both a higher demand and a higher price around $x_i = \frac{L}{4}$, producer surplus is also higher if both firms were to locate more uniformly:

$$(Q_i^{\frac{L}{4}} - Q_i^{\frac{L}{2}}) + (p_i^{\frac{L}{4}} - p_i^{\frac{L}{2}}) > 0$$

$$\left(\frac{1}{8} - \frac{L}{16}\right)\left(Lt + \sqrt{(16\alpha^2 + Lt(8c - 8\alpha + 57Lt) - 32\alpha c + 16c^2)}\right) \\ - \left(\frac{1}{2} - \frac{L}{4}\right)\sqrt{(\alpha^2 + 5L^2t^2 - 2\alpha c + c^2)} + t\frac{L^2}{16} > 0$$

Appendix B

Source Code for Q-Statistic

Listing B.1: The Q-statistic

```
%%% Import data
clear all
% FIRM 1 data - A
[~, ~, raw] = xlsread('firm1data.csv');
raw = raw(2:end,2:end);
data = cell2mat(raw);
x1 = data(:,2);
y1 = data(:,1);
xy1=unique([x1,y1],'rows');
x1=xy1(:,1);
y1=xy1(:,2);
clearvars data raw columnIndices; % clear temporary variables

% FIRM 2 data - B
[~, ~, raw] = xlsread('firm2data.csv');
raw = raw(2:end,2:end);
data = cell2mat(raw);
```

```
x = data(:,2);
y = data(:,1);
xy=unique([x,y],'rows');
x=xy(:,1);
y=xy(:,2);
clearvars data raw columnIndices; % clear temporary variables

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%% Calculating Q(1-->2)%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

[Q_12, z_12, Inference, bnd] = Q12stat(x1,y1,x,y);
Q_12
Inference

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%% Simulation based significance testing %%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% 1. Pick number of iterations
iter = 600;

% 2. Choose one of the following methods to perform the
%      simulations:

%%% Alternative1: Monte Carlo Simulation w/ varying buffers
%vb_buff=1/3;
```

```
%[Q12statsim] = vb_siminfQ12stat(x1,y1,x,y,xy,iter,bnd,vb_buff);


%%% Alternative2: Monte Carlo Simulation w/ varying buffers
%vb_buff=1/4;
%[Q12statsim] = vb_siminfQ12stat(x1,y1,x,y,xy,iter,bnd,vb_buff);


%%% Alternative3: Monte Carlo Simulation w/o buffers
Q12statsim = zeros(iter, 1);
disp('Computing...'); drawnow;
tic;
parfor simNum = 1:iter
    aminx= bnd(:,1);
    bmaxx= bnd(:,2);
    aminy= bnd(:,3);
    bmaxy= bnd(:,4);
    xsim=(bmaxx-aminx).*rand(length(x1),1)+aminx;
    ysim=(bmaxy-aminy).*rand(length(y1),1)+aminy;
    xysim=unique([xsim,ysim],'rows');
    xsim=xysim(:,1);
    ysim=xysim(:,2);
    Q12statsim(simNum) = simQ12stat(xsim,ysim,x1,y1,x,y);
end
toc;


% 3. Post Process - plot simulated distribution
figure
xvalues1=0:0.05:1;
```

```
hist(Q12statsim, xvalues1);
mean_Q12statsim = mean(Q12statsim);
std_Q12statsim = std(Q12statsim);
qnt2575_Q12statsim = quantile(Q12statsim, [0.25 0.75])

% 4. Hypothesis testing based on simulated distribution
signlevel = 5; %Choose significance level 1,5 or 10
[Inf_sim,Q12cv] = MC_inference(Q_12,Q12statsim,signlevel,iter);
Q12cv %1st row = significance levels, 2nd row = lower crit val,
       %3rd row = upper crit val
Inf_sim
```

Listing B.2: Calculating the Q-statistic and Inference

```
function [Q_12, z_12, Inference, bnd] = Q12stat(x1,y1,x,y)
%where x1 & y1 are firm 1's location and x & y firm 2's

%%%%% VoronoiLimit by Jakob Sievers (slightly modified)%%%%%
warning('off','map:polygon:noExternalContours');

if ~any(size(x)==1) || ~any(size(y)==1) ||
    numel(x)==1 || numel(y)==1
    disp('Input vectors should be single rows or columns')
    return
end

x=x(:);
```

```
y=y(:);

% Market boundaries: Find min and max x and y of both data sets
if min(x) <= min(x1) %minimum x value
    minx = min(x);
else
    minx = min(x1);
end

if max(x) >= max(x1) %maximum x value
    maxx = max(x);
else
    maxx = max(x1);
end

if min(y) <= min(y1) %minimum y value
    miny = min(y);
else
    miny = min(y1);
end

if max(y) >= max(y1) %maximum y value
    maxy = max(y);
else
    maxy = max(y1);
end
```

```
bnd=[minx maxx miny maxy]; %data bounds
crs=double([bnd(1) bnd(4);bnd(2) bnd(4);bnd(2) bnd(3);
    bnd(1) bnd(3);bnd(1) bnd(4)]); %data boundary corners

%Triangulation and creation of Thiessen polygons/Voronoi diagram
dt=DelaunayTri(x(:),y(:));
[V,C]=voronoiDiagram(dt);
%This structure gives vertices for each individual point but
%is missing all "infinite" vertices - C = cells/polygons,
%V = corresponding vertices for cell/polygons
[vx,vy]=voronoi(x,y);
%This structure includes the "infinite" vertices but provides
%everything as a completele list of vertices rather than
%individually for each point. Hence we need to add the missing
%vertices from vx and vy to the V and C structure.
vxyl=[vx(:) vy(:)];
xix=ones(size(vx));

%Eliminate spurious double entries in V(C{ij})
av=cell(length(C),1);
for ih=1:length(C)
    av{ih}=[V(C{ih},1),V(C{ih},2)];
    if size(unique(av{ih},'rows'),1)<size(av{ih},1);
        k=0;
        ctr=0;
        while k==0
            ctr=ctr+1;
```

```
            if ctr==1
                tt=unique(V(C{ih}(ctr),:)==V(C{ih}(end),:));
                if length(tt)<2 && tt==1
                    C{ih}(end)=[];
                    k=1;
                else
                  tt=unique(V(C{ih}(ctr),:)==V(C{ih}(ctr+1),:));
                  if length(tt)<2 && tt==1
                      C{ih}(ctr+1)=[];
                      k=1;
                  end
                end
            else
                tt=unique(V(C{ih}(ctr),:)==V(C{ih}(ctr+1),:));
                if length(tt)<2 && tt==1
                    C{ih}(ctr+1)=[];
                    k=1;
                end
            end
        end
    end
end
lV0=length(V);

%Find missing pts that should be added to existing V/C structure
for ii=1:length(vxyl)
    fix=find(V(:,1)==vxyl(ii,1));
```

```
    if ~isempty(fix)
        if any(V(fix,2)==vxyl(ii,2))
            xix(ii)=0;
        end
    end
end

mix=find(xix==1)./2; %index of missing values
lmix=length(mix);
mvx=vx(2,mix); %missing vx
mvy=vy(2,mix); %missing vy
mv=[mvx',mvy'];
cpx=vx(1,mix); %connector point x
%(connects btw outer missing pts & inner existing pts in V/C)
cpy=vy(1,mix); %connector point y
%(connects btw outer missing pts & inner existing pts in V/C)

ctr=0;
mv2=[];
cpVixt=cell(lmix,1); %connector points, index in V structure
for ii=1:lmix
    if any(V(:,1)==cpx(ii) & V(:,2)==cpy(ii))
        cpVixt{ii}=find(V(:,1)==cpx(ii) & V(:,2)==cpy(ii));
        lval=length(cpVixt{ii});
        if lval==1
            ctr=ctr+1;
            mv2(ctr,:)=mv(ii,:);
```

```
        elseif lval>1
            ctr=ctr+1;
            mv2(ctr:ctr+lval-1,:)=
            [ones(lval,1).*mv(ii,1) ones(lval,1).*mv(ii,2)];
            ctr=ctr+lval-1;
        end
    end
end
cpVixt=cell2mat(cpVixt);



V=[V;mv2]; %add pts to V structure
allVixinp=inpolygon(V(:,1),V(:,2),crs(:,1),crs(:,2));
%determine which pts in V that are within the data boundaries.

%Addition-routine:
%add missing pts (mvx,mvy) to individual vertice-polygons(C)
for ij=1:length(C)
    if any(C{ij}==1)
        ixa=find(cpVixt==C{ij}(2));
        ixb=find(cpVixt==C{ij}(end));
        if length(C{ij})<3 %corner point detected
            C{ij}(1)=lV0+ixa(1);
            C{ij}=[C{ij},lV0+ixa(2)];
        else
            if length(ixa)==1 && length(ixb)==1
                C{ij}(1)=lV0+ixa;
```

```
            C{ij}=[C{ij},lV0+ixb];
        elseif length(ixa)==2 && length(ixb)==1
            C{ij}=[C{ij},lV0+ixb];
            [~,minix]=
            min(sqrt((V(C{ij}(end),1)-V(cpVixt(ixa),1)).^2
            +(V(C{ij}(end),2)-V(cpVixt(ixa),2)).^2));
            C{ij}(1)=lV0+ixa(minix);
        elseif length(ixa)==1 && length(ixb)==2
            C{ij}(1)=lV0+ixa;
            [~,minix]=min(sqrt((V(C{ij}(1),1)
            -V(cpVixt(ixb),1)).^2+(V(C{ij}(1),2)
            -V(cpVixt(ixb),2)).^2));
            C{ij}=[C{ij},lV0+ixb(minix)];
        elseif length(ixa)==2 && length(ixb)==2
            [~,minix1]=min(sqrt((x(ij)
            -V(lV0+ixa,1)).^2+(y(ij)-V(lV0+ixa,2)).^2));
            [~,minix2]=min(sqrt((x(ij)
            -V(lV0+ixb,1)).^2+(y(ij)-V(lV0+ixb,2)).^2));
            C{ij}(1)=lV0+ixa(minix1);
            C{ij}=[C{ij},lV0+ixb(minix2)];
        end
      end
    end
end

%[V(:,1),V(:,2)]=poly2ccw(V(:,1),V(:,2));
```

```
%Polybool for restriction of polygons to domain.
C1=C;
%Do this analysis based on old vertice descriptions
%to avoid problems
for ij=1:length(C)
    if sum(allVixinp(C{ij}))~=length(C{ij})
        [xb, yb] = polybool('intersection',crs(:,1),crs(:,2),
        V(C1{ij},1),V(C1{ij},2));
        ix=nan(1,length(xb));
        for il=1:length(xb)
            if any(V(:,1)==xb(il)) && any(V(:,2)==yb(il))
                ix1=find(V(:,1)==xb(il));
                ix2=find(V(:,2)==yb(il));
                for ib=1:length(ix1)
                    if any(ix1(ib)==ix2)
                        ix(il)=ix1(ib);
                    end
                end
                if isnan(ix(il))==1
                    lv=length(V);
                    V(lv+1,1)=xb(il);
                    V(lv+1,2)=yb(il);
                    allVixinp(lv+1)=1;
                    ix(il)=lv+1;
                end
            else
                lv=length(V);
```

```
                V(lv+1,1)=xb(il);
                V(lv+1,2)=yb(il);
                allVixinp(lv+1)=1;
                ix(il)=lv+1;
            end
        end
        C{ij}=ix;
    end
end



%%%%% Creating Co-location Areas by Isabelle Nilsson %%%%%
n = max(cellfun('length',C)) +1;
m = length(C);
x_mid = NaN([m,n]);
y_mid = NaN([m,n]);

for i = 1: length(C)
    vxi = V(C{i},1);
    for j = 1:length(vxi)
        mxi(j) = (x(i) + vxi(j))/2;
        x_mid(i,j) = mxi(j);
    end
end

for k = 1: length(C)
    vyi = V(C{k},2);
```

```
    for g = 1:length(vyi)
        myi(g) = (y(k) + vyi(g))/2;
        y_mid(k,g) = myi(g);
    end
end

xcoloc = reshape(x_mid.',1,[]);
ycoloc = reshape(y_mid.',1,[]);
[x_coloc, y_coloc] = removeExtraNanSeparators(xcoloc, ycoloc);


%%%%% inpolygons(x,y,xv,yv) by Kelly Kearney %%%%%

% Check inputs

if size(x1) ~= size(y1)
    error('x and y must have the same dimensions');
end

if ~isvector(x_coloc) || ~isvector(y_coloc) ||
    length(x_coloc) ~= length(y_coloc)
    error('xv and yv must be vectors of the same length');
end

% Find number of and starting indices of polygons

[x12, y12] = polysplit(x_coloc, y_coloc);
```

```
[xsplit, ysplit] = poly2cw(x12, y12);
isCw = ispolycw(xsplit, ysplit);
mainPolyIndices = find(isCw);
nHolesPer = diff([mainPolyIndices;length(isCw)+1]) - 1;

% Test if points are in each polygon

originalSize = size(x1);
x1 = x1(:);
y1 = y1(:);

isIn = zeros(length(x1), length(mainPolyIndices));
for ipoly = 1:length(mainPolyIndices)
    isInMain = inpolygon(x1, y1, xsplit{mainPolyIndices(ipoly)},
    ysplit{mainPolyIndices(ipoly)});
    if nHolesPer(ipoly) > 0
        isInHole = zeros(length(x1), nHolesPer(ipoly));
        for ihole = 1:nHolesPer(ipoly)
            isInHole(:,ihole) = inpolygon(x1, y1,
            xsplit{mainPolyIndices(ipoly)+ihole},
            ysplit{mainPolyIndices(ipoly)+ihole});
        end
        isIn(:,ipoly) = isInMain & ~any(isInHole,2);
    else
        isIn(:,ipoly) = isInMain;
    end
end
```

```
in = any(isIn, 2);
in = reshape(in, originalSize);

if nargout == 2

    index = num2cell(zeros(size(x1)));
    for ipoint = 1:length(x1)
        loc = find(isIn(ipoint,:));
        if ~isempty(loc)
            index{ipoint} = loc;
        end
    end

    index = reshape(index, originalSize);
end



%%%%% Visualization %%%%%
figure
box on
hold on
plot(x,y,'.k')
voronoi(x,y,'w')
for id=1:length(C)
    h0=plot(V(C{id},1),V(C{id},2),'-b');
end
```

```
dx=(bnd(2)-bnd(1))/10;
dy=(bnd(4)-bnd(3))/10;
axis([bnd(1)-dx bnd(2)+dx bnd(3)-dy bnd(4)+dy])
%title('The Q-Statistic') %optional
hold on
h1=plot(x_coloc, y_coloc, '-g');
%hold on
h2=plot(x, y, '.g');
%hold on
h3=plot(x1(in),y1(in),'ro');
%hold on
h4=plot(x1, y1,'.k');
%legend([h0 h1 h4 h2 h3],{'Voronoi decomposition',
%'Co-location areas','A','B','A within B Attraction Area'},
%'Location','NorthEastOutside') %optional



%%%%% Q-statistic %%%%%
A = abs(maxx - minx)*abs(maxy - miny);

a = zeros(1,length(mainPolyIndices));
for f = 1:length(mainPolyIndices)
    areas = polyarea(xsplit{mainPolyIndices(f)},
    ysplit{mainPolyIndices(f)});
    a(:,f) = areas;
end
```

```
Q_12 = sum(in)/length(x1);

%%%%% Inference %%%%%
% For Large Data Sets - z-stat
EQ_12 = 1/4;
std_Q12 = sqrt(((EQ_12)*(1-EQ_12))/length(x1));
z_12 = (Q_12 - EQ_12)/std_Q12;

if z_12 > 1.96
    Inference = 'Attraction at 5% significance level';
elseif z_12 < -1.96
    Inference = 'Avoidance at 5% significance level';
else
    Inference = 'Independence at 5% significance level';
end
```

Listing B.3: Simulated Distribution

```
function [Q_12sim] = simQ12stat(xsim,ysim,x1,y1,x,y)
%where x1(xsim) & y1(ysim) are firm 1's loc and x & y firm 2's

%%%%% VoronoiLimit by Jakob Sievers (slightly modified)%%%%%

if ~any(size(x)==1) || ~any(size(y)==1) ||
    numel(x)==1 || numel(y)==1
    disp('Input vectors should be single rows or columns')
    return
```

```
end

x=x(:);
y=y(:);

% Market boundaries: Find min and max x and y of both data sets
if min(x) <= min(x1) %minimum x value
    minx = min(x);
else
    minx = min(x1);
end

if max(x) >= max(x1) %maximum x value
    maxx = max(x);
else
    maxx = max(x1);
end

if min(y) <= min(y1) %minimum y value
    miny = min(y);
else
    miny = min(y1);
end

if max(y) >= max(y1) %maximum y value
    maxy = max(y);
else
```

```
    maxy = max(y1);
end

bnd=[minx maxx miny maxy]; %data bounds
crs=double([bnd(1) bnd(4);bnd(2) bnd(4);bnd(2) bnd(3);
    bnd(1) bnd(3);bnd(1) bnd(4)]);

%Triangulation and creation of Thiessen polygons/Voronoi diagram
dt=DelaunayTri(x(:),y(:));
[V,C]=voronoiDiagram(dt);
[vx,vy]=voronoi(x,y);
vxyl=[vx(:) vy(:)];
xix=ones(size(vx));

%Eliminate spurious double entries in V(C{ij})
av=cell(length(C),1);
for ih=1:length(C)
    av{ih}=[V(C{ih},1),V(C{ih},2)];
    if size(unique(av{ih},'rows'),1)<size(av{ih},1);
        k=0;
        ctr=0;
        while k==0
            ctr=ctr+1;
            if ctr==1
                tt=unique(V(C{ih}(ctr),:)==
                V(C{ih}(end),:));
                if length(tt)<2 && tt==1
```

```
                    C{ih}(end)=[];
                    k=1;
                else
                    tt=unique(V(C{ih}(ctr),:)==V(C{ih}(ctr+1),:));
                    if length(tt)<2 && tt==1
                        C{ih}(ctr+1)=[];
                        k=1;
                    end
                end
            else
                tt=unique(V(C{ih}(ctr),:)==V(C{ih}(ctr+1),:));
                if length(tt)<2 && tt==1
                    C{ih}(ctr+1)=[];
                    k=1;
                end
            end
        end
    end
end
lV0=length(V);

%Find missing pts that should be added to existing V/C structure
for ii=1:length(vxyl)
    fix=find(V(:,1)==vxyl(ii,1));
    if ~isempty(fix)
        if any(V(fix,2)==vxyl(ii,2))
            xix(ii)=0;
```

```
        end
    end
end

mix=find(xix==1)./2; %index of missing values
lmix=length(mix);
mvx=vx(2,mix); %missing vx
mvy=vy(2,mix); %missing vy
mv=[mvx',mvy'];
cpx=vx(1,mix); %connector point x
cpy=vy(1,mix); %connector point y

ctr=0;
mv2=[];
cpVixt=cell(lmix,1); %connector points, index in V structure
for ii=1:lmix
    if any(V(:,1)==cpx(ii) & V(:,2)==cpy(ii))
        cpVixt{ii}=find(V(:,1)==cpx(ii) & V(:,2)==cpy(ii));
        lval=length(cpVixt{ii});
        if lval==1
            ctr=ctr+1;
            mv2(ctr,:)=mv(ii,:);
        elseif lval>1
            ctr=ctr+1;
            mv2(ctr:ctr+lval-1,:)=[ones(lval,1).*mv(ii,1)
                ones(lval,1).*mv(ii,2)];
            ctr=ctr+lval-1;
```

```
            end
        end
    end
    cpVixt=cell2mat(cpVixt);



    V=[V;mv2]; %add points to V structure
    allVixinp=inpolygon(V(:,1),V(:,2),crs(:,1),crs(:,2));


    %Addition-routine:
    for ij=1:length(C)
        if any(C{ij}==1)
            ixa=find(cpVixt==C{ij}(2));
            ixb=find(cpVixt==C{ij}(end));
            if length(C{ij})<3 %corner point detected
                C{ij}(1)=lV0+ixa(1);
                C{ij}=[C{ij},lV0+ixa(2)];
            else
                if length(ixa)==1 && length(ixb)==1
                    C{ij}(1)=lV0+ixa;
                    C{ij}=[C{ij},lV0+ixb];
                elseif length(ixa)==2 && length(ixb)==1
                    C{ij}=[C{ij},lV0+ixb];
                    [~,minix]=min(sqrt((V(C{ij}(end),1)
                    -V(cpVixt(ixa),1)).^2+(V(C{ij}(end),2)
                    -V(cpVixt(ixa),2)).^2));
                    C{ij}(1)=lV0+ixa(minix);
```

```
            elseif length(ixa)==1 && length(ixb)==2
                C{ij}(1)=lV0+ixa;
                [~,minix]=min(sqrt((V(C{ij}(1),1)
                -V(cpVixt(ixb),1)).^2+(V(C{ij}(1),2)
                -V(cpVixt(ixb),2)).^2));
                C{ij}=[C{ij},lV0+ixb(minix)];
            elseif length(ixa)==2 && length(ixb)==2
                [~,minix1]=min(sqrt((x(ij)-V(lV0+ixa,1)).^2
                +(y(ij)-V(lV0+ixa,2)).^2));
                [~,minix2]=min(sqrt((x(ij)-V(lV0+ixb,1)).^2
                +(y(ij)-V(lV0+ixb,2)).^2));
                C{ij}(1)=lV0+ixa(minix1);
                C{ij}=[C{ij},lV0+ixb(minix2)];
            end
        end
    end
end

%Polybool for restriction of polygons to domain.
C1=C;
for ij=1:length(C)
    if sum(allVixinp(C{ij}))~=length(C{ij})
        [xb, yb] = polybool('intersection',crs(:,1),crs(:,2),
        V(C1{ij},1),V(C1{ij},2));
        ix=nan(1,length(xb));
        for il=1:length(xb)
            if any(V(:,1)==xb(il)) && any(V(:,2)==yb(il))
```

```
                ix1=find(V(:,1)==xb(il));
                ix2=find(V(:,2)==yb(il));
                for ib=1:length(ix1)
                    if any(ix1(ib)==ix2)
                        ix(il)=ix1(ib);
                    end
                end
                if isnan(ix(il))==1
                    lv=length(V);
                    V(lv+1,1)=xb(il);
                    V(lv+1,2)=yb(il);
                    allVixinp(lv+1)=1;
                    ix(il)=lv+1;
                end
            else
                lv=length(V);
                V(lv+1,1)=xb(il);
                V(lv+1,2)=yb(il);
                allVixinp(lv+1)=1;
                ix(il)=lv+1;
            end
        end
        C{ij}=ix;
    end
end

%%%%% Creating Co-location Areas %%%%%
```

```
n = max(cellfun('length',C)) +1;
m = length(C);
x_mid = NaN([m,n]);
y_mid = NaN([m,n]);

for i = 1: length(C)
    vxi = V(C{i},1);
    for j = 1:length(vxi)
        mxi(j) = (x(i) + vxi(j))/2;
        x_mid(i,j) = mxi(j);
    end
end

for k = 1: length(C)
    vyi = V(C{k},2);
    for g = 1:length(vyi)
        myi(g) = (y(k) + vyi(g))/2;
        y_mid(k,g) = myi(g);
    end
end

x_coloc = reshape(x_mid.',1,[]);
y_coloc = reshape(y_mid.',1,[]);


%%%%% inploygons(x,y,xv,yv) by Kelly Kearney %%%%%
```

```
% Check inputs

if size(xsim) ~= size(ysim)
    error('x and y must have the same dimensions');
end

if ~isvector(x_coloc) || ~isvector(y_coloc)
    || length(x_coloc) ~= length(y_coloc)
    error('xv and yv must be vectors of the same length');
end

% Find number of and starting indices of polygons

[x12, y12] = polysplit(x_coloc, y_coloc);
[xsplit, ysplit] = poly2cw(x12, y12);
isCw = ispolycw(xsplit, ysplit);
mainPolyIndices = find(isCw);
nHolesPer = diff([mainPolyIndices;length(isCw)+1]) - 1;

% Test if points are in each polygon

originalSize = size(xsim);
xsim = xsim(:);
ysim = ysim(:);

isIn = zeros(length(xsim), length(mainPolyIndices));
for ipoly = 1:length(mainPolyIndices)
```

```
    isInMain = inpolygon(xsim, ysim,xsplit{mainPolyIndices(ipoly)},
    ysplit{mainPolyIndices(ipoly)});
    if nHolesPer(ipoly) > 0
        isInHole = zeros(length(xsim), nHolesPer(ipoly));
        for ihole = 1:nHolesPer(ipoly)
            isInHole(:,ihole) = inpolygon(xsim, ysim,
            xsplit{mainPolyIndices(ipoly)+ihole},
            ysplit{mainPolyIndices(ipoly)+ihole});
        end
        isIn(:,ipoly) = isInMain & ~any(isInHole,2);
    else
        isIn(:,ipoly) = isInMain;
    end
end

in = any(isIn, 2);
in = reshape(in, originalSize);

if nargout == 2

    index = num2cell(zeros(size(xsim)));
    for ipoint = 1:length(xsim)
        loc = find(isIn(ipoint,:));
        if ~isempty(loc)
            index{ipoint} = loc;
        end
    end
```

```
    index = reshape(index, originalSize);
end

%%%%% Q-stat %%%%%
A = abs(maxx - minx)*abs(maxy - miny);

a = zeros(1,length(mainPolyIndices));
for f = 1:length(mainPolyIndices)
    areas = polyarea(xsplit{mainPolyIndices(f)},
    ysplit{mainPolyIndices(f)});
    a(:,f) = areas;
end

Q_12sim = sum(in)/length(x1);
```

Listing B.4: Inference from Simulated Distribution

```
function [Inf_sim, Q12cv] =
MC_inference(Q_12, Q12statsim, signlevel, iter)
srtQ12sim=sort(Q12statsim);
Q12crit1=[srtQ12sim(0.005*iter),srtQ12sim(0.995*iter)];
Q12crit5=[srtQ12sim(0.025*iter),srtQ12sim(0.975*iter)];
Q12crit10=[srtQ12sim(0.05*iter),srtQ12sim(0.95*iter)];

if signlevel == 1
    if Q_12 <= Q12crit1(:,1)
```

```
        Inf_sim = 'Avoidance at 1% significance level';
    elseif Q_12 >= Q12crit1(:,2)
        Inf_sim = 'Attraction at 1% significance level';
    else
        Inf_sim = 'Independence at 1% significance level';
    end
elseif signlevel == 5
    if Q_12 <= Q12crit5(:,1)
        Inf_sim = 'Avoidance at 5% significance level';
    elseif Q_12 >= Q12crit5(:,2)
        Inf_sim = 'Attraction at 5% significance level';
    else
        Inf_sim = 'Independence at 5% significance level';
    end
else
    if Q_12 <= Q12crit10(:,1)
        Inf_sim = 'Avoidance at 10% significance level';
    elseif Q_12 >= Q12crit10(:,2)
        Inf_sim = 'Attraction at 10% significance level';
    else
        Inf_sim = 'Independence at 10% significance level';
    end
end

levels = [0.01 0.05 0.1];
Q12critval_av = [Q12crit1(:,1) Q12crit5(:,1) Q12crit10(:,1)];
Q12critval_co = [Q12crit1(:,2) Q12crit5(:,2) Q12crit10(:,2)];
```

```
Q12cv = [levels;Q12critval_av;Q12critval_co];
```

Listing B.5: Simulated Distribution with Varying Buffer

```
%%% Remove points that are too close to each other %%%
function [Q12statsim] =
vb_siminfQ12stat(x1,y1,x,y,xy,iter,bnd,vb_buff)
Q12statsim = zeros(1, iter);
for simNum = 1:iter
    aminx= bnd(:,1);
    bmaxx= bnd(:,2);
    aminy= bnd(:,3);
    bmaxy= bnd(:,4);
    xsim=(bmaxx-aminx).*rand(length(x1),1)+aminx;
    ysim=(bmaxy-aminy).*rand(length(y1),1)+aminy;
    xysim=unique([xsim,ysim],'rows');
    for k = length(x1)
        ptk = [xysim(k,1), xysim(k,2)]; %Firm 1's location k
        [idx,d] = knnsearch(xy,ptk,'k',1);
        %find k's NN Firm 2 locations
        NNxy = xy(idx,:); %The nearest Firm 2 loc
        remxy = xy;
        remxy(idx,:)=[]; %The all but the NNxy
        [idx1,d1] = knnsearch(remxy,NNxy,'k',3);
        %Find the 3 NN Firm 2 loc to NNxy
        buff = (vb_buff)*(sum(d1)/3);
        %Buffer = X average dist btw 3 NN Firm 2 locations
```

```
            for l = 1:length(x1)-1
                dist = sqrt(bsxfun(@minus,xysim(:,1),xysim(:,1)').^2
                + bsxfun(@minus,xysim(:,2),xysim(:,2)').^2);
                distk = dist(k,:);
                distk(:,k)=[];
                    while distk(:,l)<buff
                        xysim(k,:) = [(bmaxx-aminx).*rand(1,1)+aminx,
                            (bmaxy-aminy).*rand(1,1)+aminy];
                        dist = sqrt(bsxfun
                        (@minus,xysim(:,1),xysim(:,1)').^2
                        + bsxfun(@minus,xysim(:,2),xysim(:,2)').^2);
                        distk = dist(k,:);
                        distk(:,k)=[];
                    end
            end
        end
        xsim=xysim(:,1);
        ysim=xysim(:,2);
        Q12statsim(simNum) = simQ12stat(xsim,ysim,x1,y1,x,y);
end
```

Appendix C

Additional Simulation Results

Table C.1: Case 2 ($n_A > n_B$), $\frac{1}{3}$ and $\frac{1}{4}$ Varying Buffer

$n_A > n_B$	$Q_{A\to B}$	$z^{\mathbf{Inf}}_{Q_{A\to B}}$	Mean (Std)	$Q^*_{A\to B}$ (lower, upper)$^{\mathbf{Inf}*}$	Quartiles $(1^{st}, 3^{rd})$
Varying buffer ($\frac{1}{3}$)					
$10 > 5$	0.2	$-0.3651^{\mathbf{I}}$	0.2463 (0.1348)	$(0, 0.5)^{\mathbf{I}}$	(0.1, 0.3)
$20 > 10$	0.3	$0.5164^{\mathbf{I}}$	0.2481 (0.0969)	$(0.1, 0.45)^{\mathbf{I}}$	(0.2, 0.3)
$40 > 20$	0.2	$-0.7303^{\mathbf{I}}$	0.2503 (0.0664)	$(0.125, 0.4)^{\mathbf{I}}$	(0.2, 0.3)
$80 > 40$	0.225	$-0.5164^{\mathbf{I}}$	0.2471 (0.0496)	$(0.15, 0.3375)^{\mathbf{I}}$	(0.2125, 0.2875)
$160 > 80$	0.3	$1.4606^{\mathbf{I}}$	0.25 (0.0359)	$(0.1812, 0.325)^{\mathbf{I}}$	(0.225, 0.275)
$320 > 160$	0.2875	$1.5492^{\mathbf{I}}$	0.2485 (0.025)	$(0.2, 0.2969)^{\mathbf{I}}$	(0.2313, 0.2656)
Varying buffer ($\frac{1}{4}$)					
$10 > 5$	0.2	$-0.3651^{\mathbf{I}}$	0.2482 (0.14)	$(0, 0.5)^{\mathbf{I}}$	(0.1, 0.3)
$20 > 10$	0.3	$0.5164^{\mathbf{I}}$	0.2452 (0.0954)	$(0.05, 0.4)^{\mathbf{I}}$	(0.2, 0.3)
$40 > 20$	0.2	$-0.7303^{\mathbf{I}}$	0.2489 (0.0677)	$(0.125, 0.375)^{\mathbf{I}}$	(0.2, 0.3)
$80 > 40$	0.225	$-0.5164^{\mathbf{I}}$	0.253 (0.048)	$(0.1625, 0.35)^{\mathbf{I}}$	(0.2125, 0.2875)
$160 > 80$	0.3	$1.4606^{\mathbf{I}}$	0.2537 (0.0332)	$(0.1875, 0.3187)^{\mathbf{I}}$	(0.2313, 0.275)
$320 > 160$	0.2875	$1.5492^{\mathbf{I}}$	0.2485 (0.0246)	$(0.2031, 0.2969)^{\mathbf{I}}$	(0.2313, 0.2656)

Note: The table shows the $Q_{A\to B}$, $z^{\mathbf{Inf}}_{Q_{A\to B}}$ (where $^{\mathbf{Inf}}$ shows the inference – attraction $^{\mathbf{A}}$, independence $^{\mathbf{I}}$ and avoidance $^{\mathbf{R}}$ – at the 5% significance level), the 5% critical values ($Q^*_{A\to B}$) with inference $^{\mathbf{Inf}}$ and the quartiles $(1^{st}, 3^{rd})$ from the empirical distribution generated through Monte Carlo simulations.

Table C.2: Case 3 ($n_A < n_B$), $\frac{1}{3}$ and $\frac{1}{4}$ Varying Buffer

$n_A < n_B$	$Q_{A\to B}$	$z^{\mathbf{Inf}}_{Q_{A\to B}}$	Mean (Std)	$Q^*_{A\to B}$ (lower, upper)$^{\mathbf{Inf}*}$	Quartiles $(1^{st}, 3^{rd})$
Varying buffer ($\frac{1}{3}$)					
$5 < 10$	0.2	$-0.2582^{\mathbf{I}}$	0.254 (0.1882)	$(0, 0.6)^{\mathbf{I}}$	(0.2, 0.4)
$10 < 20$	0.2	$-0.3651^{\mathbf{I}}$	0.2588 (0.1362)	$(0, 0.5)^{\mathbf{I}}$	(0.2, 0.4)
$20 < 40$	0.2	$-0.5164^{\mathbf{I}}$	0.2507 (0.0997)	$(0.05, 0.45)^{\mathbf{I}}$	(0.2, 0.3)
$40 < 80$	0.275	$0.3651^{\mathbf{I}}$	0.275 (0.0702)	$(0.125, 0.4)^{\mathbf{I}}$	(0.2, 0.3)
$80 < 160$	0.2625	$0.2445^{\mathbf{I}}$	0.2625 (0.0509)	$(0.15, 0.3625)^{\mathbf{I}}$	(0.2125, 0.2875)
$160 < 320$	0.2313	$-0.5477^{\mathbf{I}}$	0.2516 (0.0349)	$(0.1812, 0.3187)^{\mathbf{I}}$	(0.2313, 0.275)
Varying buffer ($\frac{1}{4}$)					
$5 < 10$	0.2	$-0.2582^{\mathbf{I}}$	0.252 (0.1974)	$(0, 0.6)^{\mathbf{I}}$	(0.2, 0.4)
$10 < 20$	0.2	$-0.3651^{\mathbf{I}}$	0.2562 (0.1387)	$(0, 0.5)^{\mathbf{I}}$	(0.2, 0.3)
$20 < 40$	0.2	$-0.5164^{\mathbf{I}}$	0.2502 (0.0935)	$(0.05, 0.4)^{\mathbf{I}}$	(0.2, 0.3)
$40 < 80$	0.275	$0.3651^{\mathbf{I}}$	0.2499 (0.0691)	$(0.125, 0.4)^{\mathbf{I}}$	(0.2, 0.3)
$80 < 160$	0.2625	$0.2445^{\mathbf{I}}$	0.2508 (0.0491)	$(0.1625, 0.35)^{\mathbf{I}}$	(0.2125, 0.2875)
$160 < 320$	0.2313	$-0.5477^{\mathbf{I}}$	0.2313 (0.0341)	$(0.1812, 0.3187)^{\mathbf{I}}$	(0.225, 0.275)

Note: The table shows the $Q_{A\to B}$, $z^{\mathbf{Inf}}_{Q_{A\to B}}$ (where $^{\mathbf{Inf}}$ shows the inference – attraction $^{\mathbf{A}}$, independence $^{\mathbf{I}}$ and avoidance $^{\mathbf{R}}$ – at the 5% significance level), the 5% critical values ($Q^*_{A\to B}$) with inference $^{\mathbf{Inf}}$ and the quartiles $(1^{st}, 3^{rd})$ from the empirical distribution generated through Monte Carlo simulations.