U.S.N.A. --- Trident Scholar project report; no. 323 (2004)

An Empirical Investigation of Product Differentiation in the Retail Gasoline Industry

by

Midshipman 1/c James J. Light, Class of 2004
United States Naval Academy
Annapolis, Md

____
(signature)

Certification of Advisers Approval

Assistant Professor Pamela M. Schmitt
Economics Department

____
(signature)

____
(date)

Associate Professor Suzanne K. McCoskey
Economics Department

____
(signature)

____
(date)

Acceptance for the Trident Scholar Committee

Professor Joyce E. Shade
Deputy Director of Research and Scholarship

____
(signature)

____
(date)

USNA-1531-2

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATE COVERED
	6 May 2004	

4. TITLE AND SUBTITLE
An empirical investigation of product differentiation in the retail gasoline industry

5. FUNDING NUMBERS

6. AUTHOR(S)
Light, James J. (James Joshua), 1981-

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
US Naval Academy
Annapolis, MD 21402

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
Trident Scholar project report no. 323 (2004)

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
This document has been approved for public release; its distribution is UNLIMITED.

12b. DISTRIBUTION CODE

13. ABSTRACT: This Trident project constructed and estimated a model of product differentiation in the retail gasoline industry. Retail gasoline stations product differentiate by choosing location and offering amenities such as pay at the pump, full service, car washes, service stations and food markets. Using daily price data from the Minneapolis retail gasoline market, the effect for spatial and quality differentiation was empirically investigated. Spatial differentiation was measured by creating three degrees of neighbors (competing stations) around each individual gasoline station. Each successive neighborhood had an area of increasing size around the station; this differentiated between closer directly competitive neighbors and indirectly competitive neighbors that were away from a station. Quality differentiation was measured by a binary variable when gasoline stations offer additional amenities. The estimation procedure accounted for spatial autocorrelation and market characteristics specific to the retail gasoline industry. The results indicated that spatial location of firms had the largest and most significant effect on firm price. The spatially lagged autoregressive coefficient was found to be positive and significant; indicating that gasoline firm prices move together and firms that are closer in neighborhood degree have a more significant effect on price. In addition, stations decreased price when the number of competitors in its closest neighborhood increased and when the number of independent stations (stations not connected with a major refiner) increased in the closest neighborhood. Price also increased when stations of the same brand overlapped into each other's neighborhood areas. Quality differentiation had less effect on price than spatial competition. The two most significant measures of quality that affected price were convenience stores and service stations. Possessing a service station caused stations to increase price while possessing a convenience store caused stations to decrease price. An increase in the number of pumps also caused stations to decrease price. This result can be attributed to a downward sloping demand curve if stations were increasing their quantity supplied and therefore charging a lower price as they moved down the demand curve. The model constructed in this study represented a clear departure from the standard microeconomic analysis due to the inclusion of space. Space embodies an important element that must be accounted for when describing price competition in any market and this analysis supported that. Space was continually significant throughout the models that were constructed for the retail gasoline industry.
.

14. SUBJECT TERMS:
Spatial Differentiation; Product Quality; Spatial Autocorrelation

15. NUMBER OF PAGES
67

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT

An Empirical Investigation of Product Differentiation in the Retail Gasoline Industry

This Trident project constructed and estimated a model of product differentiation in the retail gasoline industry. Retail gasoline stations product differentiate by choosing location and offering amenities such as pay at the pump, full service, car washes, service stations and food markets. Using daily price data from the Minneapolis retail gasoline market, the effect for spatial and quality differentiation was empirically investigated. Spatial differentiation was measured by creating three degrees of neighbors (competing stations) around each individual gasoline station. Each successive neighborhood had an area of increasing size around the station; this differentiated between closer directly competitive neighbors and indirectly competitive neighbors that were away from a station. Quality differentiation was measured by a binary variable when gasoline stations offer additional amenities. The estimation procedure accounted for spatial autocorrelation and market characteristics specific to the retail gasoline industry.

The results indicated that spatial location of firms had the largest and most significant effect on firm price. The spatially lagged autoregressive coefficient was found to be positive and significant; indicating that gasoline firm prices move together and firms that are closer in neighborhood degree have a more significant effect on price. In addition, stations decreased price when the number of competitors in its closest neighborhood increased and when the number of independent stations (stations not connected with a major refiner) increased in the closest neighborhood. Price also increased when stations of the same brand overlapped into each other's neighborhood areas.

Quality differentiation had less effect on price than spatial competition. The two most significant measures of quality that affected price were convenience stores and service stations. Possessing a service station caused stations to increase price while possessing a convenience store caused stations to decrease price. An increase in the number of pumps also caused stations to decrease price. This result can be attributed to a downward sloping demand curve if stations were increasing their quantity supplied and therefore charging a lower price as they moved down the demand curve.

The model constructed in this study represented a clear departure from the standard microeconomic analysis due to the inclusion of space. Space embodies an important element that must be accounted for when describing price competition in any market and this analysis supported that. Space was continually significant throughout the models that were constructed for the retail gasoline industry.

Keywords: Spatial Differentiation; Product Quality; Spatial Autocorrelation

Table of Contents

I. Introduction

Retail gasoline is one of the most analyzed products in the world and more so in the United States because of people's reliance on cars. The price of gasoline is an important issue to consumers because it is consumed on an almost daily basis by all drivers. The actual price is determined by many different factors including crude oil prices, taxes at all levels of government including federal, state and local, refinery costs and local distribution, and marketing costs. The price of gasoline is routinely analyzed in relation to national or regional price changes that usually occur due to changes in supply, in crude oil prices and in seasonal demand. However, this study takes a different approach and attempts to explain *local* variations in gasoline prices according to product differentiation. The use of market cross section data rather than time series allows this study to eliminate changes in crude oil prices and other long term factors that may affect price. The effect of spatial competition on price can be empirically investigated.

Product differentiation among firms occurs in two dimensions: spatial location of firms and product quality. Product differentiation is a mechanism which allows firms to price above marginal cost and earn an economic profit when firms simultaneously choose price. Goods that are perfectly homogeneous in all aspects and that are produced and sold by two different firms cannot be priced above marginal cost. This Nash equilibrium result based on Bertrand's theory implies that when two firms simultaneously set price, treating quantity as given, each firm has an incentive to decrease its price and capture the entire market. This price undercutting continues until each firm's price equals the marginal cost level. Neither firm has an incentive to deviate from this price and thus it is a Nash equilibrium.

Firms attempt to differentiate their products in various dimensions in order to make them non-homogenous. Non-homogenous products can be priced above marginal cost because they are viewed as different in the eyes of consumers and thus consumers have different demand curves for them. Retail gasoline within a specific grade is a homogenous product without product differentiation at the retail gasoline station level. To avoid the Bertrand paradox, retail gasoline stations product differentiate by choosing to locate at different intersections and by having amenities such as pay at the pump, car washes, convenience stores and service stations. To examine the effects of spatial and quality differentiation on the retail gasoline market, this study empirically examines the price of gasoline stations against the spatial and quality characteristics in the market.

The empirical market that is studied is the Minneapolis-St Paul metropolitan area. This market contains 234 retail gasoline stations. Spatial differentiation is measured by mapping firm locations and dividing competitors up into neighbor markets according to the *Neighbor Rules*.[1] This differentiates between direct and indirect competitors. Vertical differentiation is measured by additional amenities that the gasoline station provides beyond gasoline, such as pay at the pump, car washes, convenience stores and service stations. The characteristics of the gasoline industry, specifically the homogenous nature of gasoline and low consumer search costs due to posted prices, could lead to the extremes of intense price competition where each firm tries to undercut the other and gain a larger market share, or collusion where the posted prices act as a signaling device and firms are able to maintain artificially high prices in the market. Price is expected to decrease as stations locate closer together if they are operating in the former price competitive environment. If this effect is evident, then stations are expected to add quality characteristics to differentiate themselves from their competitors and gain market power. The addition of quality characteristics, in particular convenience stores and service stations, represents a form of product bundling for the gasoline firm. Bundling may allow the firm to price discriminate in order to extract the consumers' surplus and convert this into firm profits.

[1] For complete *Neighbor Rules* see Appendix A.

The results show that price decreases as the number of closest spatial competitors increases. This price decrease seems to indicate a degree of price competition in the gasoline industry among a station's closest competitors. As the number of indirect competitors increases, the effect on price is positive or negative depending upon the model specification. The effects of the quality variables defined in the model are ambiguous. Adding some quality characteristics, such as a service station, leads to an increase in price while adding other characteristics, such as a convenience store, leads to a decrease in gasoline price at the station.

The paper proceeds as follows. Section II discusses the relevant theoretical and empirical economic literature. Section III provides the data and methods. Section IV describes the OLS model and preliminary results from that specification while Section V discusses problems associated with estimation using space and a re-estimated model using Maximum Likelihood (ML) Estimation. Section VI concludes.

II. Literature Review

Theoretical models of spatial differentiation fail to reach a consensus as to whether firms minimally or maximally differentiate. In Hotelling (1929), firms choose to minimally differentiate their location in order to lure customers from their competitors and gain the largest market share. Conflicting models, including d'Aspremont et al (1978), conclude firms maximally differentiate to reduce price competition. d'Aspremont et al (1978) assume quadratic rather than Hotelling's (1929) linear transportation costs for consumers Shaked and Sutton (1982) introduce vertical differentiation to the theoretical literature. Shaked and Sutton (1982) conclude that firms maximally differentiate in product quality to minimize price competition. Theoretical models a la Hotelling (1929) and his predecessors are abundant yet collectively inconclusive. This necessitates an empirical approach that will be applied in this study.

Empirical analysis of spatial differentiation is scarce in economics literature. The relevant studies include Slade (1992), Borenstein and Netz (1999), and Netz and Taylor (2002).

A. Theoretical Spatial Differentiation

Hotelling (1929) explains the location choices of firms by establishing the theory of spatial differentiation. Hotelling creates a one dimensional market, a line, that contains two firms labeled A and B. The firms have identical marginal costs of production that are equal to zero. Consumers are distributed uniformly along the line and their demand is perfectly inelastic. The market is covered; that is, each consumer uys a unit of the good. Consumers' preferences for the good are based upon price plus a linear transportation cost. Consumers' utility is maximized when they purchase the good for the minimum price plus transportation cost. Firms first choose a profit maximizing location with the knowledge that they will then compete in prices. Hotelling assumed that firms could costlessly relocate their locations. Firms then choose price based upon their location.

Hotelling finds that firms choose identical locations at the center of the market, minimally differentiating. Each firm wants to maximize its market share and because firms can costlessly relocate, they do so until they are both at the center of the market. Firm A choosing any position away from the center of the market allows Firm B to capture more than fifty percent of the market and this decreases A's profits. The same is true for Firm B and therefore both firms locate at the center. Because both firms are located at the center of the market, they engage in Bertrand competition in prices until both prices are the same and equal to the marginal cost of production. In the gasoline market if two firms locate in the same location, or nearly the same location such as adjacent street corners since exact co-location is impossible in a real market, Hotelling's model can be empirically examined. Gasoline firms that locate extremely close together do not substantiate Hotelling's predicted result. They still price above marginal cost.

The reasons for this, which will be addressed in later theoretical models, are that Hotelling's model fails to account for specific tastes and preferences of consumers, product differentiation in dimensions other than space, markets with more than two firms and the distribution of consumers in space.

An interesting addition to Hotelling's model is to have the two firms' locate on the linear city where they minimize transportation costs for consumers. This solution represents a socially optimal market. In Hotelling's equilibrium solutions he maximized firm profits without regard to consumers' costs. The socially optimal outcome is that each firm locates at the quartiles charging equal prices. This is the socially optimal outcome in the market because it minimizes the consumers' transportation costs. The maximum distance a consumer has to travel in the market is ¼.[2] Despite being socially optimal, this price-location is not an equilibrium; each firm has the incentive to deviate (move closer to the other firm) and increase its own profits.

Hotelling introduces location as a tool for firms to differentiate themselves from their competitors. Although his conclusions on how firms choose to locate are based upon a simplistic model of one dimension and only two firms, he creates a framework to study spatial differentiation that can be expanded upon both theoretically and empirically.

Teitz (1968) introduces the idea of firms that have multiple locations but profit-maximize collectively; each distinct location represents a *branch* of the firm. The branches of the firm together are known as chains and each chain of firms competes against other chains of firms or single firms with only one branch. Teitz investigates equilibria for two chains of firms, A and B, competing in Hotelling's linear city. When chain A has more branches than chain B, chain A's branches locate at the socially optimal positions in the market (two branches would locate at the quartiles of the market). The location choices for chain A are made regardless of where chain B

[2] In the Hotelling linear city with both firm locating at the center of the market, the maximum distance a consumer has to travel is ½.

locates its branches. The gasoline market contains firms competing against each other and chains of firms that are strategically allied. An example of this is that chains of Shell stations are competing against chains of Mobil stations but individual Mobil stations in general do not compete against other Mobil stations for customers. Individual Mobil stations may be independently owned and want to compete with each other but the refiner that supplies them would have the incentive for them not to compete so that price would remain high.[3] Thus, the refiner's goal for the Mobil stations is to steal consumers from other chains or independent firms in the market and not their own stations. Teitz's theoretical assumptions lead us to the conclusion that chains of stations will attempt to locate efficiently in the market so they are not competing for the same costumers. If chains are unable to efficiently locate due to outside forces such as zoning laws, it is expected that any price interactions that occur will be collusive in nature rather than price competitive.

Eaton (1972) modifies Hotelling's (1929) linear city model by examining the affect of market size on the location choice of firms. Eaton shows that when the market line is sufficiently large, firms choose to locate at the quartiles and act as spatial monopolists. In this situation, the products of the firms are no longer perfect substitutes for each other, as in the original Hotelling model, and therefore the market demand is not perfectly inelastic. The transportation costs incurred by consumers allow each firm to have an absolute advantage over all other firms in a certain geographical area surrounding the firm and therefore firms can price above marginal cost to the consumers in that area. Each firm earns a positive profit which is greater than the minimum differentiation profits from locating at the center of the market and engaging in Bertrand price competition. As the market size decreases firms symmetrically locate closer and closer to the center of the market and at a certain market size they reach Hotelling's

[3] See Section IV.A. for more on gasoline stations vertical relationships and incentives for firms of the same chain to keep prices high.

equilibrium. In the large market, each firm has no incentive to steal customers from its competitors. They are profit maximizing spatial monopolists, given quartile locations in the market. This equilibrium changes as the market becomes smaller and each firm can increase profits by stealing customers from its competitors. This is why the firms locate closer and closer to the center of the market. Eaton demonstrates that Hotelling's minimum differentiation is not robust for all market sizes. In large markets, maximum differentiation is more prevalent and as market size decreases the result of minimum differentiation becomes more robust. In this study's empirical gasoline model, if the market is not covered by the firms due to locational constraints as suggested by Eaton, then a positively signed coefficient is expected for the number of competitors in a market. This theoretically follows from Eaton's model if one assumes that firms cannot locate with perfect efficiency in the real-world market and thus they compete in prices.

Eaton and Lipsey (1975) evaluate Hotelling's equilibrium with respect to the number of firms and dimension of the market as another expansion of the theoretical spatial differentiation literature. They change the assumptions of Hotelling's model by increasing the number of firms in the market and by making the market two dimensional rather than a one dimensional line. These changes more accurately characterize the empirical market in this study. The Nash equilibrium for two firms along the Hotelling line (one dimension) is minimum differentiation at the center, but when the market size increases to three firms no Nash equilibrium exists. Eaton and Lipsey conclude that in order for any equilibrium to exist when the number of firms is greater than one, each peripheral firm[4] must be paired. In the three firm market, each peripheral firm attempts to pair itself with the interior firm, minimally differentiating and leaving the interior firm a market share of zero. The interior firm thus has the incentive to deviate its

[4] Eaton and Lipsey (1975) define a peripheral firm as a firm whose market boundary is an exterior boundary over some of its range.

position in this arrangement and thus no equilibrium can be reached. The four-firm market has a Nash equilibrium of two firms paired off at each quartile of the market. The addition of a fifth firm produces a unique Nash equilibrium of two firms each paired off at 1/6 and 5/6 and a fifth firm located at the center of the market. The six firm equilibrium is no longer unique, but the peripheral firms are always paired off while the center firms can be paired off or maximally differentiated.

In general, Eaton and Lipsey (1975) conclude that a combination of minimum and maximum differentiation produces equilibrium conditions when the number of firms increases; pairs minimally differentiate amongst themselves and maximally differentiate from other pairs of firms. They find that no distinct equilibrium exists in a two dimensional space for market sizes greater than two firms. Despite the fact that no equilibrium exists, they do observe a general principle of locational clustering, (firms pairing off) rather than strict minimum differentiation, to be present in this market. The retail gasoline market being examined is also a two dimensional model with constraints such as zoning and the existing road system that limits the potential locations that gasoline stations can occupy. To test the significance of Eaton and Lipsey's (1975) conclusions about two dimensional markets, the map locating all the gasoline stations in the market can be examined visually. A more rigorous examination could include testing how many competitors affect a center station's price and thus are paired with that station.

Prescott and Visscher modify the Hotelling problem to ensure that the demand and profit functions for the firms are continuous; this avoids the problems noted by Eaton (1976) and d'Aspremont (1978) with Hotelling's (1929) model. Firms compete in prices and waiting times that are representative of location. Waiting times represent a level of productive defectiveness, and are, therefore, inversely related to firm's prices (as price decreases waiting time increases). The introduction of waiting times to the model causes the firm's profit and demand functions to be continuous. Consumers are uniformly distributed in the market with a specific valuation of waiting time.

The unique equilibrium in the two firm model given these new assumptions is that firms locate far apart. The intuition of this result is that waiting times (locations) too close together cause price cutting to appear to be lucrative. Each firm anticipates that it can gain a large share of the market by cutting its price a small amount and thus intense price competition results between the two firms. The combination of both firms cutting prices results in lower profits for both of them. By choosing waiting times (locations) further apart, price cutting appears to be less beneficial and this leads to higher prices and profits for the firms. The Prescott-Visscher model produces no distinct equilibrium for three firms on a Hotelling line. The equilibrium that Prescott-Visscher reach is unique and contradicts Hotelling's model but it is difficult to apply to this study's empirical model because it only produces an equilibrium for the two firm market. While the specific theory discussed might not directly apply to this paper's empirical model, the fact that it contradicts earlier theories and leads to a degree of ambiguity in the spatial differentiation literature (changing small assumptions carries the risk of extreme changes in firm behavior) causes its inclusion here.

Prescott and Visscher (1977) also critique Hotelling's model because it does not account for the costs of relocation. They assume that relocation in any product dimension is quite costly and propose an alternative solution to Hotelling's model with the introduction of relocation costs. They introduce relocation costs to the model by assuming firms face a fixed cost of locating which allows a firm to be profitable the first time they locate, but causes subsequent relocating to be unprofitable. The gasoline industry carries extremely high relocation costs for firms. If a gasoline station decides to relocate it will most likely add another station at a new location rather than relocating the original station.[5]

Prescott and Visscher conclude that the fixed cost of entry affects the location choices of firms, their profitability, and the number of entrants in the market when it is added as an assumption to the model. The fixed cost acts as a barrier to entry; as it increases, the number of firms in the market decreases. The fixed cost also affects the profits of the firms already in the market. As fixed cost decreases, the closer the second entrant into the market has to locate to the first entrant to forestall entry of a third firm. This closer location results in increased price competition, leading to lower prices charged by the firms and lower profits. Conversely, as fixed cost increases, two firms in the market can locate further apart and still forestall entry of a third firm while earning higher profits. The increased revenues that result from the ability to charge higher prices outweighs the increased fixed cost of entry the firms pay initially and thus their profits are higher. There are fixed costs of entry in the gasoline station industry and they are expected to lead to increased spatial differentiation among the firms because they do not need to locate as closely in order to forestall future entry.

Prescott and Visscher (1977) attempt to construct a spatial differentiation model more applicable to the real world. They introduce the ideas of relocation costs and continuous demand functions to the spatial differentiation literature. The theoretical conclusions they reach seem to be intuitively true in the real world, but actual empirical investigations of these ideas is the only way to prove their validity. Prescott and Visscher also show that Hotelling's minimum differentiation is not robust to changes in the model involving relocation costs and continuous demand functions.

---

[5] Gasoline firms will increase in the number of firms in the market rather than leaving N (the number of firms) fixed as Prescott and Visscher (1977) assume in their theoretical model.

d'Aspremont et al (1978) claim that Hotelling's minimum differentiation for two firms is invalid because distinct pure price equilibriums do not exist at all pairs of points where the two firms might locate along the Hotelling line with a linear transportation cost function. The nonexistence of price equilibrium occurs when the firms are symmetrically located inside the quartiles of the market. The nonexistence is caused by the discontinuous nature of the firms' demand and profit functions. The discontinuity means that each firm has the incentive to undercut the other's price when they are sufficiently close together and gain the entire market share. Dasgupta and Maskin (1986) claim that a mixed strategy price equilibrium can exist in the linear model. The nonexistence of price equilibrium means that no solution for the model can exist because as soon as the firms are located inside the quartiles of the market nothing further can be said about their respective actions.

d'Aspremont et al (1978) propose an alternative solution to Hotelling's model that produces price equilibriums for any pair of firm locations along the Hotelling line by introducing a quadratic transportation cost. The equilibrium in this model is that each firm maximally differentiates along the linear city because they have negative profit functions. Therefore, Hotelling's minimum differentiation is not robust to changes in the transportation cost function. In the gasoline market, consumers have a linear transportation cost for the gasoline they consume. The amount of gasoline consumed when driving to the gasoline station increases linearly as the distance to that station increases. But consumers also incur transportation costs such as time which decreases their utility when they have to travel further in order to buy gasoline. Time could have a linear or quadratic function (or another function entirely) depending upon the consumer.

Gabszewicz and Thisse (1996) address location theory as formulated by Hotelling (1929) and apply additional assumptions to make it more relevant to real world situations. The authors define the concept of an industry as being a dispersed number of firms. Each firm competes with its closest neighbors (direct competitors), but is also affected by firms located beyond their neighbors (indirect competitors). The difference between direct and indirect competitors is that the product provided by direct competitors is more substitutable for the product sold by the center firm. The degree of substitutability decreases as indirect competitors are located further in the spatial and product dimensions from the center firm. In the gasoline industry, the perfect pair of substitutes is represented by two gasoline stations located right next to each other selling the same brand of gasoline with the same quality amenities. In reality, no two firms are perfect substitute but direct competitors are still good substitutes and should have a greater effect on price than indirect competitors. This study tests the affect of both direct and indirect competitors on each others prices to encompass the complete scope of possible firm interactions.

Gabszewicz and Thisse (1996) conclude that firms gain market power through space; the less their market overlaps both their direct and indirect competitors' markets the greater their market power. This spatial differentiation among firms contrasts Hotelling's (1929) result of minimum differentiation. In most cases, the authors reason that firms differentiate spatially to relax price competition and gain market power but they also recognize that locational theory is inconclusive as a whole. Despite their inability to provide us with one unambiguous theory, Gabszewicz and Thisse (1996) provide a basis for defining a spatial relationship between firms in an industry in terms of direct and indirect competitors. To empirically test Gabszewicz and Thisse's (1996) assumptions about types of competitors, a neighbor matrix of direct and indirect competitors is constructed in this empirical study of the Minneapolis-St Paul market.

Hotelling's (1929) theory of minimum differentiation for two firms along a one dimensional line holds true for a very restrictive set of assumptions. Hotelling's original result holds if demand in a market is fairly concentrated or in the absence of price competition in the market. Without these factors, minimum differentiation is not robust. Eaton (1972) shows that changes in market size cause firms to spatially differentiate themselves. Eaton and Lipsey (1975) determine that as the number of firms increases, firms tend to pair off and minimally differentiate with their closest competitor while maximally differentiating with competing pairs. d'Aspremont et al (1978) show that Hotelling's model is inconclusive given his assumptions and an alternative model with quadratic transportation costs produces a result of maximum differentiation between two firms. Prescott and Visscher (1977) establish that minimum differentiation is not robust when relocation costs are introduced into the model; firms tend to maximally differentiate. Teitz (1968) shows that branches of the same firm differentiate themselves across the market optimally if they have more branches than their competitor. Gabszewicz and Thisse (1996) construct a spatial model of firms using the ideas of minimum and maximum differentiation and show that firms have both direct competitors (neighbors) and indirect competitors. In summary, maximum differentiation is the more robust result when the additional assumptions are made in the spatial literature but it is not robust across all sets of assumptions.

B. Theoretical Vertical Differentiation

Shaked and Sutton (1982) examine the effect that quality differentiation has on price competition. In this model vertical (quality) differences are similar to horizontal (location) differences among firms except that quality is measured on a vertical scale; a higher quality firm's product is viewed as absolutely better by all consumers. In spatial models, consumer preferences depend upon the transportation cost incurred so consumers prefer different firms depending upon their specific location in the market. This contrasts vertical models which value one good as having higher quality and thus it is preferred by all consumers in the market.

In the Shaked and Sutton (1982) model firms play a three stage game. In the first stage firms simultaneously decide whether or not to enter the market, in the second they simultaneously choose product quality, *s*, and in the third stage they simultaneously choose price. The ordering of the game implies that price can be costlessly varied while quality cannot be. The firm must modify its production facilities to change its quality, so this modification is more costly than changing price. This model is applicable to the gasoline industry because it is easier and cheaper for a gasoline station to change its posted price than to change its quality amenities.

The primary divergences from the horizontal model are that consumers have a utility function modeled by

$$U = \theta * s_i - p_i \tag{1}$$

where s is the product quality, θ is the consumer's taste for quality, and p is the price of the good at quality s. The taste parameter θ is equivalent to location in the horizontal model and it is uniformly distributed across the market. The importance of the utility function is that it accounts for higher quality firms being absolutely better than lower quality firms (in the horizontal model no firm is absolutely better than another; that firm is just preferred by consumers closer to it). The entire market is represented by a vertical line rather than a horizontal line to capture this characteristic.

Shaked and Sutton (1982) impose maximum differentiation between firms' respective qualities and solve the model using backwards induction. They conclude that the higher quality firm earns a higher profit when marginal costs are not dependent on quality. This result is not particularly robust due to the "finiteness result" that the authors demonstrate. The authors claim that only a finite number of firms can exist in an industry regardless of the size of demand and entry costs; this result contrasts with market size in the horizontal model found in Eaton and Lipsey (1975), which acts as a determining factor for the number of profitable firms that can exist in the market. In the vertical model, the number of firms is limited by price competition among high quality firms that drives lower quality firms out of the market; consumers choose the high quality product when all products are sold at marginal cost. The absence of low quality firms eliminates the possibility of maximal differentiation in the market.

Vertical differentiation is the second means that firms use to relax price competition in a competitive market. In the gasoline industry, firms increase their quality by adding services such as pay at the pump, convenience stores, car washes and service stations. The robustness of Shaked and Sutton's (1982) results can be tested empirically in this investigation of the Minneapolis-St Paul retail gasoline industry.

C. Empirical Spatial Differentiation

The spatial differentiation literature is extensive in the theoretical realm but empirical investigations of actual markets do not encompass such breadth and depth of literature. The airline industry and gasoline are two such markets that have been investigated empirically. Borenstein and Netz (1999) analyze the airline industry. Slade (1992) and Netz and Taylor (2002) examine the retail gasoline industry in Vancouver and the Los Angeles basin respectively.

Borenstein and Netz (1999) examine spatial differentiation in the airline industry using flight departure times representing firm location. Each consumer is located at a "most preferred departure time" rather than a distinct point in space. The consumer's utility is represented by the price of their ticket plus the cost in time that the actual departure time differs from their most preferred departure time on a 24-hour clock. In choosing flight departure times, airlines have the incentive to either minimally differentiate to gain market share and steal consumers from their rivals *or* to maximally differentiate to gain market power and charge higher prices. Two distinct historical periods of departure times are used in study, one in the regulated era (pre-1978) when fares were determined exogenously by the government, and one in the non-regulated era (post-1978) when firms chose price endogenously.

Borenstein and Netz (1999) conclude that airlines want to maximally differentiate departure times from their competitors along the same route but tend towards minimum differentiation due to positive externalities in the industry. They establish the tendency of maximum differentiation through a comparison of pre- and post-regulation departure times. After regulation when airlines could choose price endogenously, they attempted to differentiate on routes where the positive externalities of minimum differentiation were smallest. These attempts to differentiate did not appear to exist in the pre-regulation flights. The externalities that mitigate departure time differentiation include less scheduling flexibility at major hubs where the volume of traffic and slotting allowances influence arrival and departure times. Airlines also have less scheduling flexibility in cross country flights because consumers' preferred departure times are concentrated rather than randomly distributed around the 24-hour clock due to changes in time zones and because of airline considerations for flight connections.[6] These factors tend to overpower the airlines' spatial tendency towards maximum differentiation in the authors' analysis.

Slade (1992) studies gasoline stations in Vancouver during a three month price war period. An econometric model of the price war period is developed by the author to study the dynamics of how firms attempt to tacitly collude in the market. Slade finds that prices remain relatively stable during most periods but unexpected demand shocks cause price wars. The price cuts are usually precipitated by independent (non-major) gasoline retailers and occur because of

[6] For instance, if Airline X's flight from Baltimore to Chicago carries mostly passengers who switch planes in Chicago to continue their trip westward, then Airline X schedules the flight to coincide with other flights from the east coast that arrive in Chicago at the same time. It is costly for Airline X to reschedule the Baltimore to Chicago flight and product differentiate (by changing the departure time) so it can raise the price on the Baltimore-Chicago route because it will lose connecting passengers because their layovers become too long or short in Chicago.

demand shocks in the market[7]. The price wars end when major gasoline retailers signal they are over by initiating price increases. The price response of competitors to the independent retailers decreasing price is found to be stronger than the response when independents increase price. In contrast to the independents, Slade finds that competitors respond more strongly to price increases from major retailers than to price decreases. These two facts seems to support the notion that independents precipitated the price wars in the Vancouver market and major suppliers signaled the end to price wars by raising their prices to a normal level. Slade notes that pre- and post price war rules for determining the average market price seemed to change after each price war.

Netz and Taylor (2002) empirically investigate the Los Angeles basin retail gasoline market. They test whether firms tend to locate products close to their competitors to gain a larger market share (market share effect) or differentiate their product to reduce price competition (market power effect). Netz and Taylor (2002) define their model in terms of the degree of spatial differentiation in the market. They examine how changes in competition in the market, specifically the introduction of new entrants, affect the spatial relationships of firms. The two distinct market characteristics are: 1) if the market share effect dominates an increase in competition leads to clustering; or 2) if the market power effect dominates an increase in competition leads to more dispersed locations.

Spatial differentiation is modeled with respect to the degree of competition in the market, the vertical differentiation among stations, demand conditions, and entry costs. The degree of competition is measured by the total number of stations in a defined market. Competitor degrees are defined by circles of radius ½ mile, 1 mile and 2 miles around all stations. In addition the

[7] Demand shocks in the Canadian market in the earlier 1980s occurred due to a reversal in the Canadian-American relative price for gasoline. Before 1980 Canadian prices were lower than American prices, but American prices began to fall and by 1983 American gasoline was cheaper. Canadians crossed the border to buy American gasoline leaving the Canadian market with excess capacity.

proportion of stations in a market that are non-branded (independent) and the proportion of stations that are the same brand as the center station are measured. In this study degree of neighbor (instead of competitor) is defined using the general guidelines of Netz and Taylor (2002) but include more specific rules which account for situations such as firms being located close together but on the opposite sides of divided highways.[8] Netz and Taylor (2002) measure vertical differentiation by attributes the station offers in addition to gasoline; this includes pay at the pump, full service, car washes, convenience stores, or service stations. The vertical attributes are modeled by a binary variable of "1" if the station possesses the characteristic and "0" if they do not. Vertical characteristics are modeled using a similar method in this study. Demand conditions are modeled by identifying stations that are close to or on a major road (within ¼ mile) and the percentage of the center station's competitors that are on a major road. Entry costs are quantified by measuring two independent variables: 1) the proportion of gasoline stations that require prepayment[9], and 2) the median value of housing in the market and the median household income in the market.

The authors examine approximately 4,000 gasoline stations in the Los Angeles basin from 1992 to 1996. They define "entry" stations as stations that are new to the market in any given year and "stablemarket" stations as stations that are present each year of the study. The results of the study are particularly robust across the market sizes and types of markets (entry or stable) in terms of competition. The authors find that stations tend to locate further from competitors (market power effect dominates) when the number of stations, the fraction of the market served by nonbranded stations, and the fraction of the market that is the same brand as

[8] The actual *Neighbor Rules* are discussed in Section III.A. For full reference see Appendix A.

[9] The proportion of stations requiring prepayment, which represents the customer paying before they pump gasoline, is predicted to be correlated with the crime rate in an area. Stations in higher crime areas would be expected to require prepayment more often than stations in lower crime areas because stations in high crime areas would expect more customers to try to drive off without paying.

the center, increase in a given market. There is a weak relationship between entry costs and increased spatial differentiation. Stations are less spatially differentiated when they are close to major roads; this supports the theory that stations will locate where the demand for gasoline is located. Netz and Taylor (2002) conclude that stations that are vertically differentiated from their competitors still attempt to spatially differentiate themselves from competitors. Their findings contrast with the theoretical literature that concludes firms maximally differentiate in one dimension while minimally differentiating in all other dimensions.

Netz and Taylor's paper provides an empirical study in the same industry as this paper. The important differences between the two investigations are that while Netz and Taylor explain spatial differentiation, this paper examines how firms choose price based upon their quality and location characteristics. The robustness of this study's results to those of Netz and Taylor can be examined since they are obtained through different approaches.

III. Data

A. Data Collection

To conduct an empirical study, price and station level characteristics were gathered from the Minneapolis-St Paul metropolitan area. This area of the country was chosen because of the availability of price data. The area includes the Minneapolis-St Paul metropolitan area and surrounding suburban areas with a boundary established where the number of gasoline stations decreased appreciably per area on the map. At the edges of the defined market there could be additional interactions with gasoline firms outside the boundary but these are accepted as part of the error inherent to any real-world model.

Once the area was defined, stations were identified and price data from May 23rd to May 26th, 2003, was gathered. Station identification and price data was obtained from

Twincitiesgasprices.com.[10] Prices are recorded by individuals who log onto the website and enter the price of regular unleaded gasoline for a particular station referenced by the station's name and address. If more than one entry exists, because two individuals enter the price for a given station, then the most recent entry was used for the given period.[11] All prices represent the lowest grade of gasoline available. The data consists of approximately 1200 observations recorded for 234 retail gasoline stations. Prices were gathered over a four day survey period to ensure that if a station's price was not entered one day it would be entered on subsequent days. The prices recorded for an individual gasoline station were then averaged together so there was one averaged price for each station in the market. This study is a cross section analysis that does not investigate changes in the market across time. In addition, to avoid missing any additional stations, stations were cross-checked using commercial gasoline distributor websites.

To define the spatial differentiation of the firms, it was first necessary to physically plot all 234 firms in the Minneapolis-St Paul retail market onto a map of the entire metropolitan area. The entire market area is approximately 200 square miles. Each firm was plotted and labeled on the map using numbered, color coded pins. Firms operating under the same chain, such as all the Super America, Mobil or Holiday gasoline stations, were coded as the same color to differentiate competitors from same branded stations all operating under the same major refiner. The number written on each pin corresponds to the address of each specific firm in the market. An example of a map of the market appears in Figure 1.

[10] Twincitiesgasprices.com finds that approximately 95% of the submitted prices for gasoline are correct.
[11] Prices are categorized by whether they were entered in the last 8, 12 or 24 hours.

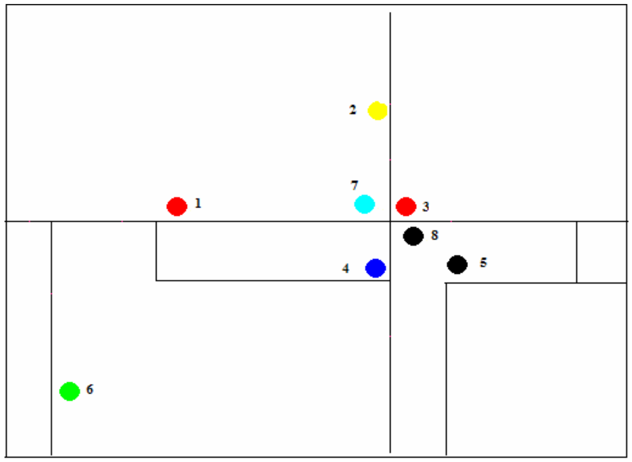

Figure 1: Example Map

After mapping the firms, a *Neighbor Matrix* was constructed to capture the spatial relationship that exists between the firms. Each firm was individually defined as a *center station* and the spatial relationship between this *center station* and all of its competitors was established. This procedure was repeated for all 234 firms using the physical map and a defined set of *Neighbor Matrix Rules* (for specific rules, see Appendix A) that allowed identification for each *center station's* first, second and third degree neighbors. The first degree neighbors represent each *center station's* set of closest competitors. It follows that the second degree and third degree neighbors are sets of competitors that compete less directly with the *center station.* The assumption of degrees of neighbors follows from Gabszewicz and Thisse (1996). The set of *Neighbor Matrix Rules* is constructed from a combination of rules in the literature[12] and intuition of real world arrangements of gasoline stations.

The neighbor relationships were recorded in a 234x234 *Neighbor Matrix*. Each row in the *Neighbor Matrix* represents a single *center station*. A single cell a_{ij} is coded as "*1*" if *station j* is a first degree neighbor of *center station i.* If *station j* is a second degree neighbor with *center*

[12] Shepard (1991), Barron, Taylor and Umbeck (2000b), and Netz and Taylor (2002) all use circular markets around the *center station* of radius one-half mile for first degree neighbors, one mile for second degree neighbors, and two miles for third degree neighbors.

station i then a_{ij} is coded as "*2*" and if *station j* is a third degree neighbor with *center station i* then a_{ij} is coded as "*3*". If *station j* has no relationship with *center station i* then a_{ij} is coded as "*0*". The diagonal in the *Neighbor Matrix* is also coded as "*0"* because stations are not neighbors to themselves. The *Neighbor Matrix* is not symmetric because the coding of cell a_{ij} is determined independently from cell a_{ji}.[13] An example of the *Neighbor Matrix* for the example map in Figure 1 appears in Figure 2.

		1	2	3	4	5	6	7	8
1		0	3	2	2	0	3	1	2
2		3	0	1	2	0	0	1	2
3		2	2	0	1	2	0	1	1
4		2	2	1	0	2	0	1	1
5		0	0	2	2	0	0	2	1
6		3	0	0	0	0	0	0	0
7		2	2	1	1	2	0	0	1
8		2	2	1	1	2	0	1	0

Figure 2: Example Neighbor Matrix

Two matrices of distance and time were created using the *Neighbor Matrix*. The distance in driving *distance* and *driving time* between the center station and its first, second and third degree neighbors was inputted into these matrices using Mapquest.com. The distances were recorded in units of miles and estimated minutes, respectively. The development of the *Distance Matrix* and the *Driving Time Matrix* follows from the *Neighbor Matrix* because only cells in the *Neighbor Matrix* that contained "*1", "2"* or *"3"* were inputted into Mapquest.com to determine the driving distance and time between the two competing stations represented by that cell. The *Distance Matrix* and *Driving Time Matrix* are two weight matrices that may be symmetric in

[13] An example of the nonsymmetrical properties of the *Neighbor Matrix* is demonstrated by the map of eight gasoline stations and the accompanying 8x8 *Neighbor Matrix* that appears in Figures 2 and 3. Cell (7,2) of the *Neighbor Matrix* codes Firm 2 as a second degree neighbor of Firm 7. There are three firms that are closer to Firm 7 than Firm 2 so those three firms code as Firm 7's first degree neighbors. Conversely, cell (2,7) of the *Neighbor Matrix* codes Firm 7 as a 1st degree neighbor of Firm 2 since Firm 7 meets the maximum distance requirements to be a first degree neighbor and Firm 7 is one of the three closest firms to Firm 2.

nature because the distance from *station i* to *station j* is the same as the distance from *station j* to *station i*. However, if *station i* is any degree of neighbor in relation to *station j* while *station j* is not a neighbor to *station i* then the matrices may not be symmetric. Stations that do not have any spatial relationship have zeroes coded into the unfilled cells in the two matrices. Both the *Distance* and *Driving Time Matrix* also code zeroes along the diagonal because stations have no spatial relationship with themselves. Figures 3 and 4 represent examples of the *Distance Matrix* and *Driving Time Matrix* for the map of eight firms in Figure 1, respectively. The *Distance Matrix* has cells that code the distance in miles between competing stations, while the *Driving Time Matrix* has cells that code driving time in minutes between competing stations.

		1	2	3	4	5	6	7	8
1		0	1.2	.8	.7	0	1.5	.5	.9
2		1.2	0	.4	.7	0	0	.4	.6
3		.8	.4	0	.3	1.5	0	.2	.2
4		.7	.7	.3	0	1.8	0	.3	.2
5		0	0	1.5	1.8	0	0	1.7	1.3
6		1.5	0	0	0	0	0	0	0
7		.5	.4	.2	.3	1.7	0	0	.3
8		.9	.6	.2	.2	1.3	0	.3	0

Figure 3: Example Distance Matrix

		1	2	3	4	5	6	7	8
1		0	5	3	3	0	6	2	3
2		5	0	2	3	0	0	2	3
3		3	2	0	2	8	0	1	1
4		3	3	2	0	10	0	2	1
5		0	0	8	10	0	0	9	7
6		6	0	0	0	0	0	0	0
7		2	2	1	2	9	0	0	2
8		3	3	1	1	7	0	2	0

Figure 4: Example Driving Time Matrix

The weakness in the Neighbor Matrix Rules is demonstrated by Firm 5 in Figures 1 and 2. The unique arrangement of the streets in Figure 1 causes Firm 5 to be a second degree competitor to Firms 3, 4 and 7 and a first degree competitor to Firm 8. This is due to rules 1E.ii and 2C.ii (See Appendix A) respectively, which stipulate that firms be designated as first or second degree competitors, respectively, if they are within a certain circular radius of the center firm. The effect on this study's analysis of the market due to unique street arrangement such as this is eliminated by screening for firms that exceed the distance requirements after they are inputted into Mapquest and redefining the *Neighbor Matrix* accordingly. An example of the recalculated *Neighbor Matrix* is shown in Figure 5.

		1	2	3	4	5	6	7	8
1		0	3	2	2	0	3	1	2
2		3	0	1	2	0	0	1	2
3		2	2	0	1	3	0	1	1
4		2	2	1	0	0	0	1	1
5		0	0	3	0	0	0	0	3
6		3	0	0	0	0	0	0	0
7		2	2	1	1	0	0	0	1
8		2	2	1	1	3	0	1	0

Figure 5: Example Recalculate Neighbor Matrix

Quality differentiation of the firms was measured by direct survey of all 234 retail gasoline stations in the market. Direct survey involved calling all 234 retail gasoline stations and asking them six questions about their station attributes. An example of the form used for collecting the firm data can be found in Appendix B. The station attributes that are used as a measure of quality[14] for a station are whether or not it has a food market, a car wash, a service station, pay at the pump option, full service option and the number of pumps. This is similar to the quality attributes used by Netz and Taylor (2002). The first five attributes are represented by binary variables with a "*1*" indicating that a station possesses the specific characteristic, such as having pay at the pump option or a food market or a "*0*" indicating that station does not possess

[14] The lowest octane grade of gasoline at each station was always considered so this was not used as a quality differentiating characteristic.

the characteristic. The binary variables are represented in the regression by FOOD for a food market; WASH for a car wash; SERVICE for a service station; PUMP for pay at the pump option, and FULL for full service option. The number of pumps is also included for each station as an exogenous quality variable. This variable is represented in the regression by NUMPUMP.

B. Competition Variables

Three station level exogenous variables are constructed from the data to quantify the degree of competition in the market area surrounding a *center station.* The first variable, NUM1 represents the number of competitors that are first degree in the *Neighbor Matrix* for a given *center station*. NUM2 and NUM3 represent the number of competitors that are second and third degree in the *Neighbor Matrix* for a given *center station* respectively.

The relationship of a *center station* and whether its neighbors are competitors or part of the same chain is established as a second exogenous station variable. This variable quantifies competition in the market because it is expected that stations of the same brand will interact differently with each other than stations of competing brands. Stations of the same brand are closer substitutes for each other than stations of competing brands. Teitz (1968) predicts that same branded stations will locate at the socially optimal locations of the market to maximize their market share. The same branded relationship is represented by the station specific variable SAME1, which is a ratio of the number of stations that are the same brand out of the first degree neighbors compared to the total number of stations that are first degree neighbors. The variables SAME2 and SAME3 represent the ratio of like branded stations over the total number of stations that are second and third degree neighbors respectively.

The relationship of a *center station* and whether its neighbors are an independent retailer or a major branded station is established as a third exogenous station variable.[15] Barron et all (2000a) find that independent stations charge lower prices than major branded stations and this may lead to different interactions between independents and branded stations in the market. The independently branded relationship is represented by the station specific variable INDEP1, which is the ratio of the number of stations that are independently branded out of the first degree neighbors compared to the total number of stations that are first degree neighbors. The variables SAME2 and SAME3 represent the ratio of like branded stations over the total number of stations that are second and third degree neighbors respectively.

IV. OLS Model

A. Predicted Signs

The regression estimating the effects of station attributes and space on price is analyzed first through OLS regression techniques. The explanatory variables and expected signs can be found in Table 1. The dependent variable in this case is $\ln P_i$ where P_i is the average price of the center station in the market over the four day survey period. Price is used as the dependent variable in this study to show how spatial and quality differentiation affect price. The natural log of the price is taken so that each estimator in the regression represents a percent change in the price of the station rather than an absolute change in the price of gasoline.

In the stable market analysis that represents the basis for this study, each individual gasoline station views its spatial location and quality amenities as fixed in the short term.[16]

---

[15] The major branded stations in the market were defined as *Super America, Mobil, Holiday, Marathon, Amoco, Citgo, Shell, Texaco, Phillips66, Exxon, Conoco* and *Sinclair*. All other stations were defined as independent suppliers.

[16] This corresponds to the spatial differentiation literature (Hotelling 1929, etc) which solve for spatial differentiation by backwards induction with price as the final decision that firms make after choosing location and quality.

Therefore each station can only maximize its revenue[17] and thus its profits through its market price. The first series of OLS regressions, 1(A) – 4(A), model the natural log of the price in terms of station level quality attributes and competitors in the market.

The quality variables for *center station i*, including PUMP, WASH, SERVICE, FOOD and FULL are expected to have positively signed coefficients. Stations possessing these characteristics are assumed to have higher quality and according to the Shaked and Sutton (1982) model, firms with higher quality charge higher prices. The NUMPUMP variable could have a positively or negatively signed coefficient depending upon the specific characteristics of the market. If an increase in the number of pumps causes the gasoline station to be more efficient and serve customers faster, thus providing higher quality, then a positive sign is expected. If NUMPUMP simply represents demand in the market, and if the demand for gasoline is downward sloping, then a negatively signed coefficient is expected.

The number of competitors in the market, NUM, are expected to have a negatively signed coefficient because as the number of competitors in a given market area increases, price competition results and price decreases. This does not account for demand being non-uniformly distributed throughout the market and firms locating close to the demand. There was not direct access to the demand in the market, specifically the traffic density on all streets in the Minneapolis-St Paul market. Despite this fact, due to the nature of gasoline stations and posted prices, as the number of stations increases price should decrease. Netz and Taylor (2002) explain this by examining an individual driving the same route to work each day and observing gasoline prices. If a new station enters the market along the route, the individual observes another price with no additional search cost. This means that the individual has one more option to consider when purchasing gasoline and the additional station leads to more competition in the

[17] Maximizing total revenues represents a reasonable approximation for maximizing total profits for the firm in this model although there are variable costs that are not accounted for in this analysis.

market. Stations can not easily modify their quality or location as discussed above, thus competition takes place in prices and price decreases.[18]

The expected sign of the coefficient for the SAME variable is positive. Major (branded, see footnote #4) gasoline stations operate under a number of contractual arrangements with refiners; owned and operated by the refiner, leased by a single station owner from the refiner, or owned independently at the local level and contracting with the refiner to carry the brand name. In all three agreements, the refiner[19] has the incentive to maximize the total market share of all stations and it attempts to do this by locating them at the socially optimal positions in the market (as described by Teitz (1968)) regardless of the positions of its competitors. The refiner wants to attract consumers from competing stations while not attracting consumers away from its own stations. In the non-optimal case where stations are not strategically spaced throughout the market and competition between like branded stations occurs, station prices would be expected to follow a collusive rather than price competitive outcome. Like-branded stations can use signaling to successfully drive up prices in a market.

The expected sign of the coefficient for the INDEP variable can be positive or negative depending upon specific assumptions of consumers in the market. Barron et al (2000a) find that branded stations charge significantly higher prices than independent stations. As stated in Netz and Taylor (2002), consumers may perceive that independent stations have inferior quality gasoline when compared to gasoline sold at branded stations. If consumers perceive that independent stations are inferior, then the lower prices that independents set will not increase price competition in the market because the branded stations will already be differentiated in terms of quality and thus be able to charge higher prices. This case would lead to the expected

[18] Barron, Taylor and Umbeck (2000b) find that average price is lower in gasoline markets with more stations.
[19] I assume that when the refiner does not own or operate the station it solves the vertical integration problem by extracting franchising fees related to the stations profit so in all three cases the refiner wants to maximize station level profits because this in turn maximizes its own profits.

sign of INDEP being positive because the presence of more independents in the market decreases price competition since independents are differentiated from branded stations through perceived quality differences. Conversely, if consumers perceive that independent stations sell the same quality gasoline as branded stations, then a higher ratio of independents in the market will increase price competition since independents charge lower prices. This leads to an expected negative sign on the variable INDEP.

B. OLS Estimation

The OLS estimation is done for four regression models, Regressions 1(A), 2(A), 3(A) and 4(A). The results of the four regression models are listed in Table 2 and a summary of the significant results for the OLS models appears in Table 3. The first four regression models build upon each other with the data that is collected from the stations. Regression 1(A) provides a baseline reference for the effect of only the center station's quality variables on its own price. Regressions 2(A), 3(A) and 4(A) then begin to build a system of analysis to distinguish between direct and indirect competitors to that center station. It is expected the closest, first degree competitors will be representative of direct competitors as defined by Gabszewicz and Thisse (1996) while the second and third degree competitors will represent indirect competitors to the center station. Regression 2(A) builds upon 1(A) by adding only the first degree, direct competitor variables while second and third degree variables are added in Regressions 3(A) and 4(A), respectively. The effect of indirect competitors is distinguished by changes with the values, sign and significance of the estimators between the 2(A), 3(A) and 4(A). In Regression 4(A) the persistence of the competition variables across the degrees of neighbors can be identified. It is possible that the effect of NUM, SAME and INDEP could have different levels of persistence. A third degree competitor may have a significant effect on the center station's

Table 1: Explanatory Variables and Expected Signs

Variable	Description	Expected Sign
ln(Pi)	Natural log of station i's price for gasoline average over the four day survey period	n/a
FOODi	Qualitative measure of whether a food mart exists, 1=yes, 0=no	+
SERVICEi	Qualitative measure of whether a service station exists, 1=yes, 0=no	+
WASHi	Qualitative measure of whether a car wash exists, 1=yes, 0=no	+
PUMPi	Qualitative measure of whether pay at the pump option exists, 1=yes, 0=no	+
FULLi	Qualitative measure of whether full service option exists, 1=yes, 0=no	+
NUMPUMPi	Quantitative measure of the number of pumps	+ or -
NUM1	Quantitative measure of the number of 1st degree competitors in a station's 1st degree market area	-
NUM2	Quantitative measure of the number of 2nd degree competitors in a station's 2nd degree market area	-
NUM3	Quantitative measure of the number of 3rd degree competitors in a station's 3rd degree market area	-
SAME1	Ratio of the number of same branded competitors divided by the total number of competitors in a station's 1st degree market area	+
SAME2	Ratio of the number of same branded competitors divided by the total number of competitors in a station's 2nd degree market area	+
SAME3	Ratio of the number of same branded competitors divided by the total number of competitors in a station's 3rd degree market area	+
INDEP1	Ratio of the number of independently branded competitors divided by the total number of competitors in a station's 1st degree market area	-
INDEP2	Ratio of the number of independently branded competitors divided by the total number of competitors in a station's 2nd degree market area	-
INDEP3	Ratio of the number of independently branded competitors divided by the total number of competitors in a station's 3rd degree market area	-
ln(P)i(est)1	Estimated natural log of price using average quality variables of the 1st degree competitors	+
ln(P)i(est)2	Estimated natural log of price using average quality variables of the 2nd degree competitors	+
ln(P)i(est)3	Estimated natural log of price using average quality variables of the 3rd degree competitors	+

price due to it being the SAME brand as the center station while its effect as an increase in NUM of third degree competitors may not be significant.

The binary quality variable SERVICE is positive and significant across all four regressions which indicates that if a gasoline station has a service station then they will charge a higher price. The overall effect of the service station is to increase price by approximately 2.26 – 2.99% depending upon the other variables included in the regression. The binary quality variable FOOD is negative and significant in all four regressions which indicates that if a gasoline station has a food market then they will charge a lower price. This result contradicts the predicted sign for this quality variable. However, this result can be explained due to the fact that when a station offers higher quality by possessing a food mart, and can bundle goods together in order to price discriminate among consumers. Gasoline stations with convenience stores have the incentive to charge lower prices at the pump to gain more consumers. The station then earns a higher profits because it price discriminates consumers who then enter the food market and purchase items that have a high markup. The exogenous station variable NUMPUMP is negative and significant in Regressions 2(A), 3(A), and 4(A). This means that as the number of pumps supplying gasoline increases the price of gasoline decreases. This result seems to indicate that gasoline stations have a downward sloping demand curve so that an increase in supply causes a decrease in price. This involves the strategic choice the gasoline stations have between offering higher quantity at a low price or lower quantity at a high price.

The competition variable NUM1 is negative in all three regression models in which it appears and is significant in two of them. This implies that as the number of first degree competitors in the market increases, the price of gasoline decreases. This is expected due to the characteristics of the gasoline market that lead to price competition, including a relatively homogeneous product and posted prices. This result is consistent with the theories presented by

Eaton (1972), that the gasoline market must be sufficiently large for the gasoline stations to locate as spatial monopolists and not compete in prices. In this model, the market size is not large enough for such an outcome and the firms compete in prices because this is the easiest way to gain market share in the short run.

The competition variable NUM2 is negative in Regressions 3(A) and 4(A) and significant in 4(A). Regressions 3(A) and 4(A) only includes the quality variables for the center stations and the competition variables and does not account for the quality characteristics of *center stations*' competitors. The logic for this sign follows from the same line of reasoning presented for NUM1; that in this market, firms compete in prices in order to gain the largest market share and as the number of competitors increases, price decreases. The competition variable NUM3 is significant and positive in Regression 4(A). This contradicts the expected sign outlined in the hypotheses. An explanation for this result could be that the geographic boundaries (neighborhood degrees) defined in this model do not correspond to the actual geographic boundaries of direct and indirect competitors in the market.

The competition variable SAME1 is positive and significant in all three regressions that are modeled using it. This means that as the ratio of firms that are the same brand as the *center station* in the first degree neighborhood increases, price increases as well. This result confirms the expectation that gasoline stations of the same brand attempt to minimize price competition between themselves. The competition variable SAME2 is positive, but not significant in Regressions 3(A) and 4(A) while the competition variable SAME3 is positive and not significant in Regression 4(A).

The competition variable INDEP1 is negative and significant in all three regressions that are modeled using it. This implies that as the ratio of the number of independent gasoline stations in the first degree market compared to the total number of stations in the first degree market increases, price decreases. The sign of this variable substantiates the hypothesis that consumers perceive the gasoline sold by independents to be inferior to the quality of gasoline sold by major brands. Gasoline sold by independents is a not a perfect substitute for gasoline sold by major brands so independents must price lower in order to mitigate the perceived quality difference of their gasoline in the eyes of the consumers.[20] The competition variables INDEP2 and INDEP3 are negative but not significant in any of the regressions in which they appear.

C. Easterly-Levine Regressors in OLS Estimation

The OLS models in Regressions 1(A)-4(A) above predict the price of the *center station* in the market with respect to that station's own quality characteristics and the number and type of competitors surrounding that station in the market. However, there are other factors that affect price. First, quality and competition characteristics of competing firms have not yet been included in this model. Second, the impact of space needs to tested on the model.

The effect that quality differentiation of station's competitors has on the price of the *center station* is captured using a quality based regressor similar to the "weighted predicted growth" variable that Easterly-Levine (1998) use. The variable, $ln(P)_i(est)$, is found by substituting the *center station's* first, second or third degree neighbors' average quality variables into the model estimated by Regression 1(A). For example, to obtain the value for $ln(P)_i(est)1$, the average quality variables ($PUMP_i$, $WASH_i$, etc) of the first degree neighbors are multiplied together with their specific OLS coefficient estimates that are derived from running Regression 1(A). These values are then summed together to create a quality based estimator $ln(P)_i(est)1$ for each individual firm.[21] This estimated value then becomes an independent variable in a new series of OLS Regressions: 2(B), 3(B) and 4(B).

---

[20] Barron et al (2000a) find that branded stations charge significantly higher prices than independent stations.
[21] This method is similar to using 2SLS (two stage least squares) to estimate the coefficients.

Table 2: OLS Estimation Results

	1 (A)	2 (A)	3 (A)	4 (A)	2 (B)	3 (B)	4 (B)
Constant	0.385***	0.391***	0.3910***	0.388***	0.186*	0.0373	-0.0898
	(0.007)	(0.007)	(0.008)	(0.008)	(0.10)	(0.12)	(0.17)
PUMP	0.0066	0.0060	0.0052	0.0056	0.0066	0.0070	0.0072
	(0.0047)	(0.0046)	(0.0046)	(0.0046)	(0.0046)	(0.0046)	(0.0046)
WASH	-0.0034	-0.0030	-0.00261	-0.0021	-0.0018	-0.0016	-0.0016
	(0.0042)	(0.0041)	(0.0041)	(0.0041)	(0.0041)	(0.0041)	(0.0041)
SERVICE	0.0241***	0.0226***	0.0299***	0.0233***	0.0224***	0.0229***	0.0233***
	(0.0044)	(0.0043)	(0.0044)	(0.0044)	(0.0043)	(0.0043)	(0.0043)
FOOD	-0.0109**	-0.0102*	-0.0102*	-0.0104*	-0.00882	-0.008523	-0.00874
	(0.0055)	(0.0054)	(0.0054)	(0.0053)	(0.0054)	(0.0054)	(0.0054)
FULL	0.0047	0.0053	0.0053	0.0066	0.0049	0.0061	0.0067
	(0.0046)	(0.0045)	(0.0045)	(0.0045)	(0.00446)	(0.0045)	(0.0045)
NUMPUMP	-0.00072	-0.0010*	-0.0009*	-0.0010*	-0.0011**	-0.0010*	-0.0010*
	(0.00054)	(0.00053)	(0.00053)	(0.00053)	(0.00053)	(0.00053)	(0.00053)
NUM1		-0.0029*	-0.0025	-0.0051**	-0.0022	-0.0023	-0.0023
		(0.0015)	(0.0016)	(0.0022)	(0.0016)	(0.0016)	(0.0016)
NUM2			-.0039	-0.0032*		0.0023**	0.0030**
			(0.0061)	(0.0016)		(0.0014)	(0.0014)
NUM3				0.0024*			-0.0004
				(0.0013)			(0.0007)
SAME1		0.0179*	0.0188*	0.0206**	0.0207**	0.0221**	0.0203**
		(0.0097)	(0.0097)	(0.0097)	(0.0097)	(0.0097)	(0.0097)
SAME2			0.0102	0.0083		0.0100	0.0083
			(0.0074)	(0.0075)		(0.0073)	(0.0074)
SAME3				0.0171			0.0178*
				(0.010)			(0.010)
INDEP1		-0.0136*	-0.0132**	-0.0146**	-0.0147**	-0.0163**	-0.0163**
		(0.0063)	(0.0065)	(0.0066)	(0.0063)	(0.0063)	(0.0065)
INDEP2			-0.0005	-0.0018		-0.0007	-0.0007
			(0.0062)	(0.0063)		(0.0062)	(0.0063)
INDEP3				0.0088			0.0067
				(0.0078)			(0.0078)
LNP1est					0.531**	0.473*	0.421
					(0.26)	(0.26)	(0.26)
LNP2est						0.432**	0.458**
						(0.21)	(0.21)
LNP3est							0.355
							(0.28)
R^2	0.2450	0.2949	0.3087	0.3202	0.3077	0.3325	0.3484
adj R^2	0.2255	0.2665	0.2712	0.2734	0.2767	0.2898	0.2938
Observations	234	234	234	234	234	234	234

*** : 1%
**: 5%
*: 10%
Standard Errors in parentheses.

Table 3: OLS Model Analysis

Model	Specification	Results (significant variables)
1 (A)	Center station quality variables	SERVICE (+) FOOD (-)
2 (A)	Center station quality variables and 1st degree competition variables	SERVICE (+) FOOD (-) NUMPUMP (-) NUM1 (-) SAME1 (+) INDEP1 (-)
3 (A)	Center station quality variables, 1st and 2nd degree competition variables	SERVICE (+) FOOD (-) NUMPUMP (-) SAME1 (+) INDEP1 (-)
4 (A)	Center station quality variables, 1st, 2nd and 3rd degree competition variables	SERVICE (+) FOOD (-) NUMPUMP (-) NUM1 (-) SAME1 (+) INDEP1 (-) NUM2 (-) NUM3 (+)
2 (B)	Center station quality variables, 1st degree competition variables and 1st degree quality based regressor	SERVICE (+) NUMPUMP (-) SAME1 (+) INDEP1 (-) *ln(P)i(est)1* (+)
3 (B)	Center station quality variables, 1st and 2nd degree competition variables, and 1st and 2nd degree quality based regressors	SERVICE (+) FOOD (-) SAME1 (+) INDEP1 (-) NUM2 (+) *ln(P)i(est)1* (+) *ln(P)i(est)2* (+)
4 (B)	Center station quality variables, 1st, 2nd and 3rd degree competition variables and 1st, 2nd and 3rd degree quality based regressors	SERVICE (+) FOOD (-) NUMPUMP (-) SAME1 (+) INDEP1 (-) NUM2 (+) SAME3 (+) *ln(P)i(est)2* (+)

The variable *ln(P)$_i$(est)2* is found by taking the average of the quality variables for the second degree neighbors and then estimating Regression 1(A) with these values and the variable *ln(P)$_i$(est)3* is found by taking the average quality variables of the third degree neighbors and then estimating Regression 1(A) with these values. Regressions 2(B), 3(B) and 4(B) include the quality based regressors and utilize the same general idea as Regressions 2(A), 3(A) and 4(A). Regression 2(B) consists of the quality variables for the center station, the competition variables for the first degree neighbors and the quality based regressor for the first degree neighbors. Regression 3(B) consists of the quality variables for the center station, the competition variables for the first and second degree neighbors and the quality based regressors for the first and second degree neighbors. Regression 4(B) consists of the quality variables for the center station, the competition variables for the first, second and third degree neighbors and the quality based regressors for the first, second and third degree neighbors. The system of analysis constructed between Regressions 2(B)-4(B) allows one to distinguish between direct and indirect competitors in the market and also determine the persistence of the quality based regressor across the model.

D. Predicted Signs and Results for OLS with Quality Based Regressor

The quality based regressor *lnP$_i$(est)* estimates the natural log of price of competing stations based upon each station's own quality characteristics. The expected signs for the coefficients for the quality based regressors are positive because station prices are expected to move together. Additionally, the first degree quality based regressor should be greater in magnitude and significance than the second and third degree quality based regressors because the first degree competitors are in competition with the *center station* due to the *Neighbor Rules.* This indicates that a change in quality for a first degree competitor has a greater effect on the *center station's* price than a change in quality for a second of third degree competitor.

The quality based regressor also should have an effect on the magnitude of the other variables in the model. One possible effect could be that the quality based regressor drowns out the other estimators in the model. If this domination is present then it motivates the use of further techniques to correct for the use of the quality based regressor. Since earlier studies such as McCoskey (2004) have found that the weighted neighbor variable drowns out the significance of the other variables, it is predicted that the magnitude of the other estimators will decrease and some will lose their significance. There should be a greater effect on the competition variables, in particular NUM, than on the *center station* quality variables. This prediction assumes that the quality based regressor will somehow include NUM in its estimation of price and thus drown out NUM's own significance in the model.

The quality based regressor $lnP_i(est)1$ is positive and significant in Regressions 2(B) and 3(B), but positive and not significant in Regression 4(B). As the first degree competitor's average natural log of price increases the *center station's* natural log of price increases as well. The percent change in the price of the *center station* in Regression 2(B) is .531% per 1% increase in the quality based regressor. It is interesting to note that the price change is inelastic rather than unit elastic.[22] A change in a first degree competitor's price does not have a one to one affect on the center station's price. This quality based regressor effect is still significantly larger in magnitude than the effects of the other independent variables in the regression. This implies that the quality characteristics of the surrounding gasoline stations have a meaningful affect on the price of the *center station.* The quality based regressor $lnP_i(est)2$ is positive and significant in Regressions 3(B), and 4(B) and the quality based regressor $lnP_i(est)3$ is positive but not significant in Regression 4(B). The magnitude of the coefficient of first degree quality based regressor, $lnP_i(est)1$, is greater than the magnitude of the coefficient of the second degree

[22] The hypothesis test of whether or not the coefficient for the first degree quality based regressor was equal to 1 rejected the null hypothesis to the 10% level of significance.

quality based regressor in Regression 3(B). This implies that the quality characteristics of the first degree competitors have a larger effect on the *center station's* price than the quality characteristics of the second degree competitors.

In Regression 4(B) the first degree quality based regressor has a value of .421 and is no longer greater in magnitude than the second degree quality based regressor, which has a value of .458. The second degree quality based regressor is significant while the first degree is not. This result was troubling in light of the predictions and results confirmed in Regression 3(B). The coefficient for the quality based regressor $lnP_i(est)2$ increases in magnitude, from a value of .432 in 3(B) to a value of .458 in 4(B)), with the introduction of the variable $lnP_i(est)3$ to the model. The result in Regression 4(B) does not seem to intuitively make sense because it implies that the quality characteristics of the second degree competitors have a greater effect on the price of the *center station* than the quality characteristics of the first degree competitors. This cannot be attributed to the greater number of stations in the second degree neighborhood causing this effect because in Regression 3(B), the magnitude of the coefficient for $lnP_i(est)1$ is greater than the magnitude of the coefficient for $lnP_i(est)2$. The inclusion of the third degree quality based regressor seems to be altering the results in some way. More study is needed to understand the phenomena that occurred in Regression 4(B). One possibility is that the actual geographic boundaries differ from the geographic boundaries that were constructed in this study using the Neighbor Rules.

The introduction of the quality based regressors to the model affects the magnitude and significance of the other estimated variables in the model. The binary quality variable SERVICE is not significantly different in Regressions 2(B), 3(B) and 4(B) from the earlier results in 2(A), 3(A) and 4(A), but the binary variable FOOD is negative, yet no longer significant in 2(B), 3(B) and 4(B). One possible reason for the insignificance of FOOD is that once the quality based regressor is added, quality differentiation of competing firms has a much greater effect on price than adding a food market does, which would drown out the significance of FOOD. Such a result is noted in the predictions for this model, although a more significant drowning out effect is expected on the competition variables. The exogenous quality variable NUMPUMP is still negative and significant in Regressions 2(B), 3(B) and 4(B).

Some of the competition variables are also affected by the addition of the quality based regressor to the model in Regressions 2(B), 3(B) and 4(B). NUM1 is negative but no longer significant in these regressions compared to Regressions 2(A) and 4(A), where it is significant. One reason for this is that the number of first degree competitors is accounted for by the quality based regressor.[23] This result alone is not remarkably disturbing until the results of NUM2 are examined. NUM2 is *positive* and significant in Regressions 3(B) and 4(B) compared to Regression 4(A) where it is negative and significant. The positive sign implies that as the number of second degree competitors to a *center station* increases, the *center station's* price increases. Additionally, the significance of NUM2 while NUM1 is insignificant seems to contradict the theory of persistence across the market; firms closer in proximity to the *center station* should have a larger effect on price. The quandaries of the NUM2 result are addressed one at a time: first the switch of the sign from positive to negative, and then the rationale for a positively signed coefficient.

Regression 4(A) has a lower adjusted R^2 than Regressions 3(B) and 4(B) implying that Regressions 3(B) and 4(B) more accurately model the independent variables that effect price of the gasoline stations. The lack of inclusion of the quality based regressor in Regression 4(A)

[23] The average number of first degree neighbors for the model was 1.07. The first degree quality based regressor averaged the qualities of the first degree neighbors together and since on average, there is one first degree neighbor, this could correlate with NUM1 and drown out its significance.

potentially biases the results (omitted variables bias) causing NUM2 to have a negatively signed coefficient that is significant when in reality it should have a positively signed coefficient.

The fact that NUM2 seems to have a positively signed coefficient can be explained by restrictions, such as zoning laws that may exist in high demand areas so that not enough gasoline firms can locate to adequately supply the high demand. In such a scenario, second degree competitors represent firms that *want* to locate as first degree competitors in the market but can not due to zoning restrictions. The *center station* represents a station that is located in the high demand area. An increase in the number of second degree competitors indicates high demand around the *center station*, but the second degree competitors are unable to directly supply gasoline to that market of consumers and the *center station* is able to raise its price since quantity demanded of gasoline is greater than quantity supplied at the market price. This result also relates to the theory proposed by Eaton and Lipsey (1975) when addressing the equilibrium conditions for a number of firms in a two dimensional market. Eaton and Lipsey propose a general principle of clustering with firms pairing off.

The descriptive statistics for the NUM1 and NUM2 indicate an average 1.07 first degree competitors and 1.51 second degree competitors. The first degree competitors represent the firms that pair off with each other while the second degree competitors represent the firms that locationally cluster but do not engage in intense price competition with the first degree pair. They simply serve as indicators of demand in the first degree market. In such a scenario as the number of second degree competitors increases, price of the center station would increase as well.

The introduction of the quality based regressor does not affect the significance of SAME1 or INDEP1 in Regressions 2(B), 3(B) and 4(B). SAME1 remains positive and significant in the new model while INDEP1 remains negative and significant. The quality based regressor does affect SAME3, which is positive and significant in Regression 4(B) compared to being positive and insignificant in Regression 4(A). Incidentally, the competition variable SAME2 is positive and insignificant in Regressions 3(B) and 4(B). This result is unexpected because it implies that same branded stations in the second degree neighborhood have less of an effect on the natural log of price of the *center station* compared to same branded stations in the third degree neighborhood. Second degree neighbors are closer in proximity to the *center station* and intuitively their effect is expected to be greater. An explanation for this result is that the number of stations in the third degree neighborhood is much greater than the number of stations in the second degree neighborhood since the third degree neighborhood is larger in size. The overall average number of stations in a third degree neighborhood is 3.78 while the average number of stations in a second degree neighborhood is 1.51. A larger *number* of third degree, same-branded stations, might be causing a more significant upward movement in natural log of price of the *center station* than a smaller number of second degree, same-branded stations despite the fact that the second degree stations are closer in competitive degree to the *center station*. This leads to the variable SAME3 being significant while SAME2 is not. This effect is not witnessed with SAME1 being insignificant as well because the nearly direct proximity of the first degree competitors to the *center station* offsets the greater number of same-branded competitors in the second and third degree neighborhoods.

E. Persistence Across Space in OLS Models

The persistence of results across space can be analyzed using the seven OLS Regressions. In this model, persistence is defined as how far from the *center station* (to what neighborhood degree, first, second or third) is a variable still significantly affecting the *center station's* price. Persistence can be analyzed for the four variables that are defined across the neighborhoods:

NUM, SAME, INDEP, and *lnP_i(est)*. It is interesting to note that the persistence for each of the variables is not robust across the models. NUM is persistent to the second degree neighborhood. This means that the numbers of gasoline stations in the first and second degree neighborhoods of a center station are having a significant affect on the *center station's* price. The number of stations in the third degree neighborhood seem to have enough distance between themselves and the *center station* so that they have no significant effect on price. SAME is persistent to the third degree neighborhood. In this case, firms as far away as the third neighborhood from the center station that are the same brand do have an effect on the *center station's* price. INDEP is persistent to only the first degree neighborhood. Independent firms beyond the first degree neighborhood of a *center station* do not seem to affect the *center station's* price. *lnP_i(est)* was persistent across both the first and second degree neighborhoods. The quality characteristics of firms in the first and second degree neighborhoods affect the price of the *center station.*

V. Spatial Autocorrelation

The use of the quality based regressor to weight the average qualities of the competitors in the market is only the first step in the spatial analysis of the Minneapolis-St. Paul retail gasoline market. The significance of the quality based regressor and competition variables give evidence that space matters to the gasoline station. The next step in this analysis is to test for potential spatial autocorrelation and then continue with re-estimation in the potential presence of spatial autocorrelation. Spatial autocorrelation occurs in any model where location affects firm behavior. In this model it is hypothesized that location is an extremely important price factor determinant for the gasoline station and thus spatial autocorrelation is expected to be present.

In this model, the spatial arrangement of firms will be captured using a ρ parameter if spatial autocorrelation is found be to be present. This term will capture the spatial arrangement of the firms in the model. Independent variables in the model that could have a significant effect on space, but are not included in this model, include demographic factors that are non-uniformly distributed in the metropolitan area. The non-uniform distribution of these demographic factors across space may cause prices to be higher in certain areas of the market and lower in others and this could cause the non-uniform distribution of error terms in the model. Specific demographic factors that may have a significant impact include median income in an area, traffic density, crime rates and land value.

A. Tests for Spatial Autocorrelation

Spatial autocorrelation represents a two dimensional form of correlation that occurs in spatial models. The simpler problem serial correlation occurs in one dimensional models, across time series data. If spatial autocorrelation is present in the model then the error terms of the OLS model are not distributed:

$$\varepsilon_i \sim (iid)\ N(0,\sigma^2) \qquad (2)$$

This leads to problems with the estimated variances of the OLS estimators so that they may not be BLUE (the best linear unbiased estimators). They may not have minimum efficiency and thus their t-statistics may be invalid. Their inference may be biased.

The tests for spatial autocorrelation are described in detail in Anselin and Hudak (1992). The tests for spatial autocorrelation all involve the use of a weight matrix. The *Distance Matrix* and *Driving Time Matrix* are constructed to be used as weight matrices in the spatial autocorrelation tests once their rows have been normalized.[24] The *Distance* and *Driving Time Matrices* are row normalized by first taking the inverse of each cell in the matrix. This is done

[24] The reason for the row normalization of the weight matrix, according to Anselin and Hudak (1992), is so that the weight matrix represents a weighted average of observations of neighboring locations.

so that shorter distances and driving times are weighted as having a greater magnitude in the actual matrix. Firms that are closer together have a greater interaction with each other than firms that are further apart according to the theoretical literature and this transformation of the matrix allows us to model this. The rows are then normalized by summing each row individually and then dividing each cell by its row sum.

The two tests that are used to test for the presence of spatial autocorrelation are the Lagrange Multiplier Lag (LM Lag) and Lagrange Multiplier Error (LM Error) tests. Lagrange multiplier tests create a log-likelihood function and then hypothesis test whether or not the slope of this log-likelihood function is significantly different from zero at the given regression coefficients for the estimated model. The specific LM Lag and LM Error tests here correspond to the problem of spatial autocorrelation in the estimated model and therefore the tests use the spatial weight matrix and the OLS residuals as components for the log-likelihood function. LM Error test is based upon the hypothesis that in the following specification,

$$\varepsilon = \lambda W \varepsilon + \mu :$$

$$H_0 : \lambda = 0$$

$$H_A : \lambda \neq 0 \qquad (3)$$

This test is based upon a null hypothesis of no spatial dependence against an alternative hypothesis of spatial dependence. Additionally, if H_A is true the OLS estimators no longer have minimum variance among all linear unbiased estimators and therefore their *t* and *F* tests may not be valid. The invalidation of the *t* tests means that one can no longer accurately test for significance in the OLS estimators. This means that the OLS models need to be re-estimated to ensure that the *t* tests are valid.

The second test for spatial autocorrelation, the LM Lag test, is based upon the hypothesis that in the:

$y = \rho Wy + X\beta + \mu$:

$$H0 : \rho = 0$$

$$HA : \rho \neq 0 \qquad (4)$$

Similarly to the LM Error test, the null hypothesis in the LM Lag test implies no spatial dependence against an alternative hypothesis of spatial dependence in the model.

The LM Error and LM Lag tests are similar in structure but they capture the correlation across space in different ways. The LM Error test captures correlation in the residuals while the LM Lag test captures correlation in lagged dependent variables. Both the LM Error and LM Lag tests use the OLS residuals and the two normalized *Distance* and *Driving Time Matrices.* The tests are asymptotically distributed as chi-square with one degree of freedom. If spatial autocorrelation is found to be present in the model, the most common remedy is to re-estimate the model using Maximum Likelihood (ML) Estimation techniques.

B. LM Error and LM Lag Test Results

Spatial autocorrelation tests are run on all seven regression models using LM Lag and LM Error on each model. A total of four tests for spatial autocorrelation are used on each model since both the *Distance Matrix* and the *Driving Time Distance* are normalized and used as weight matrices for both the LM Lag and LM error tests. The results are that all test statistics for the seven regression models are statistically significant with p-values of less than one percent. The results of the spatial autocorrelation tests are displayed in Table 4. The presence of spatial autocorrelation in the model means that there is some part of the spatial arrangement of the firms in the market that is not captured in the previous regressions. The spatial relationship is captured in the error terms of the OLS models, rather than in the independent variables themselves.

Although all the tests are significant to the 1% level, the level of actual significance differs. The model that best captures space can be distinguished by comparing p-values of the test statistic for the LM Error and LM Lag tests.[25] Comparison of the p-values across the regressions from 1(A) to 4(A) displays decreases in the level of significance for both tests. Remember that in each Regression from 1(A) to 4(A) another level of competition variables is added to the model. Therefore, adding degrees of competitors to the model is accounting for some of the spatial interactions among firms.

The values of the p-values are found to be less in Regression 2(B) compared to Regression 4(A). This means that Regression 2(B) is more likely to have a problem of spatial autocorrelation than Regression 4(A). Regression 2(B) adds the quality based regressor to its model but only has the first degree competition variables and the first degree quality based regressor. As the degrees of neighbors are added in Regressions 2(B)-4(B) the p-values for LM Lag are well behaved and increase as is expected. Conversely, the p-values for LM Error are not well behaved and decrease in Regression 3(B) before increasing in Regression 4(B). The comparison between Regressions 3(B) and 4(B) indicates that 4(B) is more likely to be better modeling the spatial interactions among the firms due to its inclusion of the third degree neighbor variables.

[25] The test statistics are still random variables so comparison of their p-values is not very persuasive in determining which model best models space in the problem. The comparisons are included because they may lend some credence to the assertion that space is best captured when including all three degrees of neighbors in the model, but the method is by no means foolproof.

Table 4: Spatial Autocorrelation Testing Results

Model	1 (A)	2 (A)	3 (A)	4 (A)	2 (B)	3 (B)	4 (B)
Distance Matrix							
LM Lag	13.223***	12.372***	12.279***	9.872***	11.112***	9.145***	7.824***
p-value	0.0003	0.0004	0.0005	0.0017	0.0009	0.0025	0.0052
LM Error	51.767***	37.654***	38.652***	31.654***	38.127***	40.316***	34.359***
p-value	6.3e-013	8.5e-010	5.1e-010	1.8e-008	6.6e-010	2.2e-010	4.6e-009
Driving Time Matrix							
LM Lag	12.217***	11.595***	10.178***	8.986***	10.419***	8.497***	7.248***
p-value	0.0005	0.0007	0.0014	0.0027	0.0012	0.0036	0.0071
LM Error	59.684***	45.005***	44.661***	39.400***	44.389***	46.741***	40.872***
p-value	1.11e-014	2.0e-011	2.3e-011	3.4e-010	2.7e-011	8.1e-012	1.6e-010

*** : 1%
**: 5%
*: 10%

C. Re-estimation of the Model Using Maximum Likelihood (ML) Estimation

Anselin and Hudak (1992) suggest that the best alternative to OLS estimation in the presence of spatial autocorrelation is ML Estimation. ML Estimation allows us to derive the coefficients, βs, for the independent variables in the model as well as a spatial correlation parameter ρ. Maximum Likelihood Estimation is used because it is a more versatile estimation method than OLS estimation in that it allows for the estimation of the ρ parameter. ML Estimation involves the creation of a probability density function that assumes a specific distribution for the error term which need not be iid[26]; in this case the error term is assumed to follow a normal distribution modified by the inclusion of the spatial weight matrix which allows for cross-correlations among the error terms.[27] The maximizing of the log-likelihood function is in essence finding the value of the unknown parameters that maximizes the probability of the function returning the observed prices in the model. The concentrated log-likelihood model

[26] iid refers to the error terms being identically independently distributed.
[27] This is the reason for the inclusion of the first term in the log-likelihood function, $\ln|I - \rho W|$.

estimates the value of ρ first and then using this value estimates the value of the *β(hat)s*. This differs from the unconcentrated log-likelihood model which estimates the values of ρ and the *β(hat)s* simultaneously in the log-likelihood function.

The log-likelihood function to be maximized for the concentrated lag model is given by:

$$L_C = \ln|I - \rho W| - (n/2) * \ln(\pi) - (n/2) * \ln(1/n)(e_0 - \rho e_L)'(e_0 - \rho e_L) \quad (5)$$

where *n* equals the number of observations, and:

$$e_0 = y - X\beta_0 \quad (6)$$

$$e_L = Wy - X\beta_L \quad (7)$$

β_0 represents a column vector of the OLS estimators from the models run in regression 1 (A) and 4 (A) while β_L represents a column vector of the OLS estimators for the lagged model.[28] From Anselin (1988), once the *ρ(hat)* that maximizes the L_C is computed,

$$\beta(hat) = (\beta_0 - \rho\beta_L) \quad (8)$$

and

$$\sigma(hat)^2 = (1/n) * (e_0 - \rho e_L)'(e_0 - \rho e_L) \quad (9).$$

The concentrated lag likelihood model differs from the un-concentrated model because it estimates the value of ρ initially and then uses this value to determine the other coefficients in the model. The un-concentrated lag likelihood model maximizes the value of ρ and the coefficients jointly.

D. Maximum Likelihood Estimation Results

The results of the Maximum Likelihood Estimation are provided in Table 5. The Maximum Likelihood results are obtained using a total of three different weight matrices. Regressions 1(C) and 4(C) both use the concentrated lag model and just a first degree weight

[28] The lagged model has the dependent variables equal *W*y,* which is a [n x 1] vector, and the independent variables equal *X*, which is the [n x k] matrix.

matrix which coded “1” for first degree neighbors and “0” for everything else. Regressions 1(D) and 4(D) use the normalized *Distance Matrix* as *W.* The spatial lag coefficient for all four models is significant but in 1(C) it is negative. This result is problematic because it indicates the prices of first degree competing stations do not move together but rather have an inverse relationship. There is not spatial clustering in the market in terms of prices but rather spatial separation.

The other three Regressions, 1(D), 4(C) and 4(D) all have positive, significant spatial lag coefficients indicating spatial clustering in terms of price in the model. In other words, prices of competing stations move together. All four regressions have positive significant SERVICE coefficients and negative significant FOOD coefficients. This corresponds to the findings in the earlier OLS Regressions. NUMPUMP is negative and significant in Regressions 4(C) and 4(D) and this matches the result in 4(A). It is interesting to note that the inclusion of spatial lag does not drown out the significance of FOOD as it does in Regression 4(B) with the inclusion of the quality based regressor.

The competition variable NUM1 is negative and significant in Regressions 4(C) and 4(D) while both NUM2 and NUM3 are insignificant. This corresponds to the sign that was hypothesized earlier. SAME1 is positive and significant in both 4(C) and 4(D) and SAME3 is positive and significant in 4(C). This result is similar to the result found 4(B) where the third degree competition variable is significant but the second degree one is not. One possible explanation for this result, as discussed earlier, is the number of competitors in the second compared to the third degree neighborhood on average in the market. The overall average number of stations in a third degree neighborhood is 3.78 while the average number of stations in a second degree neighborhood is 1.51. A larger *number* of third degree, same-branded stations, may be causing a more significant upward movement in natural log of price of the *center station* than a smaller number of second degree, same-branded stations despite the fact that the second degree stations are closer neighbors to the *center station*. This leads to the variable SAME3 being significant while SAME2 is not. This effect is not witnessed with SAME1 being insignificant as well because the nearly direct proximity of the first degree competitors to the *center station* offsets the greater number of same-branded competitors in the second and third degree neighborhoods. Another explanation for this phenomenon could be that the arbitrary boundaries established for this study differ from the actual geographic boundaries that exist in the market. Finally, the competition variable INDEP1 is negative and significant in 4(C) and 4(D) which corresponds to our earlier OLS results.

The maximum likelihood model confirms the analysis that persistence in the market occurs for the number of competitors and the independent number of competitors only to the first degree neighbor. There does not seem to be a significant affect on price from indirect competitors in the second and third degree neighborhoods except for SAME branded stations, which are persistent to the third degree neighborhood depending upon the weight matrix used in the maximum likelihood estimations.

The importance of the weight matrix used is illustrated by 4(E). This regression uses an entirely derived weight matrix and weights all first degree neighbors as 2, second degree neighbors as .5 and third degree neighbors as .125. This is a linear specification for the neighborhood relationship rather than a quadratic one and the choice of a slope of four[29] is also no accident because it produces the best model for our data accounting for significant

[29] A number of different "weighted" matrices were run in our preliminary Maximum Likelihood Estimations to try to determine the best relationship among the neighbors. Different matrices including weighting the neighbors quadratically (1st degree weighted as 16, 2nd degree as 4 and 3rd degree as 2), weighting them logarithmically (1st degree weighted as 3, 2nd degree as 2 and 3rd degree as 1) and weighting them with different linear slopes.

coefficients, maximum log-likelihood value and maximum AIC.[30] Regression 4(E) is interesting because of the changes in variables' significance from Regression 4(C) and 4(D). NUM1, NUM2 and NUM3 are *all* negative and significant in Regression 4(E) while SAME1 is positive but not significant. The magnitudes of the coefficients for NUM1, NUM2 and NUM3 decrease from -.0592 to -.0134 to -.0039. First degree neighbors decrease a *center station's* price approximately four more times than second degree neighbors. Second degree neighbors decrease a *center station's* price approximately three more times than third degree neighbors. The new weight matrix seems to be pulling out persistence in the form of number of competitors having a large, significant effect on price. This effect occurs with both direct (first degree neighbors) and indirect (second and third degree neighbors). It also seems to drown out the effect of SAME branded stations on price even in the first degree market. The effect of INDEP first degree competitors is not changed by a significant amount as it remains negative and significant in the regression.

The new 'weight' matrix is based upon using the *Neighbor Rules* that were developed to separate the firms into degrees of neighbors. The strong significance of the three NUM variables and their decreasing magnitude is suggestive of strength in the *Neighbor Rules* compared to simply using a formula of firm distance from the *center station*. The *Neighbor Rules* account for such effects as divided highways, intersections and maximum number of firms in a neighborhood compared to a system of simply distance that does not account for these factors.

Comparison of these results to the results of Netz and Taylor (2002) is difficult since they used a different dependent variable, spatial differentiation, than this study which used a dependent variable of the natural log of price. Netz and Taylor (2002) find that increases in the

---

[30] AIC is the Akaike Information Criterion, described in Anselin (1988)) where AIC = -2L + 2K/ L is the log-likelihood and K is the number of explanatory variables. Models with the lowest AIC give the best compromise between raw fit and parsimony; similar to the use of adjusted R^2 in OLS models.

number of stations, the fraction of the same brand as the center station, and the fraction of the market served by independents lead stations to locate further from their competitors. This study finds that an increase in the number of stations and the fraction of independents leads to a decrease in price of the center station while an increase in the fraction of same brand as the center station leads to an increase in price of the center station. The findings of this study seem to confirm the results found by Netz and Taylor; the price decrease when the number of competitors and fraction of independents increases indicates that stations want to spatially differentiate in order to gain market power. Same branded stations act differently in the market; even though an increase in the number of same branded stations causes an increase in price, these types of stations do not want to cluster. They want to prevent themselves from becoming too good a substitute for their same branded neighbor despite the fact that they can price collude. They would rather spatially differentiate and cover a larger market area.

Table 5: Maximum Likelihood Estimation Results

	1 (C)	1 (D)	4 (C)	4 (D)	4 (E)
Weight Matrix	1st degree W	Normalized W	1st degree W	Normalized W	Graded Matrix
Spatial Lag	-0.0088**	.1090***	0.1450***	0.0970***	0.0770***
	(0.0038)	(.0303)	(0.0231)	(0.0294)	(0.0091)
Constant	0.3899***	.3454***	0.3895***	0.3545***	0.3901***
	(0.0073)	(0.0131)	(0.0073)	(0.0127)	(0.0071)
PUMP	0.0060	0.0076*	0.0042	0.0062	0.0033
	(0.0046)	(0.0045)	(0.0042)	(0.0043)	(0.0040)
WASH	-0.0032	-0.0020	-0.0013	-0.0009	-0.0008
	(0.0041)	(0.0040)	(0.0038)	(0.0038)	(0.0038)
SERVICE	0.0236***	0.0235***	0.0202***	0.0225***	0.0190***
	(0.0043)	(0.0042)	(0.0040)	(0.0041)	(0.0038)
FOOD	-0.0111**	-0.0136**	-0.0124**	-0.0129**	-0.0127***
	(0.00053)	(0.0053)	(0.0049)	(0.0051)	(0.0047)
FULL	0.0051	0.0026	0.0049	0.0045	0.0045
	(0.0045)	(0.0045)	(0.0041)	(0.0043)	(0.0040)
NUMPUMP	-0.0008	-0.0007	-0.0009*	-0.0009*	-0.0009*
	(0.0005)	(0.0005)	(0.0005)	(0.0015)	(0.0005)
NUM1			-0.0559***	-0.0029*	-0.0592***
			(0.0086)	(0.0015)	(0.0068)
NUM2			0.0014	0.0018	-0.0134***
			(0.0013)	(0.0013)	(0.0022)
NUM3			-0.0004	-0.0009	-0.0039***
			(0.0006)	(0.0006)	(0.0007)
SAME1			0.0153*	0.0161*	0.0124
			(0.0089)	(0.0092)	(0.0086)
SAME2			0.0083	0.0075	0.0078
			(0.0068)	(0.0071)	(0.0065)
SAME3			0.0158*	0.0137	0.0152
			(0.0095)	(0.0098)	(0.0091)
INDEP1			-0.0143**	-0.0132**	-0.0140**
			(0.0060)	(0.0062)	(0.0058)
INDEP2			-0.0015	-0.0016	-0.0012
			(0.0056)	(0.0059)	(0.0054)
INDEP3			0.0086	0.0069	0.0079
			(0.0071)	(0.0073)	(0.0068)
Iterations	12	13	19	13	19
Likelihood	605.7738	609.9016	626.095	620.8436	634.02
AIC	-1197.5476	-1205.8032	-1220.190	-1209.6872	-1236.04

VI. Conclusion

The theoretical literature is inconclusive on whether firms maximally or minimally differentiate in a competitive market. This study contributes to the empirical literature, which has found cases of both firm clustering and spatial differentiation, by studying price competition in the retail gasoline industry. Price competition is measured in terms of spatial locations of the firms in the market and firm quality.

Spatial location of firms had the largest and most significant effect on firm price in this study. The positive significance of the spatially lagged autoregressive coefficient in the maximum likelihood estimations means that gasoline firm prices move together and firms that are closer in neighborhood degree have a more significant effect on price. Competition in the market in the form of same branded competitors tends to increase price. This confirms the hypothesis that gasoline refiners are successful to a certain degree in vertically integrating individually owned stations of their "brand" of gasoline so that they do not engage in price competition. Independent competitors decrease the price of their competitors, indicating the consumers do not view independents as inferior. They are in fact driving down prices when they compete in the market because consumers view them as adequate substitutes. It is interesting to note that while the number of competitors in the market affect price to the third degree neighborhood in some specifications, independently branded stations only affect price when they were in the first degree neighborhood. Same branded stations have a few instances where the third degree same branded neighbors are significant, but in general its affect is only persistent to the first degree neighborhood as well. Overall, the persistence in types of competition varied across the market.

Quality differentiation among firms has less effect on price than spatial competition. The two most significant measures of quality that affected price are convenience stores and service stations. Possessing a service station causes stations to increase price while possessing a convenience store causes stations to decrease price. An increase in the number of pumps also causes stations to decrease price. This result can be attributed to a downward sloping demand curve for gasoline.

The model constructed in this study represents a clear departure from standard microeconomic analysis due to the inclusion of space. Space embodies an important element that must be accounted for when describing price competition in any market and this analysis proves that. Space is continually significant throughout the models that are constructed for the retail gasoline industry. Future study in this area could include optimization of a gasoline market to maximize social welfare for consumers and firms rather than allowing firms to only be profit maximizing entities.

VII. Glossary

Adjusted R^2 **–** A version of R-squared that has been adjusted for the number of independent variables in the model.

Bertrand Paradox – Two firms simultaneously set price treating quantity as given; the Nash Equilibrium is that both firms set price equal to the marginal cost of production.

Binary Variable – A qualitative variable that assumes a value of 1 or 0 depending on the presence or absence of an attribute, respectively.

Endogenous Variable – Dependent variable; a variable whose value is determined within the model from the values of other variables.

Exogenous Variable – Predetermined variable; a variable whose value is determined outside the model.

Maximum Likelihood Estimation – A method used to estimate parameters that maximizes a likelihood function; the likelihood function assumes a probability distribution for the error terms in the regression.

Nash Equilibrium – If there is a set of strategies with the property that no firm can benefit from changing their strategy while the other firms keep their strategies unchanged, then that set of strategies and the corresponding payoffs constitute the Nash Equilibrium.

Ordinary Least Square (OLS) - A method used to construct a sample regression functions and estimate parameters (coefficients) by minimizing the sum of the squared error terms.

P-value – The probability of committing a Type I error (rejecting the null hypothesis when it is true).

Spatial autocorrelation – Problem with the residuals (error terms) of the OLS model where they are not identically, independently distributed in a normal distribution with a mean of zero and a variance of sigma squared.

Two-Stage Least Square (2SLS) – A method used to construct sample regressions functions and estimate coefficients using successive applications of OLS.

VIII. Bibliography

d'Aspremont, Claude, Jean Jaskold Gaszewicz, and Jacques-Francois Thisse, "On Hotelling's 'Stability in Competition,'" *Econometrica* 47 (1979), 1145-1150.

Anselin, L., "Model validation in spatial econometrics; a review and evaluation of alternative approaches," *International Regional Science Review* 11 (1988), 279-316.

Anselin, L. and Hudak, S, "Spatial Econometrics in Practice: A Review of the Software Options," *Regional Science and Urban Economics* 22 (1992), 509-536.

Barron, John M., Beck A. Taylor and John R. Umbeck, "A Theory of Quality-Related Differences in Retail Margins: Why There is a 'Premium' on Premium Gasoline," *Economic Inquiry* 38:4 (2000a).

Barron, John M., Beck A. Taylor and John R. Umbeck, "Search Costs, Prices and Price Dispersion: A Theoretical and Empirical Investigation," working paper (2000b).

Borenstein, Severin and Janet Netz, "Why do all flights leave at 8 AM?: Competition and departure time differentiation in airline markets," *International Journal of Industrial Organization* 17:5 (1999), 611-640.

Douglas, G.W. and J.C Miller III, "Economic regulation of domestic air transport: theory and policy," Washington D.C.: Brooking Institute, 1974.

Easterly W. and Levine, R., "Troubles with the Neighbors: Africa's Problem, Africa's Opportunity," *Journal of African Economices* 7 (1998), 120-142.

Eaton, Curtis B., "Spatial competition revisited, *Canadian Journal of Economics* 5 (1972), 268-278.

Eaton, Curtis B. and Richard G. Lipsey, "The non-uniqueness of equilibrium in the Loschian location model," *American Economic Review* 66 (1976), 77-93.

Eaton, Curtis B. and Richard G. Lipsey, "The Principle of Minimum Differentiation Reconsidered: Some New Developments in the Theory of Spatial Competition," *Review of Economic Studies* 42 (1975), 27-49.

Gabszewicz, Jean Jaskold and Jacques-Francois Thisse, "Spatial Competition and Location of Firms," *Regional and Urban Economics* 26 (1996), 39-109.

Hotelling, Harold, "Stability in Competition," *The Economics Journal* 39 (1929), 41-57.

LeSage, James P., *Applied Econometric using MATLAB,* University of Toledo, October 1999.

LeSage, James P., *The Theory and Practice of Spatial Econometrics,* University of Toledo, February 1999.

McCoskey, Suzanne K., "Spatial Spillovers in the Adult Incidence of HIV-AIDS in Sub-Saharan Africa," Working Paper (2004), United States Naval Academy.

Netz, Janet S. and Beck A. Taylor, "Maximum or Minimum Differentiation? Location Patterns of Retail Outlets," *The Review of Economics and Statistics* 84 (2002), 162-175.

Prescott, E.C. and M. Visscher, "Sequential location among firms with foresight," *Bell Journal of Economics* 8 (1977), 378-393.

Shaked, Avner and John Sutton, "Relaxing Price Competition Through Product Differentiation," *Review of Economics Studies* 49 (1982), 3-13.

Shepard, Andrea, "Price Discrimination and Retail Configuration," *Journal of Political Economy* 99 (1991), 30-53.

Slade, Margaret, "Vancouver's Gasoline-Price Wars: An Empirical Exercise in Uncovering Supergame Strategies," *Review of Economic Studies* 59 (1992), 257-276.

Teitz, M., "Locational Strategies for Competitive Systems," *Journal of Regional Science* 8, (1968), 135-148.

IX. Appendices

Appendix A: Neighborhood Matrix Rules

Rules are in order of precedence; i.e. (1B) cannot supersede the relationship already established in (1A).

1st Degree Neighbors

1A. Stations on opposite sides of divided highways are not 1st degree neighbors.
1B. Minimum of 0 stations → station could be spatial monopolist.
1C. All stations on the same intersection are 1st degree competitors.
1D. Maximum of three closest stations to center station can be 1st degree neighbors.
1E. Stations within distance of:
  i. .5 miles of center station if on same road.
  ii. circle with radius of .375 miles.

2nd Degree Neighbors

2A. Stations within 1st Degree Rules but not included as 1st degree neighbors are 2nd degree neighbors.
2B. Maximum number of four closest stations to center station can be 2nd degree neighbors.
2C. Stations within distance of:
  i. 1 mile of center station if on same road.
  ii. circle with radius .75 miles.
2D. If a *station i* is 2nd degree neighbor with *station j*, then all of *station j's* 1st degree competitors are 2nd degree neighbors with *station i*.
2E. Stations on the opposite side of divided highways within 1 mile are 2nd degree neighbors.

3rd Degree Neighbors

3A. Stations within 2nd Degree Rules but not included as 2nd degree neighbors are 3rd degree neighbors.
3B. Maximum number of seven closest stations to center station can be 3rd degree neighbors.
3C. Stations within distance of:
  i. 1.5 miles of center station if on same road.
  ii. circle with radius of 1.125 miles.
3D. If a *station i* is 3rd degree neighbor with *station j*, then all *station j's* 1st degree neighbors are 3rd degree competitors with *station i*.
3E. If a *station i* is 2nd degree neighbor with *station k*, then all *station k's* 2nd degree neighbors are 3rd degree neighbors with *station i*.

Appendix B: Quality Variable Data Collection Sheet

Station Number ____________________
Phone Number ____________________
Pay at the Pump? ____________________
Car Wash? ____________________
Service Station? ____________________
Conv Store? ____________________
Full Service? ____________________
Number of Pumps? ____________________

Station Number ____________________
Phone Number ____________________
Pay at the Pump? ____________________
Car Wash? ____________________
Service Station? ____________________
Conv Store? ____________________
Full Service? ____________________
Number of Pumps? ____________________

Station Number ____________________
Phone Number ____________________
Pay at the Pump? ____________________
Car Wash? ____________________
Service Station? ____________________
Conv Store? ____________________
Full Service? ____________________
Number of Pumps? ____________________

Station Number ____________________
Phone Number ____________________
Pay at the Pump? ____________________
Car Wash? ____________________
Service Station? ____________________
Conv Store? ____________________
Full Service? ____________________
Number of Pumps? ____________________

Station Number ____________________
Phone Number ____________________
Pay at the Pump? ____________________
Car Wash? ____________________
Service Station? ____________________
Conv Store? ____________________
Full Service? ____________________
Number of Pumps? ____________________

Station Number ____________________
Phone Number ____________________
Pay at the Pump? ____________________
Car Wash? ____________________
Service Station? ____________________
Conv Store? ____________________
Full Service? ____________________
Number of Pumps? ____________________

Station Number ____________________
Phone Number ____________________
Pay at the Pump? ____________________
Car Wash? ____________________
Service Station? ____________________
Conv Store? ____________________
Full Service? ____________________
Number of Pumps? ____________________

Station Number ____________________
Phone Number ____________________
Pay at the Pump? ____________________
Car Wash? ____________________
Service Station? ____________________
Conv Store? ____________________
Full Service? ____________________
Number of Pumps? ____________________

Station Number ____________________
Phone Number ____________________
Pay at the Pump? ____________________
Car Wash? ____________________
Service Station? ____________________
Conv Store? ____________________
Full Service? ____________________
Number of Pumps? ____________________

Station Number ____________________
Phone Number ____________________
Pay at the Pump? ____________________
Car Wash? ____________________
Service Station? ____________________
Conv Store? ____________________
Full Service? ____________________
Number of Pumps? ____________________

Appendix C: MATLAB Functions and Examples

C.1 OLS Function

```
function results = OLS(y,X)
% PURPOSE: returns OLS estimators and test statistics given
%   independent X-regressors and dependent y variables
% ---------------------------------------------------------------
% USAGE: result = OLS(y,X);
% where: y = dependent variable vector (n x 1)
%        X = explanatory variable matrix (n x k)
% ---------------------------------------------------------------
% RETURNS: (k x 1) vector of the coefficient estimators that
%          sum of the squared errors
% ---------------------------------------------------------------
%  REFERENCES:
%   Gujarati, Damodor N.  "Basic Econometrics: Four Edition."
%       New York: McGraw-Hill Higher Education, 2003.
% ---------------------------------------------------------------
%  WRITTEN BY:
%   J. Light, Trident Scholar, US Naval Academy, 2004.

[n,m] = size(y);
bhat = X\y;
[k,l] = size(bhat);
ybar = mean(y);
rsq = ((bhat'*X'*y)-(n*ybar^2))/(y'*y-(n*ybar^2));
r2adj = 1-(1-rsq)*((n-1)/(n-k));
sig2 = ((y'*y)-bhat'*X'*y)/(n-k);
varco = sig2*inv((X'*X));
var = sqrt(diag(varco));
tstat = bhat./var;
pvalue = 2*(1 - tcdf(abs(tstat),n-k));

results.bhat = bhat;
results.rsq = rsq;
results.r2adj = r2adj;
results.tstat = tstat;
results.pvalue = pvalue;
```

C.2 OLS Example

```
% ------ Regression using OLS Function
load y.data;
load X.data;
results = OLS(y,X);
```

C.3 LM Error Function

```
function results = LMerror (y,X,W)
% PURPOSE: Calculates the test statistic for the LM Error Test
% ----------------------------------------------------------------
% USAGE: results = LMError (y,X,W)
% where:  y    = dependent variable vector (n x 1)
%         X    = explanatory variables matrix (n x k)
%         W    = spatial weight matrix
% ----------------------------------------------------------------
% RETURNS: a scalar equal to the test statistic for the LM Error test
% ----------------------------------------------------------------
% REFERENCES:
% Gujarati, Damodor N.  "Basic Econometrics: Four Edition."
%      New York: McGraw-Hill Higher Education, 2003.
% Anselin, L. and Hudak, S, "Spatial Econometrics in Practice: A
%      Review of the Software Options," Regional Science and Urban
%      Economics 22 (1992), 509-536.
% ----------------------------------------------------------------
%  WRITTEN BY:
%   J. Light, Trident Scholar, US Naval Academy, 2004.

% LM Error Test for Spatial Autocorrelation
% ehat = L*W*e + u
%   ehat = estimated error vector (n x 1)
%   L = spatial autoregressive coefficient scalar
%   W = normalized weight matrix (n x n)
%   u = homoskedastic uncorreleated error vector (n x 1)
% Hypothesis Test:
%   Ho: L = 0
%   Ha: L <> 0
% Test Statistic:
%   LMerr = {e'W*e/sig2}^2/trace[W'W + W^2]
% Rejection Criteria (5%):
%   Reject Ho if LMerr > 3.84146
%   The p-value for the test statistic is calculated in a chi-squared
%   distribution with 1 degree of freedom.

ehat = y - X*bhat;
LMerr = (e'*W*e/sig2)^2/(trace(W'*W + W^2));
pLMerr = 1 - chi2cdf(LMerr,1);

results.LMerr = LMerr;
results.pLMerr = pLMerr;
```

C.4 LM Error Example

```
% ------ Example using LM Error Function
load y.data;
load X.data;
load W.data;
results = LMerror(y,X,W);
```

C.5 LM Lag Function

```
function results = LMlag(y,X,W)
% PURPOSE: Calculates the test statistic for the LM Lag Test
% ---------------------------------------------------------------
% USAGE: results = LMlag (y,X,W)
% where:  y    = dependent variable vector (n x 1)
%         X    = explanatory variables matrix (n x k)
%         W    = spatial weight matrix
% ---------------------------------------------------------------
% RETURNS: a scalar equal to the test statistic for the LM Lag test
% ---------------------------------------------------------------
% REFERENCES:
% Gujarati, Damodor N.  "Basic Econometrics: Four Edition."
%       New York: McGraw-Hill Higher Education, 2003.
% Anselin, L. and Hudak, S, "Spatial Econometrics in Practice: A
%       Review of the Software Options," Regional Science and Urban
%       Economics 22 (1992), 509-536.
% ---------------------------------------------------------------
%  WRITTEN BY:
%   J. Light, Trident Scholar, US Naval Academy, 2004.

[n,m] = size(y);
bhat = X\y;
[k,l] = size(bhat);
sig2 = ((y'*y)-bhat'*X'*y)/(n-k);
ehat = y - X*bhat;

% LM Lag Test for Spatial Autocorrelation
% y = p*W*y + X*betas + e
%   p is the autoregressive coefficient
%   Wy is the spatially lagged dependent variable
% Hypothesis Test
%   Ho: p = 0
%   Ha: p <> 0
% Test Statistic
%   LMlag = ((e'*W*y/sig2)^2)/((W*X*betas)'*M*W*X*betas/sig2 + trace(W'*W +
%   W^2)) where M = eye(n) - X*inv(X'*X)*X'
% Rejection Criteria (5%):
%   Reject Ho if LMlag > 3.84146
```

```
% The p-value for the test statistic is calculated in a Chi-squared
% distribution with 1 dof.

M = eye(n) - X*inv(X'*X)*X';
LMlag = ((e'*W*y/sig2)^2)/((W*X*bhat)'*M*W*X*bhat/sig2 + trace(W'*W + W^2));
pLMlag = 1 - chi2cdf(LMlag,1);

results.LMlag = LMlag;
results.pLMlag = pLMlag;
```

C.6 LM Lag Example

```
% ------ Example using LM Lag Function
load y.data;
load X.data;
load W.data;
results = LMlag(y,X,W);
```