

Spatial Competition in the Retail Gasoline Market: An Equilibrium Approach Using SAR Models*

Sang-Yeob Lee[†]

Department of Economics, The Ohio State University, 1945 N. High Street, Columbus, OH 43210

Job Market Paper 1

August 2007

Abstract

This paper investigates the nature of competition in the retail gasoline market using a two year panel data of weekly prices for gas stations in San Diego county. The primary dimension of product differentiation in the retail gasoline market is spatial in the sense that a gas station's market power depends on the locations of all other gas stations. In contrast to previous empirical studies, I explicitly model the fact that retail gasoline prices of all gas stations are simultaneously determined in a spatially competitive system. I use IV methods to estimate several spatial autoregressive (SAR) models of stations' price reaction functions after specifying spatial weights based on distance between stations. My results are consistent with the spatial competition model. I also find that retail prices are heavily influenced by station's characteristics such as brand name and amenities. By using the SAR model I am able to identify that the brand of competing stations and their relative geographic proximity to each other are important factors in explaining price variation across gasoline stations, as opposed to just the number of competing stations. I find that gas stations most intensely compete with stations less than 1 mile away and that the intensity of competition diminishes with distance.

Keywords: Retail gasoline, spatial competition, spatial autoregressive model

JEL classification: L95

*I would like to thank Matthew Lewis, Lung-fei Lee and Bruce Weinberg as well as Stephen Cosslett and Hajime Miyazaki for their valuable comments. I would also like to thank participants in the Micro Lunch and Econometrics Seminars at the Ohio State University, and at the meeting of the Midwest Economics Association.

[†]*E-mail address:* lee.1621@osu.edu.

1 Introduction

Although regular retail gasoline is physically a nearly homogenous good, gas stations differ in terms of geographical location and other station attributes. I use a theoretical model of spatial competition following Pinske et al (2002) that captures the main features of the retail gasoline market. In the model, the primary dimension of product differentiation is geographic. Travel costs or search costs lead consumers to consider nearby gas stations as close substitutes after controlling for station specific characteristics such as brand and service quality. Since spatially differentiated gas stations compete with their neighboring stations in price, the equilibrium prices of all stations are simultaneously determined in a spatially competitive system. Thus, I structurally estimate several spatial autoregressive (SAR) models of stations's price reaction function in order to more accurately study competition and pricing behavior.

Recently, researchers have studied competition and pricing behavior in the retail gasoline market (for example, Barron et al (2004), Borenstein and Shepard (1996), Eckert and West (2004), Hastings (2004), Hosken et al (2006), and others). The previous empirical analysis typically ignores the spatial effect (or spatial interaction) that could result from the spatial differentiation of the retail gas station. While some researchers consider spatial differentiation as a source of price dispersion or dynamic price patterns, they ignore the essential aspects of spatial competition in the retail gasoline market. The estimation is typically performed assuming that the price of the retail gasoline at a particular location is determined independently of competing stations' pricing behaviors. If a gas station's price is spatially correlated with competitors' prices, and hence competitors' competitors' prices and so on, an omitted variable bias problem arises.¹

For the study of competition in the retail gasoline market, properly defining the extent of the market is very important.² If the market definition is not well-specified, the measure of the effect of an additional competitor will be biased. However, the previous empirical studies typically assume the relevant market for a station as all neighboring stations within 1 or 1.5 miles in the regression analysis, and implicitly assume that all relevant markets are independent.³ Though it

¹See Lee (2007).

²Properly defining the extent of the market is very important for a merger analysis.

³See Hastings (2004), Barron et al (2004) and Hosken et al (2006).

is true that a given gas station competes directly with its close neighbors, those neighbors compete with their respective neighbors, and so, the original gas station is linked to all other stations in a geographical space. Thus, the nature of competition in the retail gasoline market is not local, but global. Ignoring the fact that competition is global will yield a bias similar to the bias generated by ignoring the spatial effect altogether.

This paper uses a census of all gas stations in the San Diego area observed in 1998, a panel data of weekly price of unleaded regular retail gasoline observed from the first week in 2000 to the last week in 2001, and market-specific information at census tract level from Census 2000 to examine how each gas station interacts with other stations. Based on a model of spatial competition, I derive a SAR model of gas stations' price reaction function and identify the proper local market for the retail gasoline market by estimating station and week fixed effects models of the price reaction function with different local market definitions. I then estimate the price reaction function with an empirically determined market definition using instrumental variables to examine the determinants for the retail gasoline price. I then show the presence of strategic interaction among spatially differentiated gas stations.

Estimating the SAR model of the price reaction functions requires one to specify a spatial weight matrix which relies on the relevant local market definition for the retail gasoline and the intensity of competition between gasoline stations. I use three different relevant market definitions for the retail gasoline and assume that the intensity of competition between gas stations depends on their relative distance. Specifically, I define the relevant market for a station as all competing stations within 0.1 mile, 0.5 mile or 1.0 mile.⁴

Using these local market definitions, I estimate SAR models of the price reaction function after specifying spatial weight matrices based on distance between stations. By taking advantage of panel data, I estimate station and week fixed effects models of the price reaction function to identify the proper local market definition for the retail gasoline. Station fixed effects control for both observed and unobserved station heterogeneity such as brand name, service quality, costs, or demand conditions. Week fixed effects control for regional wholesale price movements and

⁴Hasting (2004) assumes that stations compete with other stations within one mile. While she collects information about the relevant market definition from various dealers and trading representatives in the San Diego area, she does not empirically test her market definition.

seasonal demand movements which cause changes in average price level in San Diego County over time. Since most station characteristics does not change over time, the parameter of interest can be estimated consistently. Thus, the station and week fixed effects models allow me to identify the proper market definition for the retail gasoline.

For the local market definition for the retail gasoline, I find little evidence that stations are competing with only the closest station, which is an important finding in light of the extensive literature on the retail gasoline market that has assumed that pricing behavior of a gas station is mostly influenced by the nearby gas station.⁵ Specifically, I find that when a station raises its price, about 60% of lost sales go to stations within 0.1 mile radius. When that radius is extended to 0.5 miles, I find that 85% of lost sales are accounted for. Finally, I find that the full 100% of sales lost by the original station are captured by stations within 1 mile. Taken together, I find strong and consistent evidence that stations heavily compete with stations less than 1 mile away, and that the intensity of competition decreases with distance. This has important policy implications for other research on spatial competition in the retail gasoline market, particularly because the estimates are very sensitive in the choice of the relevant market definition.

Using an empirically determined local market definition, I examine the determinants for the retail gasoline price in a price competition setting among spatially differentiated gasoline stations. The results show that after controlling for the spatial effect, brand name becomes the most important determinant of retail prices. Interestingly, the number of nearby stations does not explain price variation across stations once the spatial effect is controlled for. These empirical findings imply that the brand name of competing stations and even the composition of those particular brands over the geographical space are more important determinants of retail prices than the number of nearby stations. For example, I find that gasoline stations have a relatively lower price when they face Arco or unbranded (independent) stations as competitors than when they face Chevron, Mobile, Shell, Texaco, Union 76 or 7-Eleven stations as competitors. This also has important implications for policy evaluation, such as a merger analysis in the sense that a particular gas station's price is differently influenced by competing stations' brand name and their spatial

⁵For example, Barron et al. (2004) and Hosken et al (2006). They use distance to the closest gas station and the number of competitors as measures of competition to explain price variation across stations.

locations.

The remainder of this paper is organized as follows. In section 2, I introduce a theoretical model of spatial price competition and derive a SAR model of stations' price reaction functions. Section 3 describes the data set used. Then, section 4 reviews estimation methods of the standard SAR model, describes an empirical model of stations' price reaction function and specifies spatial weight matrices using different local market definitions. In section 5, I discuss the results for the local market definition and show the presence of strategic interaction among spatially differentiated gas stations. Concluding remarks follow in section 6.

2 The Theoretical Model of Spatial Competition

In this section, I describe the theoretical model of spatial competition. I follow Pinske et al (2004)'s framework to formulate a static model of spatial price competition in the retail gasoline market. In the model, the primary dimension of product differentiation is geographic. Transportation costs or search costs lead consumers to consider nearby gas stations as close substitutes after controlling for station specific characteristics such as brand and service quality. Since spatially differentiated gas stations compete with neighboring stations in price, the equilibrium prices of all stations are simultaneously determined in a competitive system.

2.1 Demand

Suppose there are n spatially differentiated gas stations. No two gas stations are identical since their market power depends on the locations of all other gas stations. Let \tilde{p}_i be the nominal price of the i th gas station, $i = 1, \dots, n$. There is also an outside good that is sold at a nominal price p_o .

The consumer c purchases a vector $q_k = (q_{1c}, \dots, q_{nc})'$ of the spatially differentiated gasoline, with $q_{ic} \geq 0$, $i = 1, \dots, n$, and q_{oc} of the outside good, with $q_{oc} > 0$, $c = 1, \dots, h$. Each consumer is located at a point in a geographical space and taking into account transportation costs or search costs, he consumes a location-based optimal amount of gasoline from a gas station.

Let the indirect utility function of consumer c be denoted by $\tilde{V}_c(p_0, \tilde{P}_n, \tilde{y}_c)$, where \tilde{y}_c is consumer c 's nominal income, and $\tilde{P}_n = (\tilde{p}_1, \dots, \tilde{p}_n)'$. To derive a linear form of the demand system, a normalized quadratic indirect utility function is specified as follows:

$$V_c(1, P_n, y_c) = -[\alpha_{oc} + \alpha'_c P_n - p_o y_c (\gamma_o + \gamma' P_n) + \frac{p_o}{2} P'_n S_{n,c} P_n], \quad (2.1)$$

where $P_n = p_o^{-1} \tilde{P}_n$, $y_c = p_o^{-1} \tilde{y}_c$, and $S_{n,c}$ is an arbitrary $n \times n$ symmetric negative semidefinite matrix. The diagonal elements of $S_{n,c}$ measure own-price effects while off-diagonal elements of $S_{n,c}$ capture cross price effects.

Using Roy's identity, each consumer's demand for gas station i can be derived as:

$$q_{i,c} = -\frac{\frac{\partial V_c}{\partial p_i}}{\frac{\partial V_c}{\partial y_c}} = \frac{\alpha_{ic} + \sum_{j=1}^n s_{ij,c} p_j - \gamma y_c}{p_o (\gamma_o + \gamma' P_n)}. \quad (2.2)$$

For simplicity, the price index $p_o (\gamma_o + \gamma' P_n)$ is assumed to be one because it can be treated as a constant. The aggregate demand for gas station i can be derived as:

$$q_i = \alpha_i + s_{ii} p_i + \sum_{j \neq i}^n s_{ij} p_j - \gamma_i y, \quad (2.3)$$

where $\alpha_i = \sum_{c=1}^h \alpha_{ic}$, $s_{ij,c} = \sum_{c=1}^h s_{ij}$, and $y = \sum_{c=1}^h y_c$ is aggregate income. Note that this aggregate demand function can be valid only over a limited quantity, but these restrictive assumptions will be satisfied at all relevant equilibria. The required conditions are $\alpha_i > 0$, $s_{ii} < 0$, and $-s_{ii} > s_{ij}$ for all $i \neq j$ to ensure concavity of the utility function. If gas stations are not identical, a higher α_i reflects an absolute advantage in demand experienced by gas station i . More specifically, a gas station with high brand loyalty and better service quality will have a higher α . While s_{ij} determines substitutability between gas station i and j , γ_i will determine the income effect for gas station i .

2.2 Supply Side: The Bertrand-Nash Game

I assume that each gas station plays a static Bertrand-Nash game in which each station chooses its own price to maximize profits given prices of other gas stations. Furthermore, I assume that gas station i faces a constant marginal cost mc_i , which is a linear function of various cost factors. (i.e., $mc_i = \sum_{l=1}^n \delta'_{l,i} mc_{l,i}$), where $mc_{l,i}$ is the marginal cost associated with cost factor l , and $l = 1, \dots, n$. The cost for retail gasoline is mainly determined by wholesale prices of gasoline. A station's brand name and its vertical contract type with the refinery are also associated with marginal cost of the retail gasoline. Therefore, gas station i 's profit function can be defined as

$$\Pi_i = (p_i - \delta'_i mc_i) q_i - F_i, \quad (2.4)$$

where F_i is the station's fixed cost.

Given competitors' prices, gas station i sets p_i to

$$\max_{p_i} (p_i - \delta'_i mc_i) (\alpha_i + s_{ii} p_i + \sum_{j \neq i}^n s_{ij} p_j - \gamma_i y) - F_i. \quad (2.5)$$

2.3 Price Reaction Function

Gas station i 's price reaction function with respect to competitors' prices can be derived by solving the first order condition of the profit maximization problem. The price reaction function is:

$$p_i = -\frac{\alpha_i}{2s_{ii}} - \frac{1}{2} \sum_{j \neq i}^n \frac{s_{ij}}{s_{ii}} p_j + \frac{\gamma_i}{2s_{ii}} y + \frac{1}{2} \delta'_i mc_i. \quad (2.6)$$

The above price reaction function includes more unknown parameters than can be estimated using a cross section or a short panel of n gas stations. Therefore, I place some restrictions on the parameters in (2.6) for the estimation. I assume that $-1/2(\alpha_i/s_{ii} - \gamma_i/s_{ii}y - \delta'_i mc_i)$ is a function of station i characteristics and market characteristics. Specifically, it is assumed that $-1/2(\alpha_i/s_{ii} - \gamma_i/s_{ii}y - \delta'_i mc_i) = X_i \beta + u_i$, where X_i is a vector of observed station characteristics and market characteristics, and u_i is a random error. The random error, u_i captures the influence of unobserved station characteristics and market characteristics. For example, X_i includes the

station's brand name, amenities and local market demographics such as median income, median rent price, and others. Then, the price reaction function (2.6) can be rewritten as:

$$p_i = X_i\beta + \frac{1}{2} \sum_{j \neq i}^n d_{ij} p_j + u_i, \quad (2.7)$$

where $d_{ij} = -\frac{s_{ij}}{s_{ii}}$ for all $i \neq j$.

The slope of the gas station i 's price reaction function with respect to gas station j multiplied by two, d_{ij} is referred to as the "diversion ratio."⁶ This is the fraction of sales lost by gas station i to gas station j when gas station j raises its price. Since the product differentiation in the retail gasoline market comes out from geographic location after controlling for station heterogeneity, it is natural to assume that d_{ij} is a function of distance between gas station i and j . If station i does not directly compete with station j , then $d_{ij} = 0$, while if there are only two stations i and j , on the same location, $d_{ij} = 1$.

The price reaction for gas station i can be rewritten as follows:

$$p_i = X_i\beta + \lambda \sum_{j \neq i}^n w_{ij} p_j + u_i, \quad (2.8)$$

where $w_{ij} = d_{ij} / \sum_{j \neq i}^n d_{ij}$, and $\lambda = 1/2 \sum_{j \neq i}^n d_{ij}$. The w_{ij} 's become the relative diversion ratio and $\sum_{j \neq i} w_{ij} p_j$ becomes a weighted average price of i 's competing stations. The coefficient of competitors' weighted average price times two measures the fraction of sales lost by station i captured by its competing stations when it raises price. Note that if the sum of diversion ratios for station i is one, when station i raises price, all sales lost by station i are captured by its competing stations and λ becomes one half.⁷

Let X_n be a $n \times k$ matrix of observed station-specific characteristics and market characteristics and u_n is a vector of random errors. Then, the system of price reaction functions (2.7) can be

⁶See Shapiro (1996).

⁷This might not be the case when consumers buy outside goods instead of retail gasoline when a station raises its price. The outside goods can be public transportation, car pooling, bicycling, etc.

written as

$$p_n = \lambda W_n p_n + X_n \beta + u_n, \quad (2.9)$$

where p_n is a vector of gas stations' prices, diagonal elements of spatial weight matrix W_n are zero, and off-diagonal elements of W_n are relative diversion ratios. The system of price reaction functions in (2.8) is well known as a spatial autoregressive (SAR) model. The $W_n p_n$ can be considered as weighted average prices of competing stations. The scalar parameter λ captures the average influence of competing stations' prices on p_n . It is typically referred to as the spatial effect parameter in the literature.

Recall that this study aims to estimate the system of price reaction functions, not the system of equilibrium price functions. Nevertheless, the equilibrium price functions can be expressed with the following reduced form equation:

$$p_n = (I_n - \lambda W_n)^{-1} X_n \beta + (I_n - \lambda W_n)^{-1} u_n \quad (2.10)$$

$$= X_n \beta + \lambda W_n X_n \beta + \lambda^2 W_n^2 X_n \beta + \dots + u_n + \lambda W_n u_n + \lambda^2 W_n^2 u_n \dots \quad (2.11)$$

As shown in the above equation, the price of a specific gas station in a spatially competitive system is simultaneously determined by its own characteristics and perhaps all other gas stations' characteristics, as well as its unobservable characteristics and all other gas stations' unobservable characteristics. Under the assumption that u_n is i.i.d with mean zero and variance σ^2 , the variance-covariance matrix of p_n is

$$\sigma^2 (I_n - \lambda W_n')^{-1} (I_n - \lambda W_n)^{-1}. \quad (2.12)$$

Thus, conditional variances of prices are heterogenous and spatially correlated. An important implication of this model is that price dispersion can be observed even after controlling for stations' heterogeneity. This is not the case for all spatial competition models.

For comparison, consider the spatial error model as follows:

$$p_n = X_n\beta + u_n, \quad u_n = \rho W_n u_n + v_n, \quad (2.13)$$

where v_n is i.i.d with mean zero and variance σ^2 . This model can be applied when unobserved components are spatially correlated. Previous empirical studies often use this model to explain price variation across stations. In this model, a particular station's price can not be affected by a change in a competing station's characteristics such as service quality. This model ignores essential aspects of spatial competition in the retail gasoline market because it does not account for the fact that the price of a particular station is dependent on competing stations' pricing behavior. Therefore, a traditional pricing equation model like (2.13) may lead to a biased estimate of the total effect of a characteristic change in a station because it ignores any spatial competition effects. For example, if one estimates the pricing equation to measure the effects of mergers on the market price of retail gasoline, the estimated effect would be inconsistent.

2.4 Interpretation of the SAR Model

Let $G_n(\lambda)$ denote $(I_n - \lambda W_n)^{-1}$ and $g_{ij}(\lambda)$ denote an element of $G_n(\lambda)$.⁸ Then, the reduced form equation from (2.11) can be expressed as

$$p_n = X_n\beta + \lambda W_n X_n\beta + \lambda^2 W_n^2 G_n(\lambda) X_n\beta + G_n(\lambda) u_n. \quad (2.14)$$

The Jacobian matrix of p_n with respect to x_k can be written as

$$\frac{\partial p_n}{\partial x'_k} = \beta_k I_n + \beta_k (\lambda W_n) + \beta_k (\lambda^2 W_n^2) G_n(\lambda). \quad (2.15)$$

The coefficient β_k captures the effects on price of gas station i of a marginal change in station i 's k -th characteristics. The first term in (2.15) ($\beta_k I_n$) is a matrix of the direct effects from gas station' own characteristics and the second term ($\beta_k \lambda W_n$) is a matrix of the indirect effects from competing stations' characteristics. The last terms ($\beta_k \lambda^2 W_n^2 G_n(\lambda)$) can be referred to as induced effects in

⁸Note that $G_n(\lambda) = \sum_{i=0}^{\infty} \lambda^i W_n^i$ and $\sum_{j=1}^n g_{ij}(\lambda) = 1/(1 - \lambda)$, when $\sum_{j=1}^n w_{ij} = 1$ for all i and $|\lambda| < 1$.

the sense that spatial spill-overs are induced by the direct and indirect changes in the first and second terms. This implies that even though gas stations are not directly competing with all other gas stations, all gas stations are linked to other gas stations located over a particular geographical space. Therefore, a gas station's price is influenced by all other gas stations' characteristics. In other words, the spatial competition in this model is global in nature.

From (2.14), the equilibrium price for station i can be written as follows:

$$p_i = \sum_{l=i}^k \beta_l x_{l,i} + \sum_{l=1}^k \sum_{j \neq i}^n g_{ij}(\lambda) \beta_l x_{l,j} + \sum_{j=1}^n g_{ij}(\lambda) u_j. \quad (2.16)$$

The partial derivative of p_i with respect to $x_{k,j}$ is given by $g_{ij}(\lambda) \beta_k$. Then, the total effect on station i 's price of a simultaneous exogenous shock of x_k in all gas stations is:

$$\sum_{j=1}^n \frac{\partial p_i}{\partial x_{k,j}} = \sum_{j=1}^n g_{ij}(\lambda) \beta_k = \frac{1}{1-\lambda} \beta_k. \quad (2.17)$$

This is not station specific. Lee (2007) shows that if a positive spatial effect is ignored as in (2.13), the estimated coefficient of x_k is less than the total effect, $1/(1-\lambda) \beta_k$ and so, the total effect cannot be estimated from a traditional pricing equation.

The total effect on all gas stations' prices of an exogenous shock to station j 's characteristic k is

$$\sum_{i=1}^n \frac{\partial p_i}{\partial x_{k,j}} = \sum_{i=1}^n g_{ij}(\lambda) \beta_k. \quad (2.18)$$

This effect is typically station specific and heterogenous.⁹ This heterogeneity of the total effect can be only investigated systematically using the structural model because it depends on where the exogenous shock occurs. For example, the impact of a change in the brand name of a gas station on the prices of other gas stations should be examined in the system because the total effect of the brand-name change can not be exploited from a traditional pricing equation that ignores the spatial competition effect.

⁹This is because $\sum_{i=1}^n g_{ij}(\lambda) \neq \sum_{i=1}^n g_{ik}(\lambda)$ for $j \neq k$ in general.

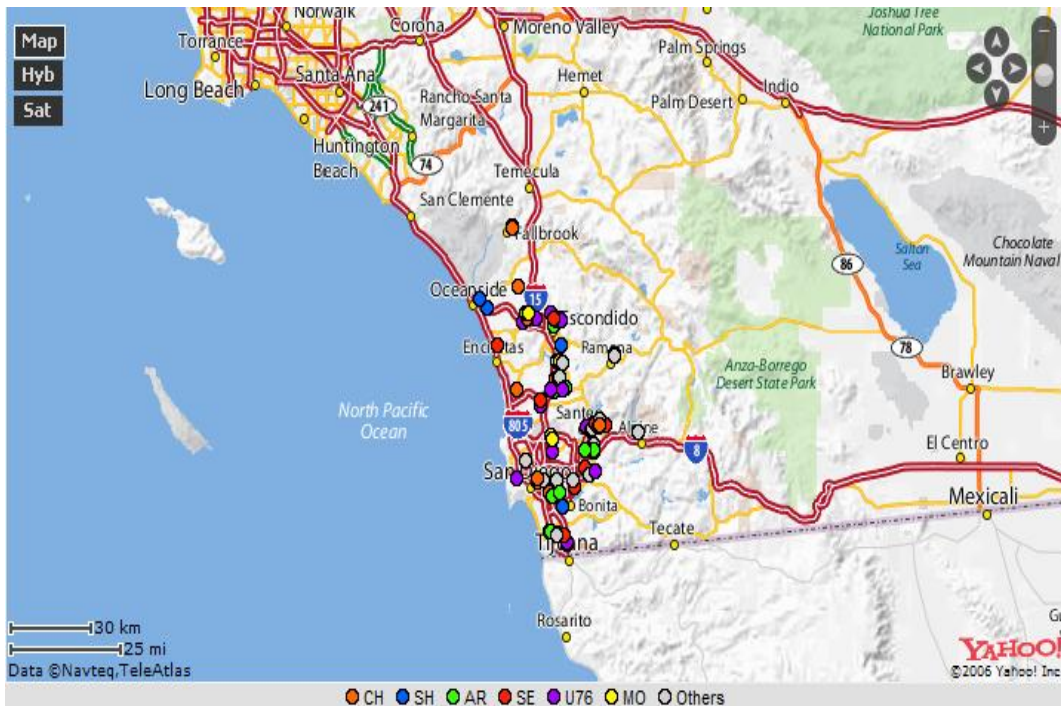


Figure 1: Map of San Diego County Area

3 Data

As shown in Figure 1, the geography of San Diego County makes it an ideal study area for this analysis. The north side of San Diego County is a large military base, the east side is a desert, the south side is the border with Mexico and the west side is the Pacific Ocean. The region has a nearly complete natural geographical boundary, which helps to avoid potential spill-over effects.

The first data set used in this study is a census of nearly all gas stations (651) in the San Diego County area which was collected by Whitey Leigh Corporation in 1998. Although there are 661 stations in the San Diego area, observations were dropped if there were no competitors within 1.5 miles. This census includes location, brand name, car wash, garage, convenience store, number of pumps and other various characteristics of each gas station.

The census also contains the type of operation and ownership for each station. There are two major types of retail gasoline station. One is branded gasoline and the other is unbranded. If a retail station is branded, the station has one of three vertical relationships with the branded refinery. The first type of branded station is a company operated station which is owned and

operated by the refiner. The second type of branded station is called a lessee-dealer. In this case, the station is owned by the refiner and is leased to a residual claimant. The lessee sets the retail price and is under contract to purchase gasoline from the refiner directly at the wholesale price. The last type of branded station is a dealer-owned station. In this case, a dealer owns the station properties and has a contract with the refiner in which the station must buy only branded gasoline from the refiner and can display the brand logo.

While branded stations are influenced by the refiner directly or indirectly when they set the retail price, unbranded stations can shop for the lowest wholesale price from any refiner at any place and set the retail price independently. I use geographic coordinates from TeleAtlas geocoding software to compute distances between gas station pairs. Then, the number of competing stations within a mile is constructed to capture variations in local market structure. A map of the San Diego City area showing the location of 100 of the gas stations used in this study is provided in Figure 2. For obvious reasons of clarity, I choose to show only 100 stations instead of the full sample of 651.

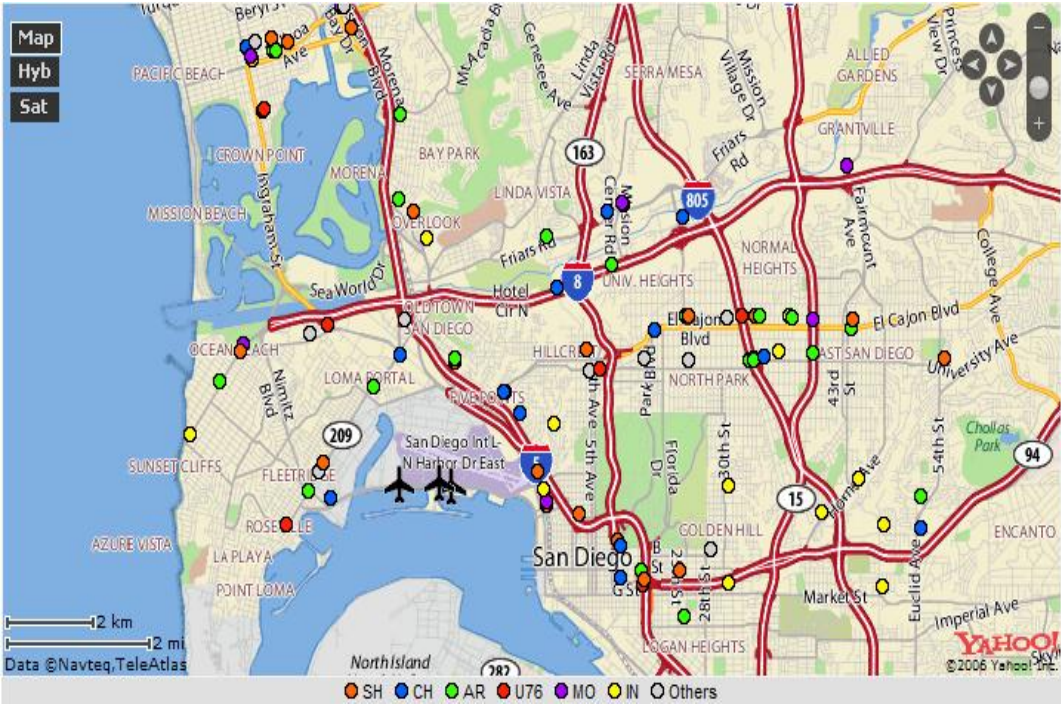


Figure 2: Map of San Diego City and Gas Stations

Table 1: Summary Statistics and Variable Definitions

Variable	Definition	Mean	S.D
<i>Brand</i>			
ARCO	Station Brand: ARCO	0.189	(0.392)
Chevron	Station Brand: Chevron	0.127	(0.334)
Exxon	Station Brand: Exxon	0.035	(0.185)
Mobile	Station Brand: Mobile	0.089	(0.285)
Shell	Station Brand: Shell	0.114	(0.318)
Texaco	Station Brand: Texaco	0.101	(0.302)
7-11	Station Brand: 7-11	0.074	(0.262)
76	Station Brand: Union 76	0.115	(0.320)
Unbranded	Independent Station	0.155	(0.362)
<i>Owner structure</i>			
Company	Company operated	0.280	(0.449)
Lessee	Lessee dealer	0.469	(0.499)
Dealer	Dealer owned company supplied	0.111	(0.314)
<i>Station Characteristics</i>			
Pumps	Number of Pumps	22.194	(10.105)
Full service	Station has Full Service	0.100	(0.300)
C-store	Station has Convenience Store	0.519	(0.500)
Car Wash	Station has Car Wash	0.144	(0.352)
Garage	Station has Auto Service	0.313	(0.464)
<i>Market Characteristics (2000 Census Tract)</i>			
Vehicle	Total Vehicles (100s)	29.664	(12.981)
Travel Times	Average travel minutes of commuters	26.602	(4.102)
Median Rent	Median Rent (\$100s)	7.600	(2.205)
Resident Density	Housing Units per residential acres	12.196	(14.354)
Median Income	Median Household Income(\$1000)	44.858	(17.193)
# of competitors	Number of competitors within 1.0 mile radius	5.868	(3.874)
Price	Regular unleaded gasoline price per gallon	168.345	(18.140)

¹⁾ Number of gas stations observed=651.

²⁾ The price is observed for 95 weeks.

The second data set contains retail prices for unleaded regular grade (87 octane). The data were collected weekly by the Utility Consumer Action Network (UCAN), a consumer advocacy group in San Diego, California. The prices are for each Monday during the years 2000 and 2001, and are measured in cents per gallon. Unfortunately, many of the stations in the sample are missing price observations for several weeks during the sample period. During the 95 weeks used in the study, retail prices of over 300 stations are recorded for every week.

The third data set includes local demand and cost variables at the census tract level from the 2000 US census. Total Vehicles (Vehicle), average travel minutes of commuters (Travel Times), median household income (Median Income) and housing units per residential acre (Resident Density) are local demand variables. Median Rent is used as a measure of local property cost. Table 1 provides summary statistics and data description for the sample used in this analysis.

4 SAR Model Estimation of the Price Reaction Function

4.1 Estimating SAR Models

In this subsection, I briefly review estimation methods for the standard SAR model with emphasis on its properties for cross section data. A standard spatial autoregressive model is specified as

$$p_n = \lambda W_n p_n + X_n \beta + u_n, \quad (4.1)$$

where p_n is a $n \times 1$ vector of dependent variables, X_n is a $n \times k$ matrix of exogenous variables, u_n is a n -vector of i.i.d disturbances with zero mean and variance σ^2 , W_n is a row normalized $n \times n$ spatial weight matrix. Multiplying both sides of the reduced form equation (2.10) by W_n , it can be shown that

$$W_n p_n = W_n (I_n - \lambda W_n)^{-1} X_n \beta + W_n (I_n - \lambda W_n)^{-1} u_n \quad (4.2)$$

and $W_n p_n$ may be correlated with u_n since $E((W_n p_n)' u_n) = \sigma^2 \text{tr}(W_n (I_n - \lambda W_n)^{-1}) \neq 0$ in general. Thus, the OLS estimates of equation (4.1) would be inconsistent due to endogeneity of $W_n p_n$.

To estimate the parameters in the SAR model (4.1) consistently, several methods have been proposed in the literature. Ord (1975) proposes the maximum likelihood estimation method under the normality assumption of the disturbance u_n . While the maximum likelihood estimators are consistent and asymptotically normal under regularity conditions, the MLE method is computationally burdensome for a large sample size. Kelejian and Prucha (1998) suggest the 2SLS method which uses $W_n X_n$, $W_n^2 X_n$ and X_n as IV matrices. Lee (2003) also suggests that the best 2SLS method should use $W_n(I_n - \lambda W_n)^{-1} X_n$ and X_n as IV matrices. While the 2SLS method is computationally simpler than the MLE method, the 2SLS method may yield less efficient estimators than the MLE method. Recently, Lee (2005) also proposed the best GMM that uses additional moment conditions together with those based on the weight matrix W_n to improve the efficiency of the 2SLS estimates. Note that these estimation methods are all specified for cross sectional data.

4.2 Empirical Specification

In this subsection, I describe the empirical model of stations' price reaction functions. I use panel data of station-level weekly prices of regular unleaded gasoline prices to control for the possible dynamic pattern of station pricing behaviors. The empirical specification is given by:

$$p_{it} = \lambda w p_{it} + X_i \beta + \sum_{t=1}^T \phi_t \text{Week}_t + \alpha_i + u_{it}, \quad (4.3)$$

where X_i contains station characteristics variables and market characteristics variables, and the $w p_{it}$'s are weighted average prices of competing stations, α_i is a station specific effect, and u_{it} is a random error. Week fixed effects capture changes in average price level over time that mostly result from regional wholesale price movements and seasonal demand movements.

Prices at a particular station are likely to be correlated across weeks due to unobserved station characteristic variables that do not change over time. Therefore, I consider two different methods to control for the correlation. One is a station fixed effects model and the other is a station random effects model with instrumental variables. Since the unobserved station specific effect α_i represents fixed factors that affect prices of the retail gasoline at all stations, it is likely that the

weighted average prices of competing stations are correlated with the station specific effect. Thus, I use a station fixed effects model to estimate the spatial effect parameter consistently. However, it would not allow me to separately identify the coefficients of station characteristics and market characteristics since all observed characteristics do not change over time.

To identify the coefficients of the time invariant station characteristics and market characteristics variable, I choose a random effects model after controlling for most of station heterogeneity and estimate the SAR model (4.3) using the IV method (Baltagi, 1981). I use weighted average characteristics of competing stations as instrumental variables to control for the endogeneity of weighted average prices of competing stations as suggested by Kelejian and Prucha (1998).

4.3 Local Market Definition and Spatial Weight Matrix

To estimate the SAR model of price reaction functions in (4.3), the spatial weight matrix must be specified to construct weighted average prices of competing stations. Common methods of specifying the weight matrix are to place equal weight on all competitors within a critical distance, a common boundary or to place equal weight on the k-nearest competitors. Since competition with the nearest station is more intensive even within a distinct area, it is not appropriate to assign the same weight to all competitors within a local area for this analysis. Thus, I place different weights on competing stations within a critical distance based on their relative distance. While Hastings (2004) obtains information about the intensity of local competition from conversation with various managers in retail gas stations and trading representatives, she does not empirically test her market definition that gasoline stations in Los Angeles and San Diego areas intensively compete with any station within a mile.¹⁰ Since pricing behavior of a gas station might be mostly influenced by the nearest station, I also consider two smaller competing groups.

I use three different values of the critical distance for the local market definition. The critical distances are 0.1 miles, 0.5 miles, and 1 mile. Using these critical distances, three measures of competitors' weighted average price are constructed as follows. The elements of the spatial weight

¹⁰Hastings(2004) asks various retail dealers and refiners about their competition group. The dealers claim that they compete mostly with any station within a mile. She also points out the fact that stations of the same brand are usually located more than a mile apart in Los Angeles and San Diego. She uses a one mile definition of competition group as her market definition.

matrix are defined as $w_{ij,cd} = \omega_{ij} / \sum_{ij}^N \omega_{ij}$, where ω_{ij} is $1/(1 + d_{ij})$ if the distance (d_{ij}) between j and i is within a critical distance and zero otherwise. To impute missing prices to construct weighted average prices of competing stations, I begin with a week fixed effects regression model of retail prices as follows:

$$p_{it} = X_i\beta + \sum_{j=1}^N w_{ij,cd}X_j\delta + \sum_{t=1}^T \phi_t \text{Week}_t + \alpha_i + u_{it}, \quad (4.4)$$

where X_i includes station characteristics and market characteristics as in (4.3), $\sum_{j=1}^N w_{ij,cd}X_j$ includes all the weighted average characteristics of competing stations, α_i is a station specific error, and u_{it} is a random error. I predict missing price using the estimated equation (4.4). Then, I construct the new price variable \tilde{p}_{it} as equal to price for stations which have observed price and as equal to the predicted price for stations which do not have an observed price. A station's weighted average price of competing stations for each week becomes:

$$wP_{it,cd} \equiv \sum_{j=1}^N w_{ij,cd}\tilde{p}_{jt}. \quad (4.5)$$

Using a 1 mile radius as a common boundary, an alternative measure of weighted average price is also constructed as

$$wP_{it,cb} \equiv \sum_{j=1}^J w_{ij,cb}\tilde{p}_{jt}, \quad (4.6)$$

where J is the number of competing stations with a mile radius and $w_{ij,cb}$ is $1/J$ if station j is within 1 mile of station i , and otherwise is zero.

5 Results

5.1 Local Market Definition

Before examining the determinants of retail gasoline prices, it is useful to look at station and week fixed effect models (5.1) of the price reaction function with different definitions of the local

market. Station fixed effects can control for both observed and unobserved station characteristics and market characteristics which do not change over time. Week fixed effects control for regional wholesale price movements and demand movements which cause changes in average price level in San Diego County over time. Since most of the station characteristics do not change over time, the estimates of the coefficients of this equation can be consistent and they are suggestive.

To determine a reasonable critical distance for local market definition, I estimate a station and week fixed effects model of price reaction functions with different local market definitions as follows.

$$p_{it} = \lambda wp_{it,cd} + \sum_{i=1}^N \beta_i \text{Station}_i + \sum_{t=1}^T \phi_t \text{Week}_t + u_{it}. \quad (5.1)$$

where $wp_{it,cd}$'s are weighted average prices of competing stations within a critical distance.

Table 2: Station and Week Fixed Effects Price Reaction Functions

Critical Distance	Price 0.1 mile	Price 0.5 mile	Price 1.0 mile	Price all
<i>Weighted Ave.Price</i>				
$wp_{0.1}$	0.291** (0.020)			0.007 (0.027)
$wp_{0.5}$		0.433** (0.0267)		0.004 (0.060)
$wp_{1.0}$			0.571** (0.028)	0.557** (0.053)
# of obs	17069	17069	17069	17069
R ²	0.8895	0.8921	0.8958	0.8958

1) Coefficients of week fixed effects and coefficients of station fixed effects are omitted.

2) † significant at 10%, * significant at 5%, ** significant at 1%.

3) Robust standard errors in parentheses.

4) Stations with at least one competitor within 0.1 mile are used for the estimation.

Table 2 contains the estimates of the spatial effect, λ , in equation (5.1) that include each weighted average price separately, as well as a specification with all measures of weighted average

price, wp . The variable $wp_{0.1}$ appears in the first specification, $wp_{0.5}$ appears in the second and so forth. The table shows that the coefficient of weighted average price, (i.e., the spatial effect) increases as the critical distance moves from 0.1 mile to 1 mile. Notice that the largest R^2 is in column 3 when a 1 mile definition of the competing group is used. These results provide little evidence that stations compete with only the closest stations. When all measures wp of weighted average prices of competing stations are included in a single equation, only the coefficient of $wp_{1.0}$ is significant at the 1% level. The estimates of the spatial effect with a 1 mile definition of the competing group is 0.571.¹¹ Recall that two times the coefficient of the spatial effect measures the fraction of sales lost by a station that are captured by competing stations when the station raises its prices. Therefore, when a station raises its price, about 60% of sales lost by the station are captured by stations within 0.1 miles; 85% of total lost sales are captured by stations within 0.5 miles; finally 100% of all lost sales are captured by stations within 1 mile. Taken together, I find strong evidence that gas stations compete with stations within a 1 mile radius, and that the intensity of competition decreases with distance.

5.2 The Determinants and Direct Effect

Now, I examine the determinants for retail gasoline prices within the spatially competitive market. Note that station fixed effects are removed and station characteristics and market characteristics variables are included in the model to control for most of station heterogeneity. With station fixed effects, the coefficients of any station characteristics and market characteristics that are time invariant cannot be identified separately from the fixed effects. Hence, I include observed station characteristics and market characteristics to control for station heterogeneity with random effects using instrumental variables to identify the effects on the retail prices of station characteristics and market characteristics.¹²

¹¹As shown in (2.7), the coefficient of the spatial effect should be around 0.5 when the local market is well defined.

¹²Notice that the estimated coefficient of the spatial effect using the station and week fixed effect model (5.1) with 1 mile as a critical distance is 0.571. The estimated coefficient of the spatial effect from the week fixed effects model (4.3) after controlling for observed station characteristics and market characteristics is 0.560. These two estimated coefficients of the spatial effect are not significantly different from each other. This implies that model (4.3) controls for most of station heterogeneity.

Table 3: IV Price Reaction Functions

Dependent Variable	Price (1)No Spatial Effect	Price (2)Common boundary 1.0 mile	Price (3)Critical Distance 1.0 mile
<i>Weighted Ave.Price</i>		0.564** (0.079)	0.560** (0.080)
ARCO	-4.4420** (0.712)	-4.525** (0.698)	-4.540** (0.699)
Chevron	2.506** (0.911)	2.582** (0.870)	2.652** (0.870)
Exxon	-0.158 (1.444)	0.057 (1.042)	0.050 (1.043)
Mobile	0.998 (0.955)	1.265 (0.880)	1.261 (0.881)
Shell	2.953** (0.966)	3.197** (0.885)	3.251** (0.886)
Texaco	1.687 [†] (0.920)	1.998* (0.863)	2.020* (0.864)
76	2.766** (0.927)	3.154** (0.839)	3.165** (0.840)
Unbranded	-2.787** (0.916)	-3.126** (0.742)	-3.090** (0.742)
# of competitors	-0.158** (0.052)	-0.026 (0.049)	-0.026 (0.049)
Pumps	-0.028 (0.021)	-0.034 [†] (0.020)	-0.034 [†] (0.020)
Car Wash	-1.252* (0.568)	-1.080* (0.531)	-1.158* (0.531)
C-store	-0.637 (0.471)	-0.784 [†] (0.439)	-0.743 [†] (0.439)
Garage	-0.484 (0.617)	-0.631 (0.489)	-0.637 (0.489)
Full Service	0.812 (0.630)	0.910 (0.641)	0.911 (0.641)
Lessee	1.250** (0.482)	1.164* (0.460)	1.158* (0.460)
Dealer	0.045 (0.749)	0.262 (0.634)	0.245 (0.635)
Median Income	-0.023 (0.016)	-0.027 [†] (0.015)	-0.026 [†] (0.015)
Residential Density	0.048 ** (0.012)	0.002 (0.014)	0.006 (0.014)
Median Rent	0.462** (0.096)	0.222* (0.109)	0.228* (0.109)
Vehicles	-0.012 (0.014)	-0.009 (0.014)	-0.008 (0.014)
Travel Times	-0.165** (0.047)	-0.083 [†] (0.043)	-0.081 [†] (0.043)
# of obs	33807	33807	33807
R ²	0.9137	0.9240	0.9244

1) Coefficients of week fixed effects are omitted ,and 7-11 brand and company operated station dummies are omitted.

2) [†] significant at 10%, * significant at 5%, ** significant at 1%

3) Robust standard errors in parentheses.

Table 3 contains results from IV estimates of the coefficient of the exogenous variables except week fixed effects. The results in column 1 are the estimates of the traditional price equation that does not account for the spatial competition effect. Column 2 reports the results for the price reaction function with average price ($w_{p_{cb}}$) and 1 mile as a common boundary. Column 3 reports the results for the price reaction function with weighted average price ($w_{p_{cd}}$) and 1 mile as a critical distance. The results demonstrate strong evidence for a spatial competition effect in pricing. First, the spatial effect estimates in columns 2 and 3 are 0.564 and 0.560, respectively and they are statistically significant at the 1% level. Second, the R^2 in column 3 is the largest. This result shows that the 1 mile as a critical distance model explains price variation across stations better than the 1 mile as a common boundary model. This result also indicates that gas stations most heavily compete with other stations within 1 mile, and that the intensity of that competition diminishes with distance.

For the determinants of the retail gasoline price, I discuss the results presented in column 3. Note that the estimated coefficient of all exogenous variables measures the direct effect from gas stations' own characteristics. The estimated coefficients of Chevron, Shell, Texaco, and Union 76 stations are positive and significant. The coefficient estimates imply that Chevron, Shell, Texaco, Union 76 station are high brand stations likely to have high brand loyalty and perceived quality. Notice that the coefficient estimate of Shell is the largest and equal to 3.251. All else equal, Shell stations have about an average 3.3 cent per gallon higher price than 7-11.¹³ In contrast with other brand stations, the estimated coefficient of Arco is negative and significant. This is consistent with the fact that Arco is widely thought of as an ultra-competitive brand with lower prices and limited service (i.e., They do not accept credit cards, etc.). All else equal, Arco stations tend to have about a 4.5 cent lower price per gallon than 7-11. In addition, unbranded stations tend to have lower prices per gallon than other branded stations except for Arco.

Coefficient estimates for several station characteristics variables are also significant. The estimated coefficient of Lessee is 1.158 and significant.¹⁴ This indicates that lessee dealers tend to have higher prices than company operated brand stations; this is consistent with the double

¹³The omitted brand dummy is 7-11.

¹⁴The omitted owner structure dummy is Company.

marginalization hypothesis. All else equal, stations with a car wash have about a 1.2 cent lower price, and stations with convenience stores have a 0.7 cent lower price on average. This could be because a station may increase its revenue by selling other goods or services via attracting more customers with lower gasoline prices. The number of fuel pumps at a station is associated with lower price. This could be because stations with more fuel pumps can sell higher volumes and still cover fixed cost at a lower price. An alternative explanation can be that a station with more fuel pumps can charge a lower price due to economics of scale due to volume discounts from the refinery.

Coefficient estimates for several market characteristics variables are also significant. The estimated coefficient of median rent is positive and significant. Price tends to be higher when median rent is higher. Price tends to be lower as median income is higher. The estimated coefficient of travel time is negative and significant. This can be because workers with longer commutes can see more gas stations on their way to work and find a gas station with a low price without additional travel or search costs.

In contrast to the specification in column 1, the number of competitors within 1 mile does not explain price variation across stations. Interestingly, these results reveal that compositions of competing stations and characteristics of competing stations are more important determinants of the retail gasoline price than the number of competitors. These findings give important implications for policy evaluation such as merger analysis in the sense that a particular gas station's price is influenced by competing stations' brand name and their spatial locations.

5.3 The Indirect Effect

Recall that the indirect effect, $\beta_k \lambda w_{ij}$ is a marginal effect on station i 's price of a change in the k -th characteristic of its competing station j . While stations who compete with Arco or unbranded stations are more likely to have a lower price, stations that compete with high brand stations such as Chevron, Mobile, Shell, Texaco, and U76 are more likely to have a higher price due to the indirect effect. For example, suppose there are two gas stations, say A and B. Station A has only an Arco station as a competitor and Station B has a Shell station as a competitor within 1 mile.

Due to the indirect effect of the Arco Station, Station A charges about 2.54 cents less than when it faces 7-11 as a competitor because its competing station, Arco is an ultra competitive brand with lower prices. But, Station B can charge about 1.82 cents more than when it faces 7-11 as a competitor due to the indirect effect of the Shell station with high brand loyalty and perceived quality. In addition, stations that compete with stations with car washes or convenience stores tends to have lower prices. Also, stations that compete with lessee dealer stations tend to have higher prices than when they compete with company operated branded stations.

5.4 The Total Effect

Recall that the total effect, $\beta_k/(1 - \lambda)$ is a marginal effect on all gas stations prices of a 1-unit change in the k -th characteristic in all gas stations. For the total effect, I discuss only continuous variables such as the number of pumps at a station, and median rent. If the number of pumps at all gas stations increases by one, prices of all gas stations will decrease by 0.07 cent. If the median rent at all gas stations increases by \$100, then prices of all gas stations will increase by 0.5 cent.

6 Concluding Remarks

In this paper, I examine the nature of competition in the retail gasoline market using a highly detailed station level data set. By applying spatial econometric techniques, I empirically identify a proper local market definition for the retail gasoline industry. Then, I more fully examine the complexity of the relationship between gas station characteristics, competing stations' characteristics, and individual station's pricing behavior. Estimating the SAR model of stations' price reaction functions allows me to make several contributions to our understanding of price variation across gas stations.

I have several results of note for the local market definition. First, I find little evidence that stations are competing with only the closest station. Second, I find strong and consistent evidence that stations heavily compete with any gas stations within 1 mile and the intensity of competition decreases with the distance between stations. This has important policy implications for other

research on spatial competition in the retail gasoline market particularly because the estimates are very sensitive to the choice of the relevant market definition.

My results are consistent with the spatial model. I also find that retail prices are heavily influenced by station's characteristics such as brand name and amenities. By using the SAR model, I identify that the brand name of competing stations and even the composition of those particular brands over the geographical space are more important determinants of the retail gasoline price than the number of nearby stations. This has important implications for policy evaluation, such as in merger analysis in the sense that a particular station's price is differently influenced by competing stations' brand name and their relative location.

The result that brand name is one of the most important determinants of gasoline price leads one to consider that a gas station interacts differently with other stations based on brand. Future research, in addition to considering geographical aspects, should take into account the possibility of brand differentiation.

Reference

- Barron, J.B., Taylor, and J. Beck. 2004. Number of Sellers, Average Prices and Price Dispersion. *International Journal of Industrial Organization* 22, 1041-1066.
- Baltagi, Badi H., 1981. Simultaneous equations with error components, *Journal of Econometrics* 17, 189-200.
- Borenstein, S., A., Shepard. 1996. Dynamic Pricing in Retail Gasoline Markets. *Rand Journal of Economics* 27, 429-451.
- Ecker, T.A., West, D.S., 2004. Retail Gasoline Price Cycles across Spatially Dispersed Gasoline Stations. *Journal of Law and Economics*, 245-273
- Hastings, J.S., 2004. Vertical Relationships and Competition in Retail Gasoline Markets. *American Economic Reviews* 94, 317-328.
- Hosken, D, R. McMillan, and C. Taylor. 2006. Retail Gasoline Pricing: What Do We Know? Federal Trade Commission, Bureau of Economics Working Paper #290
- Kelejian, H.H., Prucha, I.R., 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbance. *Journal of Real Estate Finance and Economics* 17, 99-121.
- Lee, L.F., 2003. Best spatial two-stage least squares estimators for a spatial autoregressive disturbances. *Econometric Reviews* 22, 307-335.
- Lee, L.F., 2005. GMM and 2SLS Estimation of Mixed Regressive, spatial Autoregressive Models. Manuscript, Department of Economics, OSU, Columbus, OH.
- Lee, S.Y., 2007. Bias from Misspecified Spatial Weight Matrices in SAR Models: Theory and Simulation Studies. Manuscript, Department of Economics, OSU, Columbus, OH.
- Ord, J., 1975. Estimation methods for models of spatial interaction. *Journal of the American Statistical Association* 70, 120-126.

Pinkse, J., M. E. Slade, and C. Brett., 2002. Spatial Price Competition: a Semiparametric Approach.
Econometrica 70, 1111- 1153

Shapiro, C. 1996. Mergers with Differentiated Products. Antitrust, 23-30.