A Work Project, presented as part of the requirements for the Award of a Master's degree in Management from the Nova School of Business and Economics.

THE RETAIL GASOLINE MARKET IN THE LISBON METROPOLITAN AREA – A SPATIAL COMPETITION ANALYSIS

FRANCISCO REGO MORCELA CALDEIRA FERNANDES

Work project carried out under the supervision of:

Sofia Franco

08-09-2020

Abstract

This paper analyses the determinants of the price of gasoline on the retail fuel market in the Lisbon Metropolitan Area, Portugal. Through a Spatial Autoregressive model that solves for the best response functions of gas stations, evidence was found that location and geographic boundaries of competition have effects on gasoline prices. Moreover, results also indicate that the characteristics of a gas station, that include its Brand and service offer, are more relevant indicators of price alterations than socioeconomic characteristics of the market.

Keywords: Competition, Gasoline, Spatial Analysis, Price

Introduction

Oil commodity markets are in constant change and its prices reflect multiple current trends and events. Consequently, prices on the retail gasoline market are subject to variations. However, even within the same limited region, where regulations and taxes established by governmental authorities are similar, retail prices show variations from one gas station to another gas station. It is crucial to understand how these prices behave and what are the real determinants of price on a local basis.

Consumers are only willing to go to a more distant station if the difference in price between the closest station is higher than the cost of taking the additional transportation, meaning that a decision is only made if the marginal benefit surpasses the marginal cost. If this information is relevant, location plays an important part in defining prices which translates into gas stations having a limited radius into where they compete with other gas stations. Spatial competition analysis comes into play in this setting, as it now matters to understand how spatial variables affect prices.

This paper aims to understand the effects of spatial variables on the retail gasoline market as well as understand the main determinants of price. It will explore a data set on the Lisbon Metropolitan Area (LMA). A Spatial autoregressive model was developed to explain the prices of a given gas station through the average weighted price of directing competing gas stations. It is expected to be found in some form of spatial dependency. Modern approaches to the model spatial competition are derived directly static games where decisions are taken to maximize profits and takes price as the decision variable. Pinske et. All (2002), calculates the Euclidean distance between gas stations and uses the average weighted price of competitors within a critical distance to explain the price of a given gas station. An essential assumption is that gasoline is homogeneous and that the source of product differentiation comes from the location of the gas station.

Gas stations are distinguished based on the type of ownership of the company. They can either be branded or unbranded. Branded stations are related to traditional oil companies, often with presence along the oil-fuel supply chain, with the fuel being acquired from specific refiners, also owned by the brand. On the other hand, unbranded stations also referred to as independent stations, are mainly present in the retail market, and can change suppliers to buy from the refinery that offers less price. Hastings (2004) focused on explaining how the vertical relationships between companies affected competition under the argument that refineries control for prices, as it concluded that the presence of independent stations can indeed strengthen price competition through the diminishing of prices. However, there is also evidence that this diminishing in prices, only happens to a certain extent, constraint to the effect of an unbranded name on consumers' perception of lower quality (Pennerstorf, 2009).

The consumer decision-making process plays an important part in discovering the patterns of demand, that, by itself, can influence the best response of a company in the market. Houde (2012) tested commuting patterns of consumers as their location to measure the impact of positioning and found evidence that they can more accurately predict real sales of gasoline in a specific area. This introduces a question of how consumer patterns can influence the competitiveness between firms.

Methods to compute the exact critical distance to which gas stations operate in vary, some models may define critical distance according to socioeconomic factors of the population (Pennerstorf, 2009), or using a more empirical methodology either by creating communicating directly with the gas stations (Hastings, 2004) or by finding the best distance through the results of several spatial weighted matrices (Lee, 2007). Nevertheless, there is evidence, throughout the mentioned models, that there is a spatial correlation. Firgo et. all (2015) further explored the pattern of proximity by controlling for centrality within a geographic space, finding that more central players have increased market power when compared with further away station.

Moreover, there have been recent legal changes to the commercialization of retail fuel on Portuguese law. It made obligatory for all gas stations to offer simple gasoline and simple diesel. Previously to this, the branded gas stations would not offer this type of fuel and had to change its retail offer to comply with these rules. This allowed for the homogenization of simple fuels, either if they are from branded or independent gas stations. With differentiation between stations on the final product by itself not possible, it may still happen through the characteristics of the gas stations, the remaining more premium type of fuel offered or geographical location. Whereas previous studies considered perceptions of consumers on higher fuel quality in branded gas stations to have effect on prices, obligatory simple gasoline across all gas stations may have different effects may result in different expected results. Meaning that this paper will be adding to current literature by testing the modern empirical specifications of spatial models of competition on the Portuguese market were homogenization of gasoline as a product is obligatory:

Results may help to clarify and characterize the market as well as to help identify behaviors useful to policy implementations and regulation of competition.

On the following sections, consumer and firm economics decision will . Followed The characteristics of the data set will be further explored

Theoretical Model

To further understand how spatial competition works in the retail gasoline market, a theoretical model will be built, taking as a basis Pinske et al (2002) developments and closely following Lee's (2007) frameworks of model development. The base assumption is that gas stations compete in prices and geographical location presents itself as the main source of differentiation between gas stations. Nevertheless, other characteristics of a gas station such as the quality of the service or the brand should be considered and controlled for, due to consumers that, even by considering neighboring stations closed substitutes, account for these characteristics in the decision-making process. The optimal price will be simultaneously decided between all gas stations because even if one gas station does not directly compete with another, they are indirectly linked by a network of gas stations.

Demand-side

In a market with n unique gas stations that are spatially differentiated each gas station i, $i = 1, \dots, n$ has a nominal price $\tilde{p}_i$. Additionally, an outside good, whose nominal price is p_o, will also be considered.

Each consumer, within a different geographic position, makes an optimal decision, based on location, of the quantity of gasoline to buy from a given gas station, considering the costs that he partakes in, such as transportation costs. The consumer $c, c = 1, \dots, h,$ buys $q_k = (q_{1c}, \dots, q_{nc})'$, a vector of the spatially differentiated gasoline, where $q_{1c} > 0$, $i = 1, \dots, n$, and q_{oc} of the outside good, and $q_{oc} > 0$. A consumer c presents an indirect utility function $\tilde{V}_c(p_0, \tilde{P}_n, \tilde{y}_c)$, that uses as arguments the nominal income $\tilde{y}_c$, and the vector of prices for each respective spatially differentiated gas station, $\tilde{P}_n$. According to the existing literature, it will be assumed that $\tilde{V}_c$ takes the form of a normalized quadratic utility function, that can be written as:

$$V_c(1, P_n, y_c) = -[\alpha_{oc} + \alpha'_c P_n - p_0 y_c(\gamma_0 + \gamma' P_n) + \frac{p_0}{2} P'_n S_{n,c} P_n] \tag{2.1}$$

defining $P_n = p_0^{-1}\tilde{P}_n$, $y_c = p_0^{-1}\tilde{y}_c$ and $S_{n,c}$ as a symmetric negative semi-definite matrix $n \times n$, where there can be quantified the own-price elasticity, how the demanded quantity reacts to a price alteration of the good in analysis, in its diagonal elements and on the remaining elements the cross-price elasticity of the demand function, how demand reacts to price alteration in outside goods. The quadratic form of the function enables the observation of demand as linear.

The demand for gas station for an individual consumer can be determined as the following, through Roy's identity:

$$q_{i,c} = -\frac{\frac{\partial V_c}{\partial p_i}}{\frac{\partial V_c}{\partial y_c}} = \frac{\alpha_{ic} + \sum_{j=1}^{n} s_{ij,c}\, p_j - \gamma y_c}{p_0(\gamma_0 + \gamma' P_n)} \tag{2.2}$$

The price index, $p_0(\gamma_0 + \gamma' P_n)$, as an assumption will be considered as a constant equal to one. This allows to simplify the aggregate demand function for spatially differentiated gasoline into the following equation:

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

that denotes $\alpha_i = \sum_{c=1}^{h} \alpha_{ic}$, $s_{ij,c} = \sum_{c=1}^{h} s_{ij}$ and aggregate income as $y = \sum_{c=1}^{h} y_c$. Conditional on, $\alpha_i > 0$ and $-s_{ii} > s_{ij}$ for all $i \neq j$, to ensure that the utility function is concave. α_i reflects the demand experienced by a given gas station, this means that the higher the α_i, in between different gas stations, the higher the demand. The likelihood of building a more loyal customer base, or the ability of the gas station to offer more service, increases the higher α_i is. On the other hand, s_{ij} tells how likely it is for a customer to substitute between two gas stations i and j and γ_i measures the income effect.

Supply Side

The goal of a firm is to maximize profit. I assume that gas stations play a static game of Bertrand Nash. This means that firms incorporate the likely decision of competitors when taking a decision, meaning that when a price is set the likely decision of each competitor is already considered. Moreover, each gas station i presents a fixed cost, F_i, and a marginal cost, mc_i. The marginal cost is assumed to be constant and variates from company to company as affected by different cost factors. Factors that may impact marginal costs are, for example, brands and their additional roles that along the supply chain, mainly the relation that may take with the refiners or the bargaining power with them. The profit function of a gas station i can be defined as follows:

$$\Pi_i = (p_i - \delta_i' \, mc_i)q_i - F_i \tag{2.4}$$

Computing for total demand obtained in equation (2.3), gas station i will set a price p_i, by maximizing its profit function. The prices practiced by competitors are also considered when solving for this function

$$\max_{p_i}(p_i - \delta_i' \, mc_i)(\alpha_i + s_{ii}p_i + \sum_{j \neq i}^{n} s_{ij}\, p_j - \gamma_i y) - F_i \tag{2.5}$$

Best Response Function

The price reaction function or best response function of a gas station i is obtained by calculating the first-order condition of the maximization problem of the firm. The best response function is the following:

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

It would not be possible to estimate all unknown parameters using cross-sectional data of n gas stations. Let us consider $-1/2\,(\alpha_i/s_{ii} - \gamma_i/s_{ii}y - \delta_i' \, mc_i)$ a function of a gas station i and market characteristics. To further continue it will be considered $-1/2\,(\alpha_i/s_{ii} - \gamma_i/s_{ii}y - \delta_i' \, mc_i) = X_i\beta + u_i$, that characterizes X_i as a vector of the characteristics of the market and the characteristics of a given gas station. u_i is a random error and relates to the effects of non-observed characteristics. Moreover, now $d_{ij} = \frac{s_{ij}}{s_{ii}}$, for all $i \neq j$. The function (2.6), after these modifications, is as follows:

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

The d_{ij}, is also referred in the literature as the "diversion ratio". When firm i raises the price p_i, the amount of sales that are lost from firm i to firm j is the diversion ratio. As gas station are spatially differentiated, and in this model the only source of differentiation, it can be considered d_{ij} as a function of the distance between two gas stations. If two stations i and j do not compete between each other the diversion ration is 0, whereas if they do indeed compete but they don't face any more additional competition the diversion ration will be 1, as these two stations are the only influence in the area.

The best response function will be again reformulated:

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

with $w_{ij} = d_{ij} / \sum_{j \neq i}^{n} d_{ij}$ and $\lambda = 1/2 \sum_{j \neq i}^{n} d_{ij}$. To allow for a percentage comparison of the lost amount of sales, the diversion ration is now, in the form of w_{ij} the relative diversion ratio and $\sum_{j \neq i}^{n} w_{ij}\, p_j$ relates to a weighted average of the prices of the firms that compete with i.

Let X_n represent a $n \times k$ matrix with the observed characteristics of the stations and the market. Additionally, u_n is a vector of random errors. A system of equations with the set of best response functions, from (2.8) for all gas stations observed, that takes p_n as the vector of the gas stations' optimal prices can be defined as:

$$p_n = \lambda W_n p_n + X_n\beta + u_n \tag{2.9}$$

where W_n represents the matrix of the relative diversion ratios between stations, however diagonals of W_n are zero because it is the diversion ratio of a station with itself. This system of prices is called a spatial autoregressive (SAR) model. In this model, the weighted average prices between competitors are $W_n p_n$ and λ, that can also be called the spatial effect parameter relates to the influence of the prices of competitors on p_n.

Additionally, Lee (2007) proves that the price of a given station will also be affected by the weighted average of the characteristics of the competing stations, a result that shows that prices in the model a heterogeneous and spatially correlated

Interpretation of the SAR Model

A variation of a characteristic k, of a given gas station i, will have three different effects on price. These effects can be characterized and interpreted to for the question this paper poses on.

The effect that the own characteristic k of i have on its price, is the direct effect. Secondly, we have an indirect effect, that relates to the effect that characteristic k of competing gas stations has on the price of i. Although one station is only directly competing with a restricted number of stations, these other stations are competing with other different stations. From here gas stations create spill-over effects, meaning that a given gas station can still influence the price of a gas station that is outside the direct competitor's radius. This relation, also called induced effect, allows for a common link between all gas stations over a given geographical location. If the induced effect is restricted to the boundaries of a region, meaning that there is no spatial relationship between companies inside and outside the region, the model can be characterized as global.

The total effect will encompass all effects that a characteristic k, has over the market.

Econometric Model

The initial procedure to estimate the model is to develop an OLS model:

$$p_n = X_n\beta + u_n \tag{4.1}$$

p_n is a vector, dimension $n \times 1$, of the dependent variable price. The exogenous variables are encompassed within X_n, a matrix of dimension $n \times k$, encompasses the exogenous variables, more specifically the characteristics of the station and the market, whereas u_n is the error term, i.d.d. with mean zero and variance σ^2.

Even if this model is consistent, it will need to be tested for spatial autocorrelation. Using Moran's Index for special autocorrelation and if there is evidence of spatial autocorrelation, it will be implemented the spatial autoregressive (SAR) model.

A standard SAR model can be formulated as:

$$p_n = \rho W_n p_n + X_n \beta + u_n \tag{4.1}$$

This model introduces W_n as the spatial weight matrix and $W_n p_n$ as the average weighted price of a given gas station. The ρ estimator reflects the spill-over effect that happens between gas stations. The coefficient estimates, β, on the OLS models represent the effects that a marginal increase in a characteristic k have on price. To compare this interpretation with the SAR results, it is fundamental that somehow coefficients can be similarly interpreted. The direct and indirect effects are what enables this comparison, has it comes from the theoretical, these are calculated by finding the marginal impact of a characteristic k. However in the SAR models, due to introducing spatial weighted variables, a characteristic k of a given gas station i will impact both the considered gas station i, the direct effect and its direct competitors, the indirect effect.

The spatial weight matrix measures the spatial relationship between gas stations. One way to build the spatial weight matrix would be to attribute the same weight to all competitors within a specific critical range of a given gas station or to k nearest competitors of that same station, however I assume that intensity of competition is aggravated the nearest a gas station is to another. Because of this, within the critical radius, the weight should be attributed related to the relative distance a competing station is to the central station. The method used to compute the weights is using the inverse distance between a gas station i and competitor j, $j \neq i$. The weight, w_{ij}, of a gas station j on i, will be given by $w_{ij} = 1/d_{ij}$, with d_{ij} being the Euclidian distance between the two gas stations if this distance is smaller than the defined critical range of competition, otherwise, the weight is zero.

It will be used panel data analysis, using data on the price of gasoline from the first Tuesday of each month. Controlling for fixed time and spatial effects, will allow to further improve the model, as to diminish omitted variable variables, that, without the introduction of these fixed effects would likely appear. Additionally, controlling for time effects will allow a better understanding of the behavior of prices across the year as well as to understand the presence of seasonal effects. The model will also control for Brand and City effects as to capture specifics trends and effects that these may have on price.

The final empirical model to find the price results of the gas stations will be written as follows:

$$p_{it} = \rho w p_{it} + X_i \beta + \sum_{t=1}^{T} \phi_t \, Month_t + u_{it} \tag{4.2}$$

Data

This section describes the data used to build the models. To study how firms compete it was used two data sets, one with information on all gas stations in the Lisbon Metropolitan Area (AML) and one the respective prices on the first Tuesday of each month of 2019. Both these data sets were provided by Direção-Geral de Energia e Geologia (DGEG).

Defining the geographical boundaries of the region of study helps to identify a potential spill-over effect, that happens when the prices of a gas station are also affected by prices no within the region, not making the model truly global by nature On the west side, AML makes frontier with the Pacific Ocean. However, the more peripheric cities, mainly on the furthest north and south frontiers, are directly attached to other cities outside AML. When looking more closely into what distinguishes the AML, it is observed that River Tejo divides AML in two, creating a natural barrier between two margins, that can only be transposed efficiently via 2 main bridges or by ferry. In Portugal, there are 2 refineries, one in Matosinhos on the north and one in Sines, south. Sines is the closest refinery to the AML, located to the south of it. There are also three storage units inside AML, all located in the south margin. These unique geographical characteristics can provide interesting information on how prices translate between the same region, separated by natural barriers.

The first data set contains information on all 465 gas stations in the AML, at the beginning of 2020. It includes information on location, geographic coordinates, opening hours, and brand name. Although some more specific information about the gas stations itself, for example, the number of pumps they have or what types of services they offer, could have been made useful for this study, these were unavailable, mainly because this is information is not obligatory to disclose. Location and distribution from these stations can be observed in Figure 1 and provides information on how distributed they are through the map. It shows that somehow, gas stations tend to be located on proximate spots.

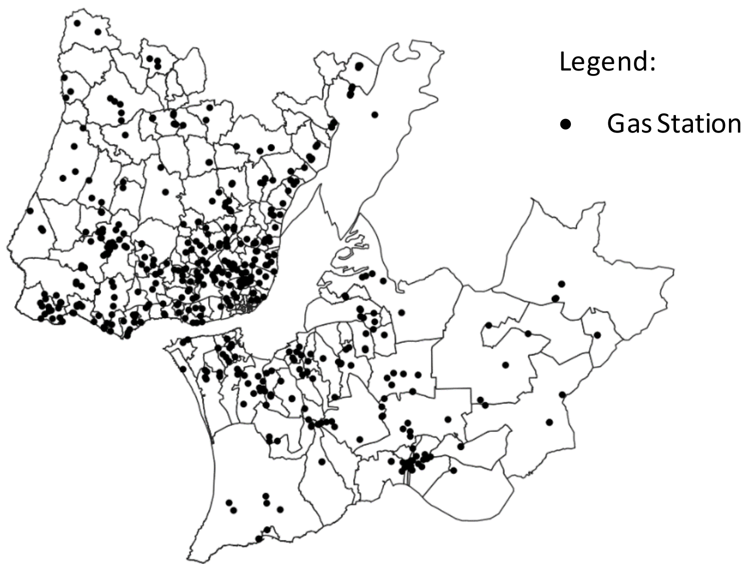

Figure 1. Distribution of gas stations on the Lisbon Metropolitan Area

In this data set, it was also computed the number of competitors each gas station faces around a given radius. From existing literature, a gas station competes on average with all stations in a distance from 1 to 3 km. However, there might happen some variations due to population density or other variables on the market that might make the radius change. so, it was calculated the number of competitors for different wider radius to control how these changes in concentration over an increasing distance may affect the price and to find an optimal radius to apply in this region. When analyzing the market of retail gas stations in Portugal there can be identified 3 different types of operators. GALP, REPSOL, CEPSA and BP are traditional oil companies that are vertically integrated, and present along the whole value chain, from refining along to the final consumer market. Ownership of these gas stations can be either from the company itself or from a licensed dealer. It is not possible, with the information available to distinguish these ownership differences.

Independent retailers are operators that are mainly or exclusively dealing with the consumer retail market of fuel, acquiring fuel to oil companies in the gross market. In Portugal, independent retailers can also be broken down into two: the companies that have their storage facilities, limited to two companies, and the remaining, that do not have their storage facilities. Finally, there are supermarket chains, such as PINGO DOCE, RECHEIO, LECLERC, AUCHAN, and INTERMACHE. These gas stations are also independent retailers but are owned by large food retail corporations. Located usually near the location of the stores, this business serves as a secondary revenue stream to attract customers to the supermarkets. The main difference between traditional oil companies and independent retailers is how they acquire the product. The cost structures of these companies may differ as there may have scale and scope economies.

The price data set was used using retail prices on the different types of fuel of gasoline, for the first Sunday of each month, also from DGEG. Although this study will only refer prices to simple gasoline of 95 octanes, it will be controlled for the types of fuel that a gas station sells as a gas station characteristic as this will allow understanding of the given offer in terms of variety fuel affects the overall price of gasoline. From the original 465 gas stations in the previous data set, 37 of them did not offer simple gasoline of 95 octanes and therefore, they were dropped. It was analyzed 428 different gas stations, and over a total of 12 months analyzed, this amounts to 4878 different price entries.

A third data set was used using information from the Portuguese census of 2011, from Instituto Nacional de Estatística (INE). These were used to identify the market characteristics, such as data on commuting times, population density, number of garage spots of a given household as well as demographic information on education, gender, and age of the population. This will help identify how demographic affects competition as well as helping to characterize the market by itself.

Table 1. Summary statistics and variable definitions

Variable	*Definition*	*Mean*	*S.D.*
Brand			
Galp	Gas station of brand Galp	0.299	(0.458)
Repsol	Gas station of brand Repsol	0.210	(0.408)
Intermarché	Gas station of brand Intermarché	0.020	(0.172)
Cepsa	Gas station of brand Cepsa	0.093	(0.291)
Prio	Gas station of brand Prio	0.075	(0.263)
BP	Gas station of brand BP	0.199	(0.399)
OZ Energia	Gas station of brand OZ Energia	0.007	(0.083)
Auchan	Gas station of brand Auchan	0.021	(0.143)
Rede Energia	Gas station of brand Rede Energia	0.019	(0.135)
Leclerc	Gas station of brand Leclerc	0.007	(0.083)
Recheio	Gas station of brand Recheio	0.002	(0.048)
Pingo Doce	Gas station of brand Pingo Doce	0.007	(0.083)
Independent	Independent gas stations	0.030	(0.172)
Station Characteristics			
Open Weekends	The station is open on Saturday	0.993	(0.083)
Total Fuel Offer	Number of different fuels a station offers	4.049	(0.998)
Distance to gas station	Distance of a gas station the closest gas station (in km)	0.689	(0.740)
# competitors 5km	Number of competitors within a radius of 5 km	1.794	(1.590)
Location			
North Margin	The station is in the north margin of Tejo River	0.652	(0.476)
Lisbon	The station is within the city of Lisbon	0.154	(0.361)
Market Characteristics	(information from 2011 census)		
# residents	Number of residents	657.435	(195.452)
# residents working they live	Number of residents working in the living zone	156.432	(67.457)
Commuting time	Average Commuting time of residents (in minutes)	25.673	(3.153)
Population density	Population density (residents per km^2)	3316.102	(3686.850)
Price	Price of simple gasoline 95 octanes (€ per liter)	1.539	(0.070)

Results

In this section, it will be analyzed the results from the models established previously.

With a comparison in mind, an OLS model was first estimated. The model was tested for the Gauss-Markov assumptions for the Ordinary Least Squares Method. Multicollinearity was tested using the VIF method. (See Appendix 1). The low level of bellow 10 on all variables shows no evidence of multicollinearity between the explainable variables. The error terms, however, are not normally distributed which, although it does not invalidate the model, lessens the strength of the estimators' standard deviations. Another implication is that it may apply an omitted variable bias on the model caused by exogenous variables not included in the model. This model is also heteroskedastic, to correct for this it was used the robust standard error method. On a final note, there was no evidence of serial autocorrelation. The results for the tests can be observed in Appendix 2.

To test for spatial autocorrelation on the model the Moran's I test was computed. The null hypothesis of no spatial autocorrelation was rejected, with a 0.01 significance level meaning that the model may evidence spatial autocorrelation. Furthermore, the positive value of the z-score (29.841) indicates the possibility there might have spatial clustering.

Table 2. OLS and SAR estimates for Price Response Functions

	Dependent variable:		
	Log(Price)		
	OLS	SAR	
	Estimators	Direct Effect	Indirect Effect
Interception	0.383^{***}	0.548^{***}	0.548^{***}
	(0.004)	(0.004)	(0.004)
Rho coefficient		0.0602^{***}	0.0602^{***}
		(0.0071)	(0.0071)
DIST_NEARESTGASSTATION	-0.001^{**}	0.0001	0.000008
	(0.0004)	(0.0003)	(0.0003)
NCOMPETITORS_5KM	-0.0002^{***}	-0.0002^{***}	-0.00005^{***}
	(0.00003)	(0.00003)	(0.00003)
N_INDIVIDUAL_RESIDENTS	-0.00001^{***}	-0.00001^{***}	-0.0000006^{***}
	(0.000001)	(0.000001)	(0.000001)
N_INDIVIDUAL_RESIDENTS_WORK_IN_LIVINGZONE	0.00001^{**}	0.00002^{**}	0.0000009^{**}
	(0.00001)	(0.000005)	(0.000005)
OPEN_WEEKENDS	0.020^{***}	0.0198^{***}	0.0013^{***}
	(0.003)	(0.002)	(0.002)
COMMUTING_TIME	0.00002	0.0001	0.000007
	(0.0001)	(0.0001)	(0.0001)
POPULATION_DENSITY	0.000003^{***}	0.000005^{***}	0.0000003^{***}
	(0.000001)	(0.000001)	(0.000001)
TOTAL_OFFER	0.003^{***}	0.0024^{***}	0.0001^{***}
	(0.0003)	(0.0002)	(0.0002)
NORTHMARGIN	-0.008^{***}	-0.008^{***}	-0.0005^{***}
	(0.002)	(0.002)	(0.002)
LISBON	0.011^{***}	0.0096^{***}	0.0006^{***}
	(0.001)	(0.001)	(0.001)
BP	0.078^{***}	0.0781^{***}	0.005^{***}
	(0.001)	(0.001)	(0.001)
CEPSA	0.045^{***}	0.0474^{***}	0.0029^{***}
	(0.002)	(0.002)	(0.002)
GALP	0.072^{***}	0.0714^{***}	0.0046^{***}
	(0.001)	(0.001)	(0.001)
INTERMARCHÉ	0.001	0.0004^{***}	0.00003^{***}
	(0.001)	(0.001)	(0.001)
LECLERC	-0.005^{**}	-0.0066^{*}	-0.0004^{*}
	(0.002)	(0.002)	(0.002)
OZ ENERGIA	0.064^{***}	0.0626^{***}	0.004^{***}
	(0.002)	(0.002)	(0.002)
PINGO DOCE	-0.006^{**}	-0.0061^{*}	-0.0004^{*}
	(0.003)	(0.003)	(0.003)
PRIO	0.037^{***}	0.0365^{***}	0.0023^{***}

	(0.001)	(0.001)	(0.001)

Table 2. OLS and SAR estimates for Price Response Functions (Continued)

	Dependent variable:		
	OLS	SAR	
	Estimators	Direct Effect	Indirect Effect
RECHEIO	-0.004*	-0.0043*	-0.0002*
	(0.002)	(0.002)	(0.002)
REDE ENERGIA	0.036***	0.0348***	0.0022***
	(0.003)	(0.003)	(0.003)
REPSOL	0.067***	0.067***	0.0043***
	(0.001)	(0.001)	(0.001)
INDEPENDENT	0.049***	0.0489***	0.0031***
	(0.003)	(0.003)	(0.003)
Observations	4878	4878	
R^2	0.899		
Adjusted R^2	0.898		
Residual Std. Error	0.015 (df = 4829)		
AIC	-27314	-27370	

Notes: Municipality and Month rows were dropped

*p<0.1; **p<0.05; ***p<0.01

The SAR models were developed using Monte Carlo estimation for the lag coefficients. It is shown that for more narrow markets, between 1 and 3 km, there is no evidence that the models given explain spatial autocorrelation. The likelihood-ratio (LR) test results, that compare the OLS and SAR models showed that the spatially lagged variables, for computed distance of competition did not provide any improvement of the SAR model over the OLS model. When comparing this result with Lee (2007), it is observed that this may not be a normal result, because evidence showed that spatial effects improve the critical in such small radius. However, as it was described by Pennerstorfer (2009), that uses a critical radius of 7.5 km due to lower populational density in central Austria, this critical distance may variate over regions depending on various market characteristics.

If the radius of competition is extended to 5 km, it is clear, through the LR test of the model that the introduction of the spatially lagged variables improved over the OLS estimation. One reason for this may be that the average distance between competitors is higher than expected, and there are few direct competitors within a radius inferior to 5 km. Another possible explanation is that Portuguese consumers are likely to exert their search to a wider radius, widening the relevant spatial effect. The defined critical radius of the competition is 5 km. To test the strength of the model, two additional tests were computed. The Wald Test that test for the overall significance of the set of explainable variables in the model, and the LM Moran test that evaluates for the significance of the spatially lagged variables. These tests results showed, with a significance level of 0.01 for both tests, that the overall set of explainable variables and the spatially lagged variables are significant in the model.

The estimators given on the SAR model are not suitable to interpret as they do not reflect the spatial impact of variables on the best response function of a given gas station. To interpret the results, it needs to be analyzed the direct and indirect effects of the characteristics over price. The table below gives a summary of these effects

The spatial coefficient ρ, rho, that relates to the spill-over effects is 0.0602. The diversion ratio, that is approximately double this value, indicates that when a gas station increases its price, 12% off loss effect on demand will be captured by a competing station.

The direct effect reflects the effect that a characteristic of a gas station has on its price. In this market, if the total offer of different types of fuel in a gas station increases by 1, keeping everything else constant, the price of gasoline in that station is expected to increase, on average 0.02%.

The indirect effect is the effect that the characteristics of competing stations have on the price of a specific gas station. For example, if a given gas station has a BP gas station as its competitor, on average, ceteris paribus, its price is expected to increase by 0.05%.

Characteristics of the market seem to have interesting effects on prices, as gas stations are competing on a 5km radius, where this distance may not be sufficient to justify substantial consumer characterization. Commuting time is not statistically relevant but evidence shows that a market with a higher number of consumers that work in the same residence areas, actually have increasing effects on price This may happen because lengthier paths of commuting, not incorporated into the model, justify for the consumer's high propensity to search outside the residential areas. On the other side, although a small marginal increment of 1 over the number of residents may have a small influence on price, a higher increase could reflect in a substantial change in price.

Gas stations in the north margin, have on average lower prices than on the south margin of the river. But interestingly, gas stations located in Lisbon, also in the north margin, have higher effects on price. This may comply with findings that the central gas stations seem to have higher prices.

With the spatial lag model, the distance to the nearest gas station loses significance but the number of competitors maintains it. It seems that the closest competitor does not affect price, but the overall of all competitors within the critical distance of the competition.

There is also a need to understand how the characteristics of the gas stations work on price, the more offer of other fuel products there is, the more the prices increase. Additionally, if a gas station is open during the weekend, its price is also expected to increase. Although gasoline by itself is homogenous, it is shown that companies can differentiate by improving or increasing on offered services.

In general, traditional oil brands such as BP, CEPSA, GALP, or REPSOL seem to have higher effects on increasing prices than other competitors. On the opposite side, independent brands related to supermarket chains, such as PINGO DOCE or RECHEIO, have opposite effects on price. The intensity of competition is higher when these independent companies are present in the market. This result may be of important consideration when making merger analysis, as it directly affects the market power of the companies.

Conclusions

It possible to conclude that, although there is evidence of spatial autocorrelation, smaller distances of competition seem to have less impact for on capturing spatial effects on price.

Although the gasoline by itself is a homogeneous product, companies have strong differentiation on the matter of internal characteristics, as the more the more substantial different effect on prices, either direct or indirect arises from brand effects have higher impacts. When in the process of policy evaluation, this may have important reflections.

Limitations for this analysis, come from the fact that a full breakthrough on all services offered by a gas station was not available due to data restrictions, may have constricted a full explanation on the effects of differentiation between offered services, despite the great results encountered on the market characteristics. Future studies could look more closely to the behavior of the consumers as to validate the suppositions created during result analysis, and further investigate effect of cross price elasticities on demand.

References

Autoridade da Concorrência. 2018. "Análise ao Setor dos Combustíveis Líquidos Rodoviários em Portugal Continental" ["Analysis to the Market of Liquid Fuels in Continental Portugal"].

Firgo, Matthias, Pennerstorfer, Dieter, and Weiss, Christoph R. 2016. "Network centrality and market prices: Empirical evidence". *Economics Letters*, 139: 79–83.

Hastings, Justine S. 2004. "Vertical Relationships and Competition in Retail Gasoline Markets: Empirical Evidence from Contract Changes in Southern California". *The American Economic Review*, 94(1): 317.

Houde, Jean-François. 2012. "Spatial Differentiation and Vertical Mergers in Retail Markets for Gasoline". *American Economic Review*, 102(5): 2147–2182.

Keleijan, Harry H., and Prucha, Ingmar R. 1999. "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model". *International Economic Review*, 40(2): 509.

Lee, Sang-Yeob. 2007. "Spatial competition in the retail gasoline market: An equilibrium approach using SAR models".

Pennerstorfer, Dieter. 2009. "Spatial Price Competition in Retail Gasoline Markets: Evidence from Austria". *Annals of Regional Science*, 43(1): 133–158.

Pennerstorfer, Dieter, and WEISS, Christoph. 2013. "Spatial clustering and market power: Evidence from the retail gasoline market". *Regional Science and Urban Economics*, 43(4): 661–675.

Pinske, Joris, Slade, Margaret E., and Brett, Craig. 2002. "Spatial price competition: A semiparametric approach". *Econometrica*, 70(3): 1111–1153.

Appendix

Appendix 1. VIF table

VIF Results

DIST_NEARESTGASSTATION	1.213
NCOMPETITORS_1.5KM	1.528
N_INDIVIDUAL_RESIDENTS	2.280
N_INDIVIDUAL_RESIDENTS_WORK_IN_LIVINGZONE	2.651
COMMUTING_TIME	1.439
POPULATION_DENSITY	1.257
TOTAL_OFFER	1.015

Appendix 2. Tests for the OLS Gauss Markov Assumptions

Tests for the OLS Gauss-Markov Assumption

	Jarque Bera Test	Breusch-Pagan test	Breusch-Godfrey
Test Value	X-squared = 4372.1	BP = 706.49	LM test = 1.0509
Df	2	47	1
p-value	2.2e-16	2.2e-16	0.3053