Universidad de Navarra

Facultad de Ciencias Económicas y Empresariales

Working Paper nº 02/19

Market prices, spatial distribution of consumers and firms' optimal locations in a linear city

Raúl Bajo Buenestado
University of Navarra

Market prices, spatial distribution of consumers and firms' optimal locations in a linear city

Raúl Bajo Buenestado
October 2019

ABSTRACT

We study a game of spatial competition in prices. In particular, we focus on the linear-city duopoly model to see what we can learn about the distribution of consumers, which is not required to be uniform —as in the original Hotelling model. Using variation in firms' prices and costs, we identify points of the distribution of consumers. Based on these points, we estimate the spatial distribution of consumers along the linear city. We apply our methodology to a dataset of prices of two gas stations on a straight highway. By estimating the distribution of consumers, we are able to find the optimal location of an entrant gas station. Using our estimated distribution of consumers and the entrant's optimal point, we simulate welfare gains under counterfactual locations of an entrant.

Raúl Bajo-Buenestado
University of Navarra

Market prices, spatial distribution of consumers and firms' optimal locations in a linear city*

Raúl Bajo-Buenestado†

October 27, 2019

Abstract

We study a game of spatial competition in prices. In particular, we focus on the linear-city duopoly model to see what we can learn about the distribution of consumers, which is not required to be uniform –as in the original Hotelling model. Using variation in firms' prices and costs, we identify points of the distribution of consumers. Based on these points, we estimate the spatial distribution of consumers along the linear city. We apply our methodology to a dataset of prices of two gas stations on a straight highway. By estimating the distribution of consumers, we are able to find the optimal location of an entrant gas station. Using our estimated distribution of consumers and the entrant's optimal point, we simulate welfare gains under counterfactual locations of an entrant.

Keywords: regulated location, linear-city model, distribution of consumers, spatial analysis, spatial price competition.

JEL Classification Numbers: L13, L5, R12, R3, R41

*I acknowledge helpful comments and suggestions by Marcus Berliant, Dodge Cahan, Gilles Duranton, Hülya Eraslan, Nicholas Frazier, Peter Hartley, Jeffrey Lin, Kenneth Medlock, Shaun McRae and Xun Tang, and seminar participants at the X EGSC Conference at Washington University in St. Louis and at the XXX Spanish Economic Association meeting. Thanks also to María García, who kindly helped me with the data. I gratefully acknowledge financial support from the Social Sciences Research Institute at Rice University and Fundación Ramón Areces. The remaining errors are solely mine. The data set and the Matlab and R codes are available upon request from the author.

†Author's contact: Department of Economics, Edificio Amigos, University of Navarra, 31009 Pamplona (Navarra) Spain. E-mail: *rbajo@unav.es*. ORCID #: 0000-0002-6997-2217.

1 Introduction

Market prices are a great source of market information. They condense, in a single parameter, information on market structure, market demand, rivals' costs, etc. In this paper we use this powerful feature of market prices in a specific context: a game of spatial competition. In particular, we focus on the celebrated linear-city model –introduced by Hotelling (1929) and refined by d'Aspremont et al. (1979)– to see what we can learn about the spatial distribution of consumers using observed market prices. We develop a strategy that, with sufficient variation in prices (and costs realizations), allows us to recover points of the distribution of consumers. Using the identified points, we get a very precise estimation of the true distribution. Once we have estimated the distribution of consumers, we show how the methodology can be used to inform policy about the optimal location of entrants in industries in which the locations of firms is regulated. We also propose some counterfactual exercises to study welfare gains/losses of alternative entrant locations.

Perhaps the most implausible assumption in the original model by Hotelling (1929) is the uniform distribution of consumers along the straight interval. In fact, since its original formulation, many authors have studied the linear-city model assuming instead that consumers are non-uniformly distributed –for instance, Neven (1986), Tabuchi and Thisse (1995), Anderson et al. (1997), Shuai (2016). As these (and many other) authors noticed, consumers usually concentrate on certain sections of a straight interval –for instance, around the center of the interval. Following the line of these authors, we deviate from the original Hotelling model by assuming a "more general" (unknown) distribution of consumers along the straight interval.

Assuming that the distribution of consumers falls within the general category of log-concave functions, we characterize the (unique) best-response price functions of the firms in the pricing-stage game with exogenously given firms' locations. Using the equilibrium conditions,[1] we develop our identification strategy to estimate the distribution of consumers. Furthermore, if we restrict such a distribution to fall within the class of beta distributions (with unknown shape parameters), we are also able to estimate the transportation cost (denoted τ). For that purpose, we estimate the shape parameters of the beta distribution via nonlinear least squares. Then, we estimate τ by minimizing the sum of squared distance (errors) between the estimated beta distribution and the identified points.

As an empirical application, we use a dataset of daily retail and wholesale prices of two gas stations located on a straight highway in Spain. The variation in the data allows us to recover the spatial distribution of consumers (drivers) –i.e. road congestion– between two points (two towns) along the highway.[2] Using our estimated distribution of drivers, we find the optimal location of an entrant gas station.

[1] To make our strategy more realistic, as Bajari and Benkard (2005) do in their seminal paper, we assume that firms choose the "right" (equilibrium) prices each period with some (zero-mean) measurement error.

[2] As we explain later, the estimated distribution of consumers along the highway does not reflect a static distribution of consumers' residence locations or addresses. It reveals rather the pattern of traffic distribution in different sections of the highway. That is, the peak of the distribution reflects the section of the highway which is usually congested. We carefully selected a straight section of a highway with entries and exits. Entries and exits along a highway generate different traffic flows and congestion zones in different sections of the highway.

Gasoline is an appropriate good for the empirical study of a model of spatial price competition. The main reason for this is that gasoline has usually been considered to be a homogeneous product –see Tappata (2009), Wang (2009), Chandra and Tappata (2011), Janssen et al. (2011).[3] Furthermore, to avoid differences due to asymmetric quality perceptions across brands, we focus on a long and straight section of a highway served by two gas stations with similar characteristics (same brand, same opening hours, etc.). Therefore, the challenges generated by product differentiation –other than firm location– or unobserved product characteristics are left aside.

To the best of our knowledge, we are not aware of any empirical paper that estimates the distribution of consumers in a linear-city model of spatial competition in prices. Some authors have previously proposed similar strategies to identify demand parameters or consumers' preferences with spatially differentiated goods. For instance, Colwell et al. (2002) develop a theoretical model to show that, in equilibrium, household residence location reveal households' preferences for recreation and amenities (e.g. the beach). Another theoretical exercise is by Moul (2015). Using Salop's circular-city model, he identifies demand parameters and firms' variable costs. However, his strategy builds upon particularly restrictive assumptions. For instance, he assumes that consumers are uniformly distributed around a unit circumference and that there are three symmetrically (and exogenously) located firms around the unit circumference.

Closely related to our paper is the work by Manuszak and Moul (2009), who examine consumers' willingness to travel to avoid gasoline and cigarette excise taxes using data on tax regions near Chicago. As we do, these authors exploit variation in market outcomes –firms' prices in our study, taxes in theirs– to estimate the economic primitives that govern the behavior of consumers and firms. Manuszak and Moul (2009) find that the willingness of a consumer to travel one additional mile to buy gasoline corresponds to a saving of approximately 7 cents per gallon.

Doganoglu (2003) and Laussel et al. (2004) use the Hotelling model to study markets with network effects. In particular, the latter paper (which is an extension of the former) studies the evolution of access prices and firms' market shares for goods with consumption externalities –i.e. congestion effects. Unlike us, they assume that consumers are uniformly distributed along a straight interval to analyze the impact of market shares on access prices. Contrary to them, we are interested in analyzing how changes in observable variables from one period to another can lead us to discern the spatial distribution of consumers.

Besides the novelty of the empirical application of the linear-city model, this paper also contributes to the spatial competition literature by introducing a relatively simple identification strategy, based on firms' best-response functions and on observable firm locations and prices (and their variation from one period to another) to recover the spatial distribution of consumers. This strategy relies on the relatively simple and well-defined structure of the linear-city model of spatial price competition.

This paper not only aims to provide a simple estimation strategy *per se*, but also to provide a methodology that

[3]The fact that gasoline is a homogeneous good is challenged by Lewis (2008). However, his argument relies on the price differences that arise between "high-brand" and "low-brand" retailers in highly dense urban areas considering the number and type of the surrounding competitors.

can be used to inform policy and market analysis. In particular, our framework can be employed in industries with spatially-regulated entry (such as the pharmacy industry or the oil and gas storage industry) to study the optimal location of an entrant.

This is of a particular interest for the case of gas stations. As is the case in many other countries, to open a new gas station on a Spanish highway it is mandatory to acquire an authorization issued and approved by the General Direction of Highways (*Dirección General de Carreteras*), which is a body of the Ministry of Public Works and Transport. The General Direction of Highways studies both the potential location along the highway and the viability of the project (among many other things), and decides whether a new gas station should (or should not) be built on the proposed location.[4] Therefore, using our methodology, we shed some light on the evaluation process of the location of an entrant firm based on consumers' welfare.

The rest of the paper is organized as follows. In Section 2, we introduce the theoretical model of price competition and we discuss some relevant theoretical aspects regarding firms' locations. Section 3 develops our empirical strategy and discusses some empirical challenges. In Section 4 we apply our estimation procedure to data from two gas stations on a straight highway in Spain and we present some counterfactual policy experiments. Section 5 concludes.

2 The model

There are two firms, $i \in \{1, 2\}$, located along the interval $[\underline{x}, \overline{x}]$.[5] We denote firm i's position as x_i and we assume that they are exogenously given and constant over time. Without loss of generality we assume that $x_1 \leq x_2$. Both firms sell a homogeneous good. There is a mass one continuum of consumers distributed over the aforementioned interval according to an a priori unknown strictly stationary distribution with density function $f(\cdot)$. We denote $F(\cdot)$ the cumulative distribution function associated with $f(\cdot)$, and we assume that $F^{-1}(\cdot)$ exists. To guarantee uniqueness of the equilibrium prices, we impose the following assumption.

Assumption 1. *$f(x)$ is log-concave on the interval $[\underline{x}, \overline{x}]$. That is, $\log f(x)$ is concave.*

Note that most of the commonly used distributions satisfy the log-concavity assumption, which is less restrictive than concavity. It also implies that both $F(\cdot)$ and $[1 - F(\cdot)]$ are log-concave, which is a direct consequence of the Prékopa-Borell Theorem.[6]

Both firms and consumers make decisions at $t = 1, 2, ..., \infty$. At time t, each firm faces a marginal cost. Following Thomadsen (2005), we assume that each firm's marginal cost is equal to a common (observable) wholesale cost –

[4] Similar requisites are required in many other countries, such as Australia (see, for instance, the technical note on the "Investigation of Service Station Sites", that describes industry best practice in the assessment of service station sites in NSW) and Italy (*Decreto legislativo 11 febbraio 1998, n. 32*).

[5] For the sake of simplicity, we normalize this interval to [0,1] in the upcoming sections. However, we keep the general form throughout this section.

[6] See Caplin and Nalebuff (1991) for a more in-depth discussion on this Theorem.

denoted θ^t– plus a zero-mean, firm-specific (unobservable) cost –denoted ϵ_i^t. That is, we assume that $c(\theta^t, \epsilon_i^t) = \theta^t + \epsilon_i^t$. These cost parameters are realized at the beginning of each period.[7] Given $c(\cdot, \cdot)$, firms simultaneously choose the price they charge to consumers, i.e. at each period t firms post the price they charge consumers, $p_i^t \in \mathbb{R}_+$.

Likewise, at period t, consumers make consumption decisions. We assume that each consumer buys one unit of the homogeneous good and pays the price charged by the firm from which they buy. Moreover, as is standard in Hotelling's linear-city framework, we assume that consumers incur a quadratic "travel cost" (τ). Therefore, the total cost for a consumer located at $x \in [\underline{x}, \overline{x}]$ of buying the homogeneous product from firm i at time t is given by $p_i^t + \tau(x - x_i)^2$.

A consumer located at $\tilde{x}^t \in [\underline{x}, \overline{x}]$ is indifferent between buying the homogeneous good from firm 1 and buying the homogeneous good from firm 2 at time t if $p_1^t + \tau(\tilde{x}^t - x_1)^2 = p_2^t + \tau(\tilde{x}^t - x_2)^2$. Given x_1 and x_2, the indifferent consumer at t is uniquely determined as follows:

$$\tilde{x}^t(p_1^t, p_2^t) = \frac{p_2^t - p_1^t}{2\tau(x_2 - x_1)} + \frac{x_2 + x_1}{2} \; ; \tag{1}$$

i.e., in period t, firm 1 serves all consumers located to the left of the indifferent consumer and firm 2 serves to all consumers located to the right of $\tilde{x}^t(\cdot, \cdot)$.

We denote firm i's profit as $\pi_i^t(p_i^t, p_{-i}^t; \theta^t, \epsilon_i^t)$, which is characterized by the following expression:

$$\pi_i^t(p_i^t, p_{-i}^t; \theta^t, \epsilon_i^t) = \int_{\tilde{x}^-}^{\tilde{x}^+} \left[p_i^t - c(\theta^t, \epsilon_i^t)\right] f(x) dx \tag{2}$$

where $\tilde{x}^- = \underline{x}$ and $\tilde{x}^+ = \tilde{x}(\cdot, \cdot)$ if $i = 1$, and $\tilde{x}^- = \tilde{x}(\cdot, \cdot)$ and $\tilde{x}^+ = \overline{x}$ if $i = 2$.

We assume that the current decision variable (today's retail price) does not impact the (evolution of the) state variable (future firms' costs, including the wholesale price) or conversely, that the firms' costs does not depend on a firm's retail price in the previous period.[8]

Since each period the firms face an identical problem, we can characterize the problem recursively. Thus, letting $\delta > 0$ be the common discount rate, the lifetime value of firm i is given by the following value function:

$$V_i(\theta, \epsilon_i; p) = \int_{\tilde{x}^-}^{\tilde{x}^+} [p_i - c(\theta, \epsilon_i)] f(x) dx + \delta V_i(\theta', \epsilon_i'; p) \tag{3}$$

[7] In our empirical application, θ is the (common) fuel wholesale price and ϵ captures gas station-specific marginal costs (labor efficiency, management issues, etc.) –see Thomadsen (2005).

[8] In other words, we assume that firms' retail price decisions do not affect next period wholesale prices and that there is no retail price inertia. In our empirical application, the wholesale price for the gas stations is the wholesale price of gasoline. On the contrary, the assumption that the current firm's retail price has an impact on the evolution of the wholesale price may be interpreted in two ways in the context of gasoline retail markets. First, by doing so, we may be assuming that the posted prices of individual gas stations have an impact on wholesale prices and the Brent crude oil price. Second, we may have in mind some kind of inertia in posted prices, potentially explained by some kind of rigidity (such as menu costs). The first explanation is implausible, since no gas station has enough power to determine the evolution of the price of fuel commodities. The second explanation is also implausible if we take into account that gasoline posted prices change every day and that they are posted electronically, so the cost of changing them is virtually zero. However, we are aware that this assumption may not be realistic if we study dominant-market firms (which may influence next period variable costs) and/or firms facing some menu costs (for which there is some price inertia).

where $\tilde{x}^-$ and $\tilde{x}^+$ are defined as above.

The solution concept we use is the (pure strategy) Markov Perfect Equilibrium (MPE). As explained by Maskin and Tirole (2001), the MPE restricts strategies to be a function of current payoff-relevant state variables (in our case, the contemporaneous firms' costs). The optimal strategies are characterized as follows. Let σ_i be a Markov strategy for firm i, (which is a mapping from the set of marginal costs to the set of prices), and fix the rival strategy σ_{-i}. A (Markov) strategy profile, denoted by $\vec{\sigma} \equiv (\sigma_i, \sigma_{-i})$ is a MPE if firm i prefers its strategy σ_i to all alternative strategies $\tilde{\sigma_i}$ for all i, i.e. $V_i(\theta, \epsilon_i; \vec{\sigma}) \geq V_i(\theta, \epsilon_i; \tilde{\sigma_i}, \sigma_{-i}) \ \ \forall i$.

Therefore, an MPE is characterized by a set of two (pricing) policies of firms' costs-contingent actions that form a Nash equilibrium for every realization of the costs at every period t.[9] At an interior solution, firm i's best response (i.e. $p_i^{t,*} > 0$) solves the following expression:

$$\left[\frac{\partial \tilde{x}^+}{\partial p_i^t} f(\tilde{x}^+) - \frac{\partial \tilde{x}^-}{\partial p_i^t} f(\tilde{x}^-)\right] \left[p_i^t - c(\theta^t, \epsilon_i^t)\right] + \left[F(\tilde{x}^+) - F(\tilde{x}^-)\right] = 0 \quad \forall t \tag{4}$$

Since we are considering the two-firm case, we can characterize equation 4 for both firms. The closed-form expressions for both firms are as follows:

$$p_1^{t,*}(\cdot) = 2\tau(x_2 - x_1)\frac{F(\tilde{x}^t)}{f(\tilde{x}^t)} + c(\theta^t, \epsilon_1^t) \tag{5}$$

and

$$p_2^{t,*}(\cdot) = 2\tau(x_2 - x_1)\frac{1 - F(\tilde{x}^t)}{f(\tilde{x}^t)} + c(\theta^t, \epsilon_2^t) \tag{6}$$

that must be satisfied in equilibrium at every t.

Proposition 1. *Under Assumption 1, $p_i^{t,*}(\cdot)$, the equilibrium price is unique for $i \in \{1, 2\}$ for all t.*

Proof. As pointed by Anderson et al. (1997) –whose expression is similar to ours– uniqueness of the price equilibrium follows from Proposition 6 in Caplin and Nalebuff (1991), which is guaranteed by Assumption 1. Strictly stationarity of $F(\cdot)$ (i.e. the cdf does not changes over time) guarantees that the equilibrium is unique at every t. $\square$

Firms' best-response functions (equations 5 and 6) depend on three key parameters. First, they depend on the shape of the consumers' distribution along the interval. Second, on the location of the firms. Third, on the transportation cost for the consumers, denoted by τ. In Figure 1 we illustrate how these parameters are crucial to determine firms' equilibrium retail prices.

[9] In other words, for the reasons explained in the previous footnote, the dynamic game that we present ends up with a static game solution (since there are no "dynamics" from one period to the next one in the model). Still, we present this game as a dynamic one to illustrate that firms solve this game and choose a strategy (the retail price) at $t = 1, 2, 3, \cdots$.

The baseline case is given in subfigure 1a.[10] In subfigure 1b we assume a lower transportation cost relative to the baseline case. As a result, equilibrium prices decrease –in the extreme case in which $\tau = 0$, firms simply reduce their prices to the highest marginal cost (Bertrand competition). Next, in subfigure 1c we consider a slightly left-skewed distribution of consumers relative to the baseline case. Firm 1 has now a market advantage, since it is closer to the point with the highest concentration of consumers. On the other hand, since the bulk of consumers is farther from firm 2, this firm needs to lower its price to attract more consumers. Finally, subfigure 1d contains a location advantage for firm 1 relative to the baseline case. In this case, firm 2's equilibrium price is lower than firm 1's equilibrium price. Notice that it also make sense to see a reduction in the equilibrium prices when both firms are closer to each other –in the extreme case in which both firms are located in the same point (i.e. $x_1 = x_2$), firms simply compete à la Bertrand.

Finally, firms' best-response functions also depend on firms' costs. In our empirical setup, these costs are allowed to change from one period to another. On the other hand, neither the distribution –due to being strictly stationary– nor the locations –exogenously given– nor the transportation costs change from one period to another. Therefore, firms' costs will give us some variation that is key in our estimation procedure.[11] However, the other (constant) parameters also play a prominent role in determining the best response functions at every t (as shown in Figure 1).

2.1 Further theoretical aspects: entrant's optimal regulated location

As explained above, one of the goals of this study is to provide a tool that helps us to determine where the optimal location is of a potential entrant in an industry with regulated location. Therefore, a social planner will decide where to locate the potential entrant firm along the interval in such a way that it maximizes consumers' "spatial" welfare. As Tabuchi (1994) and Shuai (2016) do, we assume that finding the optimal location is equivalent to minimizing consumers' total transportation cost.[12], [13]

If the entrant firm is to be located between the two incumbent firms, then we need to find the consumer $\hat{x}$ such that the sum of consumers' transportation costs to the right and to the left of it are equal. That is,

$$\int_{x_1}^{\hat{x}} \tau(x - x_1) f(x) dx = \int_{\hat{x}}^{x_2} \tau(x - x_2) f(x) dx \tag{7}$$

where $\hat{x}$ is unique if $F(\cdot)$ is strictly monotone in between x_1 and x_2.

Claim 1. *If $F(\cdot)$ is strictly monotone over the interval $[x_1, x_2]$, then $\hat{x}$ is unique.*

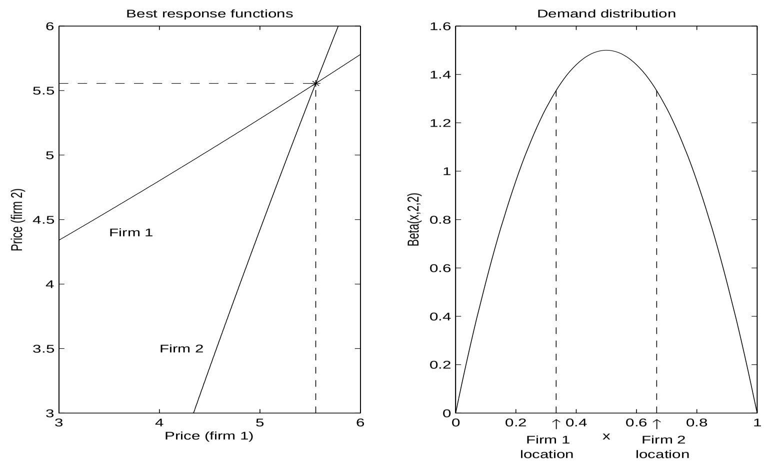
[10] For such a baseline case, we assume a symmetric beta distribution, with shape parameters $\alpha = \beta = 2$, firms' symmetric locations –in particular $x_1 = \frac{1}{3}$ and $x_2 = \frac{2}{3}$–, and transportation cost $\tau = 25$.

[11] A different question concerns what the observable and the non-observable variables are. We deal with this issue in the following section.

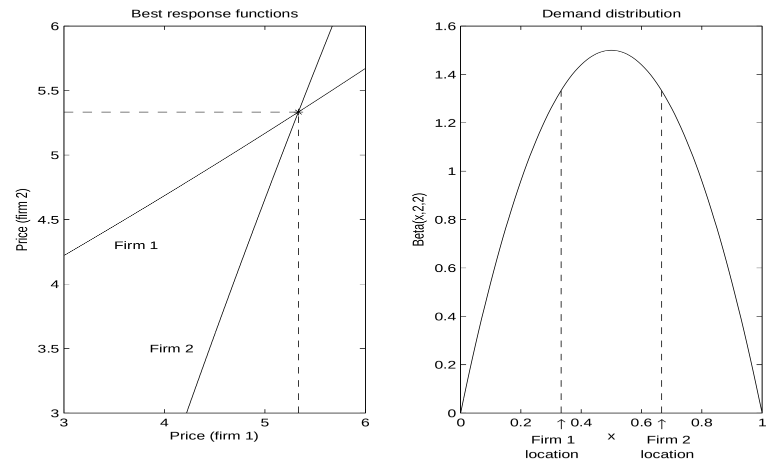
[12] More broadly speaking, Shuai (2016) explains that, if demand is sufficiently inelastic (which is the case of the gasoline demand), maximizing welfare is always equivalent to minimizing consumers transportation cost.

[13] If we think in our empirical application, the goal of the social planner (the General Direction of Highways) is to authorize the construction of a new gas station in such a way that the total probability of vehicles running out of gas is minimized.

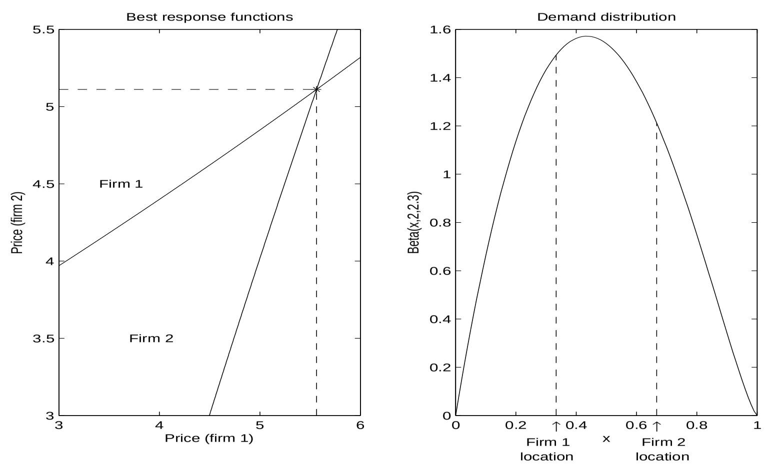
Figure 1: Equilibrium prices for different firms' locations, distributions of consumers and transportation costs



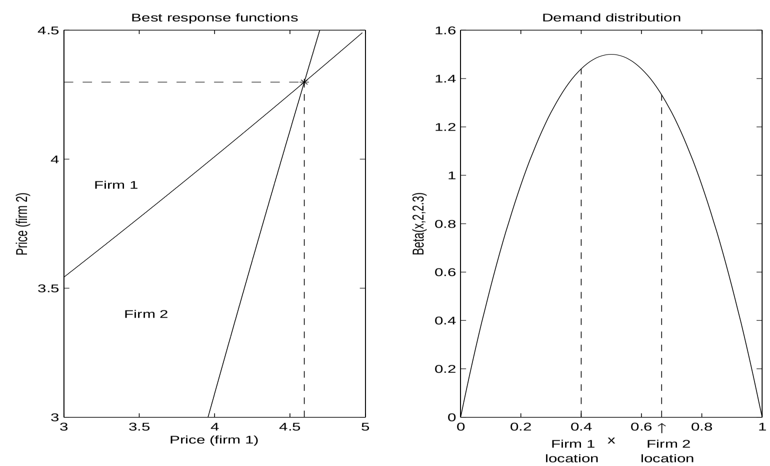
(a) Base case: symmetric distribution and locations ($\tau = 25$)



(b) Symmetric distribution and locations, lower transportation cost ($\tau = 24$)



(c) Demand advantage for firm 1: left skewed distribution, symmetric location ($\tau = 25$)



(d) Location advantage for firm 1, symmetric distribution ($\tau = 25$)

Proof Since the transportation cost exhibits the same quadratic growth for all consumers along the interval, solving $\hat{x}$ in equation 7 is similar to calculating the point $\hat{x}$ in between x_1 and x_2 such that the areas to the left of $\hat{x}$ until x_1 and to the right of $\hat{x}$ until x_2 are equal. That is,

$$\int_{x_1}^{\hat{x}} f(x)dx = \int_{\hat{x}}^{x_2} f(x)dx$$

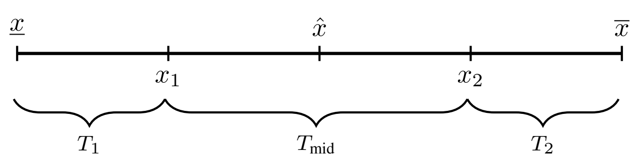
which implies that $\hat{x}$ is determined by

$$F(\hat{x}) = \frac{F(x_2) + F(x_1)}{2} \tag{8}$$

To show that $\hat{x}$ is unique, assume for the sake of contradiction that there exists $\hat{x}'$ and $\hat{x}''$ which satisfy equation 8. Without loss of generality, assume $\hat{x}' > \hat{x}''$. By strict monotonicity of $F(\cdot)$, then $F(\hat{x}') > F(\hat{x}'')$. Thus, $F(\hat{x}') = \frac{F(x_2)+F(x_1)}{2} > F(\hat{x}'')$, which is a contradiction to the fact that both $\hat{x}'$ and $\hat{x}''$ satisfy equation 8. The proof if $\hat{x}' < \hat{x}''$ is similar. □

In order to determine where to locate the potential entrant such that total transportation cost is minimized, we need to compare the total transportation cost that consumers incur for each of the three relevant areas indicated in Figure 2. That is, the total transportation cost for consumers between $\underline{x}$ and x_1, denoted by T_1; for consumers between x_1 and x_2, denoted by T_{mid}; and for consumers between x_2 and $\overline{x}$, denoted by T_2.

Figure 2: Total transportation costs for different sections of the linear-city



Analytically, the total transportation cost that consumers incur in the three areas is given by the sum of the following expressions:

$$T_1 = \int_{\underline{x}}^{x_1} \tau(x - x_1)^2 f(x)dx \tag{9}$$

$$T_{\text{mid}} = \int_{x_1}^{\hat{x}} \tau(x - x_1)^2 f(x)dx = \int_{\hat{x}}^{x_2} \tau(x - x_2)^2 f(x)dx \tag{10}$$

$$T_2 = \int_{x_2}^{\overline{x}} \tau(x - x_2)^2 f(x)dx \tag{11}$$

Bearing these expressions in mind, the spatially-optimal decision rule on entrant's location, denoted by x^*, is given as follows:

a) If $T_1 > \max\{T_{\text{mid}}, T_2\}$, the optimal location of the entrant is at $x^* = \arg\min_{x\in[\underline{x},\overline{x}]} f(x)$ such that $f(x^*) > 0$.[14]

b) If $T_{\text{mid}} > \max\{T_1, T_2\}$, the optimal location of the entrant is at $x^* = \hat{x}$.

c) If $T_2 > \max\{T_1, T_{\text{mid}}\}$, the optimal location of the entrant is at $x^* = \arg\max_{x\in[\underline{x},\overline{x}]} f(x)$ such that $f(x^*) > 0$.[15]

In the borderline cases in which there is a tie, the tie-break rule is just to toss a coin.[16]

We use these decision rules to run a welfare analysis in Section 4 by allowing a potential entrant to be located at the optimal location, in comparison to some other randomly chosen locations.

3 Empirical strategy

In the previous section we obtained a couple of expressions that characterize firms' equilibrium prices at period t. As discussed, these expressions depend on four key elements, namely the shape of the distribution of consumers, firms' locations, consumers' transportation cost, and firms' costs. Of these elements, we assumed that only the latter change from one period to another, while the others (the distribution of consumers, firms' locations, consumers' transportation cost) are assumed to be constant over time.

A different discussion regarding these variables is whether they are observable or unobservable. Our assumptions are as follows. First, we assume that the locations of the firms, which are constant over time and exogenously given, are observable. We can easily obtain the precise location of a firm using GPS technology.[17]

Second, we assume that the common component of the firms' marginal cost at every period t is also observable. Typically, in homogeneous-good markets, finding the (common) wholesale price of homogeneous raw materials is not difficult if we take into account that many of the commonly used raw materials are publicly traded, so there is a commonly known wholesale price every day. However, we assume that the gas station-specific component of the marginal cost is unobservable to the econometrician.

[14] In this case, if $f(\cdot)$ is strictly positive in all its support, the optimal location of the entrant is at $x^* = 0$.

[15] In this case, if $f(\cdot)$ is strictly positive in all its support, the optimal location of the entrant is at $x^* = 1$.

[16] Formally, tie-breaking rules are as follows:

d) (Tie-breaking rule #1) If $T_1 = T_{\text{mid}} > T_2$, the optimal location of the entrant is determined by the following rule: $x^* = \arg\min_{x\in[\underline{x},\overline{x}]} f(x)$ such that $f(x^*) > 0$ with probability $\frac{1}{2}$ and $x^* = \hat{x}$ with probability $\frac{1}{2}$.

e) (Tie-breaking rule #2) If $T_{\text{mid}} = T_2 > T_1$, the optimal location of the entrant is determined by the following rule: $x^* = \hat{x}$ with probability $\frac{1}{2}$ and $x^* = \arg\max_{x\in[\underline{x},\overline{x}]} f(x)$ such that $f(x^*) > 0$ with probability $\frac{1}{2}$.

f) (Tie-breaking rule #3) If $T_2 = T_1 > T_{\text{mid}}$, the optimal location of the entrant is determined by the following rule: $x^* = \arg\max_{x\in[\underline{x},\overline{x}]} f(x)$ such that $f(x^*) > 0$ with probability $\frac{1}{2}$ and $x^* = \arg\max f(x)$ such that $f(x) > 0$ with probability $\frac{1}{2}$.

[17] More realistic is to assume that entry and exit exists in the market instead, but we leave this issue aside. Estimation of entry and exit decisions in a (general) oligopolistic market is by Bajari et al. (2007).

Third, we assume that the distribution of consumers is unobservable. As mentioned in the introduction, many authors have previously assumed that consumers are uniformly distributed along the interval. However, we believe that this assumption is not particularly realistic. Instead, our main task is to estimate this distribution, assuming that it falls in a more general class of distributions.[18]

Finally, we assume that the transportation cost is also unobservable, since this is consumers' private information (however, in order to illustrate our strategy, this parameter is assumed to be known just in Section 3.1, for the sake of expositional clarity). Therefore, the transportation cost parameter constitutes a primitive of our model which we estimate in the upcoming subsections.

3.1 Estimation of the distribution of consumers with known transportation cost

Part of the problem we face is how to recover the (unobservable) distribution of consumers, i.e. $f(\cdot)$. For this purpose, we proceed as follows. First, we solve the equilibrium best-response price functions (equations 5 and 6) at t for $F(\cdot)$. Using the resulting condition, we rely on the existence of variation in the aforemetioned parameters.

After T periods, the changes in both variables will give us a sample of T points in $\mathbb{R}^2$ that correspond to points on the Cartesian plane of the distribution function we are interested in. Thus, using the set of T points on the Cartesian plane, we estimate the underlying distribution using basic econometric techniques.

To justify our identification strategy, two clarifications are necessary at this point. First, we assume that the variation in the firms' costs is exogenous. That is, the wholesale price and the gas station-specific marginal costs are exogenously determined. Second, firms' costs impact upon the retail price, but the neither the retail price nor the other parameters in the model impact upon the firms' costs. In other words, there is no reverse causality.[19]

Reconsider equations 5 and 6, and denote firm i's equilibrium price at t as $p_i^{t,*}$, for $i \in \{1, 2\}$. Solving these equations for $F(\cdot)$ we get:

$$f\left[\frac{p_2^{t,*} - p_1^{t,*}}{2\tau(x_2 - x_1)} + \frac{x_2 + x_1}{2}\right] = \frac{2\tau(x_2 - x_1)}{p_1^{t,*} - \theta^t - \epsilon_1^t + p_2^{t,*} - \theta^t - \epsilon_2^t} \quad \forall t \tag{12}$$

Following the seminal paper by Bajari and Benkard (2005), we assume that both firms choose the equilibrium retail price predicted by the model with some (zero-mean) measurement error.[20] Therefore, the previous condition can be rewritten as follows

$$f\left[\lambda_{\text{obs}}^t + \varepsilon_\lambda\right] = \nu_{\text{obs}}^t + \varepsilon_\nu \tag{13}$$

[18] Some authors have previously analyzed Hotelling's linear-city model assuming a more general class of distribution of consumers. See Neven (1986), Tabuchi and Thisse (1995), Anderson et al. (1997), Torrisi (2011), Shuai (2016).

[19] As we have detailed extensively in footnote 8, in our empirical application we cannot assume that individual gas stations' posted prices have an impact on wholesale prices and Brent crude oil prices: no gas station has enough power to determine the evolution of the price of fuel commodities. However, this assumption may not be realistic if there we study dominant-market firms (which may influence wholesale prices).

[20] These measurement errors could arise due to modeling error, random idiosyncratic facts and/or firm optimization error.

where ε_λ and ε_ν are i.i.d. and, as can be easily shown, have mean zero; and where $\lambda^t_{\text{obs}} \equiv \frac{p_2^{t,*} - p_1^{t,*}}{2\tau(x_2 - x_1)} + \frac{x_2 + x_1}{2}$ and $\nu^t_{\text{obs}} \equiv \frac{2\tau(x_2 - x_1)}{p_1^{t,*} - c(\theta^t) + p_2^{t,*} - c(\theta^t)}$.

Notice that if we assume that the transportation cost parameter (τ) is known, then λ^t_{obs} and ν^t_{obs} are known at every t. Let $\lambda^t \equiv \lambda^t_{\text{obs}} + \varepsilon_\lambda$ and $\nu^t \equiv \nu^t_{\text{obs}} + \varepsilon_\nu$. Following the aforementioned strategy, if we get enough observations, we will have sufficiently distinct points $(\lambda^1, \nu^1), \cdots, (\lambda^T, \nu^T)$, which constitute a sample from (around) the distribution of consumers. Using these points, we can apply the usual quantitative techniques to estimate the true underlying density function.

To illustrate how this procedure works in a specific situation, in Appendix B we estimate the underlying distribution of consumers using simulated data. In particular, we estimate the underlying distribution using a local polynomial regression applied on identified points.

3.2 Estimation of the distribution of consumers with unknown transportation cost

In the previous subsection we assumed that we know the transportation cost (τ) to illustrate how we identify points in the distribution of consumers and how we estimate the true distribution of consumers using such points. However, as we argued above, the assumption that τ is known is fairly unrealistic. The transportation cost parameter is private consumer information, and there is not an easy, nor obvious, way to empirically measure it.[21] To estimate the distribution of consumer in this case, we need to impose some parametric restrictions on the distribution of consumers.

3.2.1 Parametric restriction of the distribution of consumers

Both the RHS and the LHS of equation 12 –which is the key expression in our empirical strategy– depend on the parameter τ, which is unknown. Thus, given T observations of the equilibrium prices and firms' costs, we will have T points of the distribution that depend on the unknown τ, $[\lambda^1(\tau), \nu^1(\tau)], \cdots, [\lambda^T(\tau), \nu^T(\tau)]$. This implies that for each parameter τ, we can identify T different points on the Cartesian plane around the distribution of consumers.

In order to solve this problem and, more generally, to estimate the distribution of consumers with τ unknown, we need to impose some parametric restrictions. In particular, we assume that consumers are distributed according to a beta distribution, with shape parameters (α, β). There are several reasons to use this distribution. First, the beta distribution is log-concave as long as $\alpha, \beta \geq 1$. If so, Assumption 1 is satisfied. Second, the support of the beta distribution is the interval $[0, 1]$. This interval is usually employed in the context of the linear-city model. Third, the shape parameters of the beta distribution can produce a rich variety of shapes, yielding a symmetric, left-skewed or right-skewed distribution (with different high for peaks and length for valleys) by modifying α and β respectively.

[21] Previous authors have proposed different methodologies to estimate transportation cost parameters –see, for instance, Englin and Shonkwiler (1995). However, we need to think that τ (the transportation cost) acts as an "equilibrium parameter" that helps us to fit the model equilibrium into the observed data.

Thus, the beta distribution allows us to consider many different scenarios, including the extreme and unlikely case in which consumers are uniformly distributed.[22]

3.2.2 Estimation of the shape parameters of the distribution

In this subsection we estimate the shape parameters of the distribution of consumers, which is assumed to belong to the class of beta distributions. We denote $\hat{\alpha}(\tau)$ and $\hat{\beta}(\tau)$ the shape parameters estimators that are based on the sample of identified points $[\lambda^1(\tau), \nu^1(\tau)], \cdots, [\lambda^T(\tau), \nu^T(\tau)]$. Since our sample of identified points depends on τ, the estimators of the shape parameters do also depend on τ, as indicated.

We use the identified sample to perform a nonlinear regression given τ. In this case, the model we are interested in can be written as follows:

$$\nu^t(\tau) = f(\lambda^t(\tau), \alpha, \beta) + \varepsilon^t \tag{14}$$

where $f(\cdot, \alpha, \beta)$ is the density function of a beta distribution with shape parameters α and β and ε^t is the error term.

We can obtain an estimator of the shape parameters by performing nonlinear least squares (NLS), by minimizing the sum of squared errors $\sum_{t=0}^{T} \left[\nu^t(\tau) - f\left(\lambda^t(\tau), \alpha, \beta\right)\right]^2$. Given τ, the NLS estimator of the shape parameters, denoted $\hat{\alpha}^{\text{nls}}(\tau)$ and $\hat{\beta}^{\text{nls}}(\tau)$, solve the following system of FOCs

$$-2\sum_{t=0}^{T} \frac{\partial f(\cdot)}{\partial \alpha}\bigg|_{\hat{\alpha}^{nls}(\tau)} \left[\nu^t(\tau) - f\left(\lambda^t(\tau), \hat{\alpha}^{nls}(\tau), \hat{\beta}^{nls}(\tau)\right)\right] = 0 \tag{15}$$

$$-2\sum_{t=0}^{T} \frac{\partial f(\cdot)}{\partial \beta}\bigg|_{\hat{\beta}^{nls}(\tau)} \left[\nu^t(\tau) - f\left(\lambda^t(\tau), \hat{\alpha}^{nls}(\tau), \hat{\beta}^{nls}(\tau)\right)\right] = 0 \tag{16}$$

which can be solved using numerical methods.

3.2.3 Estimation of the transportation cost

In the previous subsection we discuss how to estimate the shape parameters for any transportation cost (τ). The task in this subsection is to select the "appropriate" τ. In other words, we need to estimate the transportation cost in our model.

To do so we rely on the fact that if the real transportation cost (τ) and the estimated one (denoted by $\hat{\tau}$) are close to each other, the identified points and the beta distribution whose shape parameters are estimated via NLS will be close to each other too. Based on this idea, we build an estimator of the transportation by minimizing sum of the square of the discrepancies between identified points and the estimated beta distribution.

[22] This happens when both shape parameters, α and β, are equal to 1.

More formally, and keeping in mind the assumption that the true distribution of consumers follows a beta distribution with shape parameters α and β, if τ is the true transportation cost parameter then:

$$\nu^t(\tau) = f(x^t, \alpha, \beta) \ \ \forall t$$

where x^t is the value in the domain of the beta distribution whose image is $\nu^t(\tau)$ under the beta distribution function.[23]

However, we do not have α and β but an estimator of them (via nonlinear least squares), denoted $\hat{\alpha}^{\text{nls}}(\hat{\tau})$ and $\hat{\beta}^{\text{nls}}(\hat{\tau})$. Therefore:

$$\nu^t(\hat{\tau}) \approx f(x^t, \hat{\alpha}^{\text{nls}}(\hat{\tau}), \hat{\beta}^{\text{nls}}(\hat{\tau})) \ \ \forall t \tag{17}$$

where x^t is as defined above. Thus, we can obtain a precise estimator of the transportation cost by minimizing the sum of the (squared) distance between the LHS and the RHS of expression 17. Notice that the properties that apply to the shape parameter estimators (including consistency) also apply for the transportation cost estimator.

To alleviate the computational burden and, more importantly, to avoid erroneous solutions, we impose $\hat{\tau}$ to belong to a set of feasible solutions, denoted by $\hat{\mathcal{T}}$. The set of feasible solutions is obtained by using both the model assumptions and the identification conditions. In particular, the set of feasible transportation costs is given by:

$$\hat{\mathcal{T}} = \left\{ \hat{\tau} \in \mathbb{R}_{++} \mid \lambda^t(\hat{\tau}) \in [0,1] \ \forall t \ \ \wedge \ \ \nu^t(\hat{\tau}) \geq 0 \ \forall t \ \ \wedge \ \ \hat{\alpha}^{\text{nls}}(\hat{\tau}) \geq 1 \ \ \wedge \ \ \hat{\beta}^{\text{nls}}(\hat{\tau}) \geq 1 \right\}$$

where the first condition requires that the domain of the function must lie between 0 and 1 –which is the domain of the beta distribution; the second condition guarantees the non-negativity of the range of the distribution of consumers; and the third and fourth conditions are imposed to satisfy Assumption 1 (log-concave distribution function).

Therefore, the estimation of the transportation cost ($\hat{\tau}$) is reduced to a local optimization problem, which can be written as follows:

$$\hat{\tau} \equiv \arg\min_{\tau \in \hat{\mathcal{T}}} \sum_{t=0}^{T} \left[\nu^t(\tau) - f(x^t, \hat{\alpha}^{\text{nls}}(\hat{\tau}), \hat{\beta}^{\text{nls}}(\hat{\tau})) \right]^2 \tag{18}$$

If there exist at least one set of identified λ's and ν's that lies in the feasible solutions set and the estimated shape parameters are greater than or equal to 1 for every candidate τ, $\hat{\mathcal{T}}$ is a closed and bounded set.[24] Therefore, the extreme value theorem guarantees the existence of a solution.

To illustrate how this procedure works in a specific situation, in Appendix B we estimate the shape parameters of the distribution of consumers and the transportation cost using simulated data.

[23] If τ is the true transportation cost parameter, then $x^t = \lambda^t(\tau)$.

[24] In particular, using the conditions above, the set is bounded below by $\max\left\{\min\left\{0, \frac{p_1^{t,*} - p_2^{t,*}}{(x_2 - x_1)^2}\right\}, \min\left\{0, \frac{p_2^{t,*} - p_1^{t,*}}{(x_2 - x_1)(2 - x_2 - x_1)}\right\}\right\}$ and the set can be bounded above (if necessary) considering that the sum of the (unique) ν's must be less than or equal to 1 –which is a requirement of the distribution itself–, i.e. $\tau \leq \frac{\sum_t p_1^{t,*} - c(\cdot) + p_2^{t,*} - c(\cdot)}{2(x_2 - x_1)}$ for every unique pair of equilibrium prices and costs.

4 The Spanish retail gasoline market data

In the present section we apply our methodology to a dataset on prices and wholesale costs of two gas stations located along a straight highway in Spain. We organize this section as follows. First, we briefly describe the main characteristics of the retail gasoline market in Spain (subsection 4.1). Then, we present the data and some descriptive statistics (subsection 4.2). Finally, we estimate the distribution of consumers using our dataset and we conduct some counterfactual welfare experiments (subsection 4.3).

4.1 Main features of the market

The Spanish retail gasoline market is heavily concentrated. According to Stolper (2016) just three companies control about 60% of the retail gas stations in Spain. These companies, which are well-recognized among Spanish drivers, also own the nine oil refineries operating in the country, and a majority stake in the national pipeline distribution network. Historical data support that the these firms have been controlling the market for many decades –see Contín-Pilart et al. (2009) and Bello and Contín-Pilart (2012). Thus, even though motor fuels have been usually considered a homogeneous product, there might be some degree of differentiation associated to different brands (firms) –see Lewis (2008). To avoid issues relating to asymmetric quality perceptions of different brands, we focus on two gas stations owned by one of these well-recognized firms with similar characteristics (same amenities, same opening hours, etc.).

As in many European countries, Spanish car buyers have a choice of two engine types: gasoline and diesel fueled-cars. Diesel cars are more expensive than gasoline cars, but have better mileage. These cars are typically chosen by professional drivers (e.g. truck drivers, taxi drivers) and relatively high-kilometer drivers. Gasoline cars are typically cheaper than diesel cars, but have worse mileage. These cars are typically chosen by occasional drivers.[25] According to The European Automobile Manufacturers' Association, the share of diesel cars among new passenger cars in Spain was 62.7% in 2015, while the share of gasoline cars among new passenger cars was 35.1%. Among the overall passenger car fleet in Spain, about 56% of cars were diesel-fueled and 42% were gasoline-fueled in 2014. Given the relevance of both diesel and gasoline, we use price data for both fuels in our experiment.[26]

There are two major markets that set diesel and gasoline wholesale prices in Spain. First, and most importantly, the Genoa market in Italy (also known as Mediterranean or MED market). The reference index in this market for unleaded gasoline is the " Premium Unleaded 10 ppm MED CIF Cargoes Mid" index, while for diesel fuel it is the "ULSD 10 ppm MED CIF Cargoes Mid". Second, the Rotterdam market in the Netherlands (also known as North-Western Europe or NWE market) is also a relevant one. For unleaded gasoline the key index is the "Gasoline 10 ppm NWE CIF ARA", while for diesel fuel the key index is the "ULSD 10 ppm NWE CIF Cargoes Mid". These indices

[25] For further details on diesel cars versus gasoline cars see Verboven (2002).

[26] Hybrid cars, electric cars and cars powered with alternative fuels are still residuals in Spain. According to The European Automobile Manufacturers' Association, just 2% among new passenger cars in 2015 were hybrid or electric, and less than 2.2% of the cars among the passenger car fleet in Spain in 2014 were powered by alternative fuels.

are well-recognized among professional in the sector and are posted on a daily basis in *Platts*, the major provider of energy and commodities information.

Gas stations adjust the retail price of unleaded gasoline and diesel fuel according to the price indices in the Genoa and the Rotterdam markets. The retail gas stations prices are posted at *Geoportal*, a website of the Spanish Ministry of Industry. Following a government mandate (Spanish Act ITC/2308/2007), all gas stations are required to send in their fuel prices to the Ministry of Industry whenever they change, and weekly regardless whether changes are made or not. The Ministry of Industry makes all gas stations' prices publicly available every day[27].

4.2 Data and descriptive statistics

To empirically test our procedure, we consider data from two gas stations located on a 35.1 miles (56.5 km) section of a straight highway in Catalonia (Spain). We tracked and recorded the fuel prices of these two gas stations from October 2, 2014 to July 20, 2015.[28] Our dataset includes daily prices for the most commonly used fuels in Spain for transportation purposes, that is unleaded gasoline and diesel fuel. These prices were obtained every day from the aforementioned *Geoportal* website.

The obtained prices are tax-inclusive. To get the tax-exclusive fuel prices, we subtract the taxes.[29] In particular, we subtract VAT –which is determined at the national level– and the excise fuel taxes –which are determined both at national and regional level. The state tranche of the excise duty is 0.42469€/l for unleaded gasoline and 0.331€/l for diesel fuel. The regional tranche is 0.048€/l for both fuels. The VAT base is the retail (tax-exclusive) price plus the excise duties. The VAT rate is 21%.

Data on wholesale costs for the same period (from October 2, 2014 to July 20, 2015) was obtained from *Platts*. Using data from both the Genoa market and the Rotterdam market, we build a representative index of the wholesale cost of the fuels for the gas stations. According to the information provided by the Spanish Association of Operators of Oil Products (AOP)[30] and the National Competition Commission (CNMC), the representative wholesale cost of unleaded gasoline for a retailer is calculated as follows[31]

$$\text{Wholesale Cost Unleaded Gasoline} = 70\% * \text{“Premium Unleaded 10 ppm MED CIF Cargoes Mid” Genoa index} +$$
$$+\ 30\% * \text{“Gasoline 10 ppm NWE CIF ARA” Rotterdam index}$$

[27] The data is available at http://geoportalgasolineras.es/#/Inicio [last access: October 27, 2019].

[28] Data is not available for October 12, 2014, which is the Spanish National day.

[29] Marion and Muehlegger (2011) show that gasoline taxes are indeed fully passed onto consumers and are incorporated fully into the tax-inclusive price, under typical supply and demand conditions.

[30] Cepsa, the owner of the gas stations considered in this study, is a partner of this organization.

[31] This formula is used by the Spanish Association of Operators of Oil Products (AOP) to calculate the wholesale cost faced by gas stations in Spain. The formula is included in the anual AOP reports –see, for instance, Asociación Española de Operadores de Productos Petrolíferos (2015). Likewise, the Spanish National Competition Commission (CNC) uses the same formula to calculate gas stations' markups. See, for instance, the 2015 special report on the Spanish retail petroleum products market –Comisión Nacional de los Mercados y la Competencia (2015). A similar weighted average of the relative importance of Genoa and Rotterdam prices (66.1% the former, 33.9% the latter) was employed by Rodrigues (2009) to study asymmetries in the adjustment of pump prices for the Spanish case.

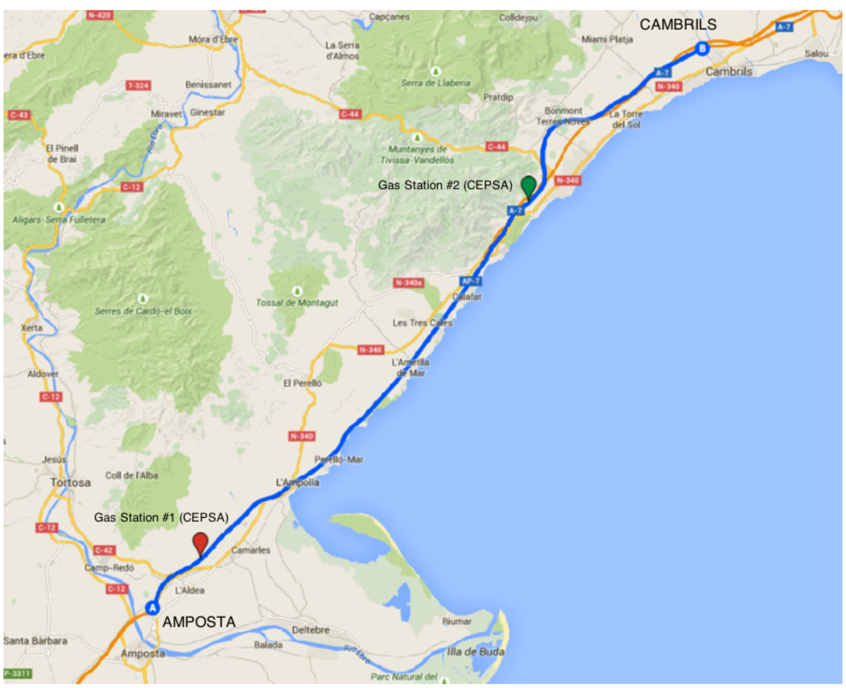
Likewise, the wholesale cost of the diesel fuel for a retailer is equal to

$$\text{Wholesale Cost Diesel Fuel} = 70\% * \text{“ULSD 10 ppm MED CIF Cargoes Mid” Genoa index} +$$
$$+\ 30\% * \text{“ULSD 10 ppm NWE CIF Cargoes Mid” Rotterdam index}$$

The parameters –Genoa and Rotterdam indices– obtained from *Platts* are in \$/metric ton. To convert from metric tons to liters, we use the reference density of unleaded gasoline (around 0.745 kg/l), and regular diesel fuel for vehicles (around 0.850 kg/l). We get a rate of 1315.78 l/metric ton and 1183.43 l/metric ton respectively. To convert from \$ to €, we use the daily exchange rate from October 2, 2014 to July 20, 2015. This data can be obtained from the Federal Reserve Bank of St. Louis and is publicly available. Finally, we also include the distribution costs as part of the gas stations' relevant costs. According to the AOP, the distribution costs are 0.12€/l for unleaded gasoline and 0.11€/l for diesel fuel.[32] A wholesaler markup is not considered, since these gas stations belong to a vertically integrated firm.

The gas stations that we consider are located on a 35.1 mile (56.5 km) straight section in the AP-7 Toll Highway in the province of Tarragona, in Catalonia (Spain). This straight section of the highway is located between the towns of Amposta and Cambrils. Figure 3 shows this section of the road and indicates the position of the two gas stations. Point A indicates the former town, while point B indicates the latter.

Figure 3: Map of the section of the AP-7 Toll Highway



Gas station #1 is located 3.1 miles away from Amposta, close to the village of L'Aldea. Between Amposta and gas station #1 the highway provides access to roads N-235 and N-340, which connect several towns along the highway,

[32] In Appendix C we include a sensitivity analysis of our main estimations for different distribution costs parameters.

such as L'Aldea. Gas station #2 is located 24.11 miles away from Amposta, close to the city of L'Hospitalet de l'Infant. Between gas station #1 and gas station #2 the highway has a few exits to several towns, such as Calafat, Les Tres Cales, and another exit to the National Park of La Rojala-Platja del Torn. Finally, from gas station #2 to Cambrils, the highway provides access to roads C-44 and T-312. These roads connect popular beaches and holiday resorts in Salou and Cambrils, and a well-known theme park called *Port Aventura*.

We normalize the interval defined by the section of the highway in between Amposta and Cambrils to $[0, 1]$. Following such normalization, gas station #1 is located at point $x_1 = 0.09$ and gas station #2 is located at point $x_2 = 0.68$. Both gas stations are operated by the same major brand (Cepsa), are open 24/7 and have a restaurant and a convenience store.

Summary statistics for the relevant variables from October 2, 2014 to July, 20 2015 are included in Table 1. All the variables are expressed in euros per liter. As we can see, all the variables present substantial variation. In fact, during the last quarter of 2014 oil prices worldwide experiment a huge decrease. This scenario is convenient for our analysis since this pattern was also observed in the Genoa and Rotterdam prices, creating variation in wholesale costs and gas stations' posted prices.

Table 1: Summary statistics

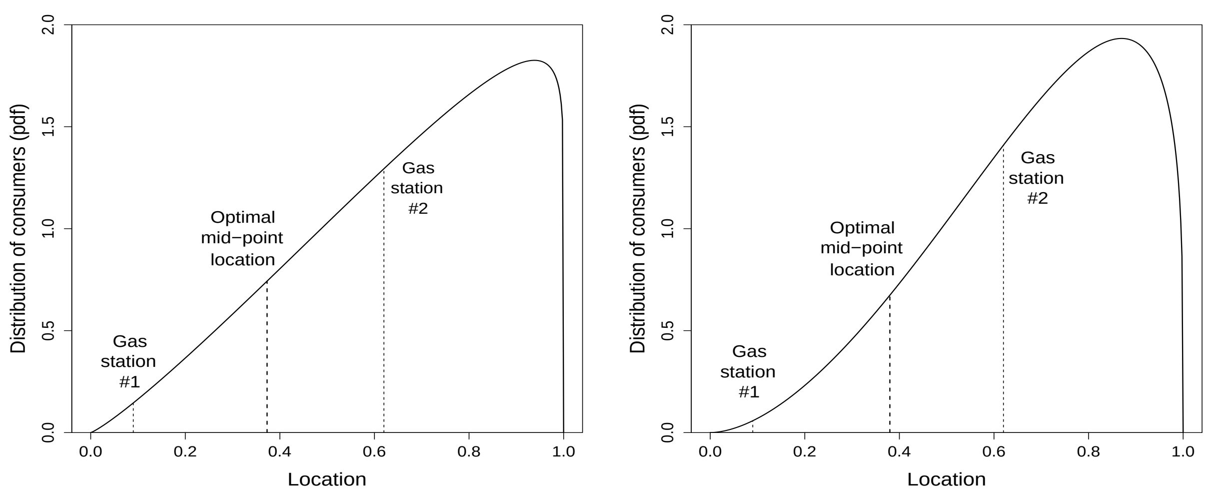
Variable	Mean	Std. Dev.	Min.	Max.
Prices, tax-inclusive (€/l)				
Unleaded gasoline, gas station #1	1.3242	0.0745	1.1690	1.4690
Unleaded gasoline, gas station #2	1.3283	0.0764	1.1690	1.4690
Diesel fuel, gas station #1	1.2374	0.0611	1.1090	1.3790
Diesel fuel, gas station #2	1.2388	0.0606	1.1090	1.3790
Prices, tax-exclusive (€/l)				
Unleaded gasoline, gas station #1	0.6217	0.0616	0.4934	0.7414
Unleaded gasoline, gas station #2	0.6251	0.0632	0.4934	0.7414
Diesel fuel, gas station #1	0.6437	0.0505	0.5375	0.7607
Diesel fuel, gas station #2	0.6448	0.0501	0.5375	0.7607
Costs data (€/l)				
Unleaded gasoline	0.4268	0.0596	0.2952	0.5354
Diesel fuel	0.4389	0.0486	0.3324	0.5462
Observations per variable				291

4.3 Main results and counterfactuals

In this subsection we use the estimation procedure described in the Section 3 to estimate the distribution of consumers using the AP-7 Toll Highway dataset. Based on this estimation procedure, we calculate the optimal location of an entrant gas station.

The main results of the estimation analysis are included in Figure 4. Subfigure 4a contains the estimated distribution of consumers of unleaded gasoline, while subfigure 4b contains the estimated distribution of consumers of diesel fuel. The distribution of consumers is left-skewed (negative skewness) for both unleaded gasoline consumers and diesel consumers, suggesting that the usually congested area in the section of the AP-7 Toll Highway is between gas station #2 and the town of Cambrils. In fact, this section of the highway between gas station #2 and Cambrils is close to very popular tourist attractions and landmarks, such as beach resorts, and the aforementioned theme park and national park. Therefore, we should expect more traffic in the section of the highway that is closer to Cambrils (town B) than in the section of the highway that is closer to Amposta (town A).

Figure 4: Estimated distribution of consumers



(a) Unleaded gasoline

(b) Diesel fuel

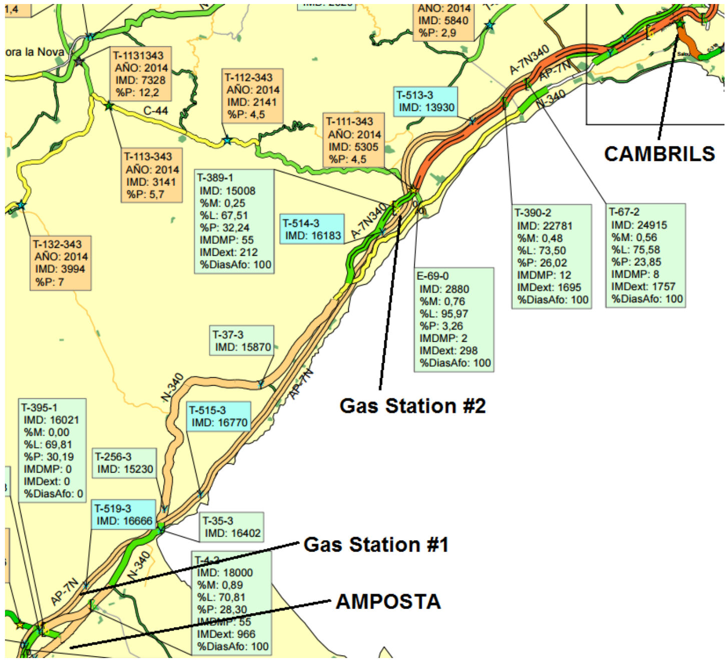
Figure 5 contains the congestion map of the section of the AP-7 Toll Highway that we consider in our study. The congestion map is based on historical data from 2014, and was obtained from the Spanish Ministry of Public Works and Transport.[33] Consistently with the estimation, the section that is usually congested within the section of the AP-7 Toll Highway that we study (indicated in a darker color) coincides with the peak of our estimated distributions in Figure 4.

Next, based on the previous estimation, we find the optimal location for an entrant gas station. For both unleaded gasoline and diesel fuel, we find that the optimal location point of an entrant is at $x = 1$, i.e. close to Cambrils. Thus, our model suggests that a new gas station serving the popular tourist areas is the optimal location. If a new gas station is to be located between gas station #1 and gas station #2,[34] our estimation using both types of fuels suggests that the optimal (mid-point) location is at point $x = 0.38$. This point is located 13.4 miles from Amposta (town A) and just close to the "Estany podrit" beach and the Ametlla camping. As expected, this point is slightly closer to gas station #2 than to gas station #1. The optimal location and the optimal mid-point location of an entrant for the AP-7 Toll

[33] Additional congestion maps are included in Appendix D. See Dirección General de Carreteras (2015) for further information and maps.

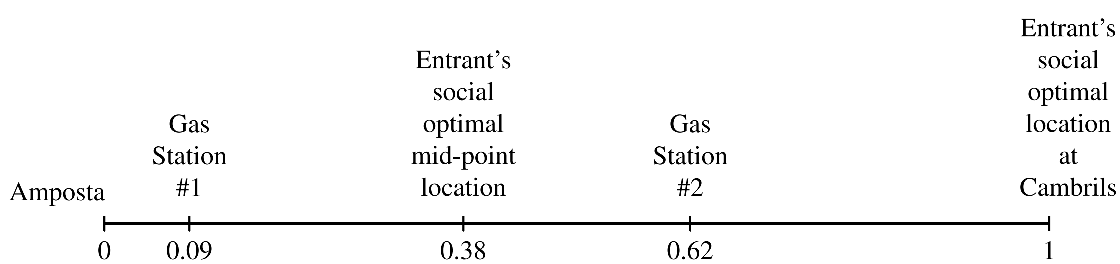
[34] We call a point between gas station #1 and gas station #2 a mid-point location.

Figure 5: Traffic congestion map in the relevant section of the AP-7 Toll Highway



Highway experiment are included in Figure 6.

Figure 6: Incumbents' locations and entrant's optimal location



Finally, using the estimated distribution of consumers and transportation costs, we study the welfare gains from locating an entrant gas station in the estimated optimal locations in comparison to random locations, chosen on the basis of geographical distances, i.e. without taking into account the underlying distribution of consumers.[35] The main results of this welfare counterfactual exercise are included in Table 2.[36]

There are welfare gains from locating an entrant gas station in the optimal location in comparison to the three proposed geographical mid-point locations. As expected, these welfare gains become lower as the alternative loca-

[35] We have considered the following as "locations based on geographical distances" for an entrant firm. First, the geographical mid-point between town A and gas station #1. Second, the geographical mid-point between gas station #1 and gas station #2. Third, the geographical mid-point between gas station #2 and town B.

[36] The results included in this table are preliminary.

Table 2: Percent of welfare gains from optimal location of gas stations vs. geographical mid-point locations

	Socially optimal location (x=1)	Socially optimal mid-point location (x=0.38)
Mid-point between town A & gas station #1 (x=0.045)	1.48%	0.20%
Mid-point between both gas stations (x=0.355)	1.14%	0.03%
Mid-point between gas station #2 & town B (x=0.810)	0.05%	-0.49%

tions get closer to the optimal one. That is, locating the entrant firm at the optimal location versus locating it at the geographical mid-point between town A and gas station #1 increases welfare by 1.48%. However, locating the entrant firm at the optimal point in comparison to the geographical mid-point between gas station #2 and town B increases welfare just by 0.05%. The reason is that the latter point is closer to the optimal point than the former one.

Column 2 in Table 2 includes welfare losses from locating an entrant gas station in the optimal point between both gas stations on both highways (the socially optimal mid-point location). The results suggests that consumer welfare would be enhanced by locating an entrant at $x = 0.81$ –i.e. between gas station #2 and town B– rather than locating it at point $x = 0.38$ –the optimal mid-point location. Again, the reason is that $x = 0.81$ is closer than $x = 0.38$ to the absolute socially optimal location, which is $x = 1$.

5 Conclusions

This paper proposes a novel procedure to recover the spatial distribution of consumers in an oligopolistic market serving horizontally differentiated products –i.e. a Hotelling linear-city model– with exogenously given firm locations. The procedure relies on a relatively simple identification strategy, from the equilibrium first order conditions. With sufficient variation in equilibrium retail and wholesale prices from one period to another, and after imposing some parametric restrictions, our methodology also allows us also to recover a precise approximation of the distribution of consumers.

This methodology is applied to data on prices and costs of two gas stations located along a straight highway in Spain. By doing so, we recover the underlying distribution of consumers between two points (towns) within this highway. Using the estimated distribution, we are able to indicate precisely where a potential entrant should be located in order to maximize consumer welfare. As is true in many other countries, opening a gas station in a Spanish highway requires an authorization issued by the Ministry of Public Works and Transport concerning the convenience of the potential location. Hence, our methodology aims to shed some light on this opaque bureaucratic process.

Besides the highway application, this methodology also has numerous potential applications. For instance, we can use it to determine the optimal position of oil and gas storage facilities along straight pipelines. In addition, we can use it to calculate the optimal location of power generation facilities throughout regions within a regional transmission organization (RTO).

Returning to the gas stations case, possible extensions should take into account the potential existence of (illegal) collusive agreements. This fact is pointed out for the case of gas stations by Borenstein and Shepard (1996). Moreover, further work should consider price inertia or rigidities in prices and costs. Although these elements do not apply to the retail gasoline market, they may play a prominent role in other markets in which we observe spatial competition or horizontal differentiation.

References

Anderson, S. P., J. K. Goeree, and R. Ramer (1997). Location, location, location. *Journal of Economic Theory 77*(1), 102–127.

Asociación Española de Operadores de Productos Petrolíferos (2015). Memoria 2014. Technical report, AOP.

Bajari, P. and C. L. Benkard (2005). Demand estimation with heterogeneous consumers and unobserved product characteristics: A hedonic approach. *Journal of Political Economy 113*(6).

Bajari, P., C. L. Benkard, and J. Levin (2007). Estimating dynamic models of imperfect competition. *Econometrica 75*(5), 1331–1370.

Bello, A. and I. Contín-Pilart (2012). Taxes, cost and demand shifters as determinants in the regional gasoline price formation process: Evidence from Spain. *Energy Policy 48*, 439–448.

Borenstein, S. and A. Shepard (1996). Dynamic pricing in retail gasoline markets. *The RAND Journal of Economics 27*(3), 429–451.

Caplin, A. and B. Nalebuff (1991). Aggregation and imperfect competition: On the existence of equilibrium. *Econometrica: Journal of the Econometric Society*, 25–59.

Chandra, A. and M. Tappata (2011). Consumer search and dynamic price dispersion: an application to gasoline markets. *The RAND Journal of Economics 42*(4), 681–704.

Colwell, P. F., C. A. Dehring, and G. K. Turnbull (2002). Recreation demand and residential location. *Journal of urban economics 51*(3), 418–428.

Comisión Nacional de los Mercados y la Competencia (2015). Estudio sobre el mercado mayorista de carburantes de automoción en España. Technical Report E/CNMC/002/15, CNMC.

Contín-Pilart, I., A. F. Correljé, and M. B. Palacios (2009). Competition, regulation, and pricing behaviour in the Spanish retail gasoline market. *Energy Policy 37*(1), 219–228.

d'Aspremont, C., J. J. Gabszewicz, and J.-F. Thisse (1979). On Hotelling's "stability in competition". *Econometrica: Journal of the Econometric Society*, 1145–1150.

Dirección General de Carreteras (2015). Mapa de Tráfico 2015. Technical Report NIPO 161-15-052-3, Centro de Publicaciones, Ministerio de Fomento (España).

Doganoglu, T. (2003). Dynamic price competition with consumption externalities. *netnomics 5*(1), 43–69.

Englin, J. and J. S. Shonkwiler (1995). Modeling recreation demand in the presence of unobservable travel costs: toward a travel price model. *Journal of environmental economics and management 29*(3), 368–377.

Fackler, P. L. and M. J. Miranda (2004). *Applied computational economics and finance*. MIT press.

Fan, J. and I. Gijbels (1995). Adaptive order polynomial fitting: Bandwidth robustification and bias reduction. *Journal of Computational and Graphical Statistics*, 213–227.

Hotelling, H. (1929). Stability in competition. *The Economic Journal 39*(153), 41–57.

Janssen, M., P. Pichler, and S. Weidenholzer (2011). Oligopolistic markets with sequential search and production cost uncertainty. *The RAND Journal of Economics 42*(3), 444–470.

Laussel, D., M. de Montmarin, and N. Van Long (2004). Dynamic duopoly with congestion effects. *International Journal of Industrial Organization 22*(5), 655–677.

Lewis, M. (2008). Price dispersion and competition with differentiated sellers. *The Journal of Industrial Economics 56*(3), 654–678.

Manuszak, M. D. and C. C. Moul (2009). How far for a buck? tax differences and the location of retail gasoline activity in southeast chicagoland. *The Review of Economics and Statistics 91*(4), 744–765.

Marion, J. and E. Muehlegger (2011). Fuel tax incidence and supply conditions. *Journal of Public Economics 95*(9), 1202–1212.

Maskin, E. and J. Tirole (2001). Markov perfect equilibrium: I. observable actions. *Journal of Economic Theory 100*(2), 191–219.

Moul, C. C. (2015). Estimating demand for spatially differentiated firms with unobserved quantities and limited price data. *Economics Letters 131*, 50–53.

Neven, D. J. (1986). On Hotelling's competition with non-uniform customer distributions. *Economics Letters 21*(2), 121–126.

Rodrigues, J. (2009). Asymmetries in the adjustment of motor diesel and gasoline pump prices in Europe. Technical report.

Shuai, J. (2016). A comment on mixed oligopoly spatial model: the non-uniform consumer distribution. *Economic Theory Bulletin*, 1–7.

Stolper, S. (2016). Who bears the burden of energy taxes? the role of local pass-through. *Discussion Paper 16-70, Harvard Kennedy School*.

Tabuchi, T. (1994). Two-stage two-dimensional spatial competition between two firms. *Regional Science and Urban Economics 24*(2), 207–227.

Tabuchi, T. and J.-F. Thisse (1995). Asymmetric equilibria in spatial competition. *International Journal of Industrial Organization 13*(2), 213–227.

Tappata, M. (2009). Rockets and feathers: Understanding asymmetric pricing. *The RAND Journal of Economics 40*(4), 673–687.

Thomadsen, R. (2005). The effect of ownership structure on prices in geographically differentiated industries. *RAND Journal of Economics 36*(4), 908–929.

Torrisi, G. (2011). The model of the linear city under triangular distribution of consumers. *Management Research and Practice 3*(2), 7–23.

Verboven, F. (2002). Quality-based price discrimination and tax incidence: evidence from gasoline and diesel cars. *RAND Journal of Economics*, 275–297.

Wang, Z. (2009). (Mixed) strategy in oligopoly pricing: evidence from gasoline price cycles before and under a timing regulation. *Journal of Political Economy 117*(6), 987–1030.

Online Appendix A: Methodology to obtain of the equilibrium prices in Figure 1

To obtain firms' best responses, and taking into account that most of the commonly used distributions –including the beta distribution– do not have a close form representation of their cdf, equations 5 and 6 are not usually tractable. Therefore, we use the collocation method proposed in Fackler and Miranda (2004) to approximate the best-response function for each firm given the other firm's best-response function.

For that purpose, the best response function, i.e. $p_i^*(\cdot)$, is approximated using a linear combination of n basis functions such that:

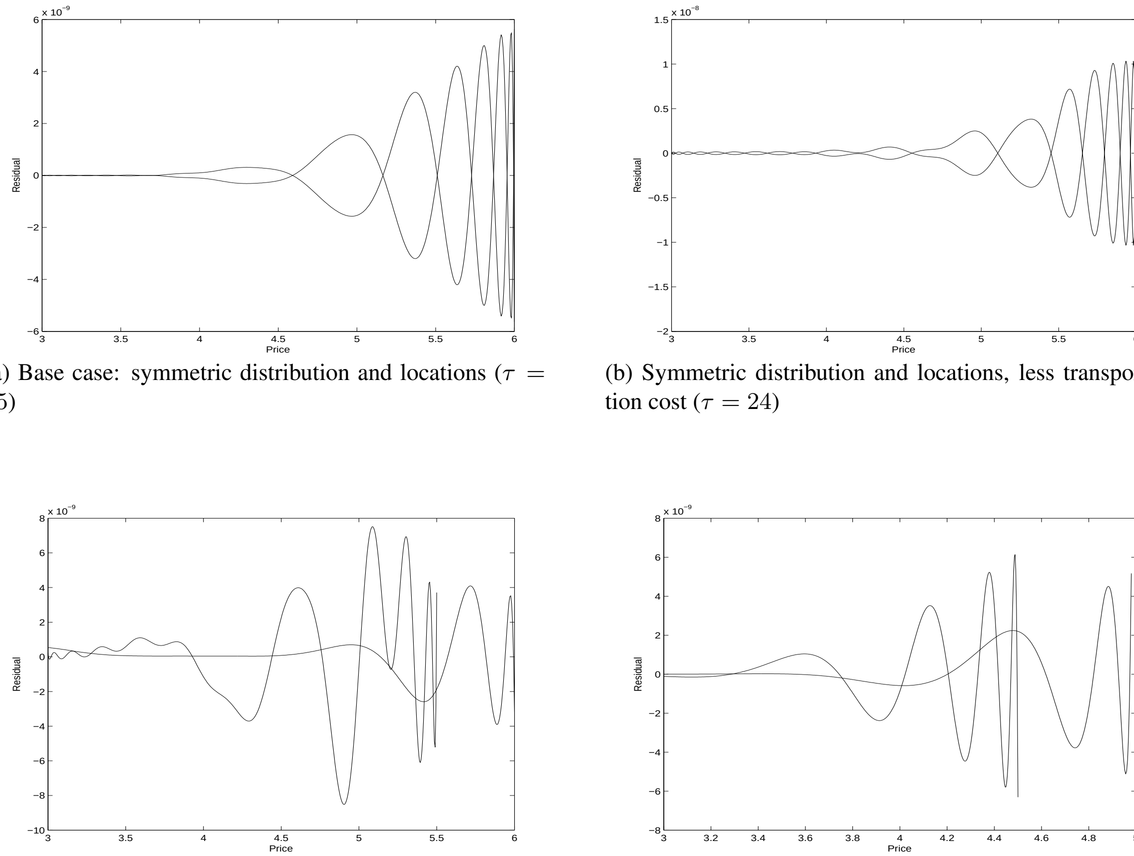
$$\hat{p}_i^*(\cdot) = \sum_{j=1}^{n} c_j \phi_j(p_{-i}) \tag{A.1}$$

for the n fixed coefficients n_j (called collocation nodes) at which the approximant is imposed to satisfy the functional form. Thus, by defining function $g : \mathbb{R}^2 \to \mathbb{R}$, such that

$$g\left(p_{-i}, \sum_{j=1}^{n} c_j \phi_j(p_{-i})\right) = 0 \qquad \text{for } \; j \in \{1, 2, \cdots, n\} \tag{A.2}$$

we can reduce the previous problem to a system of n non-linear equations with n unknowns (c_j) that can be solved using standard rootfinding techniques, such as Boydren's method or Newton's method. In our case, we use the former. Moreover, to approximate the best response functions, we use a 20-degree Chebychev polynomial as approximant (basis function). The results for different beta shape parameters and locations are displayed in Figure 1. The residuals of these approximations are always smaller than $1 * 10^{-8}$. Residual plots for each scenario are included in Figure A.1.

Figure A.1: Residual functions for Figure 1



(a) Base case: symmetric distribution and locations ($\tau = 25$)

(b) Symmetric distribution and locations, less transportation cost ($\tau = 24$)

(c) Demand advantage for firm 1: left skewed distribution, symmetric location ($\tau = 25$)

(d) Location advantage for firm 1, symmetric distribution ($\tau = 25$)

Online Appendix B: Simulation experiments

We conduct a simulation study to illustrate how we identify T points of the distribution and how we estimate the density function (the distribution of consumers) using these points. For that purpose, we consider the following parameters. First, firms' marginal costs at each period are randomly drawn from a uniform distribution between 0.7 and 1.2. That is, $c(\theta^t) \sim \mathcal{U}[0.7, 1.2]$ for all t. Second, firms' locations are $x_1 = \frac{1}{3}$ and $x_2 = \frac{2}{3}$ respectively, with the interval normalized to $[0, 1]$. We draw $2,000$ costs parameters to obtain another $2,000$ firms' equilibrium prices for each pair of variable costs drawn. Considering that $\lambda^t_{\text{obs}} \equiv \frac{p_2^{t,*} - p_1^{t,*}}{2\tau(x_2 - x_1)} + \frac{x_2 + x_1}{2}$ and $\nu^t_{\text{obs}} \equiv \frac{2\tau(x_2 - x_1)}{p_1^{t,*} - c(\theta^t) + p_2^{t,*} - c(\theta^t)}$, we obtain 2,000 points in the domain of $f(\cdot)$, λ^t_{obs}, and 2,000 points in the range of $f(\cdot)$, ν^t_{obs}.

Next, we introduce the disturbance parameters to capture measurement errors. First, we assume a disturbance parameter in λ^t_{obs}, denoted $\varepsilon_\lambda \sim N(0, \sigma_\lambda)$ i.i.d., where $\sigma_\lambda = \text{var}\,(\lambda^t_{\text{obs}})$. Second, we assume a disturbance parameter in ν^t_{obs}, denoted $\varepsilon_\nu \sim N(0, \sigma_\nu)$ i.i.d., where $\sigma_\nu = \text{var}\,(\nu^t_{\text{obs}})$. We denote $\lambda^t \equiv \lambda^t_{\text{obs}} + \varepsilon_\lambda$ and $\nu^t \equiv \nu^t_{\text{obs}} + \varepsilon_\nu$. Call $(\lambda_1, \theta_1), \cdots, (\lambda_T, \theta_T)$ the available sample of points. To get an estimation of the density function based on this sample of identified points, we use local polynomial regression between $[0, 1]$. We use both a second order polynomial approximation and a third order polynomial approximation –see Fan and Gijbels (1995).

B.1 Estimation of the distribution of consumers with known transportation cost

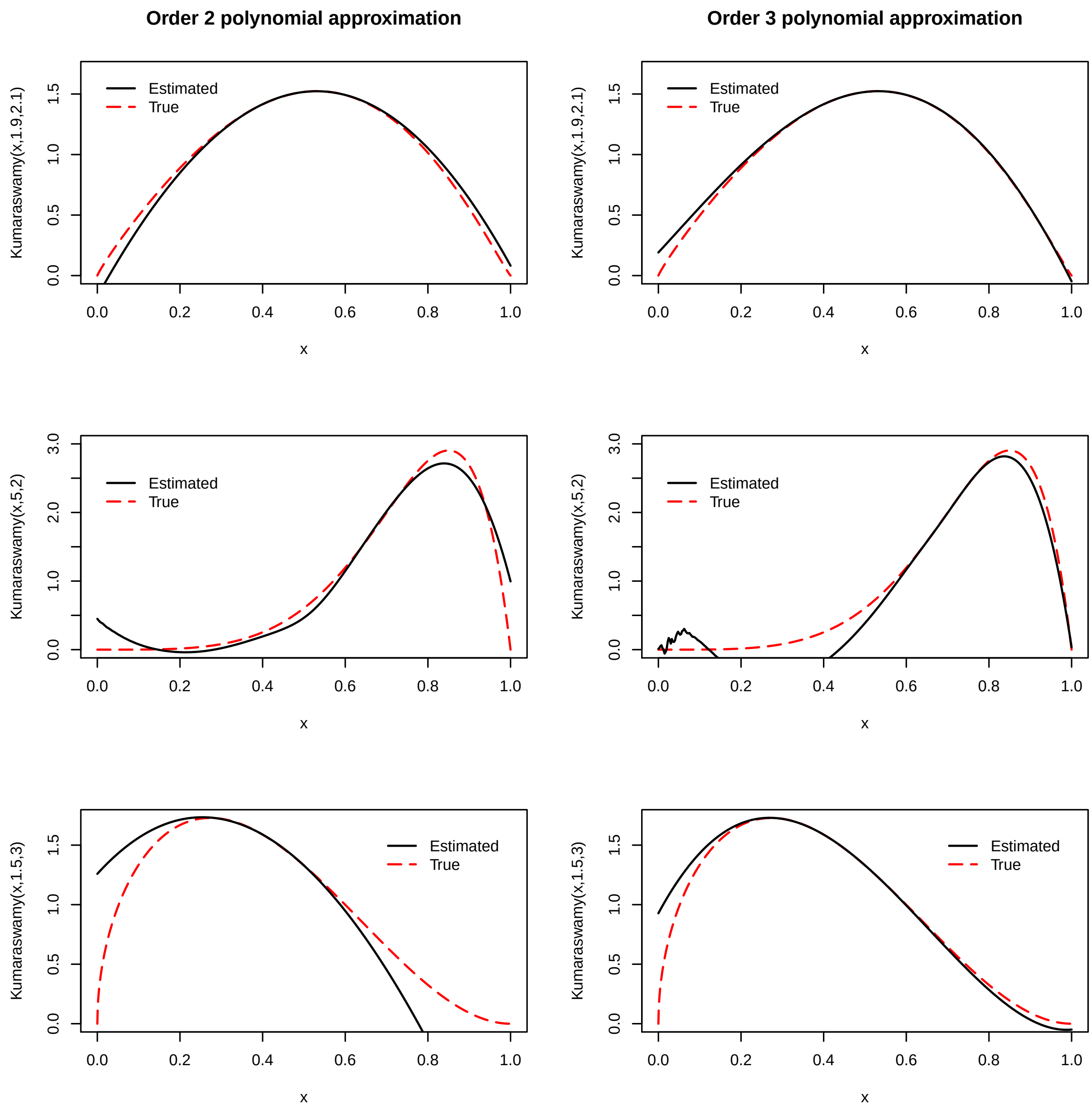
First, to illustrate the procedure described in Section 3.1, we assume that the consumer's transportation cost is known. In particular, we assume that the (known) consumers' transportation cost is $\tau = 1.5$. To show how our strategy performs for different distributions of consumers, we consider several scenarios. First, we assume that the true distribution of consumers over the interval $[0, 1]$ follows a Kumaraswamy distribution, with shape parameters $a > 1$ and $b > 1$. Second, we assume that the true distribution of consumers over the interval $[0, 1]$ follows a Beta distribution, with shape parameters $\alpha > 1$ and $\beta > 1$.

The results for the simulations with the Kumaraswamy distributions are included in Figure B.1. We consider a Kumaraswamy distribution with shape parameters $a = 2$, $b = 2$ (almost symmetric), which is plotted at the top; a Kumaraswamy distribution with shape parameters $a = 5$, $b = 2$ (negatively-skewed), which is plotted in the middle; and a Kumaraswamy distribution with shape parameters $a = 1.5$, $b = 3$ (positively-skewed), which is plotted at the bottom.

In the three cases, the methodology works reasonably well at estimating the underlying (true distribution). Despite the usual "bias-at-the-boundary" problem, our methodology accurately captures the peak of the true distribution in all cases.

Next, we assume that the true distribution of consumers over the interval $[0, 1]$ follows a Beta distribution. Again, we consider different shape parameters. The results for the simulations with the Beta distributions are included in Figure B.2. We consider a symmetric Beta distribution with shape parameters $\alpha = 2$, $\beta = 2$ (symmetric), which is plotted at the top; a Beta distribution with shape parameters $\alpha = 2.5$, $\beta = 5.5$ (positively-skewed), which is plotted in the middle; and a Beta distribution with shape parameters $\alpha = 4.5$, $\beta = 3$ (negatively-skewed), which is plotted at the bottom. Again, in the three cases considered, the methodology works reasonably well at estimating the underlying (true distribution).

Figure B.1: True and estimated distribution of consumers (true follows a Kumaraswamy distribution, τ known)

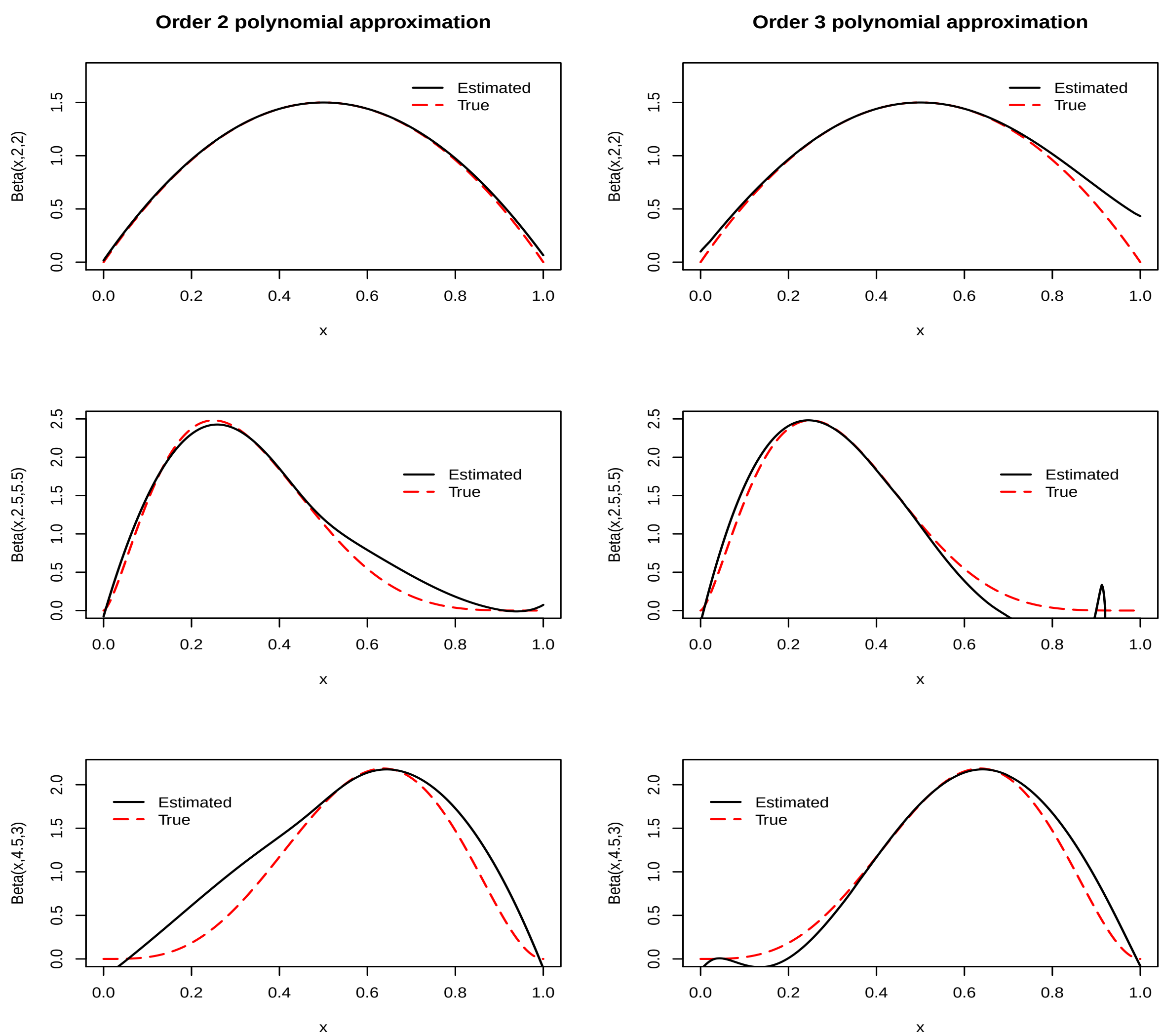


The underlying true distribution follows a Kumaraswamy distribution, with shape parameters a and b. Simulated data disturbances are $\varepsilon_\lambda \sim N(0, \text{var}\left(\lambda^t_{\text{obs}}\right))$ i.i.d. and $\varepsilon_\nu \sim N(0, \text{var}\left(\nu^t_{\text{obs}}\right))$ i.i.d. A Gaussian kernel was considered.

B.2 Estimation of the transportation cost and shape parameters (unknown transportation cost)

Next, to illustrate the procedure described in Section 3.2, we assume that the consumer's transportation cost is unknown. Our task is to estimate not only the transportation cost, but also the shape parameters of the distribution of consumers. We assume that the true (unknown) transportation cost is $\tau = 1.5$. The rest of the assumptions, including the cost parameters and firms' locations, are as in Section B.1. Again, regarding the true distribution of consumers, we consider two cases: a Kumaraswamy distribution and a Beta distribution. The shape parameters that we consider

Figure B.2: True and estimated distribution of consumers (true follows a Beta distribution, τ known)



Underlying true distribution follows a Beta distribution, with shape parameters α and β. Simulated data disturbances are $\varepsilon_\lambda \sim N(0, \text{var}\left(\lambda^t_{\text{obs}}\right))$ i.i.d. and $\varepsilon_\nu \sim N(0, \text{var}\left(\nu^t_{\text{obs}}\right))$ i.i.d. A Gaussian kernel was considered.

are the same as in Section B.1.

First, Table B.1 includes the estimated transportation costs and the estimated shape parameters for the simulation study based on 2,000 observations. The estimated parameters for the Kumaraswamy case are at the top part of the table, and the estimated parameters for the Beta distribution are at the bottom of the table. In all cases, the estimated transportation cost is, in fact, close to the true parameter.

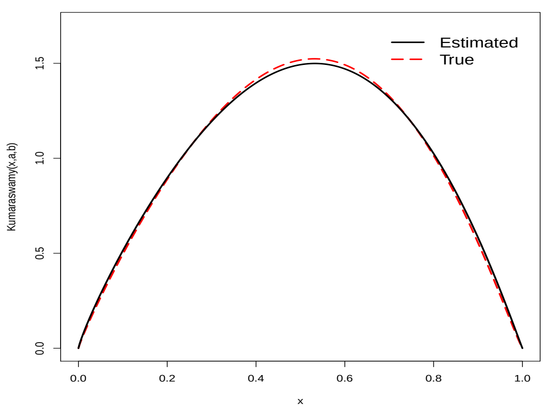
Next, we use these estimated parameters included in Table B.1 to check how close is our estimated distribution of consumer is in comparison to the true one. Figure B.3 captures the estimated and the true distribution of consumers, considering that the distribution of consumers follows a Kumaraswamy distribution. The three cases considered correspond to the three pair of shape parameters considered in Section B.1. Then, Figure B.4 captures the estimated and the true distribution of consumers, assuming that it follows a Beta distribution (again, we consider the three pair of shape parameters assumed in Section B.1). In the six cases, we can confirm that our methodology does in fact capture the shape of the underlying true distribution reasonably well.

Table B.1: True and estimated transportation costs and shape parameters for different distributions (τ unknown)

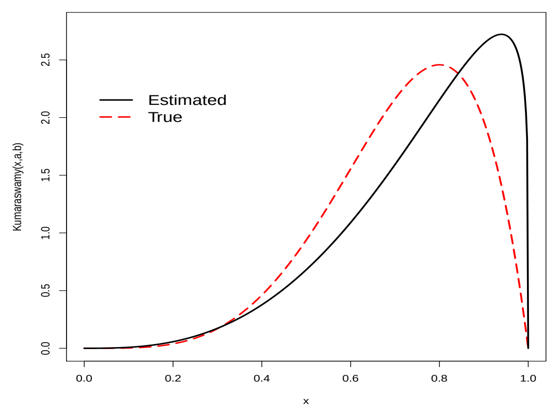
Kumaraswamy Distribution				
True parameters		$\tau = 1.5$		
		a=1.9, b=2.1	a=5, b=2	a=1.5, b=3
Estimated parameters	$\hat{\tau}$	1.5240	1.4737	1.3872
	$\hat{\alpha}^{\text{nls}}$	1.958525	3.93509	1.518468
	$\hat{\beta}^{\text{nls}}$	2.087475	1.168878	2.530797

Beta Distribution				
True parameters		$\tau = 1.5$		
		$\alpha = 2, \beta = 2$	$\alpha = 2.5, \beta = 5$	$\alpha = 4.5, \beta = 3$
Estimated parameters	$\hat{\tau}$	1.5240	1.4737	1.3872
	$\hat{\alpha}^{\text{nls}}$	1.936686	2.609055	3.869973
	$\hat{\beta}^{\text{nls}}$	1.936915	5.670833	2.544422

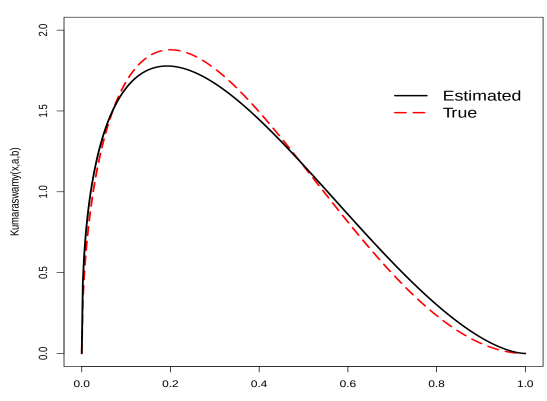
Figure B.3: True and estimated distribution of consumers (true follows a Kumaraswamy distribution, τ unknown)



(a) True shape parameters: $a = 1.9$, $b = 2.1$

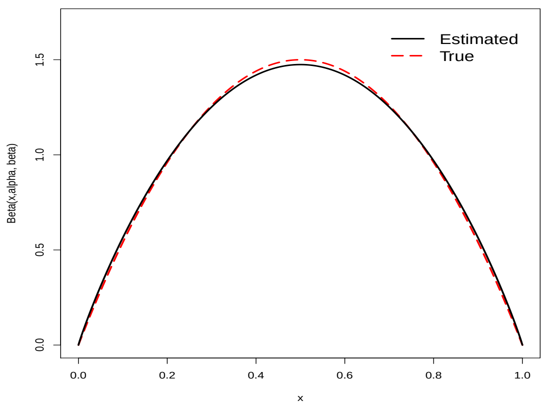


(b) True shape parameters: $a = 5$, $b = 2$

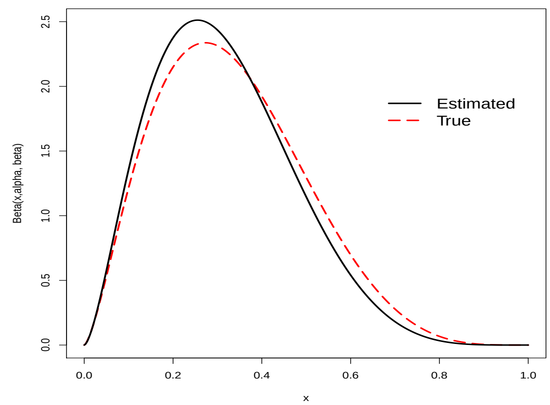


(c) True shape parameters: $a = 1.5$, $b = 3$

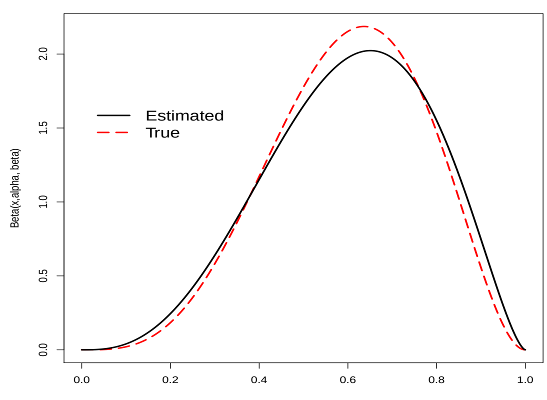
Figure B.4: True and estimated distribution of consumers (true follows a Beta distribution, τ unknown)



(a) True shape parameters: $\alpha = 2$, $\beta = 2$



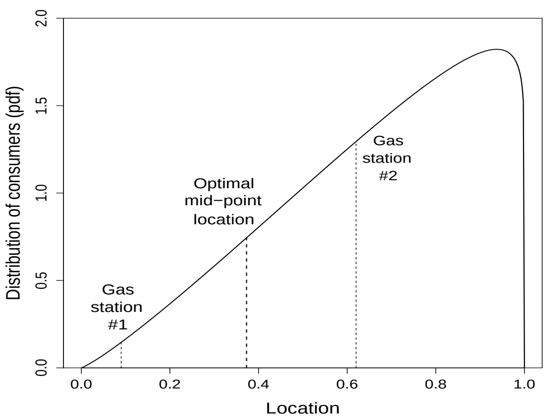
(b) True shape parameters: $\alpha = 2.5$, $\beta = 5$



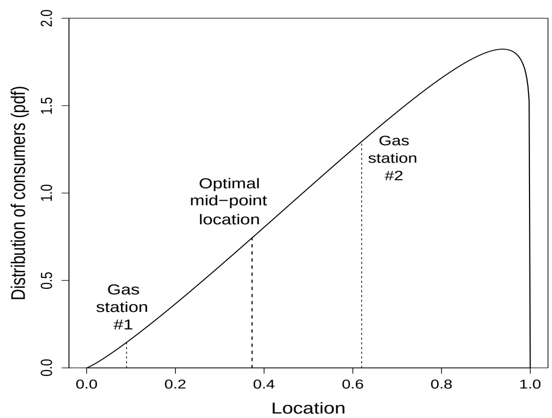
(c) True shape parameters: $\alpha = 4.5$, $\beta = 3$

Online Appendix C: Sensitivity analysis

Figure C.1: Sensitivity analysis for unleaded gasoline (Figure 4.a)

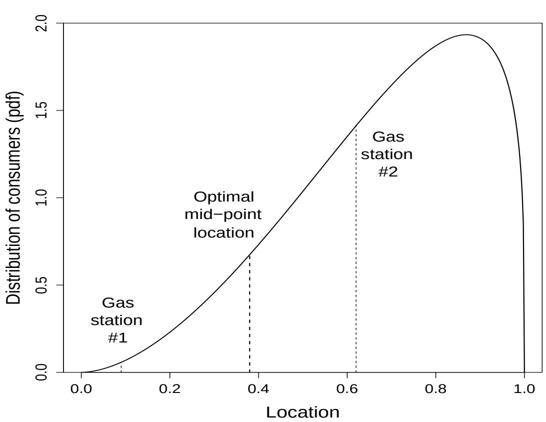


(a) Distribution cost = 0.008

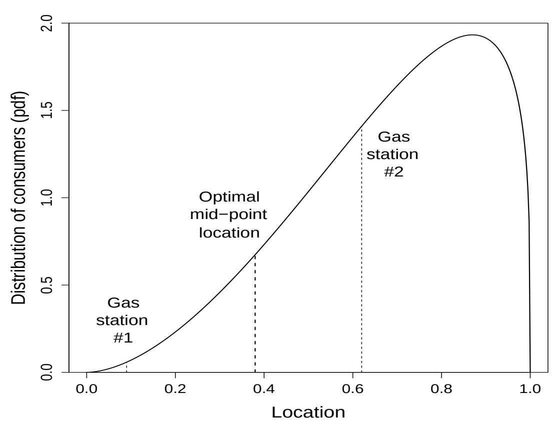


(b) Distribution cost = 0.015

Figure C.2: Sensitivity analysis for diesel fuel (Figure 4.b)



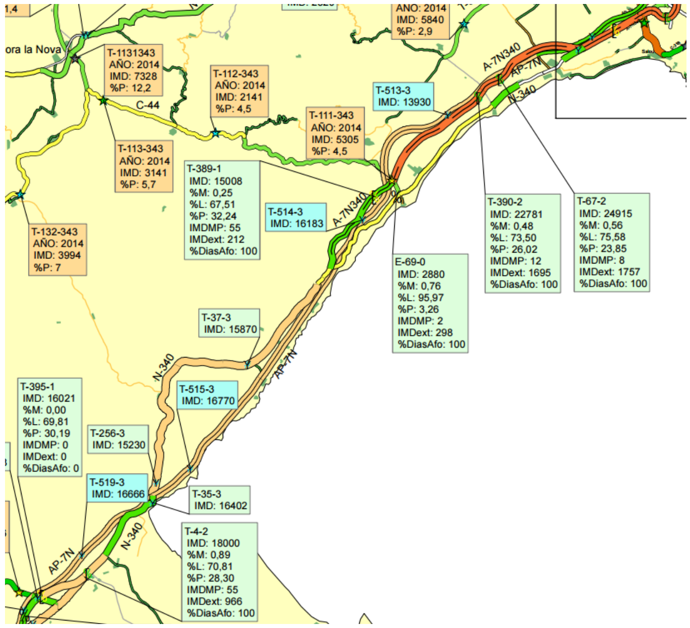
(a) Distribution cost = 0.008



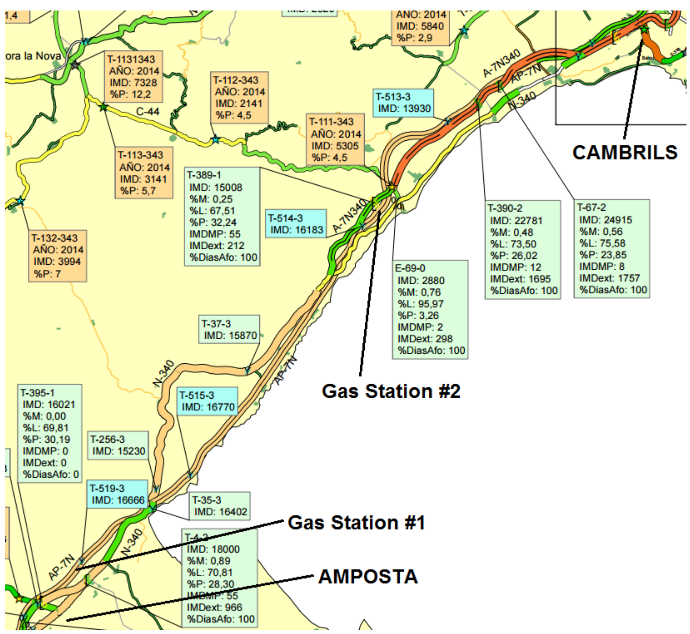
(b) Distribution cost = 0.015

Online Appendix D: Congestion maps

Figure D.1: Traffic congestion map in the relevant section of the AP-7 Toll Highway



(a) Without labels



(b) With labels

Figure D.2: Traffic congestion map in the Province of Tarragona (Catalonia, Spain) Source: Dirección General de Carreteras (2015)