Pontifícia Universidade Católica
do Rio de Janeiro

Barbara Fernandes Intropidi

Consumer Search in Brazilian Gasoline Retail

Dissertação de Mestrado

Thesis presented to the Programa de Pós–graduação em Economia, do Departamento de Economia da PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Economia.

Advisor: Prof. Leonardo Rezende

Rio de Janeiro
April 2022

Pontifícia Universidade Católica
do Rio de Janeiro

Barbara Fernandes Intropidi

Consumer Search in Brazilian Gasoline Retail

Thesis presented to the Programa de Pós–graduação em Economia da PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Economia. Approved by the Examination Committee:

Prof. Leonardo Rezende
Advisor
Departamento de Economia – PUC-Rio

Prof. Nathalie Gimenes
Departamento de Economia – PUC-Rio

Prof. João Paulo Pessoa
FGV EESP

Rio de Janeiro, April the 1st, 2022

Barbara Fernandes Intropidi

B.A. in Economics, Universidade de São Paulo (USP), 2017

Bibliographic data

Intropidi, Barbara Fernandes

Consumer Search in Brazilian Gasoline Retail / Barbara Fernandes Intropidi; advisor: Leonardo Rezende. – 2022.

52 f: il. color. ; 30 cm

Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Economia, 2022.

Inclui bibliografia

1. Economia – Teses. 2. Organização Industrial – Teses. 3. Busca do Consumidor. 4. Custos de Busca. 5. Informação Imperfeita. 6. Varejo de Gasolina. I. Rezende, Leonardo. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Economia. III. Título.

CDD: 620.11

To my parents, for their
unconditional love and support.

Acknowledgments

I would like to first thank my parents, Vilma and Francisco, for their love, patience, and support during my entire life and especially during this challenging period.

I thank my advisor, Prof. Leonardo Rezende, for his careful guidance and dedication to this project. His insights and helpful comments were essential to this research. I would also like to thank Profs. Nathalie Gimenes and João Paulo Pessoa for their useful comments. Any mistakes are my own.

I thank my dear friends for their support and company during this journey. I am also extremely grateful to my classmates, who created a friendly and collaborative environment. My experience at PUC-Rio was much more joyful thanks to them. Finally, I would like to thank the entire economics department at PUC-Rio, bright teachers who contributed to my knowledge and inspired me.

This study was partly financed by the Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq. The financial supports from PUC-Rio and Vinci Partners are gratefully acknowledged.

Abstract

Intropidi, Barbara Fernandes; Rezende, Leonardo (Advisor). **Consumer Search in Brazilian Gasoline Retail**. Rio de Janeiro, 2022. 52p. Dissertação de Mestrado – Departamento de Economia, Pontifícia Universidade Católica do Rio de Janeiro.

This paper seeks to understand consumer search patterns and whether information frictions play a role in price dispersion in Brazilian gasoline retail. In our setting, consumers must engage in costly search to gain information about the prices charged by gas stations. Empirically, we divide our analysis into two parts. In the first part, we use a structural model that permits us to estimate points of the distribution of search costs. We estimate the model using price data at the station level for multiple markets in Brazil. In the second part, in two independent analyzes, we investigate the determinants of the proportion of consumers with a low amount of search by OLS and construct an estimate for the average search cost per market by fitting our point estimates into a parametric distribution by NLS. Our findings reveal significant variation in consumer search across markets. Furthermore, our results reveal that most consumers do not compare many prices before buying gasoline. Moreover, our estimates indicate that the number of gas stations in a market, the average distance between gas stations, income, and population are important drivers of the proportion of consumers that search in only one gas station before buying. Finally, the estimated average search cost represents 3% of gasoline prices, a non-negligible proportion. Therefore, the results indicate that information frictions are important to explain price dispersion in Brazilian gasoline retail.

Keywords

Consumer Search; Search Costs; Imperfect Information; Retail Gasoline.

Resumo

Intropidi, Barbara Fernandes; Rezende, Leonardo. **Busca do Consumidor no Varejo de Gasolina Brasileiro**. Rio de Janeiro, 2022. 52p. Dissertação de Mestrado – Departamento de Economia, Pontifícia Universidade Católica do Rio de Janeiro.

Este trabalho procura entender padrões de busca do consumidor e se fricções informacionais desempenham um papel na dispersão de preços no varejo brasileiro de gasolina. Na nossa abordagem, os consumidores devem se engajar em busca custosa para obter informação sobre os preços cobrados pelos postos de gasolina. Empiricamente, dividimos nossa análise em duas partes. Na primeira, utilizamos um modelo estrutural que nos permite estimar pontos da distribuição dos custos de busca. Estimamos o modelo usando dados de preços no nível do posto para vários mercados no Brasil. Na segunda parte, em duas análises independentes, investigamos os determinantes da proporção de consumidores com baixa quantidade de busca por OLS e construímos uma estimativa para o custo médio de busca por mercado encaixando nossas estimativas pontuais em uma distribuição paramétrica por NLS. Nossas descobertas revelam uma variação significativa na busca do consumidor entre os mercados. Além disso, nossos resultados revelam que a maioria dos consumidores não compara muitos preços antes de comprar gasolina. Ademais, nossas estimativas indicam que o número de postos de gasolina em um mercado, a distância média entre os postos, a renda e a população são fatores importantes para explicar a proporção de consumidores que procuram em apenas um posto antes de comprar. Por fim, o custo médio estimado de busca representa 3% dos preços da gasolina, proporção esta não desprezível. Portanto, os resultados indicam que os atritos de informação são importantes para explicar a dispersão de preços no varejo brasileiro de gasolina.

Palavras-chave

Busca do Consumidor; Custos de Busca; Informação Imperfeita; Varejo de Gasolina.

Table of contents

1 Introduction **12**

2 Related Literature **15**

3 Data and Statistics **19**
3.1 Data 19
3.2 Price Dispersion Statistics 20
3.2.1 Different Prices Between Retailers 21
3.2.2 Temporal Dispersion 22

4 The Consumer Search Model **25**

5 Estimation **29**
5.1 First Stage: Structural Model 29
5.2 Second Stage 31
5.2.1 Linear Regression 31
5.2.2 Non-Linear Least Squares 31

6 Results **33**
6.1 Structural Estimates 33
6.2 Reduced Form Estimates 37
6.2.1 Linear Regression 37
6.2.2 Non-Linear Least Squares 39

7 Conclusion **42**

Bibliography **44**

A Appendix **47**
A.1 Structural Estimates 47
A.2 OLS Estimates - Alternative Specifications 48
A.2.1 A 49
A.2.2 B 49
A.3 NLS 50
A.3.1 NLS Fit 51
A.3.2 Alternative Specifications 51

List of figures

Figure 3.1	Residual Price Standard Deviation - Histogram	22
Figure 6.1	Average Search Cost Estimates - Histogram	41
Figure A.1	Search Cost Distribution Fitted Points - Scatter Plot	51

List of tables

Table 3.1 Descriptive Statistics 20
Table 3.2 Rank Reversal Statistics 24

Table 6.1 Maximum Likelihood Estimation Results - Proportion of consumers searching k times 34
Table 6.2 Number of Gas Stations in the Market and Average Proportion of Consumers Searching 1 and 2 Times 35
Table 6.3 Maximum Likelihood Estimation Results - Search Costs Cutoffs 36
Table 6.4 OLS Estimates 38
Table 6.5 NLS Estimates 40

Table A.1 Maximum Likelihood Estimation Results - Markets with all significant estimates 48
Table A.2 OLS Estimates - Alternative Specifications A 49
Table A.3 OLS Estimates - Alternative Specifications B 50
Table A.4 NLS Estimates - Alternative Specifications 52

List of Abreviations

ANP – Agência Nacional do Petróleo, Gás Natural e Biocombustíveis
API – Application Programming Interface
CNPJ – Cadastro Nacional da Pessoa Jurídica
IBGE – Instituto Brasileiro de Geografia e Estatística
IPCA – Índice de Preços ao Consumidor Amplo
NLS – Non-Linear Least Squares
OLS – Ordinary Least Squares
U.S. – United States

1
Introduction

Simple price competition models suggest that competition among firms that sell a homogeneous product will lead to a unique equilibrium price, the so-called law of one price. However, contrary to this theoretical outcome, price dispersion is a feature observed in many markets and, especially, in retail markets, since there are multiple sellers. This realization led Varian to write that "Economists have belatedly come to recognize that the 'law of one price' is no law at all" (Varian, 1980, p. 651).

The gasoline retail market is no exception to price dispersion. To illustrate, if we compare gasoline prices in the same week across gas stations in the municipality of São Paulo, the most populous city in Brazil, we find that in 2019, the average difference between the maximum and the minimum prices was 0.89 BRL per liter. This represents 19% of the average largest price.

Many factors could be driving this dispersion, such as differences in firms' costs, product heterogeneity, and imperfect competition. Other dimensions for differentiation, such as location, brand, and whether the gas station has a car wash or a convenience store may also play a role. In this paper, we focus on a particular source of price dispersion: information frictions.

In our setting, consumers have imperfect information on prices. Consumers know the price distribution, but they do not know which firm charges which price. To gain information about prices charged by a subset of firms, consumers must engage in costly search. It has already been proven that this costly search can sustain price dispersion even in homogeneous product markets (Reinganum, 1979; Carlson & McAfee, 1983; Burdett & Judd, 1983). In the model we adopt, for example, firms change their prices frequently to prevent consumers from identifying the lowest-priced firm, generating, thus, dispersion. In practice, this imperfect information generates market power for firms. Therefore, retailers can charge higher prices if information frictions do exist, which implies that this lack of information may reduce consumer welfare.

Thus, in this paper, we aim to understand consumer search patterns and whether information frictions may be a plausible story for price dispersion in Brazilian gasoline retail. For this purpose, we first estimate a consumer search model. The adopted model is based on Moraga-Gonzalez & Wildenbeest

(2008) and Wildenbeest (2011). On the supply side, there is a finite number of firms that employ mixed strategies in the utility space. On the demand side, a continuum of consumers that have the same preferences towards gasoline (that is, we assume gasoline is a vertically differentiated product) but have heterogeneous search costs. Consumers search non-sequentially for lower prices. Hence, before entering the market, they decide on the number of gas stations to visit and, after visiting the stations, they buy gasoline from the station providing the highest utility.

To estimate the model, we take advantage of a rich data set with gasoline prices at the level of the gas station covering 10% of the municipalities in Brazil. We estimate the model separately for each market in our data, defined as gas stations in the same 4-digit postal code in a municipality. The estimates suggest that most consumers search little before buying gasoline. The median estimates indicate that 88% of consumers search for lower prices in at most two gas stations. On the other side, there is a low proportion of consumers with intense search.

Second, in two independent analyzes, we: (a) investigate the determinants of the proportion of consumers that search once, and (b) construct an estimate for the average search cost per market. In (a), we regress the proportion of consumers searching once on market and municipality explanatory variables and find that the proportion of consumers searching once is positively related to the average distance between gas stations and the average income. On the other hand, it is negatively related to the number of gas stations in the market and the number of people in the municipality. These findings seem to corroborate the idea that the search for lower prices is related to the opportunity cost of time.

Finally, in (b), we fit the structurally-estimated points of the search cost distribution into a known parametric distribution by NLS to calculate the average search cost per market. The results indicate that the average search costs vary considerably across markets, ranging from 0.06 to 28.70 BRL cents per liter. The average across markets is 13.27 BRL cents per liter, which represents 3% of the average gasoline retail price.

In short, the main takeaways of our research are: (i) most consumers do not compare many prices before buying gasoline; (ii) the number of gas stations in a market, the average distance between gas stations, income, and population are important drivers of the proportion of consumers that search in one gas station; and (iii) the average search cost is a non-negligible proportion of gasoline prices. Therefore, the results seem to indicate that information frictions are indeed important to explain price dispersion in Brazilian gasoline retail. However, one must acknowledge that this is one potential explanation for heterogeneity in prices and that other factors can be even more important in this context, such as product heterogeneity, heterogeneity in consumer preferences, and other competition issues that we did not explore in-depth in the analysis.

Moreover, note that the search cost concept in our setting is very general. It is the cost of acquiring the price information of a firm. It can comprise several components, and we are not able to separately identify those components with the used methodology and data. For example, the cost of search may involve the opportunity cost of time in searching for lower prices, the cost of driving between gas stations to check prices, or the cost of searching online for prices. Our measure is an aggregate that indicates the total marginal cost of search and it can have multiple origins.

Lastly, it is worth mentioning that we are, to our knowledge, the first to estimate a consumer search model for Brazilian fuel retail. The only other article that studies search costs in fuel markets in Brazil, Vogt & Lucinda (2017), uses a test to infer whether costly search is relatively important to explain dispersion. However, the test does not allow one to deepen the discussion the way we do with our structural estimates, for example, by estimating the average search cost per market or identifying search patterns and relating them to observables.

The remainder of the paper is organized as follows: Chapter 2 presents a literature review. Chapter 3 introduces the data and some dispersion statistics. Chapter 4 describes the model. Chapter 5 explains the estimation procedure. We present the estimates, both from the structural model and reduced-form analysis in Chapter 6. Chapter 7 concludes.

2
Related Literature

The literature on consumer search builds on the seminal article of Stigler (1961), which is the first to formally discuss the importance of imperfect information on market equilibria and especially on equilibrium prices. In Stigler's reasoning, as a consequence of the information friction, consumers must engage in costly search to obtain information about prices in a market. This setup gives market power for the firms and can generate an equilibrium in which firms charge different prices even for homogeneous goods, that is, an equilibrium where price dispersion is observed. In Stigler's words, price dispersion is, therefore, a manifestation of "ignorance in the market" (Stigler, 1961, p. 214).

Since his influential article, there is a growing literature on search costs, composed both of theoretical and empirical papers. Focusing now on the theoretical literature, one strand of it relies on models with an information clearinghouse. The idea is that some consumers access a list provided by the clearinghouse with all prices in the market, while other consumers do not. Salop & Stiglitz (1977), Varian (1980), and Baye & Morgan (2001) fit into this framework.

The other strand, more closely related to the model we estimate in this paper, assumes consumers incur a cost for each additional price quote. Moreover, this strand is subdivided into two broad classes: sequential search (Carlson & McAfee, 1983; Stahl, 1989; Janssen et al., 2005), and non-sequential search, also called fixed sample size search (Burdett & Judd, 1983; Janssen & Moraga-Gonzalez, 2004). In models of sequential search, after each price quote, consumers compare the expected utility and the cost of search before engaging in another search. In a non-sequential search context, before searching, consumers decide on the number k of times to search, and then, after visiting k firms, they buy from the firm providing the highest utility.

Considerable effort has been made to use or modify these models so they can be estimated using aggregate market data. There are applied papers related to different product markets, whether for homogeneous or heterogeneous goods. We now briefly discuss some of these papers, focusing on those related to the strand of models in which consumers pay a cost for each additional price quote.

Hortaçsu & Syverson (2004) develop and estimate a model in which price dispersion is due both to product differentiation and costly search. Using price and quantity data, they estimate their model for the S&P 500 index funds market and find that, in this market, the estimated search costs exhibit much less dispersion than the price variation they support. Therefore, they conclude that it is indeed necessary both fund heterogeneity and search costs to explain the dispersion observed in data.

Hong & Shum (2006) develop a structural model for homogeneous goods in which search costs can be estimated using price data only. They illustrate their model estimating it for the online market of economics and statistics textbooks. Moraga-Gonzalez & Wildenbeest (2008) extend Hong & Shum (2006) to the oligopoly case and present a new method to estimate the model by maximum likelihood, applying the method to the market of computer memory chips. Wildenbeest (2011), in turn, extends Moraga-Gonzalez & Wildenbeest (2008) to the case of vertically differentiated goods and applies the model to grocery items, finding that in this market most of the price dispersion is explained by product heterogeneity rather than search frictions. Other papers that involve structural estimation of the distribution of search costs in different product markets include De Los Santos et al. (2012), Moraga-Gonzalez et al. (2013), and Honka (2014).

Moving to fuel retail applications, Yilmazkuday (2017) and Nishida & Remer (2018) estimate Wildenbeest (2011)'s model to multiple gasoline markets in the U.S.. Yilmazkuday (2017) uses daily gasoline price data for virtually every gas station in the U.S. for July 2015, while Nishida & Remer (2018) focus on the states of California, Florida, New Jersey, and Texas, using daily prices from February 27 to March 28, 2007. Both papers find expressive differences in the estimates across markets and try to investigate these heterogeneities by further linking the structural estimates to observable market-level variables.

Yilmazkuday (2017) segments the U.S. market based on the zip codes the gas stations are located at, an approach close to the one we use in this paper, and focus the analysis on the expected number of searches. The results indicate that consumers do not search much for lower prices (median of the estimates at 1.66 times). Using linear regressions, the author finds that the expected number of searches is positively related to the market area, population density, and average distance between gas stations. On the other hand, income and commuting time are negatively related to the expected number of searches. In its turn, Nishida & Remer (2018) use the concept of isolated markets and focus on the cost of search, finding that the estimated average search costs range from 0.05 to 0.45 dollars per gallon, and that income is a good predictor of the average search cost.

Additionally, some papers study costly search in fuel markets by reduced form exercises. In general, these papers test comparative static relationships implied by specific search costs models to infer whether costly search is supported by the data or not. This is the case of Marvel (1976), Lach (2002), Lewis (2008), Chandra & Tappata (2011), and Pennerstorfer et al. (2020).

For Brazil, in the consumer search literature, Vogt & Lucinda (2017) is, to our knowledge, the only application for fuel markets. They use a test proposed by Chandra & Tappata (2011) that employs the rank reversal statistic. The rank reversal compares pairs of gas stations and indicates the proportion of times in which the usually lower-priced station sets a higher price. A positive value for the statistic is evidence that mixed-strategy is being employed by firms, an assumption usually adopted in consumer search models. The rationale for using mixed strategies in contexts of imperfect information lies in the idea that companies change their prices frequently to prevent consumers from identifying the lowest-priced firm.

The test proposed by Chandra & Tappata (2011) consists of comparing the rank reversal statistics between gas stations for which imperfect information may play a role, to a control group of stations in which information frictions are absent. The control group is composed of gas stations located at the same street intersection, in which they argue information frictions are absent because a driver in the intersection observes the prices of both gas stations. If information frictions are important, then stations at the same street intersection have lower rank reversal statistics in comparison to gas stations farther away from each other since the price dispersion of the first group is driven only by product heterogeneity and not competition for informed and uninformed consumers.

Vogt & Lucinda (2017) apply the test to gasoline and ethanol markets in the municipality of São Paulo. The calculated rank reversals are higher for stations located at the same street intersection[1] than for the other pairs of gas stations in the market. They conclude, therefore, that search costs are not a relevant component in fuel retail in São Paulo. Note, however, that this type of analysis has its limitations. The test implicitly assumes that the only difference between the two groups of stations is whether or not their share the same street intersection. Nevertheless, stations can differentiate

[1]More precisely, they consider that a pair of stations is in the "same street intersection" if the distance between the stations is at most 100 meters.

themselves in multiple dimensions, such as product heterogeneity, having a convenience store, or a car wash. Additionally, these dimensions are part of the firms' choices to profit maximization, which means that they are not random. Therefore, it could be the case that search costs do exist but there are also other factors contaminating the test. Our structural approach, on the other hand, has the advantage of abstracting from this dependence on location patterns. Furthermore, it permit us to directly estimate points in the search cost distribution and calculate a measure for the search cost, a feature not possible with reduced form analysis.

Finally, the present paper also relates to the literature of competition and pricing in fuel retail markets. We now briefly describe some of these articles. For the U.S. and using reduced-form estimates, Barron et al. (2004) find evidence that a higher station density is associated with lower price levels and lower price dispersion, corroborating monopolistic competition models. Hastings (2004) investigates the effects of independent gas stations on competitors' prices, taking advantage of a sharp change in the market structure of gasoline retail in the U.S.. The results indicate that the presence of independent stations acts to decrease gasoline prices of its competitors.

Netz & Taylor (2002) analyze location patterns of gas stations in Los Angeles. Using a reduced-form approach, they find that an increase in competition is associated with a higher spatial differentiation. That is, in the face of an increase in the number of competitors, stations try to escape competition by spreading out. For Brazil, Pessoa et al. (2019) study how the diffusion of bi-fuel cars affected competition on ethanol and gasoline retail markets in Rio de Janeiro. They propose a theoretical model in which the two fuels become closer substitutes as the bi-fuel car fleet grows. Empirically, they find that retail prices and margins have fallen in response to the penetration of bi-fuel cars.

3
Data and Statistics

In this chapter, we describe the main data sources and variables used in the analysis. We then present descriptive statistics for all variables and additional dispersion statistics for gasoline prices.

3.1
Data

This paper combines data from different sources. The main data source is a survey conducted by ANP, the Brazilian regulatory agency covering the oil, gas, and biofuel industry. The survey is called *Levantamento de Preços e de Margens de Comercialização de Combustíveis* (ANP, 2021). It contains weekly data at the level of the gas station. This data set constitutes an unbalanced panel covering approximately 10% of the municipalities in Brazil. The selection of the gas stations surveyed in each municipality in a week is random and the municipalities that make up the sample are determined based on criteria such as income, car fleet, and the number of gas stations.

We use the following variables from this data set: gasoline retail prices, the date prices were collected, the postal codes and complete addresses of gas stations, and the CNPJs of gas stations, which are the identifiers of the firms. Additionally, we calculate the number of firms in a market by taking the number of distinct CNPJs in each market[1]. Our sample covers the period from January 2010 to December 2019 for 558 municipalities in Brazil. In this time interval, there are 3.3 million price observations at 25,744 gas stations.

Moreover, from the addresses of the gas stations, we construct a variable that represents the average distance between gas stations in a market. For this, we geocode the data using the HERE Geocoder API (HERE, 2021). Afterward, we compute the distances between every pair of stations in each market using the haversine formula[2]. From these measures, we calculate the average distance per market, a variable used in the post-structural estimation analysis.

[1]As we will explain later in more detail, our market definition comprises gas stations in the same 4-digit postal code in a municipality.

[2]The haversine formula determines the distance between two points on a sphere. Since the Earth is approximately an ellipsoid, the haversine formula is a better approximation of distances on Earth than the euclidean distance, which calculates the distance between two points on a plane.

For the post-structural estimation analysis, we also use complementary data sources. From the 2010 Brazilian Census (IBGE, 2010), we use the following municipal variables: average income of people aged 10 years and over, and the number of people aged 10 years and over. From the Ministry of Infrastructure (Ministério da Infraestrutura, 2010), we use data on car fleet in December 2010. From the Brazilian Institute of Geography and Statistics (IBGE, in the acronym in Portuguese), we use information on municipal area (IBGE, 2020).

Finally, all variables in BRL were deflated by the Extended Consumer Price Index (IPCA, in the acronym in Portuguese) from IBGE at December 2019 levels (IBGE, 2020).

Table 3.1 presents summary statistics.

Table 3.1: Descriptive Statistics

Variable	Mean	SD	Range	N Obs
Station Level Variables				
Retail Price	4.30	0.33	3.51	3,304,849
Market Level Variables				
# Stations	10.74	10.76	144	2,397
Avg Distance	5.95	12.04	97.35	2,274
Municipality Level Variables				
Avg Income	1,168.90	436.45	2,665.63	558
# People	185,299.30	532,345.90	9,780,530	558
Area	2,867.26	9,020.97	159,518.00	558
Car Fleet/People	0.22	0.13	0.65	558

Notes: SD represents the standard deviation of the variable. Range indicates the difference between the maximum and the minimum values observed for each variable. N Obs indicates the number of observations. Retail Price is the gasoline retail price, in BRL per liter. # Stations denotes the number of gas stations. Avg Distance is the average distance (in km) between gas stations. Avg Income denotes the average monthly income of people aged 10 years old or more in BRL. # People is the number of people aged 10 years old or more. Area is the municipal area in km^2. Car Fleet/People is the ratio between car fleet and the number of people aged 10 years old or more. All variables in BRL are deflated by IPCA to December 2019 levels.

3.2
Price Dispersion Statistics

3.2.1 Different Prices Between Retailers

Our model considers that information frictions allow firms to charge different prices for the same product. Therefore, to first investigate whether our hypothesis holds for Brazilian gasoline retail, we need a measure of price dispersion. If we compare prices in the same week across gas stations in a market, the average standard deviation of prices is 0.08 BRL per liter, with a minimum of 0 and a maximum of 1.13 BRL per liter. However, note that this simple comparison is not sufficient, since the price differences may also reflect product heterogeneity.

Thus, we do the following calculations. For each market, we run a fixed-effects regression of retail prices on firm and time fixed effects. Station fixed effects aim to control for any price differences resulting from seller heterogeneity, while time fixed effects serve as a control for movements in the general price level. Notice that each residual of this regression reveals whether the price of the gas station was above or below its expected level given the controls. Therefore, we can interpret the standard deviation of the residuals as a measure of price dispersion.

Before showing the results of these regressions, let us make clear our market definition. It comprises gas stations in the same 4-digit postal code in a municipality. This definition is similar to the adopted in Pessoa et al. (2019). It corresponds to neighborhoods in large cities and, for small cities, corresponds to the city itself. Our sample comprises 2,397 markets in 558 municipalities. On average, each municipality is composed of 4.3 markets, with minimum of one market[3] and maximum of 309 markets[4]. Additionally, the average number of gas stations per market is 10.7, and the number of gas stations ranges from 2 to 146.

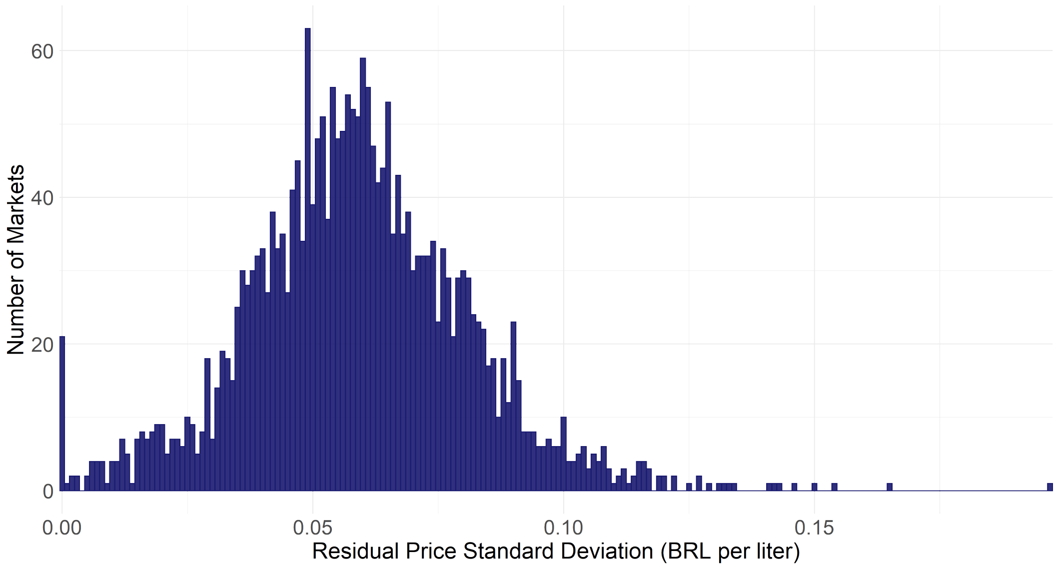
The regressions of prices on station and time fixed effects in each market yield a median R^2 of 0.94. The rest of the variation is associated with our measure of dispersion. The high R^2 value indicates that search costs are left to explain a small portion of prices in comparison to product differentiation and other supply and demand factors.

Figure 3.1 shows the distribution of the residual prices standard deviation across markets, our measure of dispersion. The average dispersion is 5.9 cents of BRL per liter and it ranges from 0 to 19.7 cents of BRL per liter. This dispersion is observed after removing gas station and time fixed effects and, therefore, is not associated with any gas station characteristics, such as brand

[3]314 municipalities in our sample are composed of only one market.
[4]The maximum number of markets, 309, is observed in the city of São Paulo.

or having a convenience store, and neither, for example, to common supply shocks that vary over time. One possible explanation for this dispersion, which we focus on in this paper, is the imperfect information consumers have on prices.

Figure 3.1: Residual Price Standard Deviation - Histogram



3.2.2
Temporal Dispersion

In this section, we investigate whether mixed strategy by firms, an assumption made in our model, is supported by the data. The idea behind this assumption is that firms change their prices over periods to prevent consumers from identifying the lowest-priced firm. Thus, if firms play mixed strategies, we should observe frequent relative price movements across stations. In other words, we should observe temporal price dispersion.

To analyze temporal price dispersion, we calculate a measure called Rank Reversal (RR), proposed by Chandra & Tappata (2011). The underlying idea of the construction of the statistic is that, if firms play a mixed-pricing strategy and we compare two firms over time, say, A and B, sometimes firm A sets the highest price of the two, and sometimes B sets the highest price of the two. The RR statistic gives a measure that indicates the extent to which these reversals occur.

Formally, let s_{ij} be a vector of the price spread between two gas stations (i, j) over T_{ij} periods in a market m, such that $p_{it} \geq p_{jt}$ is observed most of the time. As in Chandra & Tappata (2011), the price rank reversal between stations i and j is defined as the proportion of observations in which $p_{jt} > p_{it}$:

$$RR_{ij} = \frac{1}{T_{ij}} \sum_{t=1}^{T_{ij}} \mathcal{I}_{\{p_{jt} > p_{it}\}},$$

where $\mathcal{I}$ is an indicator function equal to one if $p_{jt} > p_{it}$ and zero otherwise. Thus, the price rank reversal between two firms is the proportion of observations in which the typically lower-priced firm sets a higher price. By construction, it can never be greater than 50%.

Our structural model assumes firms draw prices from a distribution. Notice, however, that given quality differences across gas stations, stations may draw prices from distributions with non-overlapping supports. In this case, even though firms are playing mixed strategies, we would not observe reversals in price ranking (that is, we would observe a RR of zero). Nonetheless, our model assumes gas stations play, on the same support, mixed strategies in utilities. Thus, even if we do not observe reversals in price rankings, we should observe reversals in utility. Taking this into account, we also calculate utility[5] rank reversals[6]. Similar to the price rank reversal, the utility rank reversal between two firms is the proportion of observations in which the firm that typically offers a lower utility sets a higher utility.

Accordingly, we compute price and utility rank reversals for each pair of gas stations in each market. To be included in the statistics, a pair of gas stations must be observed together in at least five weeks. Additionally, note that the fact that our sample is an unbalanced panel does not affect the statistic, as a RR statistic is constructed by taking the same pair of stations and analyzing their price difference in the same week over multiple weeks. We pool the RR estimates from all markets and calculate summary statistics, which are shown in Table 3.2 below.

An analysis of the price RR reveals that 81% of the pairs of gas stations reverse rankings (have strictly positive price RR). On average, the typically lower-priced firm sets a higher price in 17.13% of the time, with a standard deviation of 14.96 percentage points (p.p.). This value is slightly higher than those found by two papers that assume mixed strategies and use data for markets in the U.S.. Chandra & Tappata (2011) calculate an average RR of 13.8%, and Nishida & Remer (2018) of 8.3%[7]. Restricting our data to pairs

[5]As will be explained in more details in Chapter 5, we obtain utilities by taking the negative of the residuals in a regression of retail prices on firm and time fixed effects.

[6]This is also used in Nishida & Remer (2018).

[7]Notice, however, that the RR statistics can be influenced by the market definition. Chandra & Tappata (2011) construct the RR statistic for all possible pairs of stations separated by less than 1 mile. Nishida & Remer (2018) use isolated markets, defined in their case as a set of firms all within 1.5 miles of each other, where no other competitor is within 1.5 miles of any firm in the market. In our definition of market, the 4-digit postal code in a municipality, each market can have a different extension in area.

with positive RR, the average rises to 21.25%, with a standard deviation of 13.79 p.p..

Moreover, as expected, utility rank reversals are higher than price rank reversals. According to the table, 93% of the pairs of stations reverse utility rankings. The average utility RR is 31.35%, with a standard deviation of 14.25 p.p.. The subset of firms with positive utility RR has an average RR of 33.61%, with a standard deviation of 11.91 p.p.. Therefore, the RRs indicate there are significant relative price and utility movements between gas stations. Thus, we conclude the data seem to support the hypothesis that firms employ mixed strategies.

Table 3.2: Rank Reversal Statistics

	N obs	Mean	SD	Min	Max
Price RR					
Rank Reversals (RR)	162,605	17.13	14.96	0	50
RR, conditional on RR >0	131,047	21.25	13.79	0.22	50
Utility RR					
Rank Reversals (RR)	162,605	31.35	14.25	0	50
RR, conditional on RR >0	151,702	33.61	11.91	0.46	50

Notes: A unit of observation is a RR for a pair of gas stations in the same market. RRs are displayed in percentage terms. Pairs of gas stations in the same market must have at least five price observations on the same week to be included.

4
The Consumer Search Model

In this chapter, we present the consumer search model. The model was introduced in Wildenbeest (2011), which extends Moraga-Gonzalez & Wildenbeest (2008) by allowing vertically differentiated products[1].

Consider a market where N firms, indexed by j, sell a homogeneous good to imperfectly informed consumers. The good is bundled with firm-related services, which add value to it and permit firms to differentiate themselves in terms of quality. There is a continuum of consumers with inelastic demand, demanding each one unit of the good. Consumers share the same indirect utility, given by:

$$u_j = v_j - p_j, \tag{4-1}$$

where v_j is the valuation obtained from consuming firm j's good and p_j is its price. The following functional form for v_j is assumed:

$$v_j = x + q_j, \tag{4-2}$$

where x is the common utility derived from the homogeneous good itself, and q_j denotes firm j's quality[2].

Consumers know their valuation v_j derived from firm j's product and the distribution of prices. However, consumers do not know the specific price charged by each firm. In consequence, consumers engage in costly search to gain information about the prices charged by a subset of firms and, thus, about the utility derived from consuming from them. In each search, consumers incur search cost c. Consumers differ in terms of their search cost c, which is assumed to be a random draw from a continuous distribution $G(c)$ with support $(0, \infty)$ and density $g(c)$. In each search, the probability of finding firm j is $\frac{1}{N}$ for all j.

Additionally, we assume consumers search nonsequentially. That is, before entering the market, consumers determine the number of firms they will search, say k. After gaining information about the prices charged in these

[1] In its turn, Moraga-Gonzalez & Wildenbeest (2008) extend Hong & Shum (2006) to the case of oligopoly, and propose a maximum likelihood method to estimate the structural model. All of them are based on the theoretical non-sequential search model in Burdett & Judd (1983).

[2] We assume each firm's quality level is fixed in the short term.

k firms, consumers buy from the firm in their sample providing the highest utility[3].

Next, two assumptions are made on firm's quality q_j: (i) firms obtain quality input factors in perfectly competitive markets; and (ii) the quality production function exhibits constant returns to scale. Let's see the implications of these assumptions. Let the quality production function be denoted by $q_j(\mathbf{y})$, where $\mathbf{y} = (y_1, \ldots, y_n)$ are input factors. Let y_i denote the input used for the production of firm j's good, $i \in \{1, \ldots, n\}$. Additionally, let r_j denote firm j's cost of producing one unit of the good associated with quality q_j. Assumption (i) implies that the price paid for the quality input is equal to its marginal product value, therefore, for all $i \in \{1, \ldots, n\}$, $p_{y_i} = \frac{\partial q_j}{\partial y_i}$. Assumption (ii), by Euler's theorem[4], implies we can write the quality function as $q_j(\mathbf{y}) = \sum_{i=1}^{n} y_i \frac{\partial q_j}{\partial y_i}$. Combining the implications of (i) and (ii), we get $q_j(\mathbf{y}) = \sum_{i=1}^{n} y_i p_{y_i}$. The right-hand side of this equation is equal to the total cost of producing one unit of the good associated with quality q_j. Therefore,

$$q_j = r_j. \tag{4-3}$$

Quality q_j is determined by firms such that margin $p(q_j) - r_j$ is maximized. Using the utility function in 4-1, we can rewrite the margin as $v_j - u_j - r_j$. Therefore, for a given level u of utility, firms decide on their quality levels such that their valuation-cost markup $v_j - r_j$ is maximized. Using 4-2 and 4-3: $v_j - r_j = x + q_j - q_j = x$ for all firm j. The valuation-cost markup does not depend on firm's quality. Thus, at each offered utility level, firms are symmetric in the margin received. This allows us to focus on symmetric mixed-strategy equilibria in utility levels[5], in which firms have a common continuous utility distribution $L(u)$, with density $l(u)$.

We now analyze the equilibrium. Given firms' strategies, a consumer decides on the number k of firms to visit[6], $k \in \{1, \ldots, N\}$. For an optimal consumer behavior, it must be the case that, for a consumer searching k times, the expected utility must be greater or equal than the total cost of search,

[3]It is assumed that consumers can costlessly revisit previously searched firms.

[4]Euler's theorem for homogeneous functions says that, for a homogeneous function f of $\mathbf{x} = (x_1, \ldots, x_n)$ of degree h, it is true that $hf(\mathbf{x}) = x_1 \frac{\partial f(\mathbf{x})}{\partial x_1} + \ldots + x_n \frac{\partial f(\mathbf{x})}{\partial x_n}$.

[5]As explained in Wildenbeest (2011), mixed-strategies in equilibrium can be rationalized as follows: suppose some utility level u is set with positive probability by two firms in equilibrium. Then, offering $u + \epsilon$, $\epsilon > 0$, will give a discrete increase in profits, since some consumers search more than once, and they buy from the firm offering the highest utility in their sample. Therefore, for small ϵ, deviating, i.e., changing from setting u to $u + \epsilon$ is profitable. Thus, in equilibrium, there can be no atoms in the utility-setting strategies.

[6]Notice that, since firms choose u_j from the same utility distribution $L(u)$, they are, *ex-ante*, seem as identical to consumers in terms of expected utility, and consumers search randomly among firms.

i.e., $E\left[\max\left\{u_1,\ldots,u_k\right\}\right] \geq kc$. Moreover, the net benefit[7] of searching k times must be greater than the net benefit of searching $k-1$ and $k+1$ times.

Let c_k be the search cost of the consumer indifferent between searching k and $k+1$ times[8]. For this consumer, $E\left[\max\left\{u_1,\ldots,u_k\right\}\right] - kc_k = E\left[\max\left\{u_1,\ldots,u_{k+1}\right\}\right] - (k+1)c_k$. Rearranging, we get:

$$c_k = E\left[\max\left\{u_1,\ldots,u_{k+1}\right\}\right] - E\left[\max\left\{u_1,\ldots,u_k\right\}\right]. \tag{4-4}$$

An analogous reasoning for c_{k-1} implies that consumers searching k times must have search cost between c_k and c_{k-1}. Thus, the share of consumers searching k times is given by:

$$\gamma_k = \begin{cases} 1 - G\left(c_k\right) & \text{for} \quad k = 1 \\ G\left(c_{k-1}\right) - G\left(c_k\right) & \text{for} \quad k \in \{2,\ldots,N-1\} \\ G\left(c_{k-1}\right) & \text{for} \quad k = N \end{cases} \tag{4-5}$$

We can rewrite the search cost cutoff in 4-4 using the utility distribution[9]. Let $Z_k = \max\{u_1,\ldots,u_k\}$. Accordingly, the cumulative distribution and density of Z_k are, respectively, $F_{Z_k}(u) = L(u)^k$ and $f_{Z_k}(u) = kL(u)^{k-1}l(u)$. Then:

$$c_k = \int_{\underline{u}}^{\overline{u}} u(k+1)L(u)^k l(u)du - \int_{\underline{u}}^{\overline{u}} ukL(u)^{k-1}l(u)du. \tag{4-6}$$

We now make a change of variable that will simplify this expression. Let $y = L(u)$. Then $\frac{dy}{du} = l(u)$, $\underline{y} = L(\underline{u}) = 0$, and $\overline{y} = L(\overline{u}) = 1$. Rewriting 4-6:

$$\begin{aligned} c_k &= \int_0^1 u(y)(k+1)y^k dy - \int_0^1 u(y)ky^{k-1}dy \\ &= \int_0^1 u(y)\left[(k+1)y - k\right]y^{k-1}dy. \end{aligned} \tag{4-7}$$

Next, we analyze firms' optimal decisions. Given the search behavior of consumers and the strategies of the other firms, an individual firm chooses its level of utility u_j to maximize profits. The profit is given by:

$$\pi_j\left(u_j; L(u)\right) = \left(x - u_j\right)\left\{\sum_{k=1}^{N} \gamma_k \frac{k}{N} L\left(u_j\right)^{k-1}\right\}, \tag{4-8}$$

where the first part $(x - u_j)$ is the margin received for each unit[10] and

[7] Expected utility minus the total cost of search.

[8] We will also refer to c_k as the search cost cutoff or critical search cost.

[9] This will facilitate the estimation of the model.

[10] Remember the margin received was derived using assumptions (i) and (ii). The following equalities hold: $p_j - r_j = v_j - u_j - r_j = x + q_j - u_j - r_j = x + r_j - u_j - r_j = x - u_j$.

the second part is the expected quantities sold[11].

Given mixed strategies, in equilibrium, a firm should be indifferent between setting any utility in the support $L(u)$. Moreover, the lower bound of utility should be zero[12]. In this case, the profit equation simplifies to $\pi\left(\underline{u}\right) = x\frac{\gamma_1}{N}$. Setting this expression equal to the equilibrium profits in general gives us:

$$(x - u_j)\left\{\sum_{k=1}^{N} \gamma_k \frac{k}{N} L\left(u_j\right)^{k-1}\right\} = x\frac{\gamma_1}{N}. \tag{4-9}$$

In its turn, the upper bound of the utility distribution $\overline{u}$ can be found by setting $L(\overline{u}) = 1$ in 4-9:

$$\overline{u} = x\frac{\sum_{k=2}^{N} \gamma_k k}{\sum_{k=1}^{N} \gamma_k k}. \tag{4-10}$$

Given the expected search behavior of consumers and the expected utility distribution, firms maximize profits by choosing their utility level u_j. The first-order condition of 4-8 is then given by:

$$l(u) = \frac{\sum_{k=1}^{N} k\gamma_k L(u)^{k-1}}{(x - u)\sum_{k=1}^{N} k(k-1)\gamma_k L(u)^{k-2}}, \tag{4-11}$$

where $L(u)$ solves 4-9.

[11]It is given by the summation, over all groups of consumers (those who search one time until those who search N times), of the share of consumers searching k times γ_k multiplied by the probability $\frac{k}{N}$ that these consumers visit the firm and by the probability that in the other $k-1$ visits, the utilities provided by the other firms are no greater than u_j, that is, $L\left(u_j\right)^{k-1}$. Notice that the expected quantity sold expression implicitly assumes search with replacement.

[12]Note that a firm offering the lower bound $\underline{u}$ will only sell to consumers that search once, and surplus extracted from these consumers is maximized when $p_j = v_j$, which implies $\underline{u} = 0$.

5
Estimation

We perform the estimation in two stages. First, we estimate the structural model presented in the previous chapter for each market m in our data. For each market, the estimation yields the proportion of consumers searching k times, γ_k^m, $k \in \{1, \dots, N_m\}$, and points $(c_k^m, G^m(c_k^m))$ of the search cost distribution, $k \in \{1, \dots, N_m - 1\}$.

Then, we pool the estimates across all markets and conduct a post-structural estimation analysis, divided into two independent parts. This analysis aims to understand the differences in consumer search across markets and rationalize them with observables, as well as to construct an estimate for the average search cost in each market.

Accordingly, we first estimate an OLS, linking the proportion of consumers searching once, γ_1^m, to market and municipality variables. We focus on this type of consumer in this first analysis because low consumer search is associated with more market power for firms. Subsequently, we run a NLS estimation that fits the estimated points of the search cost distribution into a known parametric distribution. For this, we also use market and municipality explanatory variables. Fitting our point estimates into a known distribution allows us to estimate the average search cost per market.

5.1
First Stage: Structural Model

We estimate the consumer search model separately for each market m. Hence, the process explained in this section is conducted for each market independently. For simplicity, we remove the superscript (or subscript) m in the following equations.

The procedure follows Wildenbeest (2011) and Moraga-Gonzalez & Wildenbeest (2008). The first step is to estimate the utility u_j from observed prices, running the following regression by OLS:

$$p_{jt} = \delta_0 + \delta_{1j} + \delta_{2t} + \epsilon_{jt}, \tag{5-1}$$

where j denotes the firm and t the week prices are observed. The coefficient δ_0 is a constant, δ_{1j} are firm fixed effects, δ_{2t} are time fixed effects and ϵ_{jt} is the error term.

Note that we can rewrite the utility function in 4-1 as $p_{jt} = v_{jt} - u_{jt}$. Thus, the valuations v_{jt} correspond to $\delta_0 + \delta_{1j} + \delta_{2t}$. By doing this, we are using the average differences across firms' prices as a proxy for their valuations. In turn, the utilities correspond to $-\epsilon_{jt}$. Hence, we estimate utilities by the negative of the residuals $\widehat{\epsilon_{jt}}$. Additionally, the upper and lower bounds of the utility distribution are estimated, respectively, by the maximum and minimum estimated utilities.

Next, we use the density function in 4-11 to construct our log-likelihood function, which is then given by:

$$LL = \sum_{i=1}^{M} \log l(u_i), \tag{5-2}$$

where M is the total number of price observations in the market[1]. The unknowns are x, γ_1, ..., γ_N.[2]

We can reduce the number of parameters to estimate by substituting x, isolating this parameter in equation 4-10. Then, the log-likelihood is maximized with respect to parameters γ_k, $k \in \{1, \ldots, N\}$. As in Moraga-Gonzalez & Wildenbeest (2008), the numerical procedure is as follows. Firstly, we take starting values for $\{\gamma_k\}_{k=1}^{N}$. These initial guesses are arbitrary, taken randomly from a Uniform distribution on $[0, 1]$. Then, for every estimated utility, we calculate $L(u_i)$ that solves equation 4-9. In turn, this permits us to calculate $l(u_i)$. We use a trust region Preconditioned Conjugate Gradient method (PCG method), that continues to change the γ_k's until the log-likelihood function is maximized. The estimation is done with the restriction that each one of the parameters is between zero and one, inclusive. Then, we normalize the estimated parameters to sum up one. Thus, more precisely, from the log-likelihood function, we estimate the non-normalized parameters $\tilde{\gamma_k}$ and then calculate, for each k, $\gamma_k = \frac{\tilde{\gamma_k}}{\sum_{l=1}^{N} \tilde{\gamma_l}}$.

Next, we use equation 4-5 to calculate the values of the cumulative distribution of search costs at the cutoffs, and 4-7 and 4-9 to calculate the search costs cutoffs.

Standard errors of the $\tilde{\gamma_k}$'s, for $k \in \{1, \ldots, N-1\}$, are calculated by taking the square root of the diagonal entries of the inverse of the negative Hessian matrix evaluated at the optimum[3]. For the normalized estimates, γ_k,

[1]We can pool all the estimated utilities in the log-likelihood function since, by assumption, all firms draw utilities from the same distribution.

[2]Remember the number of firms in the market, N, is estimated by the number of distinct gas stations in the data.

[3]To be able to calculate the standard errors, we remove from the Hessian elements for which the corresponding parameters estimates were equal or lower than 10^{-6}.

$k \in \{1, \dots, N-1\}$, we use the Delta method. Since $\gamma_N = 1 - \sum_{k=1}^{N-1} \gamma_k$, we calculate the standard error of γ_N also by the Delta method. The standard errors of the search cost cutoffs are calculated by the Delta method as well since the cutoffs are transformations of the estimated γ_k's.

5.2 Second Stage

5.2.1 Linear Regression

In our preferred specification, we run the following regression by OLS:

$$\begin{aligned} \gamma_1^m = \psi_0 \log(Avg\,Distance_m) + \psi_1 \log(\#\,Stations_m) \\ +\psi_2 \log(Avg\,Income_m) + \psi_3 \log(Car\,Fleet/People_{mun}) \\ +\psi_4 \log(\#\,People_{mun}) + \zeta_m, \end{aligned} \tag{5-3}$$

where m is the subscript (or superscript) of variables at the market level, and mun is the subscript of variables at the municipality level. ζ_m is the error term.

5.2.2 Non-Linear Least Squares

We fit the estimated points $(c_k^m, G^m(c_k^m))$ into a mixture of log-normal distributions by Non-Linear Least Squares. The log-normal distribution is widely used in the literature of search costs - see Hortaçsu & Syverson (2004); Hong & Shum (2006); Wildenbeest (2011). We also tested other parametric distributions, for which the results are shown in Appendix A.3, but the mean squared errors of these estimates were higher. Additionally, the decision to use a mixture of distributions was based on the structural results, which showed two types of consumers, as we will explain in more detail in the next chapter. A mixture of distributions is also used in Moraga-Gonzalez & Wildenbeest (2008) and Wildenbeest (2011).

We assume the shape parameters of the log-normal distributions are invariant across markets, but the location parameter is assumed to be a linear function of market and municipality variables. We estimate the following regression:

$$G^m(c_k^m) = a\,\Phi\left(\frac{(\log c_k^m) - \mu_{1,m}}{\sigma_1}\right) + (1-a)\,\Phi\left(\frac{(\log c_k^m) - \mu_{2,m}}{\sigma_2}\right) + \eta_m, \tag{5-4}$$

where $a \in [0,1]$ denotes the weight, $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, m denotes the market, η_m is the error term, and $\mu_{1,m}$ and $\mu_{2,m}$ are given by:

$$\mu_{1,m} = \alpha_0 + \alpha_1 \, Avg\,Distance_m + \alpha_2 \,\#\, Stations_m + \alpha_3 \, Car\,Fleet/People_{mun} \tag{5-5}$$

and

$$\mu_{2,m} = \beta_0 + \beta_1 \, Avg\,Distance_m + \beta_2 \,\#\, Stations_m + \beta_3 \, Car\,Fleet/People_{mun}. \tag{5-6}$$

With the estimates of these coefficients in hand, we then calculate the average search cost for each market m using the expression:

$$\widehat{ac^m} = \int_0^\infty \left[1 - \widehat{G^m(c)}\right] dc. \tag{5-7}$$

6
Results

In this chapter, we present the results of the estimations. First, we analyze the estimates of the structural model. Then, we present, respectively, the OLS and NLS estimations that use the structural results as inputs.

6.1
Structural Estimates

We pool the estimates of the proportion of consumers searching for lower prices k times, γ_k, and construct summary statistics, which are shown in Table 6.1. For simplicity purposes, we show statistics for $k = \{1, 2, 3, N_m\}$. There is no loss of content in omitting the other estimates. In the table, # Firms and # Obs indicate, respectively, the number of firms and the number of observations. Log-lik is the value of the log-likelihood function at the solution. For each of the aforementioned statistics, we report the minimum (Min) and maximum (Max) values, as well as the 25th, 50th, and 75th percentiles.

Table 6.1 shows that there is substantial variation in estimates across markets. For example, the proportion of consumers searching once, γ_1, ranges from 19% to 100%. The median of the estimates for the proportion of consumers searching once is 59% (50th percentile column). This large portion of consumers buys from the first gas station they visit. These values are consistent with Nishida & Remer (2018), whose γ_1 for gasoline markets in the U.S. ranges from 0.4% to 94%, with average of 66%.

The table also reveals that another large percentage of consumers search in only two firms before buying gasoline (median of 29%, ranging from 0 to 56%). The fact that the median estimates indicate that approximately 88% of the consumers search in two gas stations at most before buying suggests that information frictions may be a relevant component in gasoline retail. Additionally, and although we do not model this component explicitly, this large proportion may also represent consumers who are loyal to a particular gas station.

The median estimates of $\gamma_3, \ldots, \gamma_{N_m-1}$ are all roughly zero. Finally, the median estimate indicates that 11% of the consumers search in every gas station in the market before buying.

Table 6.1: Maximum Likelihood Estimation Results - Proportion of consumers searching k times

	Percentiles				
	Min	25th	50th	75th	Max
Proportion of consumers that search k times (γ_k)					
γ_1	0.19	0.47	0.59	0.83	1
γ_2	0	0.15	0.29	0.36	0.56
γ_3	0	0	0	0	0.49
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
γ_{N_m}	0	0.02	0.11	0.18	0.49
# Firms	2	4	7	14	146
# Obs	13	258.25	675	$1,818$	$17,788$
Log-lik	-286.97	231.27	584.84	$1,480.74$	$12,619.98$

Notes: Each γ_k, $k \in \{1, ..., N_m\}$, denotes the proportion of consumers that search for lower prices k times. # Firms and # Obs indicate, respectively, the number of firms and the number of price observations. Log-lik is the value of the log-likelihood function at the solution. The total number of gas stations N_m varies with each market m. Therefore, in some cases, the same estimate is considered both in the row of the N_mth firm and in a precedent row. For example, if a market has 3 firms, then γ_3 is counted both in γ_3 and γ_{N_m} statistics.

Additionally, we note that the extreme high values for the proportion of consumers searching once, γ_1, and extreme low values for the proportion of consumers searching twice, γ_2, shown in Table 6.1, are mainly associated with markets with a small number of gas stations. To visualize this, we group markets by the number of gas stations and calculate the averages of γ_1 and γ_2 in each group. The results are shown in Table 6.2.

The table reveals that, at first, γ_1 decreases and γ_2 increases with the number of gas stations. For markets with only two gas stations, virtually all consumers search in only one gas station. For markets with 3 gas stations, on average, 96% of the consumers search in one gas station and 3% in two. Markets with number of gas stations equal to 4 or 5 have average γ_1 equal to 76% and average γ_2 equal to 20% and so on. Then, from markets with at least 12 gas stations, the average estimates of γ_1 and γ_2 seem to stabilize around 45% and 34%, respectively. For markets with 41 or more gas stations, the average proportion of consumers searching once is 42%, and the average proportion of consumers searching twice is 33%, summing up 75%.

Thus, even though the γ_k estimates are very dependent on the number of gas stations in each market, we note that even in markets with a large number of firms, the proportion of consumers searching once or twice is significant.

Table 6.2: Number of Gas Stations in the Market and Average Proportion of Consumers Searching 1 and 2 Times

	Averages	
# Stations	γ_1	γ_2
2	1	0
3	0.96	0.03
4 to 5	0.76	0.2
6 to 7	0.59	0.31
8 to 11	0.52	0.33
12 to 14	0.46	0.35
15 to 20	0.47	0.34
21 to 26	0.45	0.34
27 to 40	0.46	0.33
41 or more	0.42	0.33

Notes: γ_1 and γ_2 denote, respectively, the average proportion of consumers searching 1 and 2 times in markets with number of gas stations in # Stations.

In conclusion, the model estimates suggest that there are two groups of consumers in the Brazilian gasoline market. The first and largest group is composed of consumers who search in only one or two gas stations. These consumers are associated with higher search costs. The second group is composed of consumers who compare all prices in the market. These consumers are associated with lower search costs. The existence of these two types of consumers motivates the use of a mixture of log-normal distributions in the NLS estimation in the next section. Each log-normal is expected to be associated with each type of consumer.

In addition, notice that the presence of a large group of consumers with a low amount of search translates into market power for gas stations. Since these consumers are not comparing prices, firms can charge higher gasoline prices. The competition between gas stations occurs mainly for the consumers in $\gamma_3, \ldots, \gamma_{N_m}$.

We move to the interpretation of search costs cutoffs estimates, for which summary statistics are shown in Table 6.3. We note that there is also a great variability across markets. The highest value cutoff, that is, the search cost of the consumer indifferent between searching 1 and 2 times, c_1, ranges from 0.02 to approximately 26 cents of BRL per liter. In its turn, the search cost of the consumer indifferent between searching in $N_m - 1$ and N_m times varies from 0.01 to 15.66 cents of BRL per liter.

We analyze the median of these search costs cutoffs estimates. Table 6.3 reveals that for consumers searching only once, their search cost must be at least 7.86 cents of BRL per liter. The share of consumers who search in two gas stations, in turn, must have search costs that range from 4.05 to 7.86 cents. Furthermore, consumers who search in at least 3 gas stations but not in all stations of the market have search costs ranging from 0.8 to 4.05 cents of BRL per liter. Nevertheless, we have already seen that this type of consumer has a negligible share in the consumer population. Lastly, to rationalize the behavior of consumers who compare all prices, their search cost should be at most 0.8 cents of BRL per liter.

Table 6.3: Maximum Likelihood Estimation Results - Search Costs Cutoffs

	Percentiles				
	Min	25th	50th	75th	Max
Search Costs Cutoffs in BRL cents/liter (c_k)					
c_1	0.02	5.78	7.86	10.24	26.26
c_2	0.01	3.04	4.05	5.27	13.07
c_3	0.12	1.98	2.65	3.46	8.81
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
c_{N_m-1}	0.01	0.43	0.80	1.72	15.66

Notes: c_k denotes the search cost of the consumer indifferent between searching k and $k+1$ times, $k \in \{1, ..., N_m - 1\}$. The total number of gas stations N_m varies with each market m. Therefore, in some cases, the same estimate is considered both in the $N_m - 1$th row and in a precedent row. For example, if a market has 4 firms, then c_3 is counted both in c_3 and c_{N_m-1} statistics. Search costs cutoffs are expressed in BRL cents/liter in real terms, deflated by IPCA to December 2019 levels.

The results presented in this section show the statistics for estimates in all markets. Table A.1 in Appendix A.1 shows γ_k and c_k estimates filtering for markets in which all the estimates are statistically different from zero at the 5% significance level, considering the one-tailed test in which the alternative hypothesis is that the parameter is greater than zero. The main conclusions are unchanged.

6.2 Reduced Form Estimates

6.2.1 Linear Regression

Table 6.4 reports the linear regression estimates for the proportion of consumers searching once, γ_1^m.[1] In all of them, standard errors are clustered at the municipality level. Column (1) includes the average distance between stations in a market as a regressor. Column (2) includes the average distance between stations and the number of gas stations in a market. Column (3), in addition to the aforementioned market-level variables, includes municipality-level variables, namely the average income, the ratio between car fleet and population, and the population. All regressors are with logarithmic transformation. Our preferred specification is (3).[2]

The table confirms that # Stations explains a lot of the variation in the proportion of consumers searching once: when passing from model (1) to (2), the adjusted R^2 increases from 0.06 to 0.66. Additionally, the coefficient of # Stations is consistently negative and significant across the two last specifications. This seems intuitive: the greater the number of firms, the greater the options for consumers, who will tend to search for prices in more than one firm, decreasing γ_1.

Except for the first specification, the average distance between firms has a positive and statistically significant coefficient. Therefore, a higher average distance between gas stations is associated with a higher proportion of consumers who search only once. This seems reasonable if we conjecture that the cost of search involves the opportunity cost of time and the cost of driving between stations. Thus, the results suggest that consumers are discouraged from searching for lower prices when the distance between stations is higher.

The coefficient associated with average monthly income is positive and statistically significant. This makes sense: the higher the income, the higher the opportunity cost, which translates to a lower amount of search (higher γ_1). The variable that indicates the ratio between car fleet and people serves as a proxy for the intensity of car use in a municipality. However, its coefficient is not statistically different from zero. This may indicate that gasoline consumers in

[1] Appendix Table A.2 shows, for robustness purposes, regressions in which the dependent variable is the proportion of consumers that search once or twice, i.e., $\gamma_1 + \gamma_2$. The results in those regressions point in the same directions as the results we show in this chapter.

[2] In Appendix Table A.3, we present further regressions with additional explanatory variables.

markets that differ only on the intensity of car use do not have different search patterns.

Finally, the number of people has a negative and significant coefficient. Therefore, markets with a higher population are associated with a lower proportion of consumers searching once. One situation that could be driving this result is that, as the population increases, the heterogeneity among people also increases. As a result, we have a higher proportion of people searching in more than one gas station. We can also interpret this coefficient by thinking on the supply side. We expect that a market with a large population will also have a high number of gas stations. In this case, consumers have more options to search, thus the proportion of consumers who search once is lower.

Table 6.4: OLS Estimates

Dependent Variable:		γ_1	
Model	(1)	(2)	(3)
(Intercept)	0.691***	1.09***	0.912***
	(0.014)	(0.013)	(0.135)
Market Level Variables			
Avg Distance (Log)	-0.049***	0.012***	0.008***
	(0.005)	(0.003)	(0.002)
# Stations (Log)		-0.225***	-0.230***
		(0.009)	(0.010)
Municipality Level Variables			
Avg Income (Log)			0.059***
			(0.019)
Car Fleet/People (Log)			-0.012
			(0.011)
# People (Log)			-0.019***
			(0.005)
Observations	2,195	2,195	2,195
R^2	0.06	0.66	0.67
Adjusted R^2	0.06	0.66	0.67

Notes: Standard errors in parentheses, clustered at the municipality level.
* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

6.2.2 Non-Linear Least Squares

The results of the NLS estimation are shown in Table 6.5 below[3]. The MSE of this estimation is 0.005. Except for the Car Fleet/People variable in Log-normal 1, all the other coefficients are precisely estimated at the 5% significance level. The weight parameter a is estimated at 0.2. First, note the difference in magnitude of the estimated intercepts of the location parameter in each log-normal. Note as well the difference in the shape estimates. The first log-normal has an intercept of approximately 13 and shape estimate at 6, while the second log-normal has those estimates at 2.8 and 0.7, respectively. These differences reflect the two types of consumers we were trying to capture: one type with a low amount of search and thus high search cost, and the other type the opposite, with a high amount of search and low search cost.

Furthermore, for the two log-normals, the direction of the coefficient estimates associated with each explanatory variable seems reasonable. The average distance between stations is associated with positive coefficients, that is, the larger the average distance between gas stations, the larger the average search cost in a market. Again, this makes sense if we interpret that search costs are associated with the opportunity cost of time, and/or we believe that consumers engage in costly search by driving between stations.

The number of gas stations, in its turn, has a negative impact on the average search cost. It seems reasonable that in markets with more competitors, consumers' needed effort to search for lower prices is also lower. Finally, the ratio between car fleet and population has a coefficient with a negative sign in both specifications, although it is only statistically significant in log-normal 2.

[3]A plot with the fitted points of the search cost distribution versus the structurally estimated points is presented in Appendix A.3 Figure A.1. Appendix A.3 also shows the results of NLS estimation using different parametric distributions in Table A.4. Those estimates yielded higher Mean Squared Error (MSE) in comparison to the sum of log-normals estimates. Additionally, we tested the NLS with different explanatory variables, but the ones we show here gave us better results.

Table 6.5: NLS Estimates

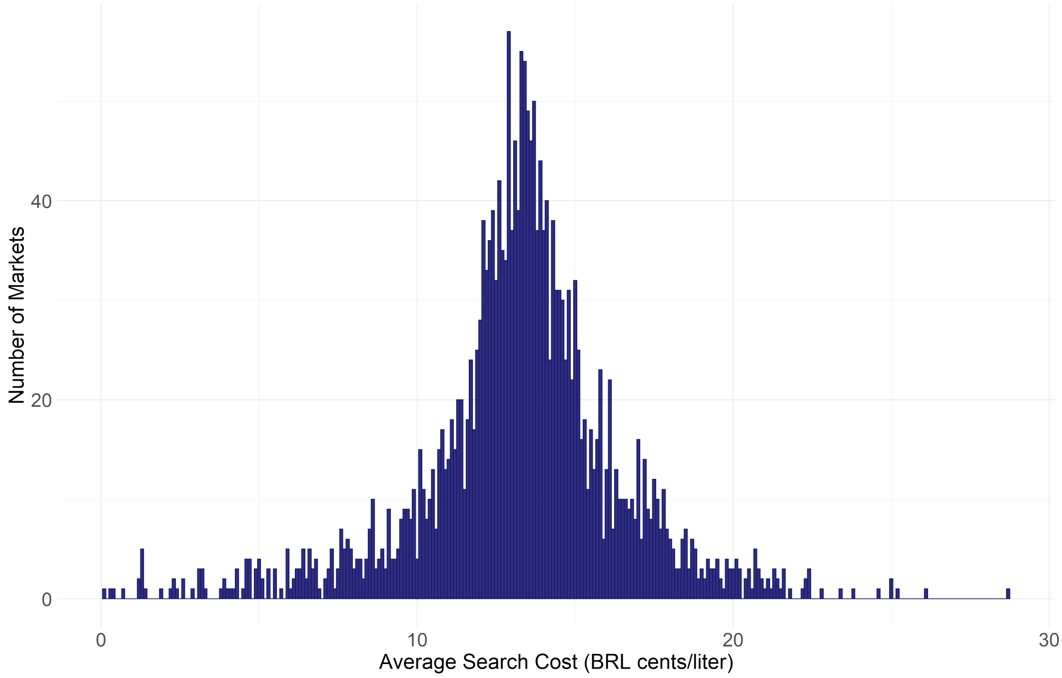
Description	Coef.	Estimate	Std. error
Weight	a	0.212***	0.001
Log-normal 1			
Intercept	α_0	13.553***	1.656
Avg Distance	α_1	0.018**	0.008
# Stations	α_2	−1.809***	0.228
Car Fleet/People	α_3	−0.384	0.597
Shape	σ_1	6.044***	0.789
Log-normal 2			
Intercept	β_0	2.809***	0.017
Avg Distance	β_1	0.001**	3.7E-04
# Gas Stations	β_2	−0.004***	4.0E-04
Car Fleet/People	β_3	−0.281***	0.038
Shape	σ_2	0.735***	0.010
Observations		21,653	
MSE		0.005	

Notes: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

With the estimates of the search cost distribution in Table 6.5, we construct an average search cost estimate for each market. A histogram of these estimates is presented in Figure 6.1. The histogram shows that there is a lot of variation in the average search cost across markets. The estimates range from 0.06 to 28.70 BRL cents per liter, with an average of 13.27 BRL cents per liter. This expressive variation shows the importance of estimating the model separately for each market. Furthermore, in relative terms, the average search cost represents 3% of the gasoline retail price[4]. Therefore, the results indicate that search frictions are relevant in the gasoline retail dynamics, and that search costs are a non-negligible proportion of prices.

[4]This takes into account that, in our data, the average price in real terms is 4.30 BRL per liter.

Figure 6.1: Average Search Cost Estimates - Histogram



7
Conclusion

This paper studies consumer search in Brazilian gasoline retail. We aim to understand consumer search patterns and whether information frictions may play a relevant role in explaining price dispersion. For this, we first estimate a non-sequential consumer search model, in which consumers differ in their search costs, and gasoline is considered a vertically differentiated product. We estimate the model using price data at the gas station level for multiple markets in Brazil, recuperating the frequency in which consumers search for lower prices, as well as points in the search cost distribution. Then, we pool the estimates across all markets and perform a post-structural estimation analysis, divided into two independent parts. In the first part, we conduct an OLS estimation, linking the proportion of consumers that search once to market and municipality variables. In the second part, we fit the estimated points of the search cost distribution into a mixture of log-normal distributions by NLS. The estimated coefficients allow us to compute the average search cost per market.

Our findings reveal significant variation in consumer search across markets. Furthermore, the results suggest that most consumers do not compare many prices before buying. Additionally, the results indicate that the number of gas stations in a market, the average distance between gas stations, income, and population are important drivers of the proportion of consumers that search in one gas station. Lastly, our estimates indicate that the average search cost is a non-negligible proportion of gasoline prices. Therefore, the results seem to indicate that information frictions are indeed important to explain price dispersion in Brazilian gasoline retail. However, one must acknowledge that this is one potential explanation for heterogeneity in prices and that other factors can be even more important in this context, such as product heterogeneity, heterogeneity in consumer preferences, and other competition issues that we did not explore in-depth in the analysis.

In that regard, for future research, it would be interesting to introduce new features in the model, such as heterogeneous preferences, although one probably would need data on quantity for estimation. In addition, it would be interesting to explore different market definitions. Note that one potential source of bias in our estimates is the existence of overlapping markets. This includes, for example, gas stations relatively close to each other that we consider in separate markets but in reality, are competing. Therefore, one possible extension to our work is to deal with isolated markets. For example, a market can comprise a set of gas stations that are all within a certain distance of each other, say, d, and there are no other competitors within the distance d from any gas station in this market. In turn, one drawback with this definition is that the researcher will probably lose multiple observations that do not satisfy the criteria for an isolated market.

Bibliography

HORTAÇSU, A.; SYVERSON, C.. **Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: A Case Study of S&P 500 Index Funds**. The Quarterly Journal of Economics, 119(2):403–456, 2004.

BARRON, J.; TAYLOR, B. ; UMBECK, J. R.. **Number of sellers, average prices, and price dispersion**. International Journal of Industrial Organization, 22(8-9):1041–1066, 2004.

BURDETT, K.; JUDD, K.. **Equilibrium price dispersion**. Econometrica, 51(4):955–69, 1983.

VOGT, C. M.; LUCINDA, C. R.. **Custos de procura e dispersão de preços de gasolina e etanol no mercado brasileiro de combustíveis: Evidências do município de são paulo**. Análise Econômica, 35(68):317–344, 9 2017.

CARLSON, J.; MCAFEE, R.. **Discrete equilibrium price dispersion**. Journal of Political Economy, 91(3):480–93, 1983.

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

DE LOS SANTOS, B.; HORTAÇSU, A. ; WILDENBEEST, M. R.. **Testing models of consumer search using data on web browsing and purchasing behavior**. American Economic Review, 102(6):2955–80, May 2012.

PENNERSTORFER, D.; SCHMIDT-DENGLER, P.; SCHUTZ, N.; WEISS, C. ; YONTCHEVA, B.. **Information And Price Dispersion: Theory And Evidence**. International Economic Review, 61(2):871–899, May 2020.

YILMAZKUDAY, H.. **Geographical dispersion of consumer search behaviour**. Applied Economics, 49(57):5740–5752, 12 2017.

VARIAN, H. R.. **A model of sales**. The American Economic Review, 70(4):651–659, 1980.

HONG, H.; SHUM, M.. **Using price distributions to estimate search costs**. RAND Journal of Economics, 37(2):257–275, 2006.

HONKA, E.. **Quantifying search and switching costs in the us auto insurance industry**. RAND Journal of Economics, 45(4):847–884, 2014.

JANSSEN, M. C. W.; MORAGA-GONZALEZ, J.. **Strategic pricing, consumer search and the number of firms**. Review of Economic Studies, 71(4):1089–1118, 2004.

JANSSEN, M.; MORAGA-GONZALEZ, J. ; WILDENBEEST, M.. **Truly costly sequential search and oligopolistic pricing**. International Journal of Industrial Organization, 23(5-6):451–466, 2005.

HASTINGS, J. S.. **Vertical Relationships and Competition in Retail Gasoline Markets: Empirical Evidence from Contract Changes in Southern California**. American Economic Review, 94(1):317–328, 3 2004.

LACH, S.. **Existence and persistence of price dispersion: An empirical analysis**. The Review of Economics and Statistics, 84(3):433–444, 2002.

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

MARVEL, H. P.. **The economics of information and retail gasoline price behavior: An empirical analysis**. Journal of Political Economy, 84(5):1033–60, 1976.

BAYE, M. R.; MORGAN, J.. **Information Gatekeepers on the Internet and the Competitiveness of Homogeneous Product Markets**. American Economic Review, 91(3):454–474, June 2001.

MORAGA-GONZALEZ, J.; SANDOR, Z. ; WILDENBEEST, M.. **Semi-nonparametric estimation of consumer search costs**. Indiana University, Kelley School of Business, Department of Business Economics and Public Policy, Working Papers, 28, 11 2013.

MORAGA-GONZALEZ, J.; WILDENBEEST, M.. **Maximum likelihood estimation of search costs**. European Economic Review, 52(5):820–848, 2008.

NETZ, J. S.; TAYLOR, B.. **Maximum or minimum differentiation? location patterns of retail outlets**. The Review of Economics and Statistics, 84(1):162–175, 2002.

NISHIDA, M.; REMER, M.. **The determinants and consequences of search cost heterogeneity: Evidence from local gasoline markets**. Journal of Marketing Research, 55, 02 2018.

PESSOA, J. P.; REZENDE, L. ; ASSUNÇÃO, J. J.. **Flex cars and competition in fuel retail markets**. International Journal of Industrial Organization, v. 63:p. 145–184, 2019.

REINGANUM, J.. **A simple model of equilibrium price dispersion**. Journal of Political Economy, 87(4):851–58, 1979.

SALOP, S.; STIGLITZ, J.. **Bargains and ripoffs: A model of monopolistically competitive price dispersion**. Review of Economic Studies, 44(3):493–510, 1977.

STAHL, D.. **Oligopolistic pricing with sequential consumer search**. American Economic Review, 79(4):700–712, 1989.

STIGLER, G.. **The economics of information**. Journal of Political Economy, 69, 1961.

WILDENBEEST, M.. **An empirical model of search with vertically differentiated products**. RAND Journal of Economics, 42(4):729–757, 2011.

ANP. **Série histórica de preços de combustíveis**. `https://bit.ly/3MPYDQX`, 2021. Accessed in 07-18-2021.

HERE. **Here Geocoder API**. `https://bit.ly/35QsOHq`, 2021. Accessed in 01-18-2022.

IBGE. **Censo Demográfico 2010**. `https://bit.ly/3IcIlOY`, 2010. Accessed in 12-28-2021.

IBGE. **Áreas Territoriais**. `https://bit.ly/3KAlizO`, 2020. Accessed in 12-12-2021.

IBGE. **Índice Nacional de Preços ao Consumidor Amplo**. `https://bit.ly/3NOOYaC`, 2020. Accessed in 08-13-2021.

MINISTÉRIO DA INFRAESTRUTURA. **Estatísticas SENATRAN**. `https://bit.ly/3MJD8Bz`, 2010. Accessed in 12-27-2021.

A
Appendix

A.1
Structural Estimates

In Table A.1, we present the structural estimates filtering for markets in which all estimates are statistically significant at 5%, considering the alternative hypothesis that the parameter is greater than zero[1].

The overall results are virtually unchanged compared to those presented in Chapter 6. There are still two types of consumers: those with low amount of search, and those with high amount of search. The median estimate of the proportion of consumers searching once is 56%, slightly lower than the estimate proportion in Table 6.1 (at 59%), whereas the median estimate of the proportion of consumers searching twice is 31% (in comparison to 29% in Table 6.1). Therefore, the median of the estimates indicates that approximately 87% of the consumers search in two gas stations at most. The rest of consumers search in every firm in the market.

The estimates of the search cost cutoffs are also similar to the ones presented in Table 6.3. We focus here on analyzing the median of the estimates. The search cost of consumers that search only once must be at least 8.21 cents of BRL per liter. The share of consumers who search in two gas stations, in turn, must have search costs that range from 4.25 to 8.21 cents. Consumers who search in at least 3 gas stations but not in all stations of the market have search costs ranging from 0.71 to 4.25 cents of BRL per liter. Finally, to rationalize the behavior of consumers who compare all prices in a market, their search cost should be at most 0.71 cents of BRL per liter.

[1]Note that there are still some zeros in the table. These estimates are not exactly zero, just rounding. This justifies, therefore, that even those are statistically significant.

Table A.1: Maximum Likelihood Estimation Results - Markets with all significant estimates

	Percentiles				
	Min	25th	50th	75th	Max
Proportion of consumers that search k times (γ_k)					
γ_1	0.19	0.46	0.56	0.71	1
γ_2	0	0.23	0.31	0.37	0.56
γ_3	0	0	0	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
γ_{N_m}	0	0.06	0.13	0.19	0.350
Search Costs Cutoffs in BRL cents/liter (c_k)					
c_1	0.10	6.18	8.21	10.51	26.26
c_2	0.11	3.36	4.25	5.46	13.07
c_3	0.13	2.07	2.71	3.52	8.81
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
c_{N_m-1}	0.08	0.41	0.71	1.38	15.66
N Firms	2	5	8	15	60
N Obs	14	364	831	$2,045$	$10,802$
Log-lik	-286.97	298.68	712.73	$1,665.98$	$10,692.85$

Notes: Each γ_k, $k \in \{1, ..., N_m\}$, denotes the proportion of consumers that search for lower prices k times. c_k denotes the search cost of the consumer indifferent between searching k and $k+1$ times. # Firms and # Obs indicate, respectively, the number of firms and the number of price observations. Log-lik is the value of the log-likelihood function at the solution. The total number of gas stations N_m varies with each market m. Therefore, in some cases, the same estimate is considered both in the row of the N_mth firm and in a precedent row. For example, if a market has 3 firms, then γ_3 is counted both in γ_3 and γ_{N_m} statistics. Search costs cutoffs are expressed in BRL cents/liter in real terms, deflated by IPCA to December 2019 levels.

A.2 OLS Estimates - Alternative Specifications

A.2.1
A

Table A.2: OLS Estimates - Alternative Specifications A

Dependent Variable:		$\gamma_1 + \gamma_2$	
Model	(1)	(2)	(3)
(Intercept)	0.914***	1.08***	0.991***
	(0.005)	(0.003)	(0.044)
Market Level Variables			
Avg Distance (Log)	-0.024***	0.002*	0.002*
	(0.002)	(0.0008)	(0.0008)
# Stations (Log)		-0.093***	-0.094***
		(0.002)	(0.002)
Municipality Level Variables			
Avg Income (Log)			0.015***
			(0.006)
Car Fleet/People (Log)			-0.004
			(0.004)
# People (Log)			-0.002*
			(0.001)
Observations	2,195	2,195	2,195
R^2	0.10	0.78	0.78
Adjusted R^2	0.10	0.78	0.78

Notes: Standard errors in parentheses, clustered at the municipality level.
* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

A.2.2
B

Table A.3 below shows alternative specifications for regressions with the proportion of consumers searching once as the dependent variable. CT - 30 to 60 min (%), CT - 60 to 120 min (%), and CT - 120 min or more (%) indicate, respectively, the proportion of people in a municipality with commuting time between 30 and 60 min, 60 and 120 min, and 120 min or more. Literacy Rate (%) indicates the percentage of people aged 10 years old or more in a municipality that can read and write. Avg Age is the average age of people in a municipality. Men (%), People in Urban Areas (%), White People (%)

indicate, respectively, the proportion of men, people in urban areas and white people in a municipality.

Table A.3: OLS Estimates - Alternative Specifications B

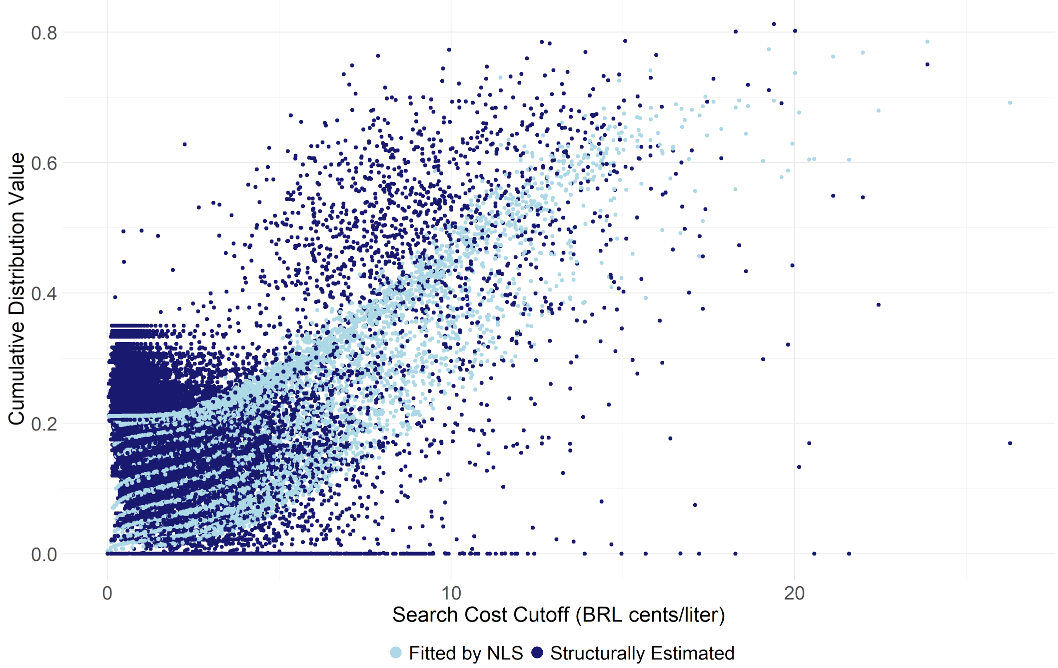
Dependent Variable:		γ_1	
Model	(1)	(2)	(3)
(Intercept)	0.985***	0.971***	0.982***
	(0.147)	(0.285)	(0.325)
Market Level Variables			
Avg Distance (Log)	0.008***	0.008***	0.007***
	(0.002)	(0.002)	(0.002)
# Stations (Log)	-0.233***	-0.230***	-0.229***
	(0.010)	(0.010)	(0.010)
Municipality Level Variables			
Avg Income (Log)	0.041**	0.059***	0.063***
	(0.020)	(0.021)	(0.020)
Car Fleet/People (Log)	-0.007	-0.013	0.006
	(0.011)	(0.012)	(0.014)
# People (Log)	-0.012***	-0.020***	-0.023***
	(0.004)	(0.004)	(0.005)
CT - 30 to 60 min (%)	-0.0003		
	(0.0008)		
CT - 60 to 120 min (%)	-0.001		
	(0.001)		
CT - 120 min or more (%)	-0.001		
	(0.003)		
Literacy Rate (%)		0.0010	
		(0.001)	
Average Age (Log)		-0.042	
		(0.071)	
Men (%)			3.41×10^{-5}
			(0.004)
People in Urban Areas (%)			0.0001
			(0.0005)
White People (%)			-0.0007**
			(0.0003)
Observations	2,195	2,195	2,195
R^2	0.67	0.67	0.67

Notes: Standard errors in parentheses, clustered at the municipality level. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

A.3
NLS

A.3.1
NLS Fit

Figure A.1: Search Cost Distribution Fitted Points - Scatter Plot



A.3.2
Alternative Specifications

Consider the parameters $\Theta_{1,m}$, $\Theta_{2,m}$, $\Theta_{3,m}$, $\Theta_{4,m}$ which are given by:

$$\Theta_{1,m} = \theta_0 + \theta_1\, Avg\, Distance_m + \theta_2\, \#\, Stations_m + \theta_3\, Car\, Fleet/People_{mun},$$

$$\Theta_{2,m} = \vartheta_0 + \vartheta_1\, Avg\, Distance_m + \vartheta_2\, \#\, Stations_m + \vartheta_3\, Car\, Fleet/People_{mun},$$

and $\Theta_{3,m} = \Theta_3$ and $\Theta_{4,m} = \Theta_4$ for all market m, i.e., these two last parameters do not vary by market. Additionally, let $a \in [0,1]$ be a weight parameter.

In the NLS estimation, we tested the following alternative specifications:

One Distribution

(1) $c^m \sim \Gamma\,(\Theta_{1,m}, \Theta_{3,m})$, where the first argument is the shape parameter of the gamma distribution and the second argument is the rate;

(2) $c^m \sim Logistic\,(\Theta_{1,m}, \Theta_{3,m})$, where the first argument is the location parameter and the second the scale;

(3) $c^m \sim Weibull\,(\Theta_{1,m}, \Theta_{3,m})$, where the first argument is the scale parameter and the second the shape;

(4) $c^m \sim Lognormal\left(\Theta_{1,m}, \Theta_{3,m}^2\right)$, where the first argument denotes the location parameter and the second the shape.

Combination of Two Distributions

(5) $c^m \sim a\Gamma\,(\Theta_{1,m}, \Theta_{3,m}) + (1-a)\,\Gamma\,(\Theta_{2,m}, \Theta_{4,m})$;

(6) $c^m \sim aLogistic\,(\Theta_{1,m}, \Theta_{3,m}) + (1-a)Logistic\,(\Theta_{2,m}, \Theta_{4,m})$;

(7) $c^m \sim aWeibull\,(\Theta_{1,m}, \Theta_{3,m}) + (1-a)Weibull\,(\Theta_{2,m}, \Theta_{4,m})$.

The results are shown in Table A.4 below. All the MSEs are higher in comparison to our chosen specification. The standard errors are shown in parentheses.

Table A.4: NLS Estimates - Alternative Specifications

	Model						
Coef.	(1)	(2)	(3)	(4)	(5)	(6)	(7)
a					0.624***	1.156***	4.31***
					(0.004)	(0.004)	(1.145)
θ_0	0.197***	15.773***	21808.377***	7.835***	10.346	11.145***	1.672***
	(0.004)	(0.148)	(5578.099)	(0.141)	(7.702)	(0.103)	(0.605)
θ_1	-0.0002***	-0.014***	-149.719***	-0.007***	0.045	-0.002	0.009***
	(2.4E-5)	(0.002)	(46.811)	(0.001)	(0.128)	(0.002)	(0.003)
θ_2	-0.001***	-0.093***	-253.315***	-0.057***	-0.064	-0.03***	0.19***
	(2.3E-5)	(0.001)	(66.198)	(0.001)	(0.042)	(0.001)	(0.023)
θ_3	0.017***	1.307***	56797.826***	0.632***	4.622	-0.096	-0.482**
	(0.003)	(0.254)	(16635.813)	(0.149)	(6.444)	(0.185)	(0.224)
Θ_3	0.00007***	8.312***	0.149***	7.646***	0.307	6.474***	0.262***
	(1.0E-5)	(0.081)	(0.004)	(0.143)	(0.278)	(0.051)	(0.006)
ϑ_0					1.871***	0.081	0.04
					(0.064)	(0.088)	(0.084)
ϑ_1					0.017***	-0.003*	0.003**
					(0.001)	(0.002)	(0.001)
ϑ_2					-0.014***	0.138***	0.102***
					(0.001)	(0.004)	(0.008)
ϑ_3					-0.102	-0.381*	-0.098
					(0.086)	(0.2)	(0.06)
Θ_4					0.734***	0.687***	0.296***
					(0.036)	(0.032)	(0.009)
MSE	0.01	0.008	0.011	0.01	0.024	0.007	0.007

Notes: MSE denotes the Mean Squared Error. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.