Price dispersion, chain heterogeneity and search in online grocery markets*

Xulia González[†]& Daniel Miles-Touya[‡]

University of Vigo

Abstract

This paper aims to identify patterns of cross-sectional and temporal price dispersion—in the Spanish grocery retail market—and also to evaluate the extend to which search costs and chain heterogeneity contribute to explain such dispersion. We build a data set comprising 836,074 prices for the most popular grocery products sold by Spanish national chains at different locations. Our results show that price dispersion is still present (albeit to a lesser extent) even after controling for chain heterogeneity and that it persists over time. We structurally estimate search costs distributions for different baskets of goods sold at several geographical markets. We find that the extent of search is low: more than 70% of consumers do *not* compare prices among supermarkets and only 4% check the prices at all stores. We also find that the products more frequently purchased have lower search costs and lower price–cost margins.

Keywords: price dispersion, retail food market, pricing strategies, search costs

JEL classification: D4, D83, L11

---

*We thank J. L. Moraga, M. J. Moral and C. Pazó for their comments and suggestions as well as Claudia Martínez-Boubeta for excellent research assistance. The paper has also benefited from presentations at Munich (EARIE 43rd annual conference) and Alicante (XXX Jornadas de Economía Industrial conference). We thank Marta Esteve and Diego Kuperman for providing data and technical assistance. Financial support from Ministerio de Economía y Competitividad (ECO2015-69334-R) is gratefully acknowledged.

[†]Facultad CC. EE y EE., Universidad de Vigo. E-mail: xgzlez@uvigo.es

[‡]Facultad CC. EE y EE., Universidad de Vigo. E-mail: dmiles@uvigo.es

1. INTRODUCTION

Price dispersion arises when stores in the same market set different prices for the same homogenous good (Hopkins, 2008). Different sources of heterogeneity among sellers as well as information frictions contribut to explain much of the price dispersion observed in many markets (Lach, 2002; Baye *et al.*, 2006). First, store differentiation (e.g., discount retailers versus well–established national chains) may lead to some stores persistently selling their products at a lower price level than others in the same market. Second, when some consumers are perfectly informed about market prices while others are not, price dispersion could arise as a mixed strategy equilibrium outcome (Stigler, 1961; Varian, 1980).[1] As a result, we might observe retailers frequently change position in the price rankings, thus making more difficult for consumers be aware of the best deal in each period.

Price dispersion is a common feature in the grocery retail sector (Zao, 2006). Supermarket chains can differs in the services offered (e.g. product assortment, shipping services, return policies, web design, etc.) and, at the same time, unannounced, short-term reductions in the prices of certain products (sales) is a frequent pricing strategy. These empirical evidences suggest that both chain differentiation and search frictions may help explain price dispersion in grocery markets.

The objective of this empirical paper is twofold. We first analize price dispersion patterns for a wide range of grocery products sold online by the main supermarket chains operating in Spain. Second, we structurally estimate search costs —while accounting for chain heterogeneity— for different baskets of goods sold in different geographical markets. We follow the approach of Hong and Shum (2006) and Wildenbeest (2011) which allow estimating search cost using only price data while account for vertical product differentiation, i.e., goods are homogeneous but chains differentiate them by offering additional services.

One should expect that *online* shopping reduces price dispersion as store differentiation

[1] Subsequent theoretical studies include Rosenthal (1980), Burdett and Judd (1983), Carlson and McAfee (1983), Stahl (1989), Dana (1994), Baye and Morgan (2001), Janssen and Moraga-González (2004), and Janssen *et al.* (2005).

turns out to be less relevant and/or consumer search costs on the Internet are significantly lower (Jansen *et al.*, 2007).[2] Nevertheless, empirical evidence shows that price dispersion remains even when consumers can easily access price information on homogeneous products from retailers' web or price comparison websites (Brynjolfsson and Smith, 2000; Baye *et al.*, 2004a, b; Ellison and Ellison, 2009). On the one hand, Brynjolfsson *et al.* (2010) highlight that retailer differentiation remain on the Internet and show the significance —for even homogeneous physical products—of nonprice factors. On the other, empirical findings suggest that search costs continue to be relevant in online grocery markets (e.g.,Wildenbeest, 2011).

In Spain, online grocery sales represent a low but growing share of the total grocery market. Most brick-and-mortar chains have already committed to this *e-grocery* segment, and online-only e-grocers (e.g., Amazon Pantry) have also entered the market. In addition, supermarket chains are increasing the number of services offered online.[3] In sum, there is growing evidence that grocery shopping are increasingly online-reliant (Hartman, 2012).

For our empirical analysis we build a novel data set consisting of prices for more than 200 products sold online by the main Spanish supermarket national chains at different locations. The data cover a period of 182 days—from 1 October 2013 through 31 March 2014—and offer a total of 836,074 price observations.

Competition among supermarket stores is highly local, even online shopping. Supermarket websites might differ among locations in terms of product availability and prices.[4] We therefore consider four distinct markets geographically distant from each other: Madrid, Barcelona, Málaga, and Vigo. Consumers usually care more about the price of a basket of

[2] Consumer access to prices through supermarket websites or price comparison sites may reduce search costs for online and offline retail consumers both, as customers can easily check prices before or even while they shop. Baye et al. (2013) examine the evolution of product search since the pre-internet era.

[3] Consumer options offered by supermarkets include: "order and deliver" as well as "order online and store pickup". Supermarket websites can save customers' shopping lists to facilitate subsequent shopping. Indeed, the consumer can use a cell phone to price-check products while inside a brick-and-mortar retail store. The expansion of delivery areas has also induced more consumers to shop for groceries online.

[4] A shopper who logs on to a supermarket's website must check to see whether delivery is available in her postal zone; since these zones are matched up to the nearest warehouse, both products and prices may differ by location.

goods than that about the price of a single item. For this reason we estimate search costs using prices of two (different) baskets of goods[5]. Although each basket contains of a set of identical products (at the bar-code level), online grocery chains remain heterogeneous in terms of reputation or level of services.[6] These characteristics, though strongly similar within a given chain's outlets, differ from one chain to another.

As a preview of our results we observe that price dispersion is still present (albeit to a lesser extent) even after we control for chain heterogeneity and that it persists over time. We show that chain heterogeneity explains much of the observed price dispersion and, at the same time, supermarkets' relative position in the price ranking does not remain constant over time. This implies that consumers must update information quite frequently if they want to know the best deal. The estimated extent of search is low: across markets, more than two thirds of consumers search just once even though estimated search costs are low. Our results also establish that the products more frequently purchased tend to have lower search costs and also lower price–cost margins. These findings accord with those in the literature on the retail food market in other countries (e.g., France, the United Kingdom).

The rest of this paper proceeds as follows. In the next section we discuss several empirical works that measure price dispersion and search cost. Section 3 describes our data, and Section 4 presents empirical evidence of price dispersion in Spanish grocery markets. In Section 5 we provide details of our model and estimation strategy; the estimation results follow in Section 6. We conclude in Section 7.

2. EMPIRICAL LITERATURE ON PRICE DISPERSION AND SEARCH COST

There are many markets in which price dispersion is observed. Empirical papers that measure price dispersion typically focus on a single product category; examples include

[5] The first basket consists of basic and frequently purchased products; the second includes different type of beverage products (spirits, alcoholic, etc...). This basket includes items that are purchased occasionally.

[6] Chains also affect price-setting behavior at the store level (Nakamura *et al.*, 2011; Elberg, 2014; Lloyd *et al.*, 2014).

orange juice (Berck *et al.*, 2008), gasoline (Barron *et al.*, 2004; Chandra and Tappata, 2011), books and CDs (Brynjolfsson and Smith, 2000; Brynjolfsson *et al.*, 2010), spare parts for cars (Delgado and Waterson, 2003), computer and electronic products (Baye *et al.*, 2004a, b; Ellison and Ellison, 2009), and plane tickets (Bachis and Piga, 2011; Orlov, 2011).[7]

Some studies compare price dispersion across different products with the goal of establishing empirical regularities that could help to identify the *sources* of price dispersion. For example, Lach (2002) compares price dispersion among goods of different price levels and finds that products with higher prices exhibit greater price dispersion. Sorensen (2000) compares prescription drugs purchased at different frequencies and reports that price dispersion is negatively correlated with the frequency of purchase.

The price data used in empirical papers come from different sources. First, data can be obtained from the price of transactions at retail outlets (gas stations, supermarkets, travel agencies, etc.). Second, some papers use the (monthly) price data collected to develop consumer or industrial price indexes. Many of these works address price rigidities and therefore analyze the frequency of price changes (for a survey, see e.g. Klenow and Malin, 2010), yet others use price data to analyze price dispersion (e.g., Lach, 2002). Third, the growth of the Internet has yielded new sources of data and thereby facilitated microeconomics research (Edelman, 2012). Thus recent years have seen increased numbers of authors using Internet-sourced data—either obtained directly from retailer websites or indirectly from price comparison websites. See, for example, Clay *et al.* (2001), Brown and Goolsbee (2002), Baye *et al.* (2004a, b), Ellison and Ellison (2009), and Cavallo (2016a, b). Still other authors combine online and traditional market data to compare prices for a given product in both markets (Orlov, 2011) or to identify consumption patterns (Pozzi, 2012).

There is a growing literature on the estimation of search costs. Burdett and Judd (1983) propose a nonsequential search model under which price dispersion can be sustained in equilibrium if some consumers observe multiple prices while other consumers observe only one price; the asymmetric distribution of price information is due to search costs. Departing from this idea, Hong and Shum (2006) propose a model for estimating search cost

[7] For a review of papers on price dispersion, see Baye *et al.* (2006).

distributions—by means of an empirical likelihood estimation procedure—when only price data are observed. Moraga-González and Wildenbeest (2008) modify Hong and Shum's approach by introducing a maximum likelihood estimator. Wildenbeest (2011) allows for vertical product differentiation, where goods are assumed to be homogeneous but stores differentiate themselves by offering additional services. Wildenbeest (2011) estimates how well vertical product differentiation and search costs explain price dispersion in the grocery retail industry; the estimation is based on a basket of grocery items from the four leading UK retailers over a 12-week period in 2008. In that paper, each firm has its own price distribution because firms are heterogeneous in terms of overall quality; this set up accounts not only price dispersion across firms but also the observation that some firms have persistently higher average prices than others. The results shows that 61% of the observed price variation is explained by supermarket heterogeneity and also that estimated search intensity is low, 71% of consumers visit only one supermarket. Using this approach, Richards *et al.* (2016) employ online grocery-price data from four large retailers in the United Kingdom to estimate search costs for consumers who engage in multi-product shopping. Their results show that 84% of consumers search only one store when searching for products in many categories at the same time. And this percentage is even higher when considering that consumers serch in only one category.

Other papers that empirically estimated search costs are, for example, Hortaçsu and Syverson (2004) that analyze the mutual funds industry; these authors assume that consumers have identical tastes but different search costs. De los Santos *et al.* (2012) employ a large data set on Web-browsing and purchasing behavior to test the extent to which consumers search in accordance with various search models. Dubois and Perrone (2014) extend the model of Hortaçsu and Syverson to allow for heterogeneous consumer preferences; thus, products are differentiated both vertically and horizontally. Dubois and Perrone use data also from observed shopping behavior, since they examine all store visits made by households within a certain period of time. Seiler (2013) proposes a structural model, with imperfect information that takes into account inventory holdings and searching. Using a consumer-level panel data set (Kantar Worldpanel UK) for laundry detergent products, Seiler reports that consumers unaware of the price of this product on 70 % of their shopping trips.

3. DATA DESCRIPTION

Our data set (summarized in Table A1 of the Appendix) consists of the prices (in euros) of 237 products as posted on the websites of the main Spanish supermarket chains (Auchan, Carrefour, El Corte Inglés, Eroski, and Mercadona) and of a regional chain (Condis) that operates only in Catalonia.[8] The aggregate revenues of the five national chains accounted for 72.2% of the net sales of main food retailing groups (CNC, 2011).[9] Daily prices are observed, from 1 October 2013 through 31 March 2014 (182 days), at four locations across Spain. The two most populous cities in Spain (Madrid and Barcelona) in addition to two medium-sized cities located in the northwest and south of Spain (Vigo and Málaga, respectively).[10] We collect a total of 836,074 price observations that include the most popular branded products in the following categories: beverages, breakfast and cereals, coffee and cocoa, milk products, pantry, and household and personal care (neither fresh nor frozen products are included). Only national brands are included in order to ensure availability in all markets. The resulting analysis affords unprecedented insight on the prices faced by supermarket shoppers in Spain.

The data set identifies products in great detail. Two products are considered to be distinct if they have different bar codes; for example, whole milk and low-fat milk of the same brand are considered different products for which separate prices are recorded. We also identify and distinguish between the same product being sold either in packs or individually. We

[8] Price data were collected by Soysuper.com, a price aggregator that gathers data from websites of the main Spanish supermarket chains.

[9] According to the Comisión Nacional de la Competencia (CNC, 2011), Mercadona is the leading operator in Spain, followed by Carrefour, Eroski, Alcampo (Auchan) and El Corte Inglés. See García and Delgado (2012) for an analysis of the competence in this sector.

[10] Madrid and Barcelona are home to (respectively) 3,165,235 and 1,602,386 inhabitants. Málaga and Vigo are, respectively, located in the south (Andalucía region) and northwest (Galicia region) of Spain and have 566,913 and 294,997 residents. Smaller cities or towns are not represented because the top five chains. Average income among the four cities differs, but not by as much: the average income are: Barcelona is €35,090; in Madrid, €36,636; in Málaga, €24,405; and in Vigo, €29,654.

identify each chain-product-store combination with a specific product code (SPC); thus, for example, a 100-gram jar of Nescafé instant coffee stocked by Auchan and Carrefour in Madrid and Barcelona is represented by four distinct SPCs, each with its own time series of daily prices. Thus identical products have different SPCs depending on the store by which they are stocked. The selection of products was based on two criteria. First, we selected the products most often included on Spanish consumers' shopping lists as ranked by the Soysuper.com popularity index.[11] The index reflects consumer searches on this comparison website, and it accounts for how often a consumer adds the product to a shopping basket and it is bought in any supermarket. Second, among these products we retained only those that were simultaneously available in at least four supermarkets throughout the period of our study.

This data set has several advantages. First, prices are the actual ones paid by consumers; that is, there are no "bait and switch" strategies whereby low prices lure consumers to a website that then steers them to a higher-priced product (Ellison and Ellison, 2009). Second, the availability of daily price observations allows us to identify short-term price movements more accurately. Third, the data set covers four geographically independent markets in which all national chains operate online; this feature enables us to assess the heterogeneity of search costs across markets. Finally, website and store prices are generally in agreement, so our findings are not limited to the online market.[12]

4. EVIDENCE OF PRICE DISPERSION

Varian (1980) was the first to distinguish between spatial and temporal price dispersion. *Spatial* dispersion implies that, even if a firm charges different prices at different times, its position in the price distribution need not change over time. In our context, any permanent

[11] We thank Soysuper.com for sharing its list of the most popular products.

[12] Some chains make the direct claim that these prices are the same. More formally, Cavallo (2016b) collects online and offline prices for more than 40 of the largest multi-channel retailers in 10 countries. This author finds that, for those retailers, "there is typically little difference between the online price collected from a website and the offline price obtained by visiting the physical store."

price differences could be explained by store differentiation because price dispersion should diminish over time if consumers can learn (from experience) that some firms consistently offer lower prices than others. *Temporal* price dispersion arises in the mixed-strategy case, when stores randomize their prices and so the store's position in a price-based ranking will change over time.

4.1. Spatial price dispersion

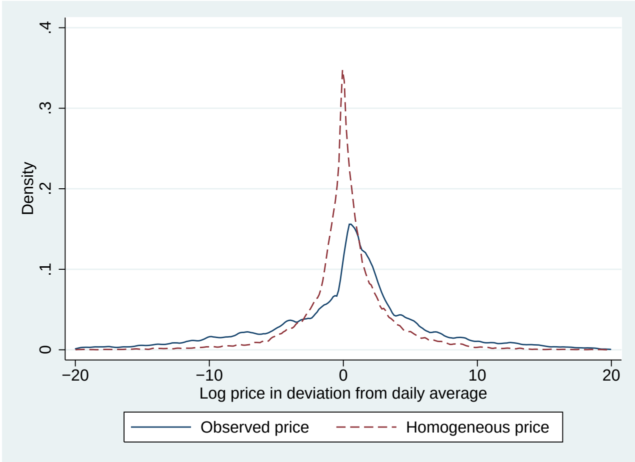
We measure price dispersion as (the log of) deviations in price from the daily mean (cf. Lach, 2002); that is, $g_{ijlt} = \log p_{ijlt} - \log \bar{p}_{ilt}$. Here p_{ijlt} denotes the price of item i in chain j at location l on day t, and $\bar{p}_{ilt}$ the *mean* price of item i at location l on day t, for $i = 1, \ldots, 237$, $j = 1, \ldots, 6$, $l = 1, \ldots, 9$, and $t = 1, \ldots, 182$.

Part of the dispersion observed in prices may be explained by product differentiation. Although we compare goods with the same physical characteristics, these goods are sold at different stores and also on different days. As a result, the products are not homogenous. We assume that identical products in different locations are *not* substitutes—since the distance between cities is great enough that we can view them as distinct markets.

We remove from prices any heterogeneity due to the supermarket chain, location, or time. For this purpose we run product-by-product regressions of prices (measured as log deviations from the daily mean) on the fixed effects of supermarket chain and location, on the interaction between those effects, and on a time effect: $g_{ijlt} = \mu + \alpha_j + \beta_l + \alpha_j \times \beta_l + \gamma_t + \varepsilon_{skt}$; here α_j is a supermarket chain fixed effect that is common across products and locations, β_l is a location fixed effect, and γ_t is a day effect. The residuals of these regressions represent the price of a homogeneous product after controlling for time-invariant, store-specific effects and for the price fluctuations common to all stores. This method of deriving homogeneous prices is now standard in the literature (see e.g. Sorensen, 2000; Lach, 2002; Zhao, 2006; and Dubois and Perrone, 2014).

Figure 1 plots the empirical density function for the spatial dispersion of observed prices (solid line) and of residual prices pooled over products, stores, and days (dashed line).[13]

Figure 1. Kernel density of spatial price dispersion



This figure reveals that prices exhibit considerable spatial dispersion. Most observed prices fluctuate between −10% and +10% of the mean and, as expected, are relatively more dispersed than are the residual prices. The empirical literature has documented that the magnitude of price dispersion varies as a function of product characteristics (price level, purchase frequency, etc.). Table 1 reports several measures of price dispersion for six product categories.

Table 1. Moments of the price dispersion distribution (by category)

	Std. Dev.		Posted price		Residual price	
Categories	g_{ijlt}	ε_{ijlt}	p5	p95	p5	p95
Beverages	6.92	4.41	−12.7	9.2	−6.7	6.4
Breakfast and cereals	8.12	3.27	−15.3	12.4	−4.0	4.4
Coffee and cocoa	6.48	4.42	−10.7	10.3	−6.9	6.2
Milk products	5.63	3.49	−9.2	7.1	−5.2	4.6
Pantry	9.54	4.44	−17.6	12.8	−6.0	6.2
Household and personal care	7.26	3.21	−12.8	12.6	−4.6	4.6
Total	7.23	3.98	−12.8	10.3	−5.6	5.6

The table's first two data columns give the standard deviations of the observed and residual prices, and the last four columns report the 5th and 95th percentiles of each distribution. For all categories, spatial price dispersion drops significantly—in terms of both the standard

[13] Because our measure of price dispersion averages to zero each day, all variation in the kernel density is cross-sectional for each day and at each location.

deviation and the quantiles—when we control for observed product heterogeneity. Nevertheless, some dispersion persists even when prices are homogeneous. Results are somewhat heterogeneous with respect to product categories. The pantry and the breakfast and cereals categories exhibit the most price dispersion, since their standard deviations are the highest; prices for milk products exhibit the least dispersion.

4.2. Temporal price dispersion

We observe temporal price dispersion when the identity of the store offering the lowest (and highest) price varies over time. If stores' relative positions in that price distribution remain constant, then dispersed prices could simply reflect store heterogeneity. However, if the ranking by prices varies over time then price dispersion could be the outcome of costly consumer search. In this latter case, firms change their prices every period to hinder buyers from identifying the cheapest store.

Temporal price dispersion is measured by comparing how stores' prices are ranked over time. We define rr as the price-rank position changes for a given product i sold at supermarket j over τ days.[14] Formally, $rr_i^j = \frac{1}{T}\sum_{t=\tau+1}^{T} I(r_{it}^j \neq r_{it-\tau}^j)$; here r_{it}^j denotes the position of the supermarket j in the price ranking of item i on day t in a given location, and $r_{it-\tau}$ is its position τ days earlier. Now we can define the *average ranking change* for supermarket j at a given location as the average of rank changes among all products: $RR^j = \frac{1}{N}\sum_{i=1}^{N} rr_i$. Although rank changes on two successive days may be small, the extent of change is increasing in the time interval.

Table 2 reports the average rank changes in prices between two observations with a time lag of 1, 10, and 30 days in Barcelona, Madrid, Málaga, and Vigo. A change in ranking position indicates a price change in at least one chain's stores. However, a given supermarket can change a price without its rank changing and can also see its rank change without changing any prices—that is, provided some competitor does.

[14] See Chandra and Tappata (2008) for a similar rank-reversal statistic.

Table 2. Average change in price ranking at four locations

	Barcelona			Madrid			Málaga			Vigo		
	RR_1	RR_{10}	RR_{30}	RR_1	RR_{10}	RR_{30}	RR_1	RR_{10}	RR_{30}	RR_1	RR_{10}	RR_{30}
Auchan	1.3	9.8	15.3	1.0	6.4	9.6	1.4	9.8	14.2	1.4	9.1	12.1
Carrefour	2.9	21.4	35.3	2.6	18.6	32.4	2.7	19.4	33.2	3.4	24.8	38.0
Condis	3.0	22.8	37.1									
El Corte Inglés	3.0	21.9	36.8	2.8	19.6	33.6	3.0	21.3	36.8	3.5	25.5	38.9
Eroski	3.3	23.1	38.5	2.8	19.9	33.9	3.1	21.5	36.3	3.5	25.8	39.4
Mercadona	2.8	21.1	35.1	2.6	18.3	32.0	2.9	20.9	34.2	3.5	25.4	39.6

The table shows that changes in rank become greater as the time lag increases. The four first columns (one under each location) show that the average percentage change in price ranking between two consecutive days is never more than 3.5%. Yet with a 10-day (resp. 30-day) time lag, the average percentage change approaches or exceeds 20% (resp. 35%). The exception to this pattern is Auchan, which is in the same position of the ranking most of the time. And as we will see later their prices are always the lowest.[15]

Price rankings that fluctuate make it more difficult for consumers to find the best deal, and chances are nearly 4 in 10 that relative prices for a given product will have changed within a month. Hence consumers must update their information about prices rather frequently if they want to continue paying the least amount possible.

5. ESTIMATING CONSUMER SEARCH WITH PRICE DATA

The empirical evidence described so far suggests that part of the dispersion observed in prices may be explained by store differentiation, although our evidence is also consistent with imperfect information about prices. To the extent that supermarkets change their prices and thus their positions in the ranking, it is more difficult for consumers to learn about prices. In this section we estimate search costs as well as the proportion of consumers who compare some prices when grocery shopping; for that purpose we use the available information about prices while accounting for supermarket chain heterogeneity.

[15] Annual reports on supermarket prices published by the Spanish consumer organization (OCU) claim that Auchan has been the lowest-price national chain in Spain in last years: https://www.ocu.org/consumo-familia/supermercados/informe/cadenas-mas-baratas and http://cincodias.com/cincodias/2014/12/26/finanzas_personales/1419594524_553238.html (both accesed 5 October 2016).

We continue to assume that consumers are more interested in the price of a basket of products than in individual product prices, and we assemble two different baskets of goods. The *basic* basket includes those branded products—from various categories—most often included on consumer shopping lists; and, the *occasional* basket includes only beverage products (alcoholic and nonalcoholic drinks). We estimate the search costs for these two baskets in four geographical markets.

5.1. The model

In estimating search costs, we use the nonsequential search model developed by Wildenbeest (2011). Suppose there are N supermarkets offering an homogeneous good—a basket of groceries—to imperfectly informed consumers at a particular location. Supermarkets sell this "good" at a unit cost of r_j. Firms compete directly in the utility space, which implies that the supermarket strategy space is reduced from two dimensions (quality and price) to a single utility dimension. This approach enables us to incorporate chain differentiation into the model.

Consumers share a common utility function and have identical preferences regarding quality; however, their search costs differ (as in Hortaçsu and Syverson, 2004). We can write

$$u_j = v_j - p_j \quad \text{for } j = 1, \ldots, N,$$

where v_j is the consumer's valuation of buying the good from supermarket j at a given location. This valuation has the additively separable structure $v(s_j) = x + s_j$; here x denotes the common consumers' valuation of the homogeneous good independent of store quality, s_j is the supermarket's level of services or quality (Wildenbeest, 2011), and p_j is the corresponding price.

Because the consumers in our setup all have the same utility function, a supermarket determines its quality level s_j by maximizing the price marginal cost margin: $p_j - r_j = p(s_j) - r(s_j) = v(s_j) - r(s_j) - u$, for a given utility level u. By Euler's theorem, the total cost of quality inputs exhausts quality-related output; that is, $r(s_j) = s_j$.[16] As a consequence, the valuation cost markup does *not* depend on store quality: $v(s_j) - r(s_j) = x + s_j - r(s_j) = x$. We can therefore focus on symmetric mixed-strategy equilibria in utility levels, where the supermarket's strategy is give by a common utility distribution function $L(u)$.

Consumers search nonsequentially. More specifically, consumers take supermarkets' strategies as given and decide on the optimal number $k \geq 1$ of stores to visit, after which they buy from the store that yields the highest utility.[17] A consumer's search cost c is assumed to be a random draw from a common and atomless distribution $G(c)$ with support $(0, \infty)$ and positive density $g(c)$.

Consumer search behavior should be optimal in this sense: the net benefit of searching k times should be greater than that of searching either $k-1$ or $k+1$ times; and the expected utility from searching should exceed its expected cost $(k \cdot c)$. The search cost of a consumer who is indifferent between searching k and $k+1$ times,

$$c_k = Eu_{1:k+1} - Eu_{1:k} \quad \text{for } \; Eu_{1:k} = E[\max(u_1, \ldots, u_k)],$$

is decreasing in k. The share q_k of consumers who sample k prices is then

$$q_k = \int_{c_k}^{c_{k-1}} g(c)\, dc = G(c_{k-1}) - G(c_k),$$

which implies the following inequalities:

$$\begin{aligned}
&q_1 = G(c_0) - G(c_1) = 1 - G(c_1),\\
&q_k = G(c_{k-1}) - G(c_k) = 1 - \sum_{k=1}^{k-1} q_k - G(c_k) \quad \text{for } \; k = 2, \ldots, N-1,\\
&q_N = G(c_{N-1}).
\end{aligned}$$

Here $G(c_0) = 1$, so every search cost is lower than c_0; and $G(c_N) = 0$, so c_N is the minimum search cost.

[16] Under perfect competition, supermarkets obtain quality input factors and the quality production function exhibits constant returns to scale.

[17] A condition that partially characterizes a symmetric equilibrium in the utility space is that some consumers search only once (at no cost) while others search more than once (see Burdett and Judd, 1983).

Supermarket j's expected profit from offering utility level u_j is

$$\pi_j(u_j; L(u)) = (x - u_j) \sum_{k=1}^{N} \frac{k}{N} q_k L(u_j)^{k-1}$$

given expected consumer behavior q_k and the distribution function $L(u)$. Here $x - u_j = p_j - r_j$ is the price–cost margin of each supermarket that implicitly sets a price p. The summation captures the expected sales quantity, which depends on: (i) the proportion q_k of consumers searching k times; (ii) the likelihood k/N of consumers observing the utility of firm j; and (iii) the probability $L(u_j)^{k-1}$ that, at utility level u_j, the firm offers the highest utility level of all the k firms searched.

In this setting, a price dispersion equilibrium is possible only when there exists a positive (though not certain) likelihood of a consumer observing only one price.[18] The characterization of the equilibrium utility distribution in mixed strategies implies that the supermarket should be indifferent about the utility level of utility to set in the support of $L(u)$ (Burdett and Judd, 1983). In particular, the supermarket should be indifferent between (a) offering a utility of zero by setting $\bar{p}_j = v_j$ and selling only to uninformed consumers (i.e., those who search just once) and (b) setting any other utility level in the support of $L(u)$, $u > 0 = \underline{u}$:

$$(x - u) \sum_{k=1}^{N} \frac{k q_k}{N} L(u)^{k-1} = x \frac{q_1}{N},$$

where the right-hand side is the expected profit when offering zero utility. In this case, if firms offer the maximum utility ($u = \overline{u}$) then $L(\overline{u}) = 1$ (all utility levels are below the maximum); hence the maximum utility is given by

$$\overline{u} = x \frac{\sum_{k=2}^{N} k q_k}{\sum_{k=1}^{N} k q_k}.$$

Our aim is to estimate the points $\{q_k, c_k\}_{k=1}^{n}$ via maximum likelihood, as in Moraga-González and Wildenbeest (2008). The equilibrium distribution of utilities, $L(u)$, can only be implicitly defined, but the density function can be derived from the first-order conditions

[18] The intuition for this claim is that, if all consumers did compare the prices at different stores, then each firm would set a price equal to their unit cost.

of expected profit maximization:

$$\frac{\partial\pi}{\partial u}=0 \quad \text{and} \quad l(u)=\frac{\sum_{k=1}^{N} kq_k(L(u))^{k-1}}{(x-u)\sum_{k=2}^{N} k(k-1)q_k(L(u))^{k-2}};$$

these conditions are then used to define the log-likelihood function $LL=\sum_{i=2}^{T}\log l(u_i)$. Here T is the total number of observations, the minimum utility is zero, and all utilities are arranged in ascending order: $\underline{u}=u_1<u_2<\cdots<u_T=\overline{u}$. Thus we have $L(\overline{u})=1$ and $L(\underline{u})=0$.

We can use this characterization of optimal searching behavior to rewrite the search cost as follows:

$$\begin{aligned}
c_k &= \int_{\overline{u},\underline{u}} (k+1)uL(u)^k l(u)\,du - \int_{\overline{u},\underline{u}} kuL(u)^{k-1} l(u)\,du \\
&= \int_{\overline{u},\underline{u}} u[(k+1)L(u)^k - kL(u)^{k-1}]l(u)\,du \\
&= \int_{\overline{u},\underline{u}} u[(k+1)L(u) - k]L(u)^{k-1}l(u)\,du.
\end{aligned}$$

Further simplification is possible if we put $y=L(u)$, so that $dy=l(u)\,du$ and $u=L^{-1}(y)=u(y)$; then

$$c_k=\int_0^1 u(y)[(k+1)y-k]y^{k-1}\,dy.$$

Now using the same change of variable in the equilibrium profit equation y yields

$$(x-u)\sum_{k=1}^{N}\frac{k\mu_k}{N}y^{k-1}=x\frac{\mu_1}{N},$$

from which the equality $u(y)=x-\frac{x\mu_1}{\sum_{k=1}^{N}k\mu_k y^{k-1}}$ follows.

5.2. Estimation strategy

The first step is to estimate utilities. Toward that end, we assume that consumers differ in their search costs but have the same preferences regarding chain characteristics.[19] Thus consumers in a given location derive utility from buying the homogeneous basket at store j

[19] This assumption implies, for example, that all consumers would prefer shopping at Eroski or El Corte Ingles than at Auchan if prices were identical across supermarket chains.

according to $u_j = v_j - p_j$, where v_j is the valuation of buying the basket at store j and p_j is the corresponding price. Utilities are then defined as prices adjusted by the heterogeneity between stores (services provided, quality,...). Consumers know their valuation of a good but do not know its price. Hence they must obtain information about basket prices at a number of supermarkets, according to their search costs.

We rewrite the preceding paragraph's equation as $p_j = v_j - u_j$, which can be estimated via a fixed-effects regression on prices. Thus $p_j = \alpha + \gamma_j + u_j$, where α is a constant, γ_j is a store fixed effect, and the (negative of) disturbance u_j represents utility.

After deriving the estimated utilities, we proceed as in Moraga-González and Wildenbeest (2008).[20] The maximum likelihood (ML) problem is given by

$$\max_{\{q_k\}_{k=1}^{N-1}} \sum_{m=2}^{M} \log l(u_m; q_1, \ldots, q_N).$$

The term M is the number of price data points at each location, and $L(u_j)$ solves

$$(x - u_j) \sum_{k=1}^{N} \frac{kq_k}{N} L(u_m)^{k-1} = x\frac{\mu_1}{N} \quad \text{for all } m = 2, 3, \ldots, M-1$$

and

$$x = \overline{u} \frac{\sum_{k=1}^{N} kq_k}{\sum_{k=2}^{N} kq_k};$$

and using the fact that $q_N = 1 - \sum_{k=1}^{N-1} q_k$. That is, the ML estimator yields estimates of the proportion q_k $(k = 1, \ldots, N)$ of consumers who are searching, and we can recover the search costs from those estimates.[21]

6. ESTIMATION RESULTS

Grocery products are usually purchased in shopping baskets, therefore, consumers are more interested in the price of a basket of goods than in a single item. For estimation purposes

[20] The estimates are derived while assuming that firms play a stationary repeated game of finite horizon, so the data in each period should reflect the equilibrium of the static game analyzed previously (cf. Moraga-González and Wildenbeest, 2008). For the estimation we select price observations every 10 days to avoid serial correlation.

[21] Maximum likelihood estimator was programmed in MATLAB.

we construct two different baskets. In what follows we describe both the "basic" and the "occasional" basket, presenting estimation results for each in turn.

6.1. Basic basket

The basic basket of goods includes the most frequently bought products (according to Soysuper.com) in each of the six product categories—beverages, breakfast and cereals, coffee and cocoa, milk products, pantry, and household and personal care—for which we have a complete series of prices across all the chains at all locations. Because all the products in this basket are branded, we can view it as a completely homogenous product. Supermarkets are assumed to be especially interested in the pricing of these highly popular products.

Spanish households spend an average of €168.76 monthly on food.[22] According to MARM (2011), consumers' shopping habits depend on which products they want to buy. Fresh products are purchased more often at specialized shops or small supermarkets, and consumers are more concerned about the quality than the price of such products. These purchases are examples of "secondary shopping", and they are made at the relatively high frequency of two or three times each week. In contrast, the standardized products purchased during "primary shopping" (beverages, dairy, cereals, etc.) are often bought in large quantities and at larger stores, such as supermarkets or superstores. Our shopping basket contains 23 products, most of which are purchased as part of primary shopping, and does not include fresh groceries; its average total cost is €79.[23]

Table 3. Price of the basic basket

	Price		
Supermarket	Mean	Max.	Min.
Auchan	74.77	78.87	70.95
Carrefour	79.69	80.47	78.45
Condis	79.89	80.80	78.58
El Corte Inglés	80.79	82.08	78.28
Eroski	80.38	81.54	78.90
Mercadona	79.17	80.62	76.01
Total	79.09	82.08	70.95

[22] This figure was obtained from the Spanish Household Budget Survey for 2013; it *excludes* fresh products.

[23] Table A2 (in the Appendix) reports the average price and the price dispersion of each product included in the basic basket.

According to Table 3, the price of the basket ranges from €70.95 (Auchan) to €82.08 (El Corte Inglés). The difference between the most and the least expensive basket is €11, but this difference varies across supermarkets. Auchan has the greatest intrafirm price dispersion, ranging between €71 and €79; the basket price at Carrefour exhibits very little variation (€78.5–€80.5). El Corte Inglés has the highest prices, which range between €78 and €82.

Figure 2. Price of the basic basket, by chain, at different locations

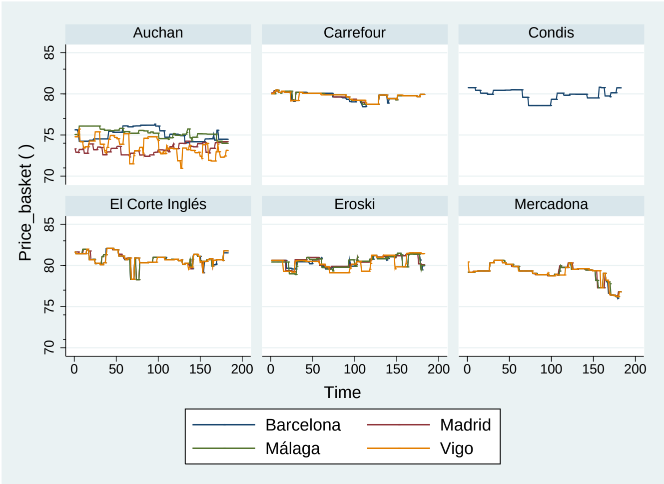
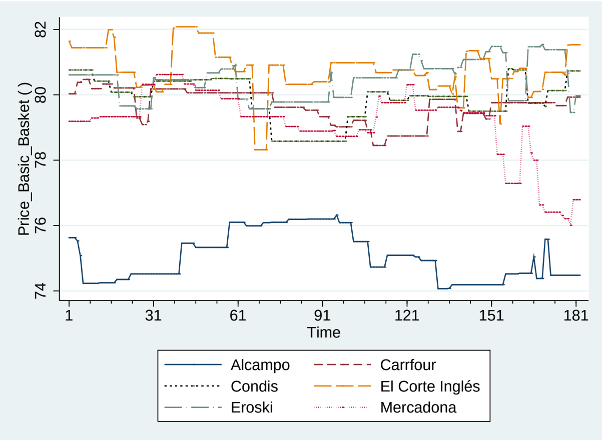


Figure 2 shows the daily price of the basket in each supermarket at the four locations. We observe substantial price differences across Auchan stores located in different cities, whereas price differences across stores in each of the four other chains are less pronounced. Differences among stores belonging to Mercadona are barely perceptible, which could indicate that most prices are centralized at the chain level.

Price dispersion of the basic basket If store characteristics do not change (at least in the short term) and if price dispersion reflects chains differentiation, then we should expect supermarkets to set prices in a manner that preserves their ranking position. Thus, high-quality supermarkets will nearly always set relatively high prices. Of course, if the position of the stores in the price distribution remains constant then it is easier for consumers to learn about prices.

Figure 3. Ranking over time, by price and estimated utility (Barcelona)

A. Price



B. Utility (est.)

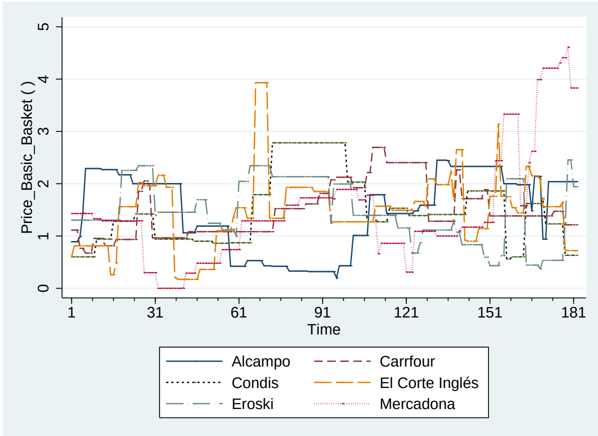


Figure 3 plots the daily price of the basket (Panel A) and estimated utility (Panel B) in Barcelona for the six supermarkets. These graphs confirm that some chains have consistently lower prices (and/or prices that change more frequently) than other chains. Auchan is clearly the supermarket with the lowest price for the basket, yet it changes prices rather frequently. The ranking of the other supermarkets varies, although El Corte Inglés usually has the highest prices and Mercadona the lowest—especially in the last month sampled, when the latter chain initiated a price-cutting campaign. Stores certainly change their prices often but do not change prices synchronously; that is, the length of time between price revisions differs across chains. Panel A reveals that price differences across chains persist even as stores adjust their price distributions. Panel B plots consumer utility calculated as the negative of the residuals of a regression of prices on store fixed effects; the resulting value can be interpreted as the price of the homogeneous good after controlling for chains heterogeneity. Figure 3B that no single supermarket yields utilities that are consistently higher or lower than others. Although the positions in the ranking of utilities seldom change daily, there is considerable fluctuation overall.

Table 4 Periods spend by prices and utilities in each quartile (Barcelona)

	Prices quartile				Utilities quartile			
Supermarket	q1	q2	q3	q4	q1	q2	q3	q4
Auchan	100	0	0	0	28.6	6.0	42.3	23.1
Carrefour	31.9	33.0	32.4	2.7	30.2	28.0	28.6	13.2
Condis	14.8	13.7	69.3	2.2	35.2	23.6	23.6	17.6
El Corte Inglés	4.4	3.3	29.1	63.2	30.8	12.1	47.3	9.9
Eroski	1.1	9.3	65.9	23.6	31.3	6.6	42.3	19.8
Mercadona	47.8	41.8	4.9	5.5	44.0	23.6	15.9	15.6

Table 4 reports the percentage of time that the price and the utility spend in each quartile (at the Barcelona location). The table's first four columns show that Auchan is the only supermarket that occupied the first quartile for the entire study period; in other words, it was always among the lowest-priced stores. For almost half the period, also Mercadona occupied the first quartile. At the other extreme, El Corte Inglés was among the most expensive supermarkets 69% of the time and among the least expensive only 4% of the time. Carrefour was located in the second quartile nearly a third of the time, and Condis and Eroski were most often located in the third quartile. Looking at the distribution of utilities (last four columns of the table), we can see that there is less concentration in particular quartiles because the distribution of utilities is more spread out. Furthermore, the utility distribution is right-skewed; that is, high levels of utility are less likely to be observed and lower levels of utility are the most frequent. This finding reflects the small share of consumers searching intensively in this market, as we shall discuss next.

These observed patterns lend support to the notion that price dispersion is due to mixed strategies in combination with chain heterogeneity. In other words, each store has its own price distribution from which to draw and—depending on the extent of firm heterogeneity—the respective supports of these distributions may well overlap.

Search cost estimates We use the basket prices to estimate the model's parameters.[24] We are mainly interested in first estimating the proportion of consumers who search and then calculating their search costs. The estimation results, which are obtained using the ML procedure described previously, are presented in Table 5.

Table 5. Estimation results

	Barcelona	Madrid	Málaga	Vigo
q_1	0.718	0.733	0.723	0.747
q_2	0.167	0.174	0.168	0.186
q_3	0.009	0.003	0.006	0.006
$\vdots$				
q_{10}	0.026	0.039	0.044	0.023
Price (mean)	79.16	78.71	79.10	78.70
Price (min.)	74.07	72.33	73.99	70.95
Price (max.)	82.08	82.08	82.08	82.08
Margin	6.73	6.91	6.98	8.44
LL	138.21	142.33	145.95	152.15
N	10	10	10	10
R^2	0.852	0.929	0.875	0.820
Observations	114	95	95	95

We estimate utilities by running the chain–fixed-effects regression of prices separately for each location. The resulting R^2 values indicate that, in all locations, more than the 80% of the variation in prices is explained by chain dummies.

The first three rows of Table 5 report the estimated proportion of consumers searching one (q_1), two (q_2), or three times (q_3). In the first case, this proportion ranges from 72% in Barcelona to 75% in Vigo. The estimated share of consumers who search once or twice is nearly 90%, yet the percentage of consumers who search *all* local stores is never more than 4%. These results show no significant differences in search costs between consumers from different locations. Search costs may reflect the opportunity cost of time (which directly affects the cost of acquiring information) and/or other consumer characteristics (e.g., education, age). Although it seems reasonable to expect a positive relationship between income level and search cost, our estimates do not confirm that expectation. One possible

[24] One assumption of the theoretical model is that price observations might be uncorrelated. We check for serial correlation and calculate the the autocorrelation function for the basket in each store–location pair. Autocorrelation between two consecutive days is, not surprisingly, quite high; however, autocorrelations calculated with a 10-day lag between price observations are not significantly different from zero. When estimating search costs, we use price observations separated by 10 days.

reason is that urban consumers differ little with respect to the opportunity cost of time.[25] On the other hand, differences across average income among these four cities may not be high enough to pick up searching differences. Dubois and Perrone (2015) find that there are not consistent differences in the proportion of people searching across income level. That account could also explain why some national supermarket chains centralize their price setting. This question merits further research; to the extent that online commerce is accessible to consumers throughout Spain, it will be possible to obtain data from localities with more heterogeneity (e.g. rural vs urban).

Our results are in line with those reported in other papers. With regard to UK grocery items, Wildenbeest (2011) finds that most of the observed price variation is explained by supermarket heterogeneity and that the estimated amount of search is low. He reports that 71% of consumers search only once, 91% search either once or twice, and only 8% of consumers compare all prices. Richards *et al.* (2016) show that 84% of consumers search only one store when searching for products in many categories at the same time. Using food data from France, Dubois and Perrone (2015) confirm that consumers observe a very limited number of prices before making a purchase.

Estimated search costs are low. We estimate that the search cost of a consumer who searches only once is €0.62, on average. Such low search costs are similar to those found by Wildenbeest (2011), who calculates that the search cost of consumers who do not search should have been at least €0.27 in order to rationalize their behavior.

The estimated price–cost margins range between €6.73 and €8.44, which translates into an average margin of between 8.5% (in Barcelona) and 11.1% (in Vigo). Estimated margins in the United Kingdom range between 8% and 9%.

Our results could be affected by the particular products selected for the shopping basket. All the products in our basic basket are among the most frequently purchased, so one can reasonably suppose that consumers have more information about them. We therefore put together a substantially different basket containing products purchased less frequently, most

[25] Online shopping is available mainly in urban areas, in fact, our prices were obtained from the central zip codes.

of which are not purchased during primary shopping. Thus the occasional basket includes alcoholic beverages (i.e., wine, beer, and spirits) in addition to nonalcoholic drinks (i.e., soda, juice, mineral water, and energy drinks).[26]

Table 6. Occasional basket: Beverages

	Price		
Supermarket	Mean	Max.	Min.
Auchan	89.8	94.4	83.0
Carrefour	98.1	101.1	94.8
Condis	97.8	100.7	93.8
El Corte Inglés	98.8	101.0	89.0
Eroski	100.4	103.4	97.4
Mercadona	98.4	101.2	93.3
Total	97.4	103.4	83.0

The cost of this basket is close to €100, and its price dispersion is greater than that of the basic food basket. As Table 6 shows, the price ranges from €83 at Auchan to €103.4 at Eroski, a difference of almost 20%. Auchan remains the supermarket with the lowest prices (just as it is for the basic basket), but now Eroski (rather than El Corte Inglés; cf. Table 3) is usually the most expensive supermarket.

Figure 4. Ranking over time: Occasional basket (Barcelona)

A. Price B. Utility

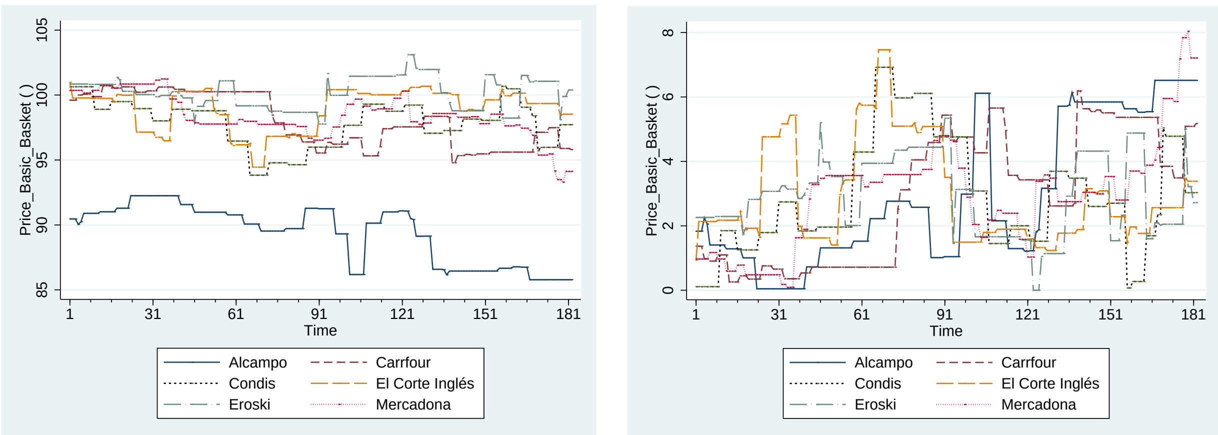


Figure 4 plots (once again, for Barcelona) the evolution of our occasional basket's price and utility. The patterns are strongly similar to those observed for the basic basket (cf.

[26]Table A3 gives the average price (and price dispersion) of each product in this basket.

Figure 3). Although Auchan is again always the least expensive supermarket, there is variability in the price ranking of the other supermarkets. Panel B of the figure shows that utility rankings (calculated as the negative of the residuals) change more frequently than do prices and that no supermarket is ranked consistently.

We regress prices on store fixed effects to obtain the utilities. For most locations, these fixed effects explain more than 80% of observed price variability. The results reported for beverages in Table 7 are similar to those (reported in Table 5) for the more general shopping basket. Here the search costs are again small—though slightly higher than for the basic basket (which includes products that are more frequently purchased).

Table 7. Estimation results: Occasional basket

	Barcelona	Madrid	Málaga	Vigo
q_1	0.775	0.744	0.754	0.647
q_2	0.150	0.197	0.193	0.218
q_3	0.017	0.022	0.004	0.012
$\vdots$				
q_{10}	0.030	0.0023	0.027	0.029
Price (mean)	97.13	96.41	97.46	96.99
Price (min.)	85.78	82.96	88.86	86.05
Price (max.)	103.11	103.37	103.37	103.38
Margin	13.97	15.00	14.24	12.50
LL	210.9	221.4	216.9	192.5
N	10	10	10	10
R^2	0.778	0.905	0.799	0.828
Observations	114	95	95	95

The estimated portion of consumers who search only once is slightly higher than in the previous case—except in Vigo, where it declines to 65% from 75%. For our occasional (beverages) basket, search costs average €1.1. Here the estimated price–cost margin is, on average, 13 %; this value is considerably higher than that for the basic basket.

So in comparison with the occasional basket, the basic basket is characterized by less price dispersion, lower search cost, and smaller margin. These results accord with those in Sorensen (2000), who established that prices for repeatedly purchased prescriptions—for which search benefits are expected to be high—exhibit relatively low price–cost margins and relatively less price dispersion.

7. CONCLUSIONS

In this paper we build a price data set of grocery and household products often included on Spanish consumers' shopping lists. Prices are gathered from the Web pages of the main supermarket chains in Spain at different locations. We identify patterns of cross-sectional and temporal price dispersion, and we measure the extent to which search cost and chain differentiation contribute to price dispersion. In order to measure search cost, we use the approach of Hong and Shum (2006) modified by Wildenbeest (2011) to account for store differentiation. This approach is specially adequate when only price data is availale, which mostly happen when price information is obtained from internet and quantity typically is not available.

Quantifying search costs is a relevant question from a competition policy perspective. As Waterson (2003) shows traditional policies that do not be aware of the importance of search costs can be less effective in enhancing competition. In fact, consumer information about prices is a necessary condition for markets to be truly competitive. If consumers engaged in no price comparison, then the monopoly price would prevail in equilibrium (Diamond, 1971); however, price dispersion will be an equilibrium in markets where at least some consumers search more than once (Burdett and Judd, 1983). Regulations that aim to promote competition must account for the distortions due to informational restrictions. For instance, Stahl (1989) shows that—in the presence of search costs—firm entry does not necessarily improve improve welfare. Similarly, Lach and Moraga-González (2012) find that consumer surplus always (albeit weakly) decreases with increased competition. Another consideration of those tasked with devising competition policy should be retailer practices that aim to confuse or mislead consumers (Ellison and Ellison, 2009).

Analyzing consumer search in online markets could yield insight into traditional, brick-and-mortar retail behavior. Consumer access to prices through supermarket websites or price comparison sites (e.g., Soysuper.com) has reduced search costs for online and offline retail consumers both. Customers can easily check prices before or even while they shop.

We find that some chains have persistently lower prices than others—even as prices change frequently. This empirical evidence suggests that both search frictions and chain differentiation help explain price dispersion in Spanish grocery markets. We estimate the distribution of consumer search costs using only price data; for that purpose we employ the model proposed by Burdett and Judd (1983) and modified (to account for vertical product differentiation) by Wildenbeest (2011). We use price data from four geographical markets of different sizes and located far from each other: Madrid, Barcelona, Málaga, and Vigo. We find that about 87% of the observed variation in prices is due to chain fixed effects and that some 70% of consumers search only once in all markets, behavior that would be economically rational only if search costs amounted to at least €0.62. These values do not differ significantly from one geographical market to the next. Our regression results also indicate that the more frequently purchased products tend to have lower search costs and lower price–cost margins.

Finally, we offer empirical support for findings previously reported in the literature on the retail food market in other countries—for instance, France and the United Kingdom. Our results comport also with findings, reported in MARM (2011), that 84% of interviewed Spanish consumers considered themselves to be fairly loyal to a supermarket and that 48.4% of them stated they always buy *without* comparing prices.

REFERENCES

Bachis, E. and C.A. Piga (2011). "Low-cost airlines and online price dispersion", *International Journal of Industrial Organization* 29(6): 655–667.

Barron, J.M., B.A. Taylor and J.R. Umbeck (2004). "Number of sellers, average prices and price dispersion", *International Journal of Industrial Organization* 22(8-9): 1041–1066.

Baye, M.R. and J. Morgan (2001). "Information gatekeepers on the Internet and the competitiveness of homogeneous product markets", *American Economic Review* 91(3): 454–474.

Baye, M.R., J. Morgan and P. Scholten (2004a). "Temporal price dispersion: evidence from an online consumer electronics market", *Journal of Interactive Marketing* 18(4): 101–115.

Baye, M.R., J. Morgan and P. Scholten (2004b). "Price dispersion in the small and in the large: evidence from an Internet price comparison site", *Journal of Industrial Economics* 52(4): 463–496.

Baye, M.R., J. Morgan and P. Scholten (2006). "Information, search and price dispersion", *Handbook of Economics and Information Systems* 1: 323-371. T. Hendershott (Ed.), Elsevier Press.

Baye, M.R., B. De los Santos, M. R. Wildenbeest (2013). "The evolution of product search" (with Mike Baye and Babur De los Santos), *Journal of Law, Economics & Policy* 9: 201–221.

Berck, P., J. Brown, J.M. Perloff and S.B Villas-Boas (2008). "Sales: Test of theories on causality and timing", *International Journal of Industrial Organization* 26(6): 1257–1273.

Brown, J.R. and A. Goolsbee (2002). "Does the Internet make markets more competitive? Evidence from the life insurance industry", *Journal of Political Economy* 110(3): 481–507.

Brynjolfsson, E. , Dick, A.A. , Smith, M.D. (2010). "A nearly perfect market? Differentiation vs. price in consumer choice", *Quantitative Marketing and Economics* 8: 1–33.

Brynjolfsson, E. and M.D. Smith (2000). "Frictionless commerce? A comparison of Internet and conventional retailers", *Management Science* 46(4): 563–585.

Burdett, K. and K. Judd (1983). "Equilibrium price dispersion", *Econometrica* 51(4): 955–969.

Carlson, J.A. and R.P. McAfee (1983). "Discrete equilibrium price dispersion", *Journal of Political Economy* 91(3): 480–493.

Cavallo, A. (2016a). "Scraped data and sticky prices", , *Review of Economics and Statistics* forthcoming.

Cavallo, A. (2016b). "Are online and offline prices similar? Evidence from multi-channel retailers", *American Economic Review* forthcoming.

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

Clay, K., R. Krishnan and E. Wolff (2001). "Prices and price dispersion on the web: evidence from online book industry", *Journal of Industrial Economics* 49(3): 521–539.

Comisión Nacional de la Competencia (CNC) (2011). "Report on the relations between manufacturers and retailers in the food sector".

Dana, J.D. (1994). "Learning in an equilibrium search model", *International Economic Review* 35(3): 745–771.

Delgado, J. and M. Waterson (2003). "Tyre price dispersion across retail outlets in the UK", *Journal of Industrial Economics* 51(4): 491–509.

De los Santos, B., A. Hortaçsu and M.R. Wildenbeest (2012). "Testing models of consumer search using data on web browsing and purchasing behavior", *American Economic Review* 102(6): 2955–2980.

Diamond, P.A. (1971). "A model of price adjustment", *Journal of Economic Theory* 3(2): 156–158.

Dubois, P. and H. Perrone (2015). "Price dispersion and search costs: the roles of imperfect information and product differentiation", mimeo.

Edelman, B. (2012). "Using Internet data for economic research", *Journal of Economic Perspectives* 26(2): 189–206.

Elberg, A. (2014). "Heterogeneous price dynamics, synchronization, and retail chains: Evidence from scanner data", *Emerging Markets, Finance and Trade* 50: 137–153.

Ellison, G. and S.F. Ellison (2009). "Search, obfuscation and price elasticities on the Internet", *Econometrica* 77(2): 427–452.

García, J.A. and J. Delgado (2012). "Análisis de la competencia en el mercado minorista de distribución en España", *Brattle Group.*

Hartman (2012). The Online Grocery Opportunity. (http://www.ruralgrocery.org/docs/HartmanGp2012OnlineGrocery.pdf). Accessed on October 18, 2016.

Hong, H. and M. Shum (2006). "Using price distributions to estimate search costs", *RAND Journal of Economics* 37(2): 257–275.

Hortaçsu, A. and C. Syverson (2004). "Product differentiation, search costs and competition in the mutual fund industry: a case study of S&P 500 Index funds", *Quarterly Journal of Economics* 119(2): 403–456.

Hopkins (2008). "Price dispersion", *The New Palgrave Dictionary of Economics.* Eds. S. N. Durlauf and L.E. Blume. Palgrave Macmillan.

Janssen, M. and J.L. Moraga-González (2004). "Strategic pricing, consumer search and the number of firms", *Review of Economic Studies* 71(4): 1089–1118.

Janssen, M., J.L. Moraga-González and M.R. Wildenbeest (2005). "Truly costly sequential search and oligopolistic pricing", *International Journal of Industrial Organization* 23(5-6): 451–466.

Janssen M., J. L. Moraga, M. R. Wildenbeest, E. Brousseau, N. Curien (2007). "Consumer search and internet markets". *Internet and Digital Economics* (Cambridge University Press, Cambridge, UK) 460–483.

Klenow, P.J. and B.A. Malin (2010). "Microeconomic evidence on price-setting", *NBER* Working Paper No. 15826.

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

Lach, S. and J.L. Moraga-González (2012). "Asymmetric price effects of competition", mimeo and Discussion Paper No. 7319, CEPR (2009).

Lloyd, T.A., S McCorriston, C.W. Morgan, E. Poen and E. Zgovu 2014. "Retail price dynamics and retailer heterogeneity: UK evidence", *Economics Letters* 124(3), 434–438.

MARM (2011). "Estudio de mercado; Observatorio del consumo y la distribución alimentaria". *Ministerio de Medio Ambiente y Medio Rural y Marino.*

Moraga-González, J.L. and M.R. Wildenbeest (2008). "Maximum likelihood estimation of search costs", *European Economic Review* 52(5): 820–848.

Nakamura, A.O., E. Nakamura and L.I. Nakamura (2011). "Price dynamics, retail chains and inflation measurement", *Journal of Econometrics* 161(1): 47–55.

Orlov, E. (2011). "How does the Internet influence price dispersion? Evidence from the airline industry", *Journal of Industrial Economics* 59(1): 21–37.

Pozzi, A. (2012). "Shopping cost and brand exploration in online grocery", *American Economic Journal: Microeconomics* 4(3): 96–120.

Richards, T., S. F. Hamilton, and W. Allender (2016). " Search and price dispersion in online grocery markets", *International Journal if Industrial Organization*, 47: 255–281.

Rosenthal, R.W. (1980). "A model in which an increase in the number of sellers leads to a higher price", *Econometrica* 48(6): 1575–1579.

Seiler, S. (2013) "The impact of search costs on consumer behaviour: A dynamic approach", *Quantitative Marketing Economics*, 11:155–203.

Sorensen, A.T. (2000). "Equilibrium price dispersion in retail markets for prescription drugs", *Journal of Political Economy* 108(4): 833–850.

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

Stigler, G. (1961). "The economics of information", *Journal of Political Economy* 69(3): 213–225.

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

Waterson, M. (2003). "The role of consumers in competition and competition policy", *International Journal of Industrial Organization* 21(2): 129–150.

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

Zhao, Y. (2006). "Price dispersion in the grocery market", *Journal of Business* 79(3):

1175–1192.

APPENDIX

Table A1. Number of products (observations) in each location and supermarket

Location	Auchan	Carrefour	Condis	El Corte Inglés	Eroski	Mercadona	Total
Barcelona	223 (40,473)	234 (42,588)	200 (36,352)	235 (42,770)	237 (43,134)	173 (31,486)	1,302 (236,803)
Madrid	221 (40,477)	233 (42,406)	—	215 (39,130)	236 (42,952)	182 (33,124)	1,089 (198,089)
Málaga	223 (40,001)	233 (42,406)	—	235 (42,770)	237 (43,134)	181 (32,942)	1,107 (201,253)
Vigo	221 (40,002)	231 (42,042)	—	236 (42,952)	235 (42,723)	177 (32,210)	1,100 (199,929)
Total	888 (160,953)	931 (169,442)	200 (36,352)	921 (167,622)	945 (171,943)	713 (129,762)	9,494 (836,074)

Table A2 Summary statistics for items in the basic basket

	Mean price	(S.D)	Minimum price	Maximum price	Coefficient of variation ($\times$100)
Aquarius lemon 1.5 l.	1.41	(0.10)	.97	1.49	6.84
Coca Cola 2 l.	1.36	(0.03)	1.01	1.38	1.97
Fanta limon 2 l.	1.18	(0.03)	0.88	1.37	2.22
Mineral Water Aquarel Nestle 5l.	1.04	(0.07)	0.84	1.21	6.29
Non alcoholic beer San Miguel 6$\times$33cc.	2.90	(0.35)	2.04	3.24	12.22
Beer Heineken pack 6x25cc.	3.26	(0.30)	1.97	3.65	9.35
Wine Rioja El Coto 75 cl.	5.18	(0.15)	3.82	5.38	2.29
Cava Brut Anna de Codorniu 75 cl.	8.16	(0.29)	5.99	8.49	3.57
Cereal Bar Chocho Crispies 6$\times$20 gr.	2.12	(0.12)	1.61	2.29	5.67
Cornk Flakes 500 gr.	1.98	(0.09)	1.37	2.25	4.76
Special K 500 gr.	2.82	(0.22)	1.87	3.00	7.78
Biscuits Principe de LU estrellas choco 225 gr.	1.97	(0.17)	1.58	3.76	5.42
Breakfast Biscuits María Fontaneda	1.99	(0.11)	1.58	3.76	5.36
Capsulas café cortado Dolce Gosto 2$\times$16u.	9.53	(0.27)	7.8	9.70	2.84
Cola Cao 800 gr.	4.90	(0.26)	3.84	5.19	5.54
Whole milk Asturiana 6$\times$1.5 l.	8.24	(0.18)	5.40	8.34	2.17
Semi-skimmed milk Asturiana 6$\times$1.5 l.	8.23	(0.23)	5.16	8.34	2.78
Deodorant Roll on Dove	2.17	(0.40)	1.58	2.95	18.52
Champoo Pantene Pro-v	2.96	(0.10)	2.00	3.19	3.52
Cillit bang 750 ml.	3.49	(0.23)	1.99	3.89	6.62
Rice Brillante 1 kg.	1.59	(0.06)	0.98	1.67	3.76
Corn Four Maizena	1.43	(0.13)	1.02	1.69	9.45
Chocolate with almond Nestlé 150 gr.	1.26	(0.12)	0.91	1.49	9.81
Whole Basket	79.08	(2.25)	70.95	82.08	2.85

Notes: listed products are some of the most popular items according to Soysuper.com. All prices are given in euros.

Table A3 Summary statistics for items in the ocasional basket

	Mean price (S.D)		Minimum price	Maximum price	Coefficient of variation (× 100)
TriNa orange 1.5 l.	1.24	(0.19)	0.95	1.39	15.2
Sprite lemon 4×33 cl.	1.41	(0.10)	1.04	1.60	7.39
Don Simon Juice pinneaple 6×1 l.	5.71	(0.37)	4.74	6.54	6.53
Bitter kas 6×20 cc.	3.82	(0.16)	2.92	3.97	4.10
Nestea Light lemon 1.5 l	1.32	(0.16)	0.89	1.44	11.84
Coca Cola caffeine-free 2 l.	1.31	(0.12)	0.75	1.37	9.30
Pascual Bifrutas Tropical 12×20cl.	4.47	(0.77)	2.00	5.30	17.31
Fanta orange 2 l.	1.17	(0.03)	0.88	1.19	2.70
Pepsi Cola 2 l.	1.11	(0.06)	0.82	1.19	5.38
Pepsi Cola light 2 l.	1.06	(0.06)	0.75	1.19	6.09
Coca Cola Zero 50 cl.	0.78	(0.03)	0.59	0.82	3.21
7 up 2 l.	1.02	(0.09)	0.63	1.74	8.60
Coca Cola Light 2 l.	1.36	(0.02)	1.01	1.37	1.79
Coca cola Zero 2.l.	1.35	(0.03)	1.01	1.37	1.96
Aquarius orange 1.5 l.	1.41	(0.10)	0.97	1.49	6.93
Aquarius lemon 1.5 l.	1.41	(0.10)	0.97	1.49	6.89
Coca Cola 2 l.	1.36	(0.03)	1.01	1.38	1.98
Fanta limon 2 l.	1.18	(0.03)	0.88	1.37	2.26
Coca cola 4 x 33 cl.	2.22	(0.07)	1.68	2.24	3.04
Coca Cola 1.5 l.	1.02	(0.04)	0.75	1.09	4.12
Mineral water Vichi Catalan 6×1 l.	6.45	(0.33)	4.56	6.84	5.19
Aquarel Nestlé mineral water 6×5 l.	6.22	(0.39)	5.04	7.26	6.33
San Miguel lemon 9×33 cl.	4.65	(0.58)	3.06	5.31	12.37
San Miguel non alcoholic beer 6×25 cl	2.21	(0.15)	1.42	2.49	6.94
San Miguel non alcoholic beer 9×33c.	4.34	(0.54)	3.06	4.86	12.45
Voll-damm beer 6×25 cl	3.60	(0.27)	2.14	3.79	7.41
Heineken beer 6×25 cl	3.25	(0.31)	1.97	3.65	9.55
Monster energetic drink 50 cl	1.08	(0.07)	0.75	1.16	6.92
El coto Rioja Wine 75 cl.	5.18	(0.15)	3.82	5.38	2.85
Viña Albali Valdepeñas Wine 75 cl.	3.18	(0.19)	2.78	3.35	5.87
Don Simon Wine 1 l.	1.14	(0.07)	0.99	1.45	6.06
Bach Cava Brut Nature	3.66	(0.30)	2.49	3.95	8.29
Freixenet Cava Carta Nevada 1 l.	4.97	(0.39)	3.56	6.59	7.89
Whisky J.B.	11.64	(0.36)	10.49	13.02	3.07
Whole Basket					

Note: listed products are some of the most popular items according with Soysuper.com. All prices are given in euros.