The Pervasive Role of Retail Chains in French Price Dispersion

Nicoletta BERARDI, Patrick SEVESTRE

Abstract

We analyse price dispersion in about 1600 French supermarkets and compare it internationally. We find that in France more than 80% of the total variance of the observed dispersion of relative prices across store and time is explained by the spatial permanent component: stores persistently sell products at relatively high (or low) prices, essentially driven by persistent heterogeneity in retail chains' pricing. The analysis of between and within retail chains' price dispersion also provides evidence consistent with a multi-stage price setting in which buying groups and local branches play a much bigger role than local stores.

Keywords: price dispersion; retail chain pricing; store heterogeneity

JEL classification: D40, L11, D22, L81

The authors would like to thank Prixing, the start-up that provided the price data used in this chapter and in particular Eric Larchevêque, the owner and CEO of the company at the time data were released. We also thank Aix-Marseille School of Economics and the Banque de France for jointly funding the LSA database. We are extremely grateful to Etienne Gagnon for very helpful comments and Roberto Berardi for his insights on the retail industry. We wish to address special thanks to Sherly Jean-Charles, Sophie Saleh, Marine Tépaut, Jonathan Thébault, Sylvie Tarrieu, Alexandre Vigneron, and Yue Zhu for their precious assistance with complementary data collection and treatment. Finally, we thank Vincent Guegan and Victor Yammouni for IT assistance. The views expressed in this chapter are those of the authors and do not necessarily represent those of their institutions.

Berardi: Banque de France, SAMIC-DEMS-DGEI, 31 rue Croix des Petits Champs, 75001 Paris, France (e-mail address: nicoletta.berardi@banque-france.fr); Sevestre : AMSE, Université Aix-Marseille, Chateau Lafarge, route des Milles, 13290 Les Milles, France, (e-mail address: patrick.sevestre@univ-amu.fr).

1 Introduction

The existence of deviations from the law of one price have been thoroughly studied. Traditionally this could be done only for very limited markets (e.g., Borenstein and Rose [1994], Asplund and Friberg [2002], Gerardi and Shapiro [2009], Sorensen [2000], Hong and Shum [2006], Moraga-González and Wildenbeest [2008], Woodward and Hall [2012], Dubois and Perrone [2015]) due to data unavailability at a larger scale. In particular, CPI price records that have become available in he last fifteen years typically don't contain the necessary information to compare exctly the same product across stores (i.e., its barcode). However, more recently other sources of data sometimes allow a precise quantification of price dispersion in the retail sector. For instance, Kaplan and Menzio [2015] analyse the structure of price dispersion in US retail sector, based on the Kilts-Nielsen Consumer Panel Data set.[1]

We exploit more than 250 millions of webscraped daily price records from more than 1600 geo-localized supermarkets in France over one year, building on previous work by Berardi et al. [2017].[2] We find that non-trivial price dispersion exists in France. On average across products, the 90^{th} percentile of relative prices is 18% higher than the 10^{th} and the mean absolute deviation with quarterly modal product prices as the measure of central tendency is 6% on average in the French retail sector. We compare it internationally and conclude that prices in the retail sector are less dispersed in France than in the US.

We also find that in France more than 80% of the total variance of the observed dispersion of relative prices across store and time is explained by the spatial permanent component: stores persistently sell products at relatively high (or low) prices, essentially driven by persistent heterogeneity in retail chains' pricing.

Finally, we investigate the determinants explaining between and within retail chains' price dispersion and find evidence consistent with a multi-stage price setting framework. Following the initial bargaining between buying groups and pro-

[1]For a more detailed review of the literature about price dispersion, please refer to Berardi et al. [2017].

[2]The original price data is the same as described in more detail in Berardi et al. [2017]. However, there prices were weekly, while in this paper they are daily. Also, part of the store data we exploit was not available at the time.

ducers, retail chains may adjust prices at the level of retail chains' local branches. Finally, prices may be fine-tuned at the store level based on its characteristics and its local market.

The remaining of the paper is structured as follows. Section 2 describes the price and store data sets, presents a measure of price dispersion, characterizes it for France and provides an international comparison. Section 3 disentangles the permanent and time-varying components of price dispersion. Section 4 investigates between and within retail chains' price dispersion. Section 5 concludes.

2 Assessing Price Dispersion in the French Retail Sector

2.1 Grocery Price and Store Data

The daily price data set we exploit contains 205 millions observations and restricts to the one thousand most widely sold products (i.e., barcodes).[3] Those are mainly food and beverages (77%), as well as personal care products (13%).[4] Prices are collected from more than 1600 geo-localized medium and large size supermarkets[5] offering 'click&collect' services.[6] Notice that each price does correspond to what the consumer would really pay for a product on the day the data was collected.

In order to improve the representativeness of the sample, as not all retailers

[3] The original data refers to more than one hundred thousand products. For more details about the cleaning of the price data, please refer to Berardi et al. [2017] (although the frequency of price records is daily in this paper, the cleaning has been kept consistent).

[4] For more details, refer to table 15.1 in Berardi et al. [2017].

[5] Medium and large size supermarkets represent in France more than 80% of grocery sales (Anderton et al. [2011]).

[6] These data were collected by Prixing, a start-up company providing consumers with a free mobile price comparator (see http://www.prixing.fr/). The crucial feature is that prices are exactly the same as those of the brick-and-mortar supermarket associated with the 'click&collect' service, which is almost always linked to a physical store (although a few stand alone drive-throughs, known as 'drives-entrepôt', exist. For more details on the characteristics and the evolution of 'click&collect', known as 'drives' in France, please refer to Berardi et al. [2017]. Notice that a few more stores are available in this paper, thanks to subsequent data mining.

developed click&collect at the same pace,[7] we weighted the data.[8] Weights have been computed to account for both the market shares of retail chains and the share of product categories in the French consumption basket.[9] In particular we assign to each observation a weight corresponding to the product of the retail chain market shares by the share of each product category divided by the number of available observations for each retail chain $\times$ product category cell.

We complemented the price data with an exhaustive data base of all medium and large supermarkets in France.[10] For more details concerning the geographical distribution and representativeness of supermarkets for which we have price records and the market share of the retail chains to which they belong, please refer to Berardi et al. [2017].

2.2 Measures of Price Dispersion at the National Level and International Comparison

In this paper we consider that price differences within a quarter for exactly the same product (i.e., barcode) sold in different supermarkets imply price dispersion.

In order to quantify price dispersion, we first compute percentage deviations of the price for product i in store s at day t (p_{ist}) from a quarterly reference price for each product (p_{iq}^{ref}):

$$p_{ist}^{rel(iq)} = (p_{ist} - p_{iq}^{ref})/p_{iq}^{ref} \tag{1}$$

We define the quarterly reference price for each product i in alternative ways. In particular, we use a rather statistical approach, as well as a more economic one.

[7] In particular, one of the major retail groups in France lagged behind regarding the opening of this type of outlet, while a smaller player offered this option in most of its supermarkets, even in small ones. Therefore, our sample of click&collect (although almost exhaustive) did not necessarily provide a representative picture of supermarket sales at the aggregate level.

[8] Unweighted results (available upon request) are not qualitatively different.

[9] Retail market shares come from Kantar Worldpanel-LSA and refer to April-May 2012. The share of product categories (COICOP-level4) in the French consumption basket are obtained from the French National Statistical Office (INSEE).

[10] The data was bought from LSA (Libre Service Actualités http://expert.lsa-conso.fr/) and includes address, retail chain and local branch of each medium ans large supermarket located in France.

Based on the former, the quarterly reference price for each product i is defined in (1) as $p_{iq}^{ref} = mean(p_{iq})$.[11] This reference price, however, most likely does not coincide with the price at which a product is sold in any store. An alternative definition for the reference price is $p_{iq}^{ref} = mode(p_{iq})$. The latter conveys a stronger economic intuition, representing the price at which a product is most often sold in a given quarter.

Figure 1 shows the distribution of relative prices in our data. The left panel is based on the mean as reference price, while the right one on the mode. The crucial conclusion emerging from the right panel of Figure 1 is that 17% of prices coincide with the product national modal price.

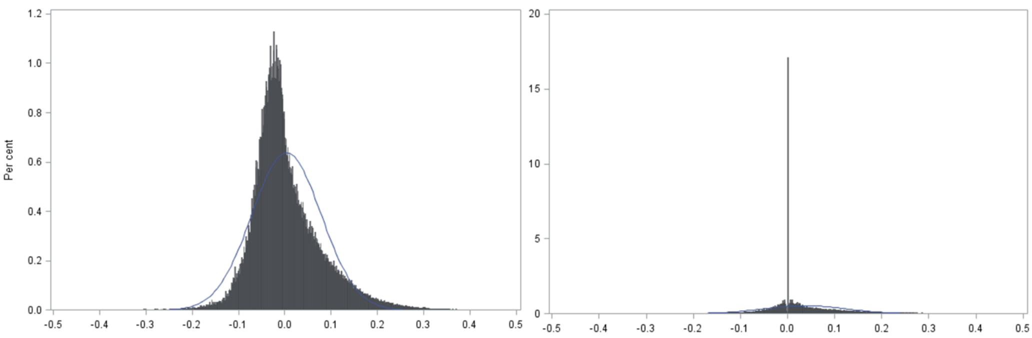

Figure 1: Distribution of the relative prices with respect to the product quarterly mean (left panel) and mode (right panel).

Although the distribution of relative prices remarkably differs in terms of prices equal to the the product reference in Figure 1, an important finding is that the summary statistics of price dispersion are very similar whether the reference price is assumed to be the product quarterly mean or mean. On average across products, the 90^{th} percentile of relative prices is respectively 16 and 18 percentage points higher than the 10^{th}. The interquartile range is 8 percentage points in both cases and the mean absolute deviation of relative prices respectively 5 and 6% in France. The latest one, the mean absolute deviation of relative prices from

[11] This is the definition adopted in Berardi et al. [2017]. It is proposed again in this article for two reasons. The first one is that this allows the comparison with the alternative measure we now prefer. The second one is that in this paper prices are daily and no longer weekly, although as expected this does not change significantly the assessment.

its quarterly modal price at the national level, is our preferred measure of price dispersion. Consistently with all the other measures, it points to a non-trivial overall price dispersion in France.

We comparing this with the only assessment of price dispersion at the barcode level available for France (Dubois and Perrone [2015]). To do so, we compute on our data the measures they propose in their paper and find similar but slightly lower price dispersion when computing their statistics with our data. On average across products, the 95^{th} percentile of observed prices is 26% higher than the 5^{th} (versus 37% computed by Dubois and Perrone [2015]). The interquartile ratios are even closer to each other (1.09 in our data, versus 1.14). This seems reasonable given that they only cover four product categories and two of them are alcoholic beverages (exibiting relatively high price dispersion in our dataset as well).

In order to compare price dispersion in France with respect to other countries, we also compute on our data the same measures of price dispersion as in other studies based on data from the US, the UK and Canada. The conclusion is that in France prices appear to be less dispersed. Indeed, price dispersion measured by Kaplan and Menzio [2015] across US brick-and-mortar stores is 19% larger than in France based on the $90^{th}/50^{th}$ centile ratio of normalized prices, and even larger for the other statistics. For instance, on average across products, the $90^{th}/10^{th}$ centile ratio of normalized prices is 38% larger in the US. Notice, however, that part of this difference could be due to the fact that Kaplan and Menzio [2015] price data concern a larger spectrum of store types than our data. Figure 2 shows that price dispersion in France is lower than in other countries based on measures provided by Gorodnichenko and Talavera [2017] for the US and Canada and by Gorodnichenko et al. [2018] for the US and te UK. However, the comparison with these studies, which rely only on online prices is not straightforward. Indeed, contrary to what consumer search theory predicts, as it has been shown that price dispersion may be larger than offline (Clay et al. [2001], Degeratu et al. [2000]).

3 Dimensions of Price Dispersion

Having assessed the overall level of price dispersion in France, we turn to disentangling its sources and exploring its spatial and temporal components.

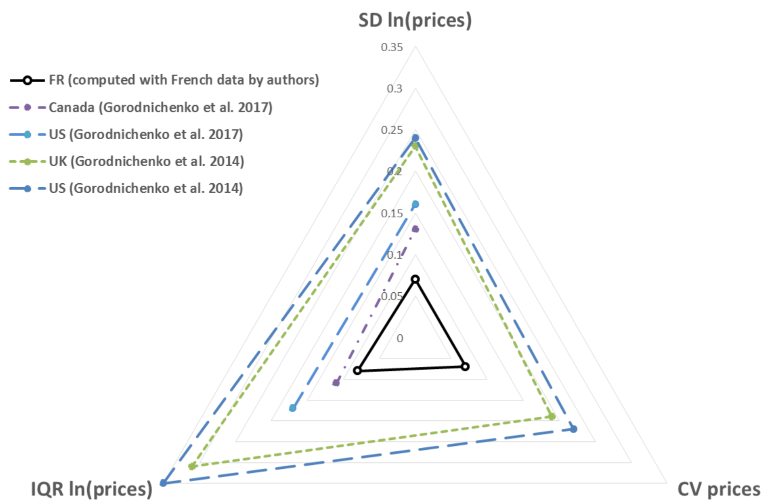

Figure 2: International comparison with measures provided by Gorodnichenko and Talavera [2017] for the US and Canada (goods sold on online price comparator betzeen November 2008 and September 2012) and by Gorodnichenko et al. [2018] for the US and the UK (goods sold on online shopping platform between May 2010 and February 2012).

3.1 Disentangling Spatial and Temporal Price Dispersion

This section contributes to the debate in the literature on price dispersion about whether price dispersion is mainly spatial (i.e., some stores persistently sell at lower price) or temporal (i.e., supermarkets vary their price over time, so that consumers cannot learn by experience which shops provide the best price). To disentangle the two dimensions of price dispersion, we estimate a fixed effect model including store fixed effects (capturing all persistent characteristics of supermarkets that affect their price setting) and daily[12] time fixed effects. We then

[12] Berardi et al. [2017] runs a fixed effect model that is similar in spirit, but where the temporal component is captured by combinations of week and localbranch fixed effects. The corresponing specification at the daily frequency however imples too many parameters to estimate.

estimate, product by product:

$$p_{ist}^{rel(iq)} = \alpha_{is} + \alpha_{it} + \varepsilon_{ist} \,, \tag{2}$$

where $p_{ist}^{rel(iq)}$ is the percentage deviation from the product quarterly modal price, α_{is} are supermarket fixed effects, α_{it} are daily time fixed effects, and ε_{ist} are the error terms.

Figure 3 shows the distribution of the overall variance explained by the above model by product. It suggests that this very simple model suffice to explain on average 87% of the observed dispersion of prices across stores and time. The model explains 51% of price dispersion in the worst case and up to 98% in the case of the product with the best fit.

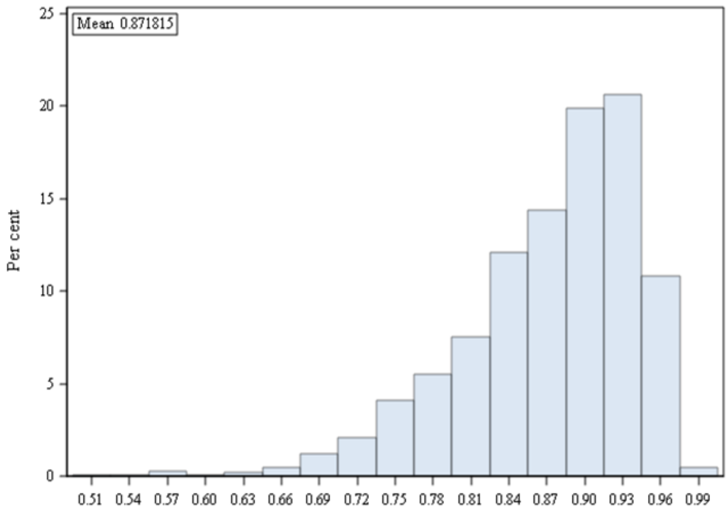

Figure 3: Distribution of the overall variance explained by the model by product.

What is more interesting, however, is that the time invariant spatial component represents by far the most important component of the total variance. Indeed, Figure 4 shows the distribution of the overall variance explained by store fixed effects α_{is} by product. Alone, they explain more than 80% of the observed dispersion of prices. Even in the case of the product for which the spatial dimension is the smallest, it represents 29% of the total variance. For the median product it explains 82% of the observed dispersion of prices. Gorodnichenko et al. [2018] similarly find

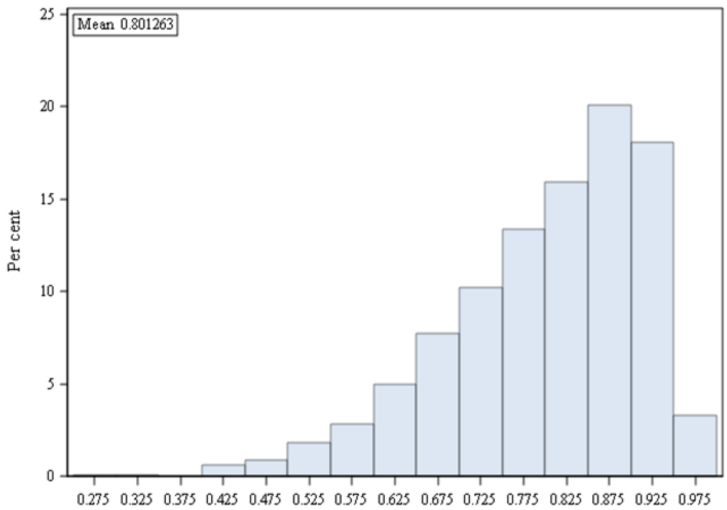

Figure 4: Distribution of the overall variance explained by store fixed effects by product.

that price dispersion appears to be best characterized as spatial rather than temporal in the US and the UK, even in online markets. Notice that the remaining variance includes, among everything else, temporary discounts, confirming the low prevalence of price sales in France (see Berardi et al., 2015).

3.2 The Role of Retail Chains in Price Dispersion

Once reached the conclusion that in France price dispersion is mainly spatial, it is possible to go one step further and try to investigate what are its main determinants. In particular, it's interesting to understand the relative role played by factors at the local and at the national level. Among the former ones, it seems likely that the urban density of the area where a supermarket is located may play a role and in particular the characteristics of local demand and competition.[13] We approximate local demand by log per capita income and population of the district m. Competition in the local market is captured by two variables. The first one is the number of supermarkets selling a product i in local market m. The second

[13]Urban density categories are based on a combination of population density and absolute population of INSEE 'canton-ou-ville'.

one is the distance (in kilometers) to the closest large supermarket, computed exploiting the geo-localization of supermarkets and the exhaustive data base of all medium and large supermarkets in France.[14] Fixed effects for the retail chain to which a store belongs account for characteristics that are in common wherever the supermarket is located.

We thus estimate the following model:

$$\begin{aligned} \alpha_{is} = \beta_i + \beta_r + \gamma_1 * \text{urban} + \gamma_2 * \text{income} + \gamma_3 * \text{pop} + \\ + \gamma_4 * \text{sameprod} + \gamma_5 * \text{closest} + \upsilon_{is} \end{aligned} \tag{3}$$

where α_{is} are the product*supermarket fixed effects estimated in the previous section, β_i are product fixed effects, β_r are fixed effects for (anonymized) retail chains to which the supermarket belongs, *urban* is a 4-level categorical variable for urban density, *income* and *pop* are respectively log per capita income and population of district m (which approximates a local market), *sameprod* corresponds to the number of supermarkets selling product i in the local market m and *closest* is the distance, in kilometers, to the closest large supermarket.

We also estimate a second model with a less restrictive definition of fixed effects, where these are specific to each product $\times$ retail chain pair:

$$\begin{aligned} \alpha_{is} = \beta_{ir} + \gamma_1 * \text{urban} + \gamma_2 * \text{income} + \gamma_3 * \text{pop} + \\ + \gamma_4 * \text{sameprod} + \gamma_5 * \text{closest} + \upsilon_{is}\,, \end{aligned} \tag{4}$$

Table 1 reports the estimated coefficients of the regressions (3) and (4). The estimation results show that the interacted fixed effect model does a better job in fitting the data than the additive fixed effect model. Indeed, the R squared of the former is 0.70 versus 0.54 of the latter, suggesting that not all products exhibit

[14] We follow the definition of competitors adopted by the French Competition Authority [2010] and assume that large supermarkets are only in competition with other large ones, while medium size supermarkets and discounts are also in competition with large supermarkets. Therefore, not all supermarkets are competing with some medium size competitor, but for all supermarkets we can compute the distance with respect to their closest large size competitor. Distances and driving time to the closest competitor are calculated using two internet applications: GoogleMap and YourNavigation. Since the computed distances are similar, in what follows we only present measures calculated from YourNavigation.

the same price dispersion across retail chains. The comparison of the regressors' contributions to the variance explained by the model reveals that, beyond product fixed effect, retail chains are by far the most relevant factor. Conditionally on all other regressors, retail chains account for 41% of the variance explained by the model in the additive fixed effect model. Not surprisingly, retail chain fixed effects show that prices in large supermarkets (denoted by the inclusion in the anonymized retail chain name in Table 1 of 'H' for 'Hypermarche' in French) tend to be lower than those in medium-size supermarkets (denoted by 'S').[15] Regarding the impact of local factors on prices, the estimates in the additive and interacted fixed effects model are very similar. First, urban density increases price levels in supermarkets. Second, supermarkets facing favorable local demand conditions in terms of larger population and per capita income also exhibit higher prices. Finally, stronger local competition tends to decrease price levels in supermarkets. Indeed, the further away is located the nearest large supermarket, the higher the prices. At the same time, if a product sold by supermarket s is also available in several other supermarkets located in the same local market, price levels decrease. However, despite their statistical significance, these local factors have a quite limited quantitative impact on prices.

The fact that in France centralized price setting strategies dominate local factors (as well as time varying ones, as shown in section 3.1) in retailing prices is the main message that can be drawn from our analysis of French price dispersion. This finding is robust to alternative definitions of the level of centralization in price setting determination. If we include buying group (instead of retail chain) dummies in model (3), the R squared decreases only to 0.50 and, conditionally on all other regressors, buying groups still account for 37% of the variance explained by the model. These results suggest that a large chunk of price setting is actually already determined at this level. Similarly, if instead we include local branch/wholesaler dummies in model (3), the R squared increases only to 0.57 and, conditionally on all other regressors, regional branches account for 46% of the variance explained by the model. Therefore, there seems to be an additional

[15]The only one exception is retail group 3, which has higher prices than some medium-size supermarket retail chains. Notice however that prices in its large supermarkets tend anyway to be lower than in its own smaller supermarkets.

Table 1: National and local determinants of spatial price dispersion

	additive FE		interacted FE	
Regressor	estimate	SE	estimate	SE
retail chain 1 H	-0.02	0.0003***		
retail chain 1 H/S	0.06	0.0004***		
retail chain 2 H	-0.04	0.0002***		
retail chain 2 H/S	0.01	0.0002***		
retail chain 3 H	0.08	0.0002***		
retail chain 3 H/S	0.12	0.0003***		
retail chain 4 H	-0.03	0.0003***		
retail chain 4 H/S	-0.12	0.0003***		
retail chain 4 S	0.07	0.0005***		
retail chain 5 H/S	-0.06	0.0002***		
retail chain 6 H/S (ref)	.	.		
rural	-0.06	0.0004***	-0.06	0.0004***
semi-urban	-0.06	0.0004***	-0.06	0.0003***
urban	-0.05	0.0004***	-0.05	0.0003***
metropolitan (ref)	.	.		
log local per capita income	0.0185	0.0002***	0.0182	0.0002***
log local population	0.0023	0.0000***	0.0024	0.0000***
n.stores selling product	-0.0007	0.0000***	-0.0007	0.0000***
closest large supermkt	0.0003	0.0000***	0.0003	0.0000***
product FE	yes		no	
product*retail chain FE	no		yes	
R squared	0.54		0.69	

Note: *** means significant at 1%.

stage of price setting happening at the regional level, but national price setting strategies are dominant.

4 Between and Within Retail Chain Price Dispersion

The retail chain to which a store belongs is a crucial dimension explaining price dispersion in France. Figure 5 shows the heterogeneity across retail chains of deviation from the national reference price for two products: one is the very in-

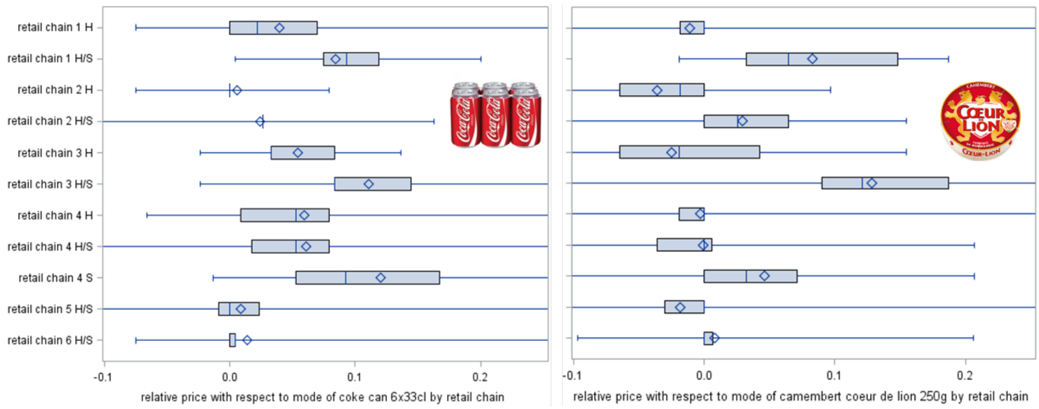

Figure 5: Boxplot of relative prices with respect to the quarterly modal price of 6x33cl Coke cans (left panel) and of 250g Coeur de Lion camembert (right panel) by retail chain.

ternational and advertised Coke can and the other one is Coeur Lion camembert, one of the brands of a very French cheese. Some characteristics of the distributions of relative prices are similar across the two products. First, the average (represented by a diamond in the figure) varies substantially across retail chains, suggesting that retail chains set different prices (something we refer to as between retail chain price dispersion). For instance, both products are about 10% more expensive in retail chains 1 H/S and 3 H/S. Second, stores belonging to the same retail chain are aligned to different extents with that average (something we refer to as within retail chain price dispersion). For example, retail chain 6 H/S exhibits very little price dispersion for both products. However, in some cases the same retail chain has very different distributions of relative prices for Coke can and Coeur Lion camembert. For instance, stores belonging to retail chains 2 H and 2 H/S sell Coke cans at very similar prices, suggesting national pricing at the retail chain level for that product. This is not the case for Coeur Lion camembert.

In order to further investigate the pricing strategies across retail chains, we define a reference price at the retail chain level (instead of the national level, like it was the case in section 2.2). More precisely, in this section we consider as reference prices p_{irq}^{ref}, the quarterly product modal price within each retail chain r, and compute relative prices at the retail chain level in the following way:

$$p_{ist}^{rel(irq)} = (p_{ist} - p_{irq}^{ref})/p_{irq}^{ref} \tag{5}$$

Between retail chain price dispersion is linked to the extent to which retail chains choose product modal prices (p_{irq}^{ref}) that are far apart. Within retail chain price dispersion is instead measured by relative prices at the retail chain level ($p_{ist}^{rel(irq)}$), whose overall distribution is represented in Figure 6. A striking finding is that, if using as reference the national quarterly product modal price 17% of prices equaled it, using as reference the chain quarterly product modal price the percentage overall doubles to 35%. In other words, in one third of the stores of a retail chain products have the same exact price. Going back to the previous product examples, Figure 7 exhibits a very concentrated distribution of relative prices for Coke cans and Coeur de Lion camembert. Almost half of price observations across stores perfectly align with the reference price at the retail chain level.

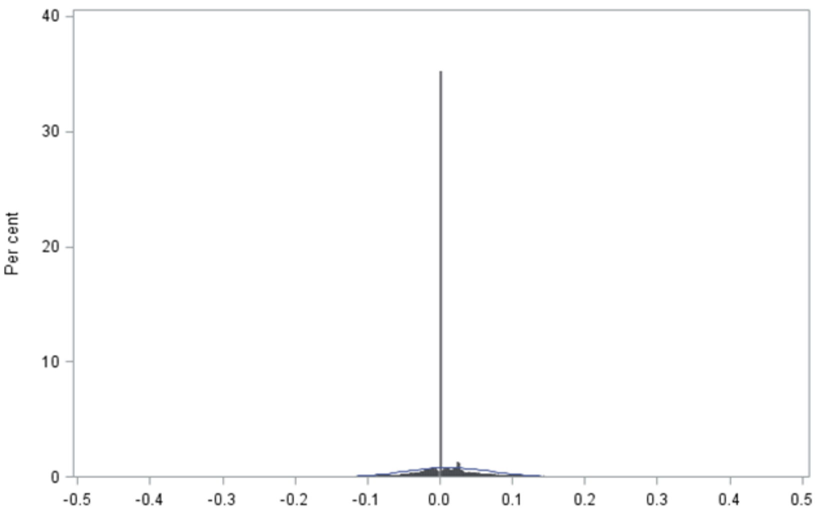

Figure 6: Distribution of the relative prices with respect to the product quarterly chain mode.

While retail chains do have strong national pricing strategies in place, there are systematic differences across retail chains in the extent those are pursued. In our two product example, for instance, Figure 8 shows that within retail chain price dispersion for Coke cans and Coeur de Lion camembert differ across retail

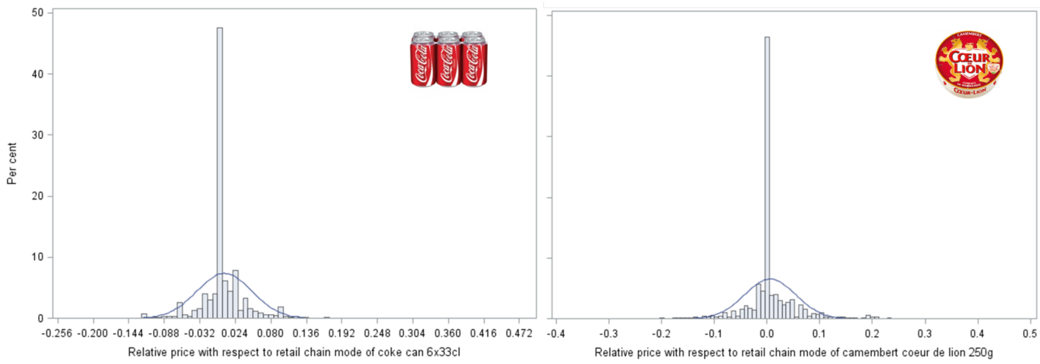

Figure 7: Distribution of the relative prices with respect to the quarterly chain mode of 6x33cl Coke cans (left panel) and of 250g Coeur de Lion camembert (right panel).

chains.

For one thing, the percentage of prices equal to the quarterly product chain modal price is larger in some retail chains and smaller in others, as shown in Figure 9.[16]

In order to assess quantitatively between and within retail chain price dispersion for each product, we compute the following two statistics. First, a measure of between chain modal price dispersion, defined as the average over retail chains and quarters of the absolute percentage deviation of product quarterly retail chain mode prices (p_{irq}^{ref}) from the corresponding national prices (p_{iq}^{ref}):

$$disp_{irq} = \left| (p_{irq}^{ref} - p_{iq}^{ref}) / p_{iq}^{ref} \right| \tag{6}$$

The left panel of Figure 10 shows its distribution. Among the products characterized by very similar modal prices across retail chains there are often alcoholic drinks,[17] while among the ones whose modal prices vary the most across retail chains there are many personal care and cleaning products. Second, a measure of within retail chain price dispersion, defined as the average over stores and quarters

[16]Notice that some retail chains (retailer 4, 5 and 6) have exclusively or mainly independent stores, while the opposite is true for other retailers (1 and 3). In the case of retail chains with both kind of stores, independent ones are less likely to align to retail chain modal prices.

[17]This may explain why price dispersion in our data is larger than in Dubois and Perrone [2015], where half of the products are alcoholic beverages.

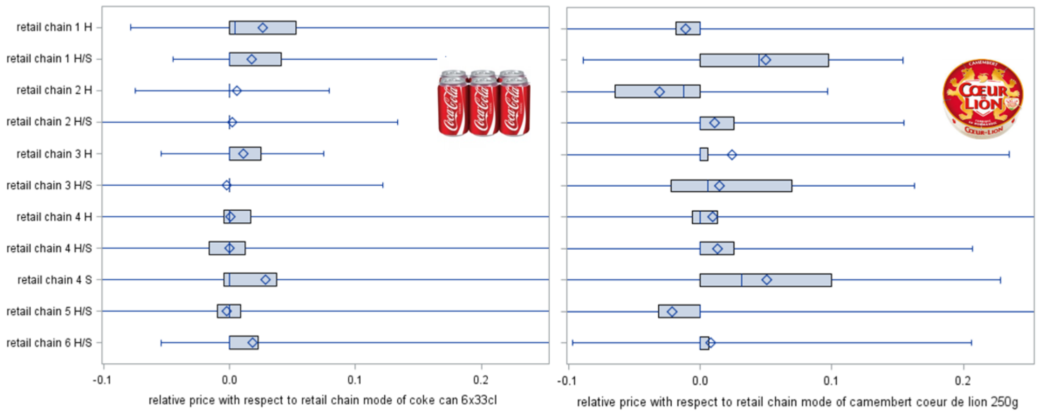

Figure 8: Boxplot of relative prices with respect to the quarterly chain modal price of 6x33cl Coke cans (left panel) and of 250g Coeur de Lion camembert (right panel) by retail chain.

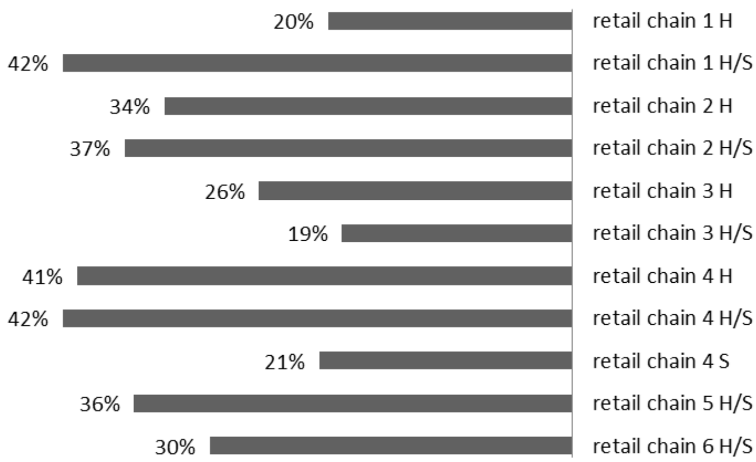

Figure 9: Percentage of prices equal to the quarterly product chain modal price by retail chain.

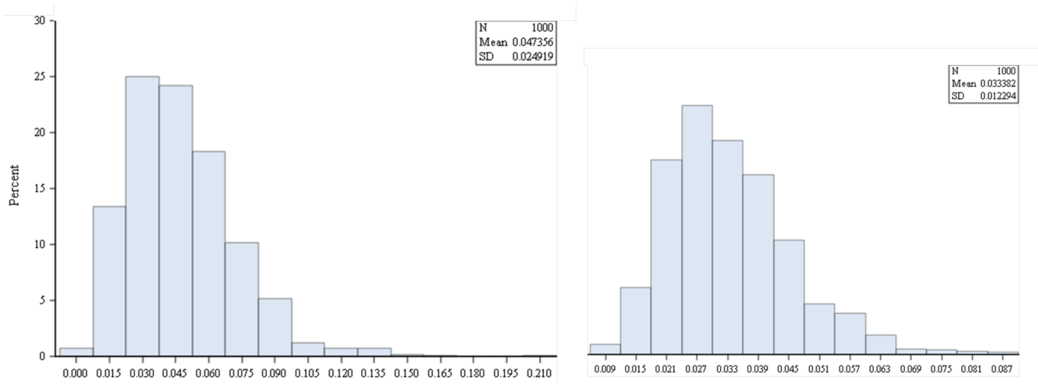

Figure 10: Distribution of between chain modal price dispersion (left panel) and distribution of within retail chain price dispersion (right panel)

of the absolute value of relative prices at the retail chain level ($p_{ist}^{rel(irq)}$). The right panel of Figure 10 plots its distribution. Interestingly, the categories of products exhibiting the smallest (respectively, the largest) variability across retail chains are also characterized by the smallest (respectively, the largest) within retail chain price dispersion. Comparing the mean and standard deviation of the two distributions suggests that between retail chain price dispersion on average dominates.

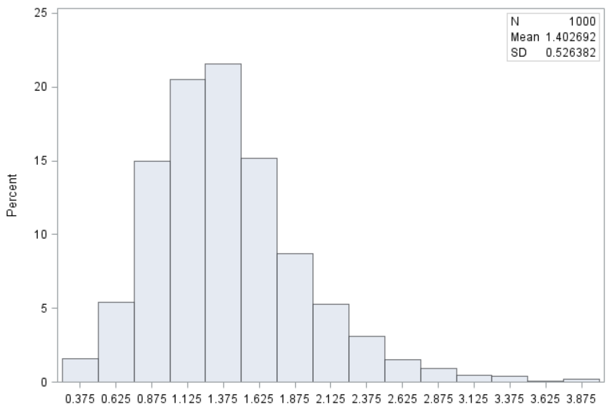

Figure 11: Distribution of the ratio of between and within retail chain price dispersion

We then take the ratio of between and within retail chain dispersion for each product. Figure 11 shows that for the bulk of products the absolute percentage deviations of retail chain modal prices from the national ones across retail chains are larger than the absolute price deviations within retail chains. Indeed, there are very few products for which the opposite is true. Among those there are several alcoholic drinks.

What explains that different retail chains set different prices for the same product? And what explains different prices within the same retail chain? Section 4.1 and 4.2 are an attempt to answer these questions respectively.

4.1 Sources of Price Dispersion Between Retail Chains

This section investigates the characteristics of retail chains that explain their modal prices. Therefore, we here abstract from price differences across stores belonging to the same retail chain and only look at the most often observed price for a product in each retail chain in a quarter. In order to describe the relative role of retail chain characteristics in their modal pricing strategy, we analyze their contribution to the overall explained variance in the following model:

$$p_{irq}^{ref} = \delta_i + \eta_1\, buyinggroup + \eta_2\, mktshares + \eta_3\, geopresence + \\ + \eta_4\, sellingareas + \eta_5\, prodranges + \zeta_{irq} \quad (7)$$

where δ_i are product fixed effects, *buyinggroup* is the buying group of a retail chain, *mktshares* is the overall market share of a retail chain, *geopresence* is the number of local markets where a retail chain is present with at least one store, *sellingareas* is the mean selling area of stores belonging to a retail chain, *prodranges* is the mean number of products in stores of a retail chain.

Disregarding product fixed effects, the main retail chain dimension in determining modal prices is its buying group, which explains 41% of modal price variance across retail chains. Market shares explain another 24% of the variance, closely followed by granular geographical presence (22% of the variance). In particular, Table 2 shows that retail chains with larger market shares and with stores in many local markets set lower modal prices. We also find that the mean selling area of stores of the retail chain decreases its modal prices, while the mean number of products sold by stores of the retail chain increases it. In other words, retail chains with larger assortment stores have higher modal prices, but larger stores per se have lower ones. The fact that the buying group and the market share of a retail chain explain alone 65% of their modal price variance suggests that an important stage of price setting is represented by the bargaining process with producers.

Table 2: Determinants of quarterly product modal prices of retail chains

Regressor	estimate	SE
market share	-224.49	5.1620***
geographical presence	-0.11	0.0026***
mean selling area	-0.002	0.0001***
mean number of products	0.01	0.0016***
product FE	yes	
buying group FE	yes	

4.2 Sources of Price Dispersion Within Retail Chains

What explains that different stores belonging to the same retail chain sell an item at different prices? This section investigates price deviations at the store level from the retail chain modal prices. We regress relative prices at the retail chain level on a number of store characteristics as follows:

$$\begin{aligned} p_{ist}^{rel(irq)} = \mu_i + \xi_1 \; localbranch + \xi_2 \; indepstore + \xi_3 \; storetype + \\ + \xi_4 \; sellingarea + \xi_5 \; prodrange + \xi_6 \; region + \xi_7 \; urban + \\ + \xi_8 \; sameprod + \xi_9 \; closest + \xi_{10} \; income + \xi_{11} \; pop + \vartheta_{ist} \end{aligned} \quad (8)$$

where μ_i are product fixed effects, $localbranch$ is the local branch of the retail chain to which a store is associated, $indepstore$ is a dummy variable taking value 1 if the store is independent and 0 if it is integrated, $storetype$ is a 4-level categorical variable for store size, $prodrange$ is the number of products in a store, $sellingarea$ is the square meter selling area of a store, $region$ is the region where a store is located, $urban$ is a 4-level categorical variable for urban density, $sameprod$ corresponds to the number of supermarkets selling the product in the district, $closest$

is the distance in kilometers to the closest large supermarket, *income* and *pop* are respectively log per capita income and population of district.

The analysis of variance defined by model (8) suggests that the main store characteristic determining price dispersion within retail chains is the local branch to which a store is associated. In the case of some retailers, stores are organized in larger areas with only half a dozen local branches nationally. Other retailers have an up to three times more granular local organization. Although retailers have a heterogeneous number of local branches, they often adjust prices at that level. Indeed, more than two thirds of relative price variance explained by the model is explained by local branch dummies. Conditional on local branches, Table 3 shows that independent stores have on average higher prices than the retail chain modal ones compared to integrated ones. On the top of local branches, also regional dummies substantially contribute to explaining price deviations with respect to retail chain modal prices. Moreover, stores located is more rural areas set slightly lower prices than the retail chain modes.

The type of store explains 4% of price variation with respect to retail chain modal prices. In particular, as one would expect, larger stores and 'drives-entrepôt' are able to set lower prices, while small supermarkets set higher prices than medium size ones. Moreover, larger selling areas are also associated with lower prices. Similarly, stores with larger assortment of products are characterized on average by prices below retail chain modal ones, probably due to economies of scale, too.

Finally, local competition and demand have a significant, although quantitatively small, effect on the prices set by stores. Indeed, stores facing stronger competition in terms of the number of stores selling the same product in the same local market set lower prices, while the further away the closest large supermarket, the higher local income and population, the higher the prices with respect to the retail chain modal prices.

In conclusion, within retail chains' price dispersion mainly depends on local branches' decisions and secondary on stores' characteristics and location.

Table 3: Determinants of relative prices with respect to the quarterly product modal prices of retail chains

Regressor	estimate	SE
independent store	0.009	0.00003***
rural	-0.019	0.00041***
semi-urban	-0.021	0.00004***
urban	-0.019	0.00004***
metropolitan (ref)	.	
drives-entrepôt	-0.012	0.00003***
large size supermarkets	-0.007	0.00001***
medium size supermarkets(ref)	.	
small supermarkets	0.036	0.00007***
selling area	-0.000001	0.00000***
number of products	-0.00001	0.00000***
n.stores selling product	-0.00004	0.00000***
closest large supermkt	0.0001	0.00000***
log local per capita income	0.001	0.00002***
log local population	0.001	0.00000***
product FE	yes	
local branch FE	yes	
region FE	yes	

5 Conclusion

Based on a large and original data set containing more than 250 millions of daily price records from about 1600 medium and large size supermarkets in France over the period October 2011 to September 2012, we characterize the overall shape and structure of price dispersion in the French retail sector. We show that temporary sales and promotions explain only little of the observed price dispersion, while the permanent component of price dispersion largely dominates.

In fact in France price dispersion across stores is essentially the result of persistent heterogeneity in retail chains' national pricing. We further explore the determinants of price setting differences across retail chains, as well as across stores belonging to the same retail chain. We find evidence of a multistage price setting process. First, retail groups bargain with producers and the market share of the corresponding retail groups affect their bargaining power. Second, retail groups set national prices at the retail chain level (i.e., retail groups owning more than one retail chains set different prices across them). Within a retail group, for instance, prices are lower in chains characterized by larger stores. More in general the average level of prices depends on the positioning and the customers' target of the retail chain. Despite this rather centralized price-setting behavior, we show that there is a third stage of price setting at the level of local branches. The last fine-tuning of prices is performed based on local conditions regarding demand or local competition between supermarkets.

References

Anderton, R., A. Meyler, L. Gattini, M. Izquierdo, V. Jarvis, R. Kaarup, M. Komzakova, B. Landau, M. Mohr, A. Page, D. Sondermann, P. Vermeulen, D. Cornille, T. Tsalinski, Z. Vladova, C. Hartmann, H. Stahl, S. Linehan, C. Balfoussia, S. Panagiotou, M. de los Llanos Matea, L. Alvarez, P. Bardet-Fremann, N. Berardi, P. Sevestre, E. Ciapanna, C. Rondinelli, D. Kapatais, E. Walch, P. Lunnemann, S. Zerafa, C. Pace, J. Kieft, F. Fritzer, F. Cardoso, M. Gabrijelcic, B. Karma, and J. Kivisto (2011). Structural features of distributive trades and their impact on prices in the euro area. European Central Bank Occasional Paper Series 128.

Asplund, M. and R. Friberg (2002). Food prices and market structure in Sweden. *Scandinavean Journal of Economics 104*(4), 547–566.

Berardi, N., E. Gautier, and H. Le Bihan (2015). More facts about prices: France before and during the great recession. *Journal of Money, Credit and Banking 47*, 1465–502.

Berardi, N., P. Sevestre, and J. Thébault (2017). *The Econometrics of Multi-Dimensional Panels*, Chapter The Determinants of Consumer Price Dispersion: Evidence from French Supermarkets. Springer.

Borenstein, S. and N. L. Rose (1994, August). Competition and Price Dispersion in the U.S. Airline Industry. *Journal of Political Economy 102*(4), 653–83.

Clay, K., R. Krishnan, and E. Wolff (2001). Prices and Price Dispersion on the Web: Evidence from the Online Book Industry. *Journal of Marketing Research 49*(4), 521–539.

Degeratu, A., A. Rangaswamy, and J. Wu (2000). Consumer choice behavior in online and traditional supermarkets: The effects of brand name, price, and other search attributes. *International Journal of research in Marketing 17*(1), 55–78.

Dubois, P. and H. Perrone (2015, October). Price dispersion and informational frictions: Evidence from supermarket purchases. CEPR Discussion Papers 10906, C.E.P.R. Discussion Papers.

French Competition Authority (2010). La concentration des zones de chalandise.

Gerardi, K. S. and A. H. Shapiro (2009, 02). Does Competition Reduce Price Dispersion? New Evidence from the Airline Industry. *Journal of Political Economy 117*(1), 1–37.

Gorodnichenko, Y., V. Sheremirov, and O. Talavera (2018). Price Setting in Online Markets: Does IT Click? *Journal of the European Economic Association 16*(6).

Gorodnichenko, Y. and O. Talavera (2017). Price setting in online markets: Basic facts, international comparisons, and cross-border integration. *American Economic Review 107*(1), 249–82.

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

Kaplan, G. and G. Menzio (2015). The morphology of price dispersion. *International Economic Review 56*(4), 1165–206.

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

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

Woodward, S. E. and R. E. Hall (2012). Diagnosing consumer confusion and sub-optimal shopping effort: Theory and mortgage-market evidence. *American Economic Review 102*(7), 3249–3276.