WORKING paper

BANQUE DE FRANCE
EUROSYSTÈME

The Determinants of Consumer Price Dispersion: Evidence from French Supermarkets

Nicoletta Berardi[1] & **Patrick Sevestre**[2] & **Jonathan Thébault**[3]

June 2017, WP # 632

ABSTRACT

We characterize the dispersion of grocery prices in France based on a large original data set of prices in more than 1500 supermarkets. On average across products, the 90th percentile of relative prices is 17 percentage points higher than the 10th and the mean absolute deviation from quarterly average product prices is 5%. We show that temporal price variations (including sales and promotions) explain only little of the observed price dispersion, while the spatial permanent component of price dispersion largely dominates. Price dispersion across stores in France essentially results from persistent heterogeneity in retail chains' pricing, while local conditions regarding demand or competition contribute to a much lower extent.[4]

Keywords: price dispersion, retail chain, wholesaler

JEL classification: E31, D40

[1] Banque de France, nicoletta.berardi@banque-france.fr
[2] Greqam-AMSE, patrick.sevestre@univ-amu.fr
[3] French Embassy in Brasil, thebault.jonathan@gmail.com
[4] 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. We wish to address special thanks to Alexandre Vigneron, Marine Tépaut, Sherly Jean-Charles, Sophie Saleh, Sylvie Tarrieu, and Yue Zhu for their precious assistance with complementary data collection and treatment. Finally, we thank Vincent Guegan and Victor Yammouni for life-saving IT assistance.

NON-TECHNICAL SUMMARY

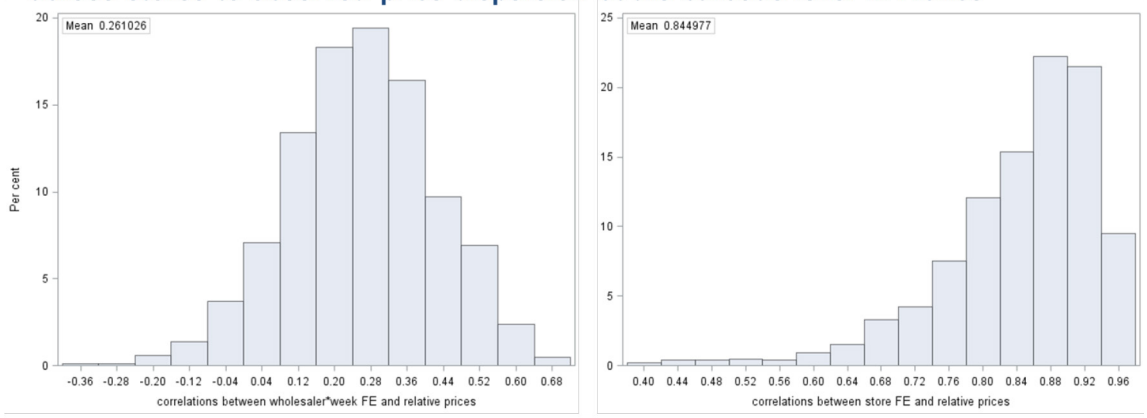
Shedding light on firms' pricing decisions is obviously key for a better understanding of the economy and of the impact of economic policies. However, the scarcity of microeconomic data on prices has been for a long time an impediment to a microeconometric investigation of firms' pricing decisions. Luckily, the last fifteen years have seen a significant improvement in researchers' access to microeconomic data on prices. In particular, access to CPI price records has allowed thorough analyses of the dynamics of consumer prices in many countries. In the meantime, price levels have been less thoroughly studied. Indeed, for confidentiality reasons, data available to researchers for studying price dynamics often do not allow to identify precisely and separately the products and the stores where they were sold. It is then impossible to consider the dispersion of prices of a precisely defined product across stores.

This paper exploits an original dataset containing price data of a large number of precisely defined products sold in different stores at different dates provided by a price comparator. More precisely, our analysis is based on a data set containing almost 40 million weekly price records regarding one thousand products (individually identified by an EAN barcode) sold in more than 1500 geolocalized medium and large size stores over the period October 2011 to September 2012.

Our contribution in this paper is twofold. First, we characterize the overall dispersion of consumer prices in the French grocery retail sector. On average across products, the 90^{th} percentile of relative prices is 17% higher than the 10^{th}. The mean absolute deviation with quarterly average product log prices as the measure of central tendency is 5% on average in the French retail sector and the standard deviation of relative prices is 7%. The comparison with price dispersion measures available for the US, the UK and Canada suggests that in France prices are less dispersed.

Second, we decompose the observed price dispersion into several components by estimating, for each product separately, a fixed effect model disentangling the relative contribution temporal price variations (including sales) and that of permanent price differences across stores to the observed price dispersion at the barcode level. We show that the permanent spatial component of price dispersion largely dominates temporal price dispersion. In a second stage we then show that the latter strongly depends on retail chains, while local markets' characteristics play a significant role but to a much lower extent. In other words, in France price dispersion across stores is essentially the result of persistent heterogeneity in retail chains' national pricing. Indeed, retail groups bargain with producers and then set prices at the level of retail chains, taking into account their positioning and customer target. For instance, prices are lower in chains with larger stores. Despite this rather centralized price-setting behavior, we show that local conditions regarding demand or local competition between supermarkets also matter for explaining observed prices on local markets, though to a much lower extent.

Relative contribution of temporary discounts and permanent price differences across stores to observed price dispersion at the barcode level in France



Note: Distribution of the correlation between relative prices and the estimated fixed effects (FE) for the combinations of wholesalers (or regional branches) and weeks, as well as for stores, product by product. The former account for temporal price dispersion (i.e., stores vary their prices over time, for instance through temporary discounts), while the latter capture spatial price dispersion (i.e., persistent characteristics of stores determining their prices).

Les déterminants de la dispersion des prix à la consommation : Analyse de la grande distribution en France

RÉSUMÉ

À partir d'une base de données constituée des prix observés dans plus de 1500 magasins de la grande distribution, nous caractérisons la dispersion des prix en France. En moyenne, tous produits confondus, le 90ème centile des prix relatifs est supérieur de 17 points de pourcentage au 10ème centile et l'écart absolu moyen par rapport au prix moyen trimestriel de chaque produit est de 5 %. Nous montrons que les variations temporelles de prix (dont les soldes et les promotions) expliquent peu la dispersion des prix observée, tandis que la composante spatiale permanente de la dispersion des prix domine largement. La dispersion des prix entre magasins en France résulte essentiellement de l'hétérogénéité persistante des prix au niveau des enseignes, tandis que les conditions locales de demande ou de concurrence y contribuent de façon beaucoup moins importante.

Mots-clés : dispersion des prix, enseigne, grande distribution, centrale d'achat

1 Introduction

Why do prices of the same good often significantly differ across shops? Deviations from the law of one price are among the most important issues that have attracted the attention of economists for a long time. However, due to the scarcity of adequate microeconomic price data, the first studies on price dispersion could only focus on specific markets. For instance, Borenstein and Rose [1994] and Gerardi and Shapiro [2009] zoom on the airline industry in the US. Sorensen [2000] analyzed the distribution of prices for several drugs across different pharmacies in two cities in upstate New York. Hong and Shum [2006] documented the distribution of prices posted by online booksellers for four academic textbooks. Moraga-González and Wildenbeest [2008] did the same for online sellers of several computer memory chips, while Woodward and Hall [2012] for mortgage brokerage services.

The last fifteen years have seen a significant improvement in researchers' access to microeconomic data on prices. While the access to CPI price records allowed thorough analyzes of the dynamics of consumer prices and have been used to establish the link between overall inflation and price dispersion,[1] the characteristics of CPI data, as released to researchers, most often do not allow a precise analysis of price dispersion[2]. Fortunately, household surveys or scanner data collected by marketing companies such as Nielsen or Kantar have recently been made available. These data sets generally contain prices of very precisely defined products in different outlets at different dates, thus allowing precise quantification of price dispersion.

This paper mainly relates with papers studying the overall shape and the structure of price dispersion in the retail sector. Asplund and Friberg [2002] use price data from a household survey implemented by the Swedish Pensioners' National Organization for five grocery products (sugar, washing-up detergent, crisp bread, spread, and cocoa) sold in about 1000 supermarkets in Sweden between 1993 and 1997 and find that most of price dispersion is explained by supermarket specific factors (like supermarket size and retail chain) and, to a lesser extent, by factors

[1] E.g., see Nakamura et al. [2016] for the US, Dixon et al. [2014] for UK, Alvarez et al. [2013] for Argentina, Moen et al. [2014] in Norway, and Eyal and Eden [2004] in Israel.

[2] One exception is Cheung and Fujii [2008] for Japan.

related to costs of sales space and transport costs. Kaplan and Menzio [2015], based on the Kilts-Nielsen Consumer Panel Data set in the US over the period 2004-2009, find that in the US the variance of prices is mainly accounted for by idiosyncratic supermarket characteristics and time variation, and only little by the retailer to which the supermarket belongs. They also find that the typical distribution of normalized prices is characterized by a unique mode and by a 19% average standard deviation for identical goods and that it is symmetric and leptokurtic. As far as the French retailing sector is concerned, Dubois and Perrone [2015] have been the only ones up to now to study price dispersion at barcode level. Their analysis is based on a household survey about consumers' purchasing of food products over 3 years (1999, 2000, and 2001). They consider information on product and supermarket characteristics, as well as household demographics. Based on four product categories (beer, coffee, cola, and whisky), they show that price dispersion is prevalent in the French food market.

This paper assesses and analyzes consumer price dispersion in France. We are able to characterize the overall shape and structure of price dispersion in the French retail sector based on an original data set containing almost 40 millions of weekly price records from geo-localized medium and large size supermarkets in France. After trimming, there are about a million and a half price trajectories regarding one thousand products (individually identified by an EAN barcode[3]) sold in more than 1500 supermarkets over the period October 2011 to September 2012.[4]

Our contribution in this paper is twofold. First, we characterize the overall dispersion of consumer prices in the French retail sector. On average across products, the 90^{th} percentile of relative prices is 17% higher than the 10^{th}. The mean absolute deviation with quarterly average product log prices as the measure of central tendency is 5% on average in the French retail sector and the standard deviation of relative prices is 7%. Second, we decompose the observed price dispersion into several components. We first assess the relative contribution of sales

[3]The EAN barcodes (originally European Article Number, now renamed International Article Number even though the abbreviation EAN has been retained) are international product identification numbers.

[4]These data were collected and made available to us by Prixing, a start-up company providing consumers with a free mobile price comparator (see http://www.prixing.fr/).

and temporary promotions and that of permanent price differences across shops to the observed price dispersion at the barcode level. In a second stage we then show that the latter strongly depends on retail chains, while local markets' characteristics play a significant role but to a much lower extent.

The remaining of the paper is structured as follows. A detailed presentation of the data sets is given in Sect. 2. Section 3 characterizes descriptively price dispersion in France and investigates its structure, while Sect. 4 disentangles the permanent and transitory components of price dispersion and further investigates the role played by national and local factors as far as the former is concerned. Section 5 concludes.

2 Data Description

Our analysis is based on the combination of two original data sets. The first one provides millions of price records regarding a thousand products sold in more than 1500 medium and large supermarkets, together with precise identifiers of both products (through their EAN barcodes) and supermarkets (through their names and addresses). The second one is an exhaustive data base of all supermarkets in France, including their names and locations.

2.1 Grocery Price Data

Our analysis of price dispersion in the grocery sector is based on millions of price records from geo-localized medium and large supermarkets in France. The price data was collected by Prixing, a start-up that offers a mobile price comparison app and website. Prixing has developed automatic procedures enabling it to collect price lists from supermarkets offering 'click&collect' services, known as 'drives' in France (which refers to the fact that customers 'drive through' close by the supermarket). This retail channel is a relatively recent service developed by French supermarkets over the past 10 years.[5] It has two distinctive features: customers

[5]There were around 500 drives in France at the end of 2010, their number doubled by the end of 2011, to reach almost 2000 by the end of 2012 and about 2700 by the end of 2013 (Dauvers [2013a] and Dauvers [2013b]).

order their shopping on the supermarket's click&collect website and then collect it at a drive-through facility.[6] More importantly for our analysis, prices are exactly the same as those of the brick-and-mortar supermarket associated with the drive-through.[7] Therefore, our price data base corresponds to grocery prices in medium and large supermarkets, which represent in France more than 80% of grocery sales (Anderton et al. [2011]).

The original data base contains slightly more than 45 millions price spells[8], corresponding to 2.3 billion daily prices of about 90 thousand products in about 1600 supermarkets, over the period from October 2011 to September 2012. Due to computational challenges, we reduced the size of the data set. First, we based our analysis on one price observation per week, chosen as the most frequently observed price over a week for each specific product sold in a specific supermarket (i.e., the mode of the weekly price distribution). This does not induce any significant loss of information regarding price dispersion, as the within week price variance is null in almost 99% of cases. Second, to keep our econometric estimation feasible,we also restricted the sample to the one thousand most widely sold products (i.e., 1000 barcodes). The resulting weekly modal price data set contains more than 37 million observations (almost 1.5 million trajectories) of the most widely sold products.[9]

[6]Most click&collect drive-throughs are associated with a supermarket, but there are also a few stand alone drive-throughs (know as 'drives-entrepôt' in French).

[7]One of the major retail chains in France at some point tested a different pricing strategy whereby a few products are cheaper when bought through the click&collect than in the associated supermarkets. However, this practice was not yet implemented at the time of the collection of the data we use here. More in general, it may happen that some discounts are available in the brick-and-mortar store and not in click&collect and vice-versa.

[8]A 'price spell' is made of the following three elements:

- the price of a precisely defined product i, identified by its barcode (e.g., a 1 liter glass bottle of brand b pure orange juice) in a given store s (e.g., the supermarket from retail chain r, located at a given address),
- the date when this price was first set (start date of the price spell),
- the date when this price was changed, even possibly only temporarily (end date of the price spell).

For instance, if the price of a product in a store decreases from 1 euro to 80 euro cents and then increases to 90 euro cents, this will define 3 price spells.

[9]We discarded spells with inconsistent start and end dates (e.g., a start date posterior to the

One advantage of our price data with respect to those collected from scanners or from household surveys is that each price does correspond to what the consumer would really pay for a product on the day the data was collected and not to an average unit value computed from several transactions recorded over a period.

Table 1 shows the category of available products based on the Classification of Individual Consumption by Purpose (COICOP). Both in terms of number of products and corresponding price observations, there is a strong predominance of food and beverages (almost 77% of barcodes); about 13% of barcodes are 'miscellaneous' goods like personal care products (e.g., toothpaste, shampoo, etc.) and about 6% 'furnishings and household equipment and routine maintenance' (e.g., washing-up liquid or dishwasher detergent) and 4% 'recreation and culture' (e.g., CDs, toys, etc.). The last two columns report the number of brands and producers by COICOP-level1 product category.

Table 1: Number and percentages of products by COICOP product categories and brand/manufacturer

COICOP-level 1	prices		products		brands	manufact.
Food & non-alc.bev.	26231005	70.6%	703	70.3%	250	76
Alc.beverages	2320956	6.3%	66	6.6%	42	22
HH eqpt.&maint.	1992881	5.4%	55	5.5%	26	8
Recreation&culture	1580217	4.2%	42	4.2%	14	3
Miscellaneous	5005850	13.5%	134	13.4%	39	12
Total	37130909	100%	1000	100%	371	121

2.2 Supermarket Data and Competition Measures

The second source of information that we exploit is an exhaustive data base of all medium and large supermarkets in France.[10] For each supermarket it includes in particular their name, location, retail chain and regional branch/wholesaler[11]. In

end date). We also dropped null prices (4 in the whole sample), as well as observations of prices whenever their log-difference to the average national price was larger than 2 in absolute value.

[10]The data was bough from LSA (Libre Service Actualités http://expert.lsa-conso.fr/).

[11]In France retail groups have regional branches, which in some cases operate as wholesalers and in other cases do not directly bargain with producers, but have a role in setting prices in their region.

the data set we identified the supermarkets for which we have price records.[12]

Table 2 shows the retail chains of the supermarkets for which we have price records. Those supermarkets belong mostly to the retail groups Carrefour, Systeme U and Les Mousquetaires with more than 18% of supermarkets each (second column of Table 2).[13] Some retail groups are over-represented with respect to their share in the total population of medium and large supermarkets. For instance, almost 30% of the sample of supermarkets for which we have price records belong to the group Systeme U and 14% to Leclerc, while at the national level Systeme U has 16.5% and Leclerc 8.4% of the supermarkets. Carrefour is instead under-represented among the supermarkets with price observations with respect to the total number of Carrefour's supermarkets in France. Notice also that, since click&collect are more frequently associated with large rather than with medium size supermarkets, retail chains of large supermarkets are over-represented with respect to those typically having smaller supermarkets. Indeed, about 40% of supermarkets for which we have price records are large supermarkets. However, this proportion is close to their market share in the French retail sector (see Anderton et al. [2011]). The last column of Table 2 reports the retail chains' market shares in 2012 as reported by Kantar Worldpanel-LSA.

The supermarkets for which we have price records are a rather representative sample in terms of their geographical distribution (see Fig. 11 in the Appendix as compared to the universe of medium and large supermarkets in France represented in Fig. 10). We also consider their location in smaller geographical areas or local markets, which correspond to French 'arrondissements' (which may roughly be translated into English as districts). Prices in our data set concerns supermarkets located in 286 local markets (out of 330 in mainland France). On average there about 13 local markets per region and, in each local market, 6 supermarkets for which we have price records.

[12] In about 50 cases finding the right correspondence proved to be problematic. We then discarded price observations relative to those stores, in order to avoid making mistakes in attributing the wrong store specific characteristics to our data set of price observations.

[13] Notice that we have omitted retail chains for which no price records are available and normalized the percentages of the remaining ones to 100%.

Table 2: Stores' retail chain/group and market shares

Retail chain	N.store (w/prices)	%	N.all stores (med.-large)	%	Mrkt shares [Kantar]
Group AUCHAN:	*92*	*5.9%*	*443*	*7.0%*	*14.5%*
Auchan (large supermkts)	55	3.5%	139	1.8%	11.2%
Simply Market (medium)	37	2.4%	304	4.0%	3.3%
Group CARREFOUR:	*286*	*18.2%*	*1733*	*27.7%*	*24.9%*
Carrefour (large)	186	11.9%	226	3.0%	14.8%
Carrefour Market (large&med.)	100	6.4%	1507	19.8%	10.1%
Group CASINO:	*126*	*8.0%*	*506*	*8.1%*	*6.7%*
Geant Casino (large)	87	5.5%	150	2.0%	3.5%
Casino (large&medium)	39	2.5%	356	4.7%	3.2%
Group SYSTEME U:	*468*	*29.8%*	*1033*	*16.5%*	*12.2%*
Hyper U (large)	55	3.5%	64	0.8%	
Marché U (medium)	7	0.5%	8	0.1%	
Super U (large&medium)	390	24.8%	717	9.4%	
U Express (medium)	16	1.0%	244	3.2%	
Group LECLERC:					
Leclerc (large)	222	14.1%	641	8.4%	24.8%
Group MOUSQUETAIRES:					
Intermarché (large&medium)	377	24.0%	1837	24.1%	16.8 %
Total	1571	100%	7596	100%	100.0%

3 Assessing Price Dispersion in the French Retail Sector

In this paper we define price dispersion as price differences within a quarter for exactly the same product (i.e., barcode) sold in different supermarkets.[14]

In order to measure price dispersion we first compute percentage price deviations from a quarterly reference price for each product i. In particular, we define the relative price of a product i in a quarter q as $p_{ist}^{rel(iq)} = \ln(p_{ist}) - \overline{\ln(p_{iq})}$, where p_{ist} is the price of a product i sold in supermarket s in week t and $\overline{p}_{iq}$ is the quarterly average log price of product i over supermarkets. Figure 1 shows the distribution of the full set of relative prices in our data.

Price dispersion can then be summarized in different ways. Among the possible measures there are:

[14]Price dispersion may also be considered using less strict definitions of a product. Some studies for instance also look at price differences for the same product categories (e.g. Kaplan and Menzio [2015]).

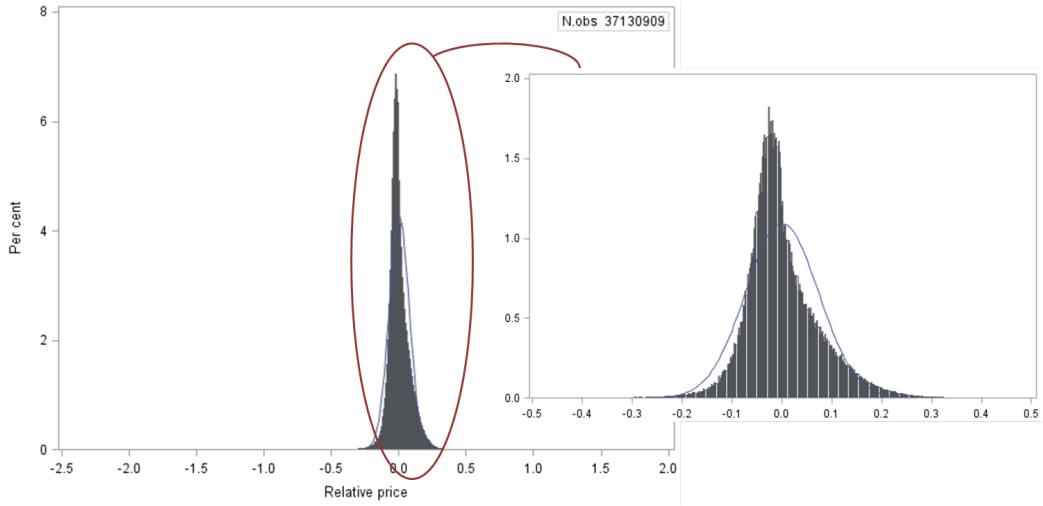


Figure 1: Distribution of the relative prices.

- the standard deviation of relative prices: $\sigma(p_{ist}^{rel(iq)})$,
- the interquartile range of relative prices: $IQ(p_{ist}^{rel(iq)}) = Q3(p_{ist}^{rel(iq)}) - Q1(p_{ist}^{rel(iq)})$,
- the interdecile range of relative prices: $ID(p_{ist}^{rel(iq)}) = P90(p_{ist}^{rel(iq)}) - P10(p_{ist}^{rel(iq)})$,
- the mean of absolute deviation : $MAD(p_{ist}^{rel(iq)}) = 1/4n \sum_{i=1}^{n} \sum_{q=1}^{4} |p_{ist}^{rel(iq)}|$.

Table 3 provides the standard deviation, interquartile range, interdecile range, and mean of absolute relative prices computed over all observations. On average across products, the 90^{th} percentile of relative prices is 17 percentage points higher than the 10^{th}. The interquartile range is 8 percentage points and the mean absolute deviation of relative prices about 5% in France, pointing at a non-trivial overall price dispersion. The distribution of the mean absolute deviation of relative prices by product is represented in Fig. 2.

For the sake of comparison, Table 4 reports several measures of price dispersion provided by other studies. The only assessment of price dispersion at the barcode level available for France is provided by Dubois and Perrone [2015]. The rest of Table 4 reports measures of price dispersion computed for the US, the UK, and Canada (Kaplan and Menzio [2015], Gorodnichenko et al. [2014], and Gorodnichenko and Talavera [2017]). The last column corresponds to our calculation of

Table 3: Different measures of price dispersion

Measures of price dispersion		
Standard deviation of relative prices	$\sigma(p_{ist}^{rel(iq)})$	0.07
Interquartile range of relative prices	$Q3(p_{ist}^{rel(iq)}) - Q1(p_{ist}^{rel(iq)})$	0.08
Interdecile range of relative prices	$P90(p_{ist}^{rel(iq)}) - P10(p_{ist}^{rel(iq)})$	0.17
Mean absolute deviation	$1/4n\sum_{i=1}^{n}\sum_{q=1}^{4}\lvert p_{ist}^{rel(iq)}\rvert$	0.05
N.obs		37,130,909
N.products		1,000

Note: Statistics are computed at the barcode-level and then averaged across barcodes.

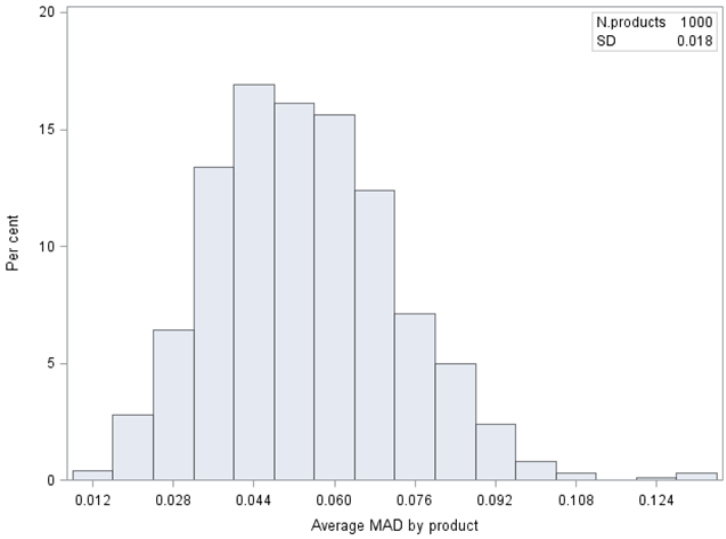


Figure 2: Distribution of MAD by products.

the same measures of price dispersion as in those studies, but based on our data set.

The first conclusion that can be drawn from Table 4 is that price dispersion in our data is quite close to that obtained by Dubois and Perrone [2015]. 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).The second conclusion emerging from comparing statistics of price dispersion computed for the US, the UK and Canada with the same statistics computed on our data is that in France prices appear to be less dispersed. In particular, the comparison with price

Table 4: Comparison with other measures of price dispersion in the retail sector

Dubois and Perrone [2015]:		
France; 1999-2001; 2 barcodes of beer, coffee, cola, whisky		with our data:
Coefficient of variation of prices	0.11	0.08
Interquartile ratio of prices	1.14	1.09
$95^{th}/5^{th}$ centile ratio of prices	1.37	1.26
Kaplan and Menzio [2015]:		
US; 2004-2009		
Standard deviation of normalized prices	0.19	0.06
$90^{th}/10^{th}$ centile ratio of normalized prices	1.72	1.13
$90^{th}/50^{th}$ centile ratio of normalized prices	1.26	1.07
$50^{th}/10^{th}$ centile ratio of normalized prices	1.35	1.06
Gorodnichenko et al. [2014]:		
US; May2010-Feb2012; goods sold on online shopping platform		
Standard deviation of log prices	0.24	0.07
Coefficient of variation of prices	0.22	0.07
Interquartile range of log prices	0.35	0.09
Gorodnichenko et al. [2014]:		
UK; May2010-Feb2012; goods sold on online shopping platform		
Standard deviation of log prices	0.23	0.07
Coefficient of variation of prices	0.19	0.07
Interquartile range of log prices	0.31	0.09
Gorodnichenko and Talavera [2017]:		
US; Nov2008-Sep2012; goods sold on online price comparator		
Standard deviation of log prices	0.16	0.07
Interquartile range of log prices	0.17	0.09
Gorodnichenko and Talavera [2017]:		
Canada; Nov2008-Sep2012; goods sold on online price comparator		
Standard deviation of log prices	0.13	0.07
Interquartile range of log prices	0.11	0.09

Note: Statistics are computed at the barcode-level and then averaged across barcodes.

dispersion across brick-and-mortar stores measured by Kaplan and Menzio [2015] suggests that US prices are 20 to 50% more dispersed than in France, depending on the statistic considered.

Price dispersion may differ across many dimensions due to factors related to the product, store or time. Among the dimensions that characterize a product i, there are its brand b, its manufacturer p, and its product category (COICOP-level4) k. A supermarket s can be characterized in terms of its regional branch/wholesaler, denoted wholesaler w as a shortcut, and retail chain r, as well as by its location in

market m and region g.

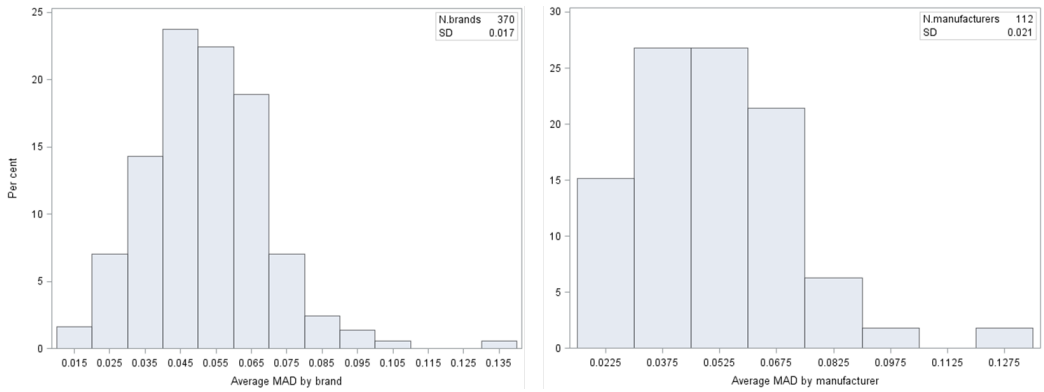


Figure 3: Distribution of MAD by products' brand and manufacturer.

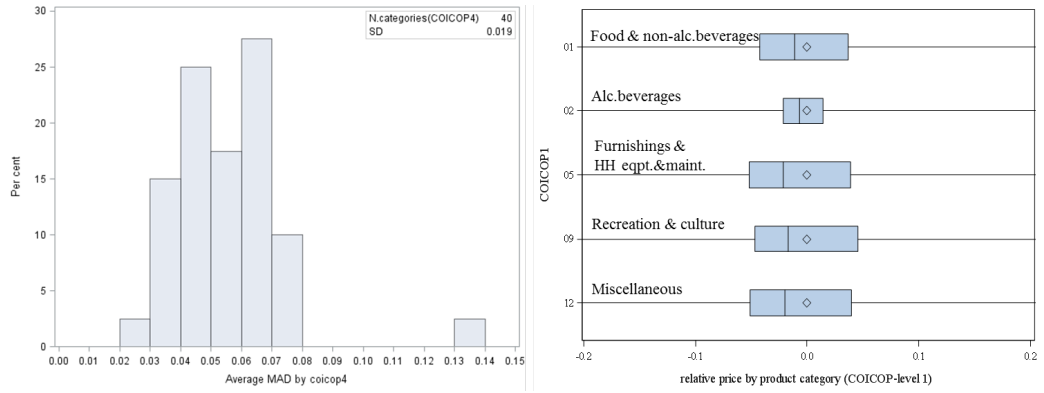


Figure 4: Distribution of MAD by products' category COICOP-level 4 and boxplot of relative prices by products' category COICOP-level 1.

In the product dimension, Fig. 3 suggests that price dispersion varies depending on the product's brand and manufacturer. The distribution of the mean absolute deviation is asymmetric, with a long right tail, indicating that products of some brands (and manufacturers) are characterized by deviations from the product's quarterly average log price of an order of more than 10%. The extent of price dispersion also differs across product categories. The left panel of Fig. 4 shows that some product categories at COICOP-level 4 are characterized by a larger price dispersion than other ones. The right panel of the same figure suggests that the median relative price is below the mean for all wide product categories (i.e., COICOP level 1).

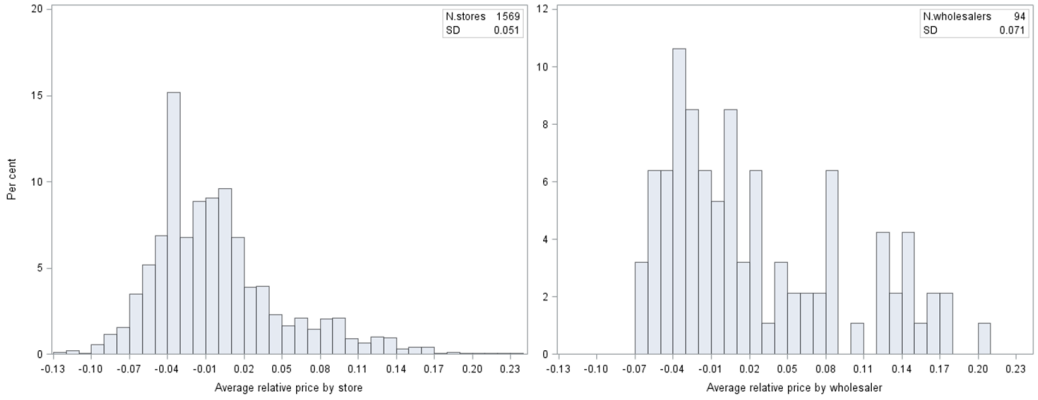


Figure 5: Distribution of average relative prices by stores and wholesalers.

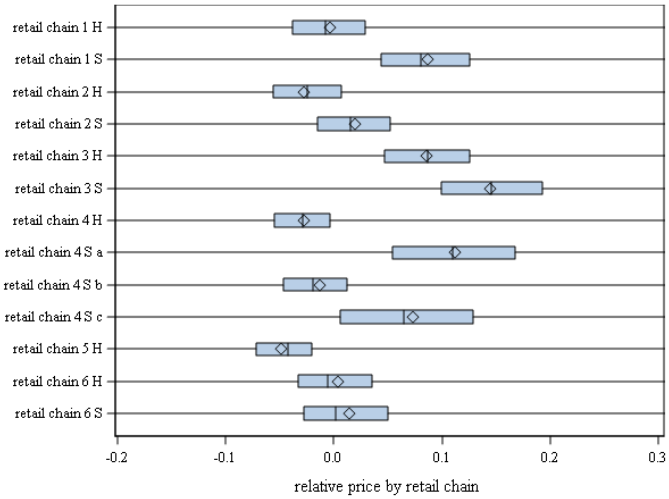


Figure 6: Boxplot of relative prices by retail chains.

Regarding the store dimension, the left panel of Fig. 5 shows that price dispersion is quite heterogenous across supermarkets. The right panel of the figure suggests that the regional branch / wholesaler affiliation of supermarkets accounts for a significant proportion of the observed price dispersion. Looking more specifically at retail chains, each line in Fig. 6 represents the boxplot of relative prices for a French retail chain. The average relative prices, as well as their dispersion, largely vary across retail chains.

Another dimension characterizing stores is their location. The long right tail

of the left panel of Fig. 7 suggests that some local markets are characterized by rather large average price dispersion and the right panel that relative prices are more dispersed in some regions, like for instance Île-de-France (region number 11, where Paris is), Corse (region number 94), or Provence-Alpes-Côte d'Azur (region number 93), than in others.[15]

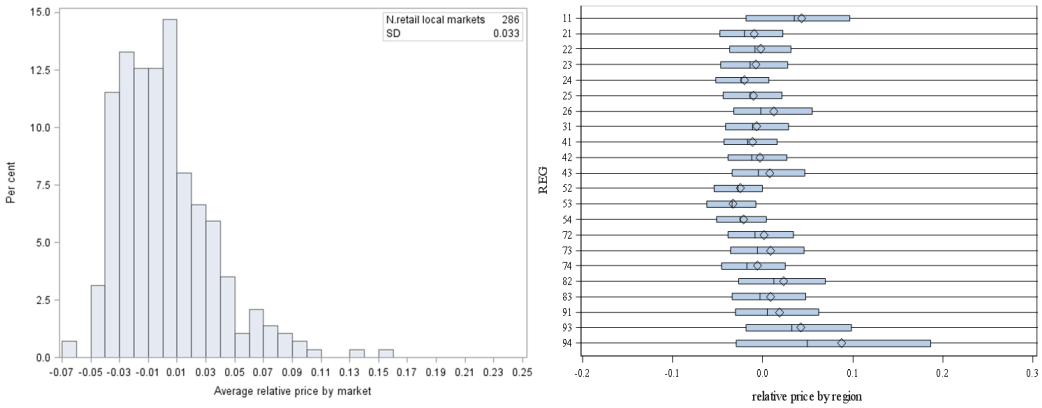


Figure 7: Distribution of relative prices by local market and boxplot of relative prices by region.

Finally, as far as time is concerned, a simple variance analysis of relative prices entailing product, store and time fixed effects suggests that the most important factors relate to stores, followed by factors related to products, while week fixed effects basically do not explain the variance of deviations from the product quarterly average log price.

4 Disentangling the Sources of Price Dispersion

The descriptive analyzes presented above may give us some indications about the potential sources of price dispersion. However, a more structured analysis is needed to properly assess their relative importance. In particular, this section

[15]French regions' code (till 2015): 11 Île-de-France, 21 Champagne-Ardenne, 22 Picardie, 23 Haute-Normandie, 24 Centre-Val de Loire, 25 Basse-Normandie, 26 Bourgogne, 31 Nord-Pas-de-Calais, 41 Lorraine, 42 Alsace, 43 Franche-Comté, 52 Pays de la Loire, 53 Bretagne, 54 Poitou-Charentes, 72 Aquitaine, 73 Midi-Pyrénées, 74 Limousin, 82 Rhône-Alpes, 83 Auvergne, 91 Languedoc-Roussillon, 93 Provence-Alpes-Côte d'Azur, 94 Corse.

provides an answer to a very debated issue in the theoretical literature on price dispersion: is price dispersion rather spatial (i.e., some shop persistently sell at lower price) or temporal (i.e., each supermarket varies its price over time, so that consumers cannot learn by experience which shops provide the best price)? In order to disentangle the two dimensions of price dispersion, we estimate, for each product separately, a fixed effect model including supermarket fixed effects, as well as fixed effects for the combinations of weeks and regional branches. The former aim at capturing all persistent characteristics of supermarkets determining price setting, while the latter account for temporary discounts. Indeed, in France bargaining with producers is typically dealt with by retail groups, while sales and promotions are chain- or regional branch-specific, and are thus implemented and advertised at the national or regional level. We then estimate the following model, product by product

$$p_{ist}^{rel(iq)} = \alpha_{is} + \alpha_{iwt} + \varepsilon_{iswt}\,, \tag{1}$$

where $p_{ist}^{rel(iq)}$ is the percentage deviation from the product quarterly mean log price, α_{is} are supermarket s fixed effects, α_{iwt} are combinations of week t and regional branch w fixed effects, and ε_{iswt} are the error terms.

Figure 8 shows the distribution of the overall variance explained by the above model by product. It suggests that store and temporary discount fixed effects suffice to explain on average 90% of the observed dispersion of prices across stores and time. The model explains 42% of price dispersion in the worst case and up to more than 98% in the case of the product with the best fit.

Moreover, Figure 9 shows that the estimated variance of supermarket fixed effects α_{is} represents by far the most important component of the total variance explained by the model. On average the correlation between store fixed effects and relative prices is 0.84, ranging from 0.39 to more than 0.98. As far as temporary discount fixed effects are concerned on average the correlation with the observed relative prices is 0.26. The latter shows that temporary discounts happen on average more frequently for products sold at a relative expensive price. However, these correlations range from -0.38 to 0.68, suggesting that discounts may nevertheless also concern products sold at a relatively low price.

Gorodnichenko et al. [2014] similarly find that price dispersion appears to be

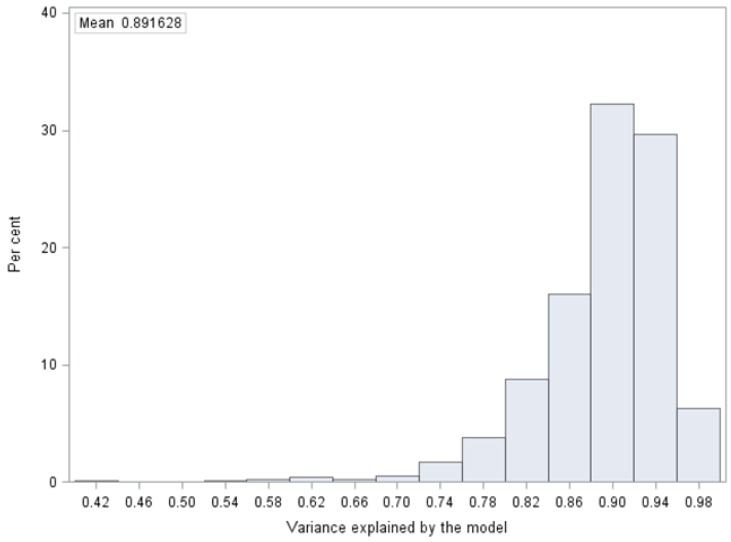


Figure 8: Distribution of the overall variance explained by the model by product.

best characterized as spatial rather than temporal in the US and the UK, even in online markets. Once reached the conclusion that in France price dispersion is mainly spatial, it is possible to go one step further and estimate a second stage that investigates its determinants. In particular, it's interesting to understand the relative role played by retail chains at the national level and factors at the local level. Among the latter 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.[16] 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 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.[17]

[16] Urban density categories are based on a combination of population density and absolute population of INSEE 'canton-ou-ville' (see Fig. 12 Appendix for details).

[17] 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 driv-

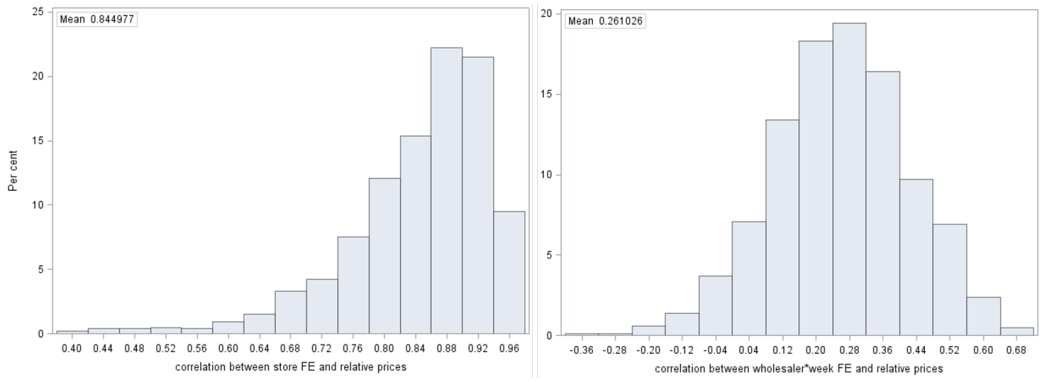


Figure 9: Distribution of correlations between store fixed effects and relative prices by product.

We thus estimate the following models:

- additive fixed effects model

$$\alpha_{is} = \delta_i + \delta_r + \gamma_1 * \text{urban} + \gamma_2 * \text{income} + \gamma_3 * \text{pop} + \\ + \gamma_4 * \text{sameprod} + \gamma_5 * \text{closest} + \upsilon_{is}\,, \quad (2)$$

- multiplicative fixed effects model

$$\alpha_{is} = \delta_i * \delta_r * + \gamma_1 * \text{urban} + \gamma_2 * \text{income} + \gamma_3 * \text{pop} + \\ + \gamma_4 * \text{sameprod} + \gamma_5 * \text{closest} + \upsilon_{is}\,, \quad (3)$$

where α_{is} are the product*supermarket fixed effects estimated in the first stage, δ_i are product fixed effects, δ_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.

ing 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.

Table 5 reports the estimated coefficients of the regressions 2 and 3. The estimation results show that the multiplicative fixed effect model does a significantly better job in fitting the data than the additive fixed effect model. The R squared is almost doubled when considering multiplicative fixed effects. The comparison of the regressors' contributions to the variance explained by the model reveals that retail chains are by far the most relevant factor. Conditionally on all other regressors, retail chains account for 76% of the variance explained by the model in the additive fixed effect model. The second most important factor is represented by product fixed effects, which account for 8% of the variance explained by the model. In the multiplicative effect model retail chain and product fixed effects represent 97% of the explained variance, suggesting that not all products exhibit the same price dispersion across retail chains. Not surprisingly, retail chain fixed effects show that prices in large supermarkets (denoted by the inclusion in the anonymized retail chain name in Table 5 of an 'H' for 'Hypermarche' in French) tend to be lower than those in medium-size supermarkets (denoted by a 'S').[18] Regarding the impact of local factors on prices, it is worth being first noticed that the estimates in the additive and multiplicative 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 a 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 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 retail group (instead of chain) dummies in model 2, the R squared decreases only to 0.30 and,

[18]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 5: National and local determinants of spatial price dispersion

	additive FE		multiplicative FE	
Regressor	estimate	SE	estimate	SE
retail chain 1 H	-0.02	0.0003***		
retail chain 1 S	0.04	0.0004***		
retail chain 2 H	-0.04	0.0002***		
retail chain 2 S	0.00	0.0002***		
retail chain 3 H	0.07	0.0002***		
retail chain 3 S	0.10	0.0004***		
retail chain 4 H	-0.03	0.0003***		
retail chain 4 S a	0.09	0.0008***		
retail chain 4 S b	-0.02	0.0002***		
retail chain 4 S c	0.05	0.0006***		
retail chain 5 H	-0.06	0.0002***		
retail chain 6 H	-0.01	0.0002***		
retail chain 6 S (ref)	.	.		
rural	-0.05	0.0005***	-0.05	0.0004***
semi-urban	-0.05	0.0004***	-0.05	0.0003***
urban	-0.04	0.0004***	-0.04	0.0003***
metropolitan (ref)	.	.		
log local per capita income	0.01480	0.0002***	0.01461	0.0002***
log local population	0.00217	0.0000***	0.00222	0.0000***
n.stores selling product	-0.00002	0.0000***	-0.00002	0.0000***
closest large supermkt	0.00030	0.0000***	0.00030	0.0000***
product FE	yes		no	
product*retail chain FE	no		yes	
R squared	0.34		0.58	

Note: *** means significant at 1%.

conditionally on all other regressors, retail groups still account for 72% 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 regional branch/wholesaler dummies in model 2, the R squared increases only to 0.37 and, conditionally on all other regressors, regional branches account for 77% of the variance explained by the model. Therefore, there seems to be an additional stage of price setting happening at the regional level, but national price setting strategies are dominant.

5 Conclusion

Based on a large and original data set containing almost 40 millions of weekly price records from more than 1500 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. First, retail groups bargain with producers and the market share of the corresponding retail groups is likely to 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) or at the regional branch level. 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 local conditions regarding demand or local competition between supermarkets also matter for explaining observed prices on local markets, though to a much lower extent.

References

Alvarez, F., M. Golzález-Rodaza, A. Neumeyer, and M. Beraja (2013). From hyperinflation to stable prices: Argentina's evidence on menu cost models. Working papers, Stanford Center for International Development.

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.

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.

Cheung, Y.-W. and E. Fujii (2008). Deviations from the Law of One Price in Japan. CESifo Working Paper Series 2275, CESifo Group Munich.

Dauvers, O. (2013a, January). Drive insights.

Dauvers, O. (2013b, December). Drive insights.

Dixon, H. D., K. B. Luintel, and K. Tian (2014, June). The impact of the 2008 crisis on UK prices: what we can learn from the CPI microdata. Cardiff Economics Working Papers E2014/7, Cardiff University, Cardiff Business School, Economics Section.

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.

Eyal, B. and B. Eden (2004, July). Price Rigidity and Price Dispersion: Evidence from Micro Data. *Review of Economic Dynamics 7*(3), 613–641.

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 (2014, December). Price Setting in Online Markets: Does IT Click? NBER Working Papers 20819, National Bureau of Economic Research, Inc.

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.

Moen, E., F. Wulfsberg, and O. Aas (2014). Dispersion in retail prices. Working paper.

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

Nakamura, E., J. Steinsson, P. Sun, and D. Villar (2016, August). The Elusive Costs of Inflation: Price Dispersion during the U.S. Great Inflation. NBER Working Papers 22505, National Bureau of Economic Research, Inc.

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.

Appendix

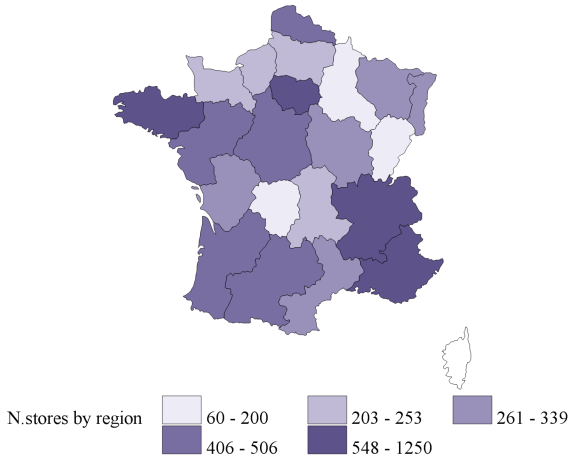


Figure 10: Geographical distribution by region of stores.
Darker colored regions correspond to larger number of stores.

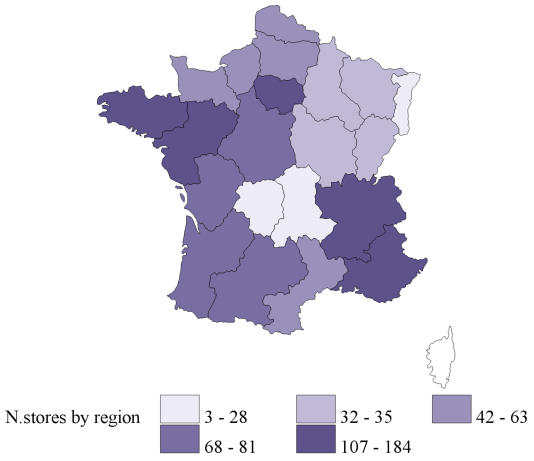


Figure 11: Geographical distribution by department of stores for which price data are available.
Darker colored regions correspond to larger number of stores.

	Population in 'canton-ou-ville'			
Pop. density	0 to 10,000	10,000 to 50,000	50,000 to 300,000	300,000 +
0 to 200	**0**	**1**	**2**	**3**
200 to 500	**1**			
500 to 3,000		**2**		
3,000 to 15,000				
15,000 +			**3**	

Figure 12: Criteria defining urban density.