Universidad de San Andrés
Departamento de Economía
Maestría en Economía

Price dispersion in Argentina

Fernanda Analía PERROTTA
34.583.792

Mentor: Facundo ALBORNOZ

Buenos Aires,
Argentina
30 de agosto, 2019

Tesis de Maestría en Economía de

Fernanda Analía PERROTTA

Price dispersion in Argentina

Abstract

We investigate how the structure of price dispersion is in Argentina. We use daily data reported by supermarkets and we measure price dispersion at the bar code level in order to analyze whether supermarkets set prices differently, whether those price differences are spatial or temporal, and whether a large devaluation shock affects pricing behavior. We use more than 9 million prices and we calculate the dispersion of weekly prices relative to the average monthly price per product. We found that on average across products the 90th percentile of relative prices is 10 percentage points higher than the 10th and the mean absolute deviation from monthly average product prices is 2%. Price dispersion across stores in Argentina is heterogeneous among retail chains but local conditions regarding demand or competition contribute significantly but they have small impact. Furthermore, we tested that supermarkets pricing behavior remains within its essence after a large devaluation shock.

Keywords: *Price, Dispersion, Argentina, Retail*

Códigos JEL: *E3, E31, L8*

Introduction

We investigate whether the theory of the law of one price holds in Argentina's supermarkets. This law states that the price of an identical good should be the same regardless of the location. If there were price differences in different locations, they should eventually be eliminated due to arbitrage opportunities. There has been extensive research on whether the law of one price holds and, in particular, Isard (1997) have already showed that in reality the law of one price is flagrantly and systematically violated by empirical data. In this paper, we analyze price dispersion in Argentina using daily data from supermarkets with the goal of finding out its characteristics and to compare our results with those obtained in a similar analysis made by Berardi et. al (2017) for the French market. Moreover, we tested whether there are changes in the pricing strategy of a supermarket after a large devaluation shock.

In the last five years, researchers have measured price dispersion with different indicators and using different databases that have different nature. Some papers, like Gorodnichenko and Talavera (2017), focus on online shopping platforms due to the fact that daily scrapping of those websites is a low barrier to create a database. Other papers like Kaplan and Menzio (2015) use Nielsen data based on a panel of consumers that scan all their purchases. Other authors take a sample of supermarket prices like Dubois and Perrone (2015) whom have constructed a price dispersion measure for eight products in a three-year period in France. And recently Berardi et. al. (2017) have gone further characterizing price dispersion in France with a high volume of supermarket price data. They have analyzed the characteristics of the distribution of a measure of price dispersion on different regions and type of supermarkets in France. In addition, they have shown that price dispersion in France has a similar distribution to other measures made in US and UK. All in all, these results have all shown that price dispersions exist to some extent and in particular, Berardi has shown that prices in France are spatially driven and the price dispersion coefficient is lower than those measured by other authors in the U.K. and U.S.

In this paper we focus first on providing a price dispersion characterization looking at its shape and structure in order to compare Argentina's results with those obtained in Berardi et.al. (2017). We check whether the price dispersion in Argentina is similar to others even if its economy is completely different from that of France, U.S. and UK. Second, we decompose price dispersion into several components. We assess the relative contribution of retailer component, temporary discounts by branch and local factors. Third, we show that price dispersion strongly depends on retail chains, while local markets' characteristics play a significant role but to a much lower extent. Lastly, we compute a ranking of retailers during 2018 before and after a large devaluation of the national currency (peso) in Argentina and we perform a test for statistical significance to see if retailers adjust quickly to large devaluation shocks.

This work is organized as follows: in section one we provide details about the databases, section two analyses the main descriptive statistics of price dispersion; section three presents the results of our econometric estimation; section four we test for differences in the rank of a store before and after a devaluation shock and finally, section five we close with some conclusion remarks.

Section 1: Data description

In Argentina retailers must provide every day information of the price of all products that are available for sale at every store to the government since May 2016. This regulation became effective with the goal of tracking daily prices in order to provide transparency to the public and with the final goal of lowering inflation. Supermarkets send more than 14 million prices per day from more than 2.000 different stores from across the country. The volume of the database has more than 7 billion prices at the moment and more than 30 terabytes of data. Therefore, it was a huge challenge to master all these data for this work. First, we decided to focus on 2018 because all supermarkets had already joined the program at that time and the number of daily prices was stable. Second, we chose the following time periods: January to April 2018 and July to September 2018 and we left out May to June because Argentina had a large devaluation during that period (of the order of 36%) which we use to do a before and after devaluation test. Third, we decided to keep only prices for the City of Buenos Aires because it represented 60% of the entire database and we have more than 2.000 different stores to compare our results. Fourth, we kept the four main retailers at the City of Buenos Aires because they have more than 25.000 products that can be compared among them and they have enough weekly frequency to be analyzed in the chosen timeframe. Lastly, we decided to reduce the dimensionality of the data from daily to weekly in order to have less data to manipulate.

We define store as the combination of retail chain, branch, and store location. A retail chain may own different branches and each one could have a different pricing strategy. In addition, store location refers to a single address where the retail-branch has a physical store. In addition, we define a product with its bar code which is the SKU classification and it is harmonized among all supermarkets in the database. Therefore, a single product has the same bar code among all stores. Then, we first calculated the average price per week per product at a given store in order to define a set of frequent products. A frequent product is the one that it is present every week at a given store throughout all the time frame analyzed for this work. For example, if a product is at one store every week from January 2018 to April 2018, then it will be kept in the database. By keeping only frequent products per store, the database was reduced from 60 million prices to approximately 10 million from 23.000 products (individually identified by its bar code) and 730 supermarkets over the period January to April 2018.

To estimate price differences, we first created a second database with the geolocation of each store to a latitude-longitude point. Thanks to the geo-localization of stores, we were able to assign each store to a neighborhood of the City of Buenos Aires. The city of Buenos Aires is organized in 15 large communes (comunas in Spanish) which are made of 48 different neighborhoods. Some communes have only one neighborhood in it and others have more than one. Second, we enriched this database by including an estimate of the private sector employment and average salary per commune from administrative labor data. Lastly, we added some local variables like the selling price per square meter per neighborhood, population, density and the distance of each store to the nearest one.

Section 2: Descriptive analysis of price dispersion in Argentina

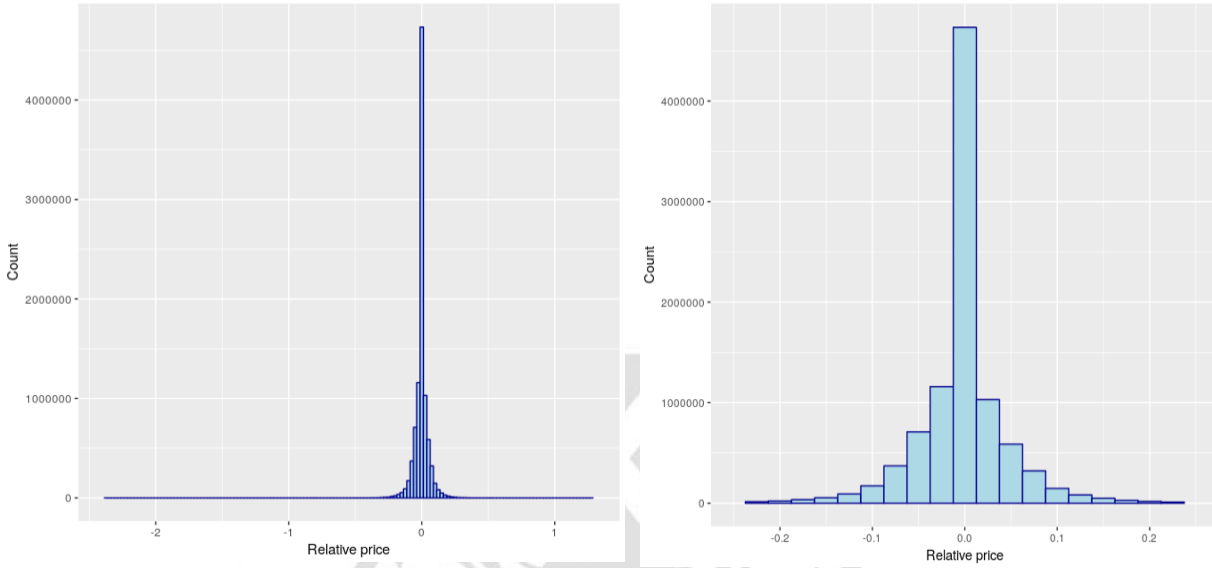
We follow a definition similar to Berardi's for price dispersion. Since Argentina's economy is more volatile than France's, we reduce the analysis of dispersion from quarters to months. Therefore, we define price dispersion as the difference of a product's price (defined by its bar code) at a given store with the average price of that product in a month in all supermarkets.

First, we calculate the average price of a product in a given week for each supermarket, defined as $p_{i,s,t}$ where i represents the product at the bar code level, s represents a store (as a combination of retail, branch and location), and t represents a week. Second, we calculate the average price of a product in a given month across all supermarkets ($p_{i,m}$) where m represents the month. Lastly, we define price dispersion for a product as a relative price, built as the difference between the log price of product i at a given store s in a given week t and the average price of that product in that month.

$$p_{i,s,t}^{rel(i,m)} = \ln\left(p_{i,s,t}\right) - \ln\left(p_{i,m}\right)$$

Price dispersion exists when the same product is offered at different prices in different stores at a given period of time. A positive relative price means that a given product is being sold higher than the average price during that month.

Graph 1: Distribution of relative prices (left) and zoom image (right)



We calculate statistical indicators in order to compare our results with Berardi et. al. (2017) and verify if our distributions are similar. In our dataset, the mean absolute price deviation of all products is 2% (see Table 1) meaning that on average prices are above the average price of a product in a given month. Moreover, the relative prices distribution is long-tailed, left-skewed, and spiked around the mean.

Table 1: Descriptive Statistics

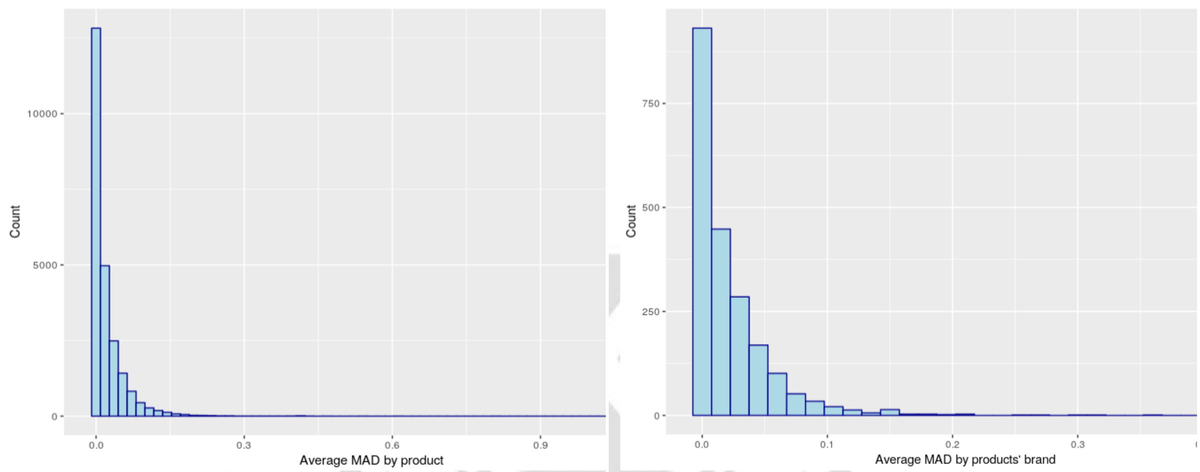
	France	Argentina
Standard deviation of relative prices	0.07	0.03
Interquartile range of relative prices	0.08	0.04
Interdecile range of relative prices	0.17	0.10
Mean of absolute deviation	0.05	0.02

Table 1 provides the calculated statistics: standard deviation, interquartile range, interdecile range, and mean of absolute relative prices computed by product (per bar code) and then averaged over all observations.

We find that on average across products the 90th percentile of relative prices is 10% higher than the 10th. The mean absolute deviation with monthly average product log prices is 2% on average in the Argentine retail sector and the standard deviation of relative prices is 3%. In addition, these estimators imply that price dispersion in our data is smaller than that obtained by Berardi N., Sevestre P., Thébault J. (2017) and Dubois and Perrone (2015). Furthermore, on average across products, the 95th percentile of observed prices is 16% higher than the 5th

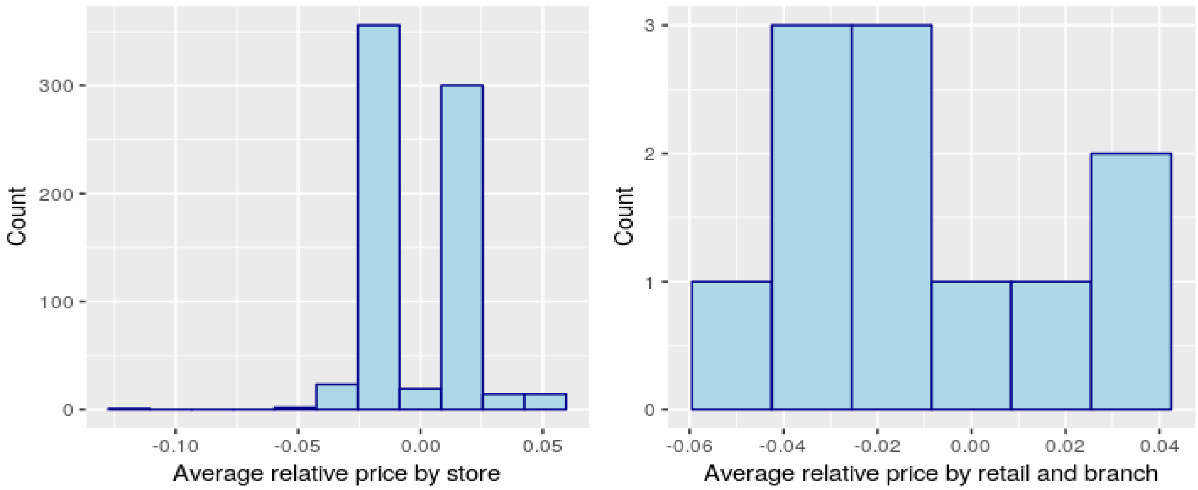
percentile (versus 37% computed by Dubois and Perrone (2015) and 26% in Berardi). The interquartile ratio is 1.03 in our data, versus 1.14 in Dubois' and 1.09 in Berardi's.
Given the main statistics, we keep on analyzing price dispersion using the different features from the data to understand which are the main variation factors related to the product, store and time. First, we characterize a point of sale of a supermarket by its branch (denoted as b for branches), retail chain (denoted by r), and commune of the city of Buenos Aires (denoted by c). The distribution of the mean absolute deviation of relative prices by product is represented in Graph 2.

Graph 2: Distribution of MAD by products (left) and by products' brand (right)



In graph 2, we see that there is a large variation by product up to 90% and that the brand of the product only account at maximum to 40% of the variation. In both cases, the distribution of the mean absolute deviation is asymmetric with a long right tail suggesting that the products of some brands are characterized by deviations from the product's monthly average log price of an order of more than 90%.

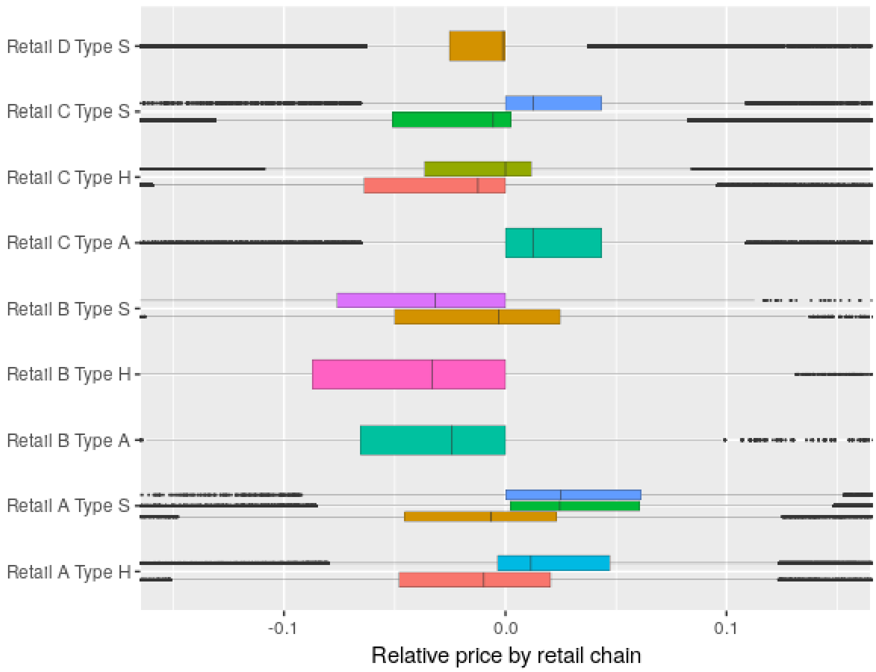
Graph 3: Distribution of average relative prices by stores (left) and retail and branch (right)



Regarding the store dimension, the left panel of Fig. 5 shows that average relative price dispersion is quite heterogeneous across supermarkets. Some stores are clearly cheaper than others and that price dispersion by stores accounts up to 5%. The right panel of Figure 5 suggests that the branch affiliation of supermarkets is well distributed on our data sample but it still accounts for a variation of price dispersion up to 4%.

We go further and we look at price dispersion at retail, branch and type of supermarket. In Argentina the type of supermarket is defined by the amount of cashier machines available and the square meters of the store. A single store could be a hypermarket (defined by H) or a supermarket (defined by S) or a self-service store (defined by A). In Graph 4 each line represents the boxplot of relative prices for the city of Buenos Aires in Argentina.

Graph 4: Boxplot of relative prices by retail chains

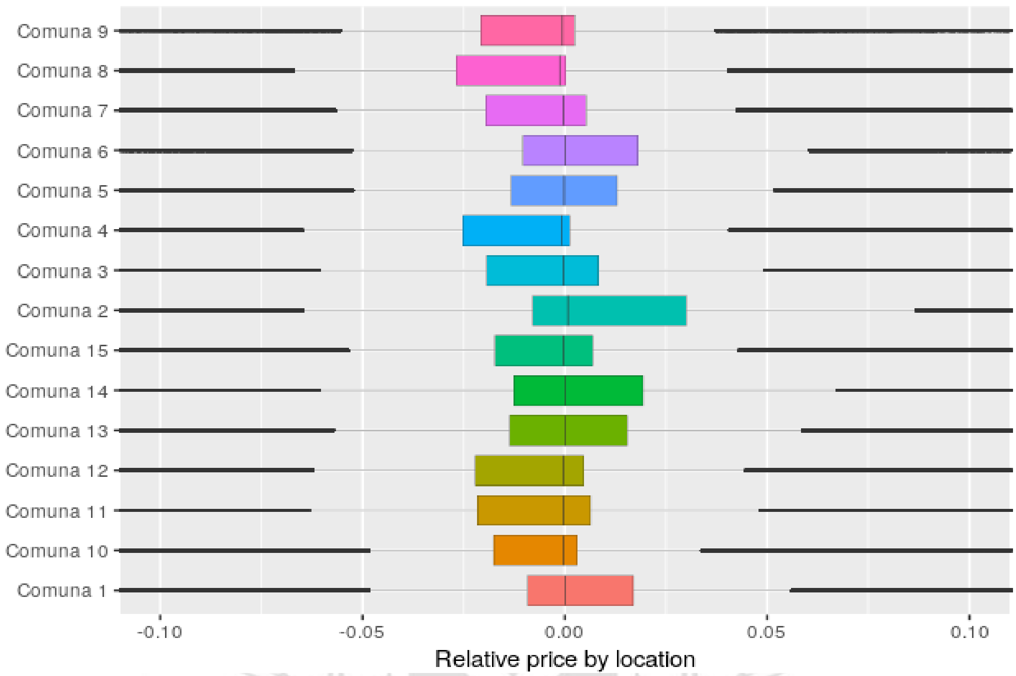


We observe that the average relative prices, as well as their dispersion, largely vary across retail chains. Some retailers have a pricing strategy below the average relative price in all its branches. In addition, some retailers are cheaper or more expensive than others in general and that their dispersion start at the zero mean marking a strategy to be clearly cheaper or more expensive than the average as it is in retail C type A.

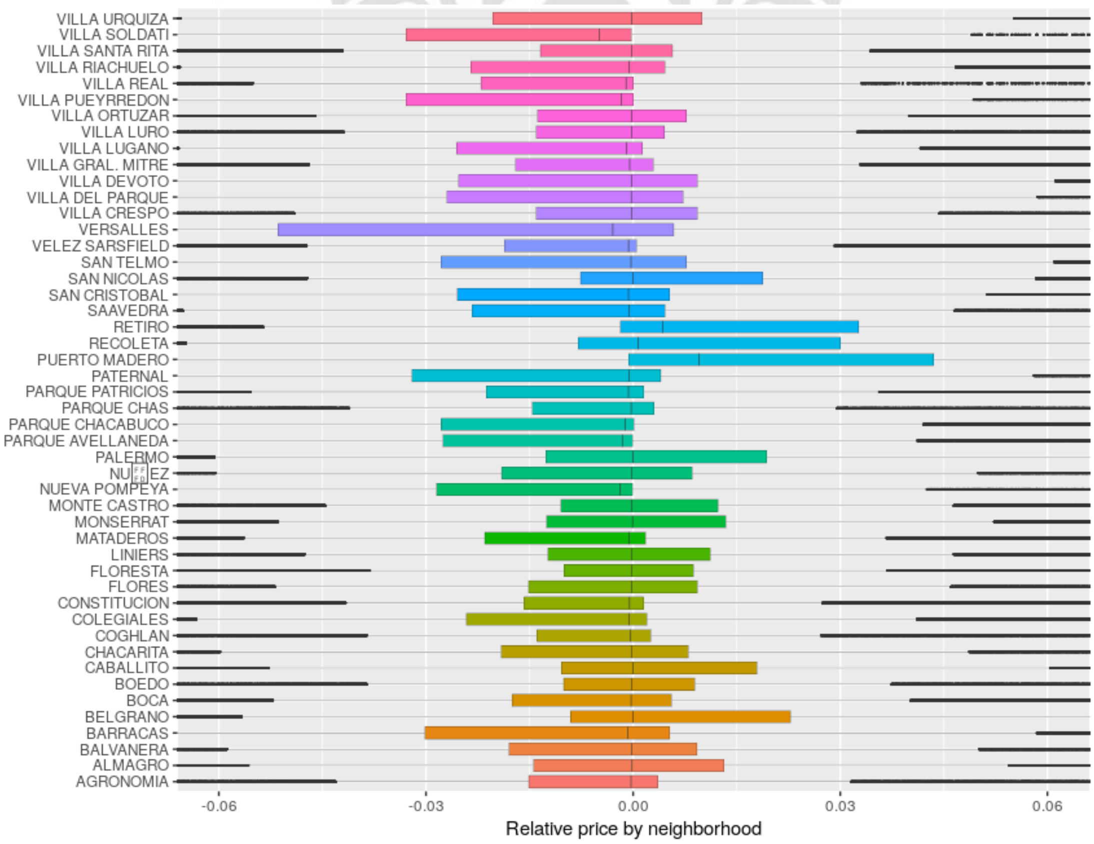
Now, we continue our analysis and we dig even deeper and look at the spatial difference of these dispersion observed by retailers' boxplot and we expand our analysis into the communes of the city of Buenos Aires and then within neighborhoods. We observe in Graph 5 that the dispersion of relative prices varies across communes just by looking at the length of the boxplot, but the average relative price is similar all around.

When we zoom into communes and we look at the differences by neighborhoods, we definitely observe in Graph 6 much larger dispersion. In particular, we see that some local markets are characterized by a rather large dispersion like for instance in Retiro, Puerto Madero and Belgrano (which in fact are the wealthiest neighborhoods in Buenos Aires). After looking at the data and its dispersion given different features, we continue with a fixed effect regression approach in order to test the statistical significance of our intuitions.

Graph 5: Boxplot of relative prices by commune of the City of Buenos Aires



Graph 6: Boxplot of relative prices by neighborhood of the City of Buenos Aires



Section 3: Components of price dispersion

In order to distinguish spatial and temporal effects of price dispersion, we estimate a fixed effects model where we control for the effect of each store and for the effects of the combination of branches and months. On one hand, we attribute the spatial effect to the fact that a single store could set prices differently given the specific location and its intrinsic characteristics (nicer store, better located, and so on). On the other hand, we attribute the temporal effect to the discounts or sales over time. In Argentina discounts are not made at the retail level but at the branch level. In other words, discounts will be captured by a fixed effect per branch per month. Lastly, we take a random sample of 60 products to conduct our regressions which we will held constant throughout all the analysis.

A retail chain can have different branches under which it sells its products. For example, a retailer c sells under three different branches: branch a, b and c. Each branch has many stores which we define as location point of sale. As we mentioned before the combination of retail, branch and location point of sell is unique. Therefore, we estimate the following:

$$p_{i,s,t}^{rel(i,m)} = \alpha_{i,s} + \alpha_{i,b,t} + \varepsilon_{i,s,b,t}$$

Where $p_{i,s,t}^{rel(i,m)}$ is the percentage deviation from the product monthly mean log $p_{i,m}$. The $\alpha_{i,s}$ represents the supermarket (s) fixed effects for product (i) and $\alpha_{i,b,t}$ are combinations of week (t) and branch (b) fixed effects, and $\varepsilon_{i,s,b,t}$ is the error term.

We estimate the first model in order to analyze how these two effects relate to relative prices and we find that for the store ($\alpha_{i,s}$) fixed effect component 97% of the stores estimates are statistically significant and the estimates for the discount effect ($\alpha_{i,b,t}$) only 27% are significant. In addition, the average correlation between store fixed effects and relative prices is 0.63 and the average correlation between branch and week fixed effects and relative prices is -0.21. Therefore, we conclude that store fixed effects are predominant and its correlation does affects largely relative prices in comparison to the discount effects. Similar to Berardi et. al. (2017) results, it seems that in Argentina relative price dispersion is driven by the store fixed effects (spatial) and not to discounts throughout time (temporal).

We continue with a second estimation to analyze the determinants of the store fixed effects. In order to do this, we use the alpha coefficients ($\alpha_{i,s}$) from the first regression and we compute a second regression to disaggregate the sore component. We aim to capture the effect of the retail chain which the store belongs to and we control for differences in communes such as density, employment, income, price of square meter and the distance to the nearest store. Hence, we define our second estimation as follows:

$$\alpha_{i,s} = \delta_i + \delta_r + \gamma_1\, density + \gamma_2\, income + \gamma_3\, employment + \gamma_4\, p_{houses} + \gamma_5\, pop + \gamma_5\, distance + \varepsilon_{i,s}$$

where we define density as population density measuring persons per one square meter per commune. Then we incorporate an income variable in order to capture a measurement of local demand defined by the log of income of private employment and we add the log of private employment as well. Then, we try to capture the amount of competition in the local market by the distance (in meters) to the closest store for each observation. In addition, we aim to capture the local market volume with the variable population defined as the log of population. Lastly, we add some characteristics to capture the expensiveness of the neighborhood like the price of selling houses measured by the selling price per square meter.

Table 2: Determinants of price dispersion

	Regression Results
	Estimates
Constant	0.133***
	(0.007)
Retail A Branch 1 S	0.008***
	(0.001)
Retail A Branch 2 S	0.038***
	(0.001)
Retail A Branch 3 H	0.034***
	(0.001)
Retail A Branch 3 S	0.035***
	(0.001)
Retail C Branch 1 H	-0.044***
	(0.001)
Retail C Branch 2 H	-0.017***
	(0.001)
Retail C Branch 2 S	-0.038***
	(0.001)
Retail C Branch 3 A	0.024***
	(0.001)
Retail C Branch 3 S	0.023***
	(0.001)
Retail D Branch 1 S	-0.026***
	(0.001)
Density	-0.00000***
	(0.000)
Income	-0.006***
	(0.001)
Employment	0.0003***
	(0.00005)
Price of Houses	0.00000***
	(0.00000)
Population	-0.003***
	(0.0003)
Distince	-0.00000***
	(0.00000)
Observations	40,920
R^2	0.959
Adjusted R^2	0.959
Residual Std. Error	0.005 (df = 40844)
F Statistic	12,674.000*** (df = 75; 40844)
Note:	*p<0.1; **p<0.05; ***p<0.01

Looking at the results, we see that our model explains 96% of the variance of the model and that the combination of retail, branch and the type of store is the most explanatory source of price dispersion. In addition, these effects vary within the same retail chain depending on the type of store. The estimates for a retailer varies on the type of store. For example, retailer A in branch 3 has a lower estimate for hypermarket and supermarket. We observed that in general supermarkets are more expensive than hypermarkets, but self-services are more expensive than supermarkets.

Regarding local factors, on the one hand, all are significant but their effects are significantly smaller. Yet, urban density, income, population and distance decrease price levels (marginally because of their low effect). On the other hand, employment and the price of selling houses increase relative prices. It makes sense to test that more expensive neighborhoods do have larger relative prices as we saw in Graph 5 and 6.
Furthermore, the results show that none of the products are statistically significant meaning that the differences in relative prices are not set at the bar code level. From the consumer point of view, as similar to what happens in France as shown by Berardi et.al. (2017), consumers can infer if a store is more expensive than the other just by looking at the retail and branch combination before going to shop.

Section 4: Strategy throughout time

Since we estimated that the combination of retail, branch and the type of store is the most explanatory source of price dispersion, we decided to rank stores throughout two periods of time: before and after a large devaluation of the peso. We perform this test because as shown by Alvarez et. al (2013) agents do respond differently when there are large shocks to an economy under large and small inflationary prices. In our case, we want to check whether stores stay or not in the same ranking given its pricing when a large devaluation shock takes place in Argentina (a country with a history of high inflation). In order to do this, we ranked stored by the average relative price during the period of January to April 2018 and we repeat this process for July to September 2018. Then we performed a paired Wilcoxon test for differences in those rankings and we find that there is no statistical difference in the rankings after a 36% devaluation of the peso that took place between May and June 2018. Therefore, pricing strategies remain even under large external shocks.

Section 5: Conclusion

We showed that the structure of price dispersion in Argentina has a spike around a zero mean and a long left tail. We find that on average across products the 90th percentile of relative prices is 10 percentage points higher than the 10th and that the mean absolute deviation from monthly average product prices is 2%. In addition, throughout the communes and the neighborhoods of the city of Buenos Aires, we find that Recoleta, Retiro, Puerto Madero and Belgrano present some of the largest dispersion of relative prices (See graph 6). Furthermore, we estimated whether price dispersion is spatial or temporal and we found that the spatial effects are more predominant among stores.
Given those results, we performed a fixed effect estimation where we controlled for the local level factors and we found that price dispersion across stores in Argentina results from persistent heterogeneity in retail chains' pricing. Meanwhile local conditions regarding demand or competition contribute to a much lower extent. In addition, we tested whether a large devaluation shocks affects pricing behavior and found that retailers adjust prices quickly enough to maintain its essence. All in all, we consider that our results are similar to those obtained by Berardi et. al (2017) in France and we consider our findings as powerful since our analysis is performed in a different time frame, country, economy and different retailers.
To wrap up, we have analyzed how price dispersion is in Argentina and we compared our results with international estimations. Having characterized price dispersion at such a granular level in a country with high inflation like Argentina, it is an important finding which provides a better understanding of the shape and determinants of pricing behavior. Since we looked at distribution among retailers, we consider that future research should focus on understanding

how those prices vary within a retailer in order to disentangle even deeper price dispersion. Future research should be devoted to understand how prices change within a firm considering spatial dimension, product´s brands, and to understand how the behavior changes for products that are in a price stabilization program or whether they belong to the basic needs basket. In this direction, changes in distribution and magnitudes of variation of prices could shed light on greater understanding.

References

Berardi N., Sevestre P., Thébault J. (2017) “The Determinants of Consumer Price Dispersion: Evidence from French Supermarkets.” In: Matyas L. (eds) The Econometrics of Multi-dimensional Panels. Advanced Studies in Theoretical and Applied Econometrics, vol 50. Springer, Cham.

Dubois, Pierre and Perrone, Helena (October 2015). “Price Dispersion and Informational Frictions: Evidence from Supermarket Purchases”. CEPR Discussion Paper No. DP10906. Available at SSRN: https://ssrn.com/abstract=2682593

Fernando Alvarez, Martin Beraja, Martín Gonzalez-Rozada, Pablo Andrés Neumeyer (February 2019) “From Hyperinflation to Stable Prices: Argentina’s Evidence on Menu Cost Models”. The Quarterly Journal of Economics, Volume 134, Issue 1, Pages 451–505, https://doi.org/10.1093/qje/qjy022

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.

Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary StatisticsTables. R package version 5.2.2. https://CRAN.R-project.org/package=stargazer

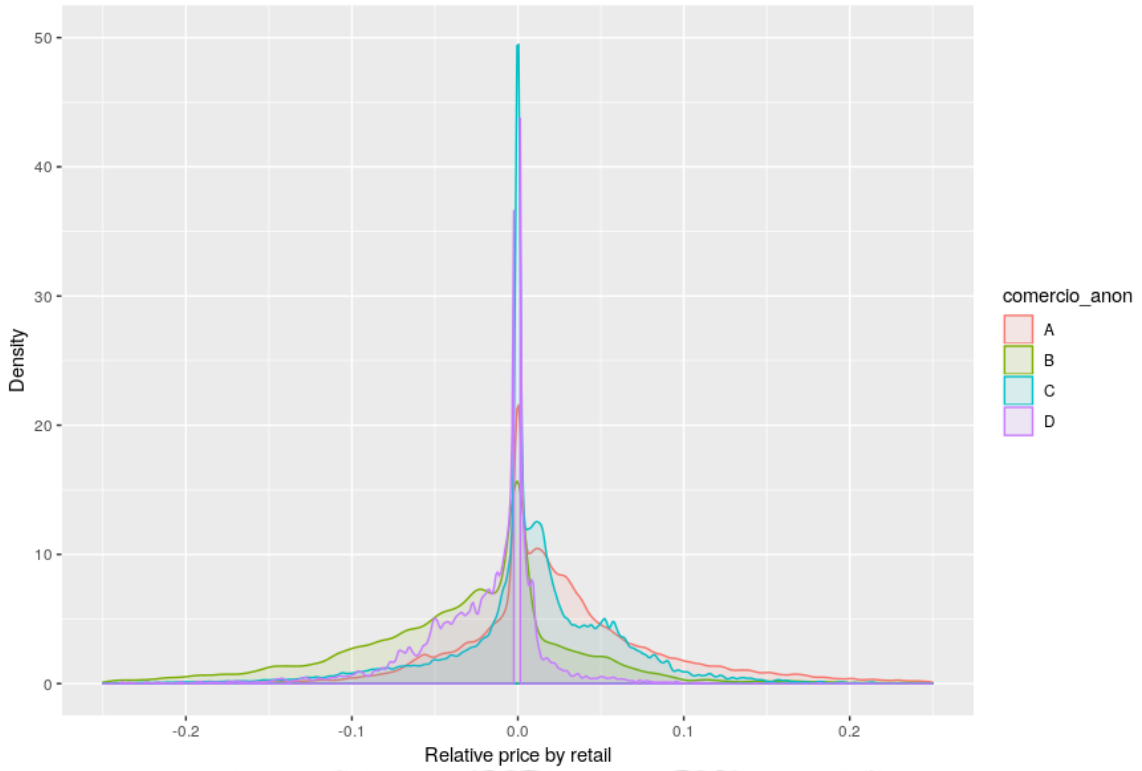
Isard, Peter. (1977). "How Far Can We Push the "Law of One Price"?," American Economic Review, American Economic Association, vol. 67(5), pages 942-948.

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

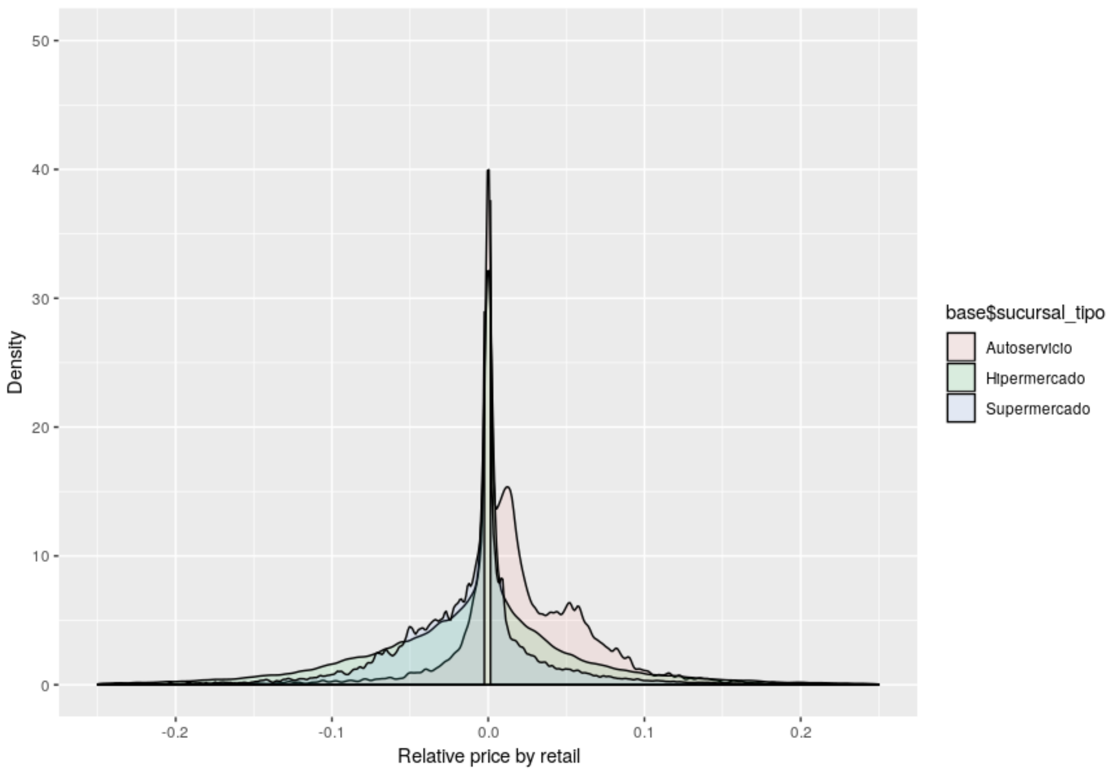
Wilcoxon, Frank (Dec 1945) "Individual comparisons by ranking methods". Biometrics Bulletin. 1 (6): 80–83

Appendix

Density of relative price by retail



Density of retail price by type of store



Density of relative price by retailer and type of store