The Scandinavian
Journal of Economics

Scand. J. of Economics 000(0), 1–26, 2019
DOI: 10.1111/sjoe.12356

Price Dispersion and the Role of Stores*

Espen R. Moen

BI Norwegian Business School, NO-0442 Oslo, Norway
espen.r.moen@bi.no

Fredrik Wulfsberg

Oslo Business School, NO-0130 Oslo, Norway
frewul@oslomet.no

Øyvind Aas

Université libre de Bruxelles, B-1050 Bruxelles, Belgium
oyvind.nilsen.aas@gmail.com

Abstract

In this paper, we study price dispersion in the Norwegian retail market for 766 products across 4,297 stores over 60 months. Price dispersion for homogeneous products is significant and persistent, with a coefficient of variation of 37 percent for the median product. Price dispersion differs between product categories and over time. Store heterogeneity accounts for 30 percent of the observed variation in prices for the median product–month, and for around 50 percent for the sample as a whole. Price dispersion is still prevalent after correcting for store heterogeneity.

Keywords: Price dispersion; retail prices; store heterogeneity

JEL classification: *D*2; *D*4; *E*3

I. Introduction

It is common knowledge that the price for a particular product or service can vary substantially between stores or outlets. One explanation for price dispersion is that stores are different. Stores can be heterogeneous in a multitude of ways, such as location, opening hours, parking facilities (e.g., Dixit and Stiglitz, 1977; Weitzman, 1994), loyalty programs (Basso *et al.*, 2009), and warranties (Grossman, 1981). In addition, idiosyncratic shocks

*We are grateful for comments from Birthe Larsen, Tore Nilssen, Magnus Söderberg, the referees, and seminar participants at the 2014 Annual Meeting of the Norwegian Association of Economists, the 8th Nordic Summer Symposium in Macroeconomics, the 9th Nordic Workshop in Industrial Organization at the University of Oslo, the 29th EEA-ESEM meeting at Toulouse School of Economics, Norwegian School of Economics, BI Norwegian Business School, Statistics Norway, and Oslo Business School. Ø. Aas gratefully acknowledges the financial support of the European Research Council under the European Union's Seventh Framework Programme (FP7/2007–2013/ERC grant agreement no. 339950).

or unexpected fluctuations in demand (Mackowiak and Wiederholt, 2009) can also yield price dispersion. Furthermore, store characteristics are often an intrinsic component of a purchase (product differentiation). For example, buying a lukewarm Coca-Cola in a supermarket in the middle of the day is different from buying a cold one from a convenience store or a petrol station in the middle of the night. Also, eating the same meal at two different restaurants might be perceived as very different depending on the characteristics of each restaurant. Thus, store characteristics can reflect different mark-ups and costs, and result in price dispersion.

In this paper, we identify the contribution of store characteristics to price dispersion by exploring monthly price observations from a wider set of products categories than in previous studies. First, we establish six stylized facts of price dispersion, as follows.

1. There is significant and persistent price dispersion in retail prices in Norway, and the median standard deviation is 33 percent of the mean price.

2. The dispersion of prices varies between products and over time as indicated by the range between the first and third quartiles of the standard deviation between 19 and 50 percent. There is less price dispersion for non-durable and durable products than for semi-durable products and services.

3. There is little variation in price dispersion between regions.

4. 84 percent of the overall variation in the standard deviation is between products while 16 percent is due to time variation.

5. The dispersion in prices increased from around 25 percent at the start of the sample to almost 40 percent at the end of the sample.

6. While the pooled distribution of normalized prices is unimodal, product-specific distribution of normalized prices is often bimodal.

Second, we identify a fixed store component of prices by observing prices of multiple products in the same store over time. Using intuitive non-parametric methods, we find that the store component accounts for about 30 percent of the price dispersion for the median product. To identify the store component for the sample as a whole, we also employ a novel parametric method, which shows that the effect of store characteristics (hereafter, the store effect) accounts for 50–60 percent of the dispersion in prices. As further evidence of the importance of store heterogeneity, we find that the ranking of stores within the price distributions is highly persistent over time.

Kaplan and Menzio (2015) use scanner data from 1.4 million grocery products across different geographical areas in the United States. They find that the quarterly average standard deviation in prices is between 19 and 36 percent, depending on the aggregation level,[1] and that store heterogeneity accounts for 10 percent of the dispersion in prices. Exploring a subset of the data and a different method, Kaplan *et al*. (2019) find that 15.5 percent of the variance in prices is due to store heterogeneity.[2] In a similar study using French grocery data from many stores belonging to a few supermarket chains, Berardi *et al*. (2017) find that the average price dispersion is 7 percent. They find that a permanent component dominates price dispersion between stores due to centralized price setting. Lach (2002) studies price dispersion for only four products[3] in Israel. He finds that store characteristics account for between 47 and 90 percent of the variation in prices. Wildenbeest (2011) investigate the price dispersion of a basket of grocery items from four retailers in the United Kingdom. He finds that store heterogeneity explains around 61 percent of variation in prices, and he attributes the rest to search frictions.

We contribute to this body of literature by covering a larger variety of products from multiple stores. We include not only food products, but also products from all 10 COICOP[4] categories such as consumer electronics, cars, petrol, apparel, restaurants, transport, and other services. This allows us to provide more detailed insight into price dispersion than in previous studies. While the above-cited studies report measures of price dispersion for the overall or pooled price distribution, we report price dispersion at the product level and the extent of variation across products. Also, we argue that the store effect might represent information about the price structure in the market.

II. The Role of Stores

Households can choose from a tremendously large set of products of different brands and qualities at different prices from different stores. The consideration of what products to buy and from where is a huge task for

[1]Products are defined by universal product code (UPC).

[2]They also decompose the variation in prices into transitory and persistent parts. The persistent component of the store–product variation in prices (which they label "relative price dispersion") constitutes 30.3 percent of the variation in prices.

[3]These are refrigerators, chicken, flour, and coffee.

[4]COICOP is an acronym for classification of individual consumption according to purpose, which is a nomenclature developed by the United Nations Statistics Division to classify and analyze individual consumption expenditures incurred by households according to their purpose.

consumers, and it is almost impossible in practice to gather and process all available information.

Suppose all variation in prices can be attributed to the store effect. This means that the ratio of the prices of any goods in two stores A and B are equal to the ratio of the average price in the two stores. Thus, knowledge of the ratio of the average in two stores then reveals the ratio of the prices of any good in the two stores, and hence also the ratio of the prices of any baskets of goods between the two stores. In effect, a multi-good search problem simplifies to a single-good search problem. If the store effect is substantial but does not explain all the price variation, then the ratio of the average price between two stores give some but not full information about the price ratios of individual goods between the two stores.

In many cases, it might be relatively easy for consumers to collect information about average price level in a store. Surveys, for instance, give information about average prices. The more that is explained by the store effect, the easier is it for the customers to be informed about relative price levels. Hence, by measuring the store effect, it is found that it is easy to obtain information for customers, which again is a stepping stone for understanding the working of the price system as a whole. Thus, the store effect can be an indication of the information structure in the market, and if the store effect is high, then it makes it more likely that the consumers are well informed and that competition works well.

There are several mechanisms that can explain why the store effect exists. Quality differences associated with location and opening hours can yield differences in mark-ups or marginal costs between stores. Sellers might have some local market power, and the elasticity of demand facing the sellers, and hence also the optimal mark-up, might vary between stores. In an environment with imperfectly informed customers, as in Burdett and Judd (1983), identical stores might charge different average prices. The sellers that charge high average prices have high mark-ups but low sales, while the reverse is true for sellers that offer low average prices, making the expected profit equal. Finally, store effects might be a result of product aggregation. As some products in our data might represent different brands and qualities across stores, this could yield a store effect. In our empirical analyses below, we do not distinguish between the different sources of the store effect.

Next, we question why store effects do not explain all dispersion in prices. It could be that the marginal cost of selling different products varies. A store with a huge freezer might have lower marginal costs of storing and selling frozen products than a store with a small freezer. Similarly, firms might specialize in selling some products, and set higher mark-ups on these products. They might specialize in expertise to help consumers choose between products of different qualities and properties, while selling more standardized products at discounted prices. A store selling cross-country skis might specialize in selecting the optimal pair of skis (with the appropriate span and stiffness) for each individual customer, and thus charge a higher price on skis while charging the same price as a general store for ski wax.

Within-store price dispersion can also be explained by price discrimination. Kaplan *et al*. (2019) analyze a two-goods variant of the model of Burdett and Judd (1983). Customers differ in the cost of shopping time. By selling the two goods at different prices, the store might be able to sell the bundle at a high average price to customers with a high cost of shopping, and in addition sell the low-priced good to customers with a low cost of shopping. This leads to within-store dispersion.

Lal and Matutes (1989) study a very similar mechanism. They consider a duopoly framework in which both sellers sell the same two goods, while customers differ in their transportation cost (the cost of going to both stores). In one equilibrium (out of several), sellers sell one of the goods at a discount, which is a good that differs between the stores. The customers with a high transportation cost buy both products in the same store, while those with a low transportation cost buy one good at each store.

Sellers with market power that serve a given customer pool might also engage in temporal price discrimination. This is particularly relevant for durable goods. Most of the time, prices are set high, and the sellers sell to high-valuation customers only. However, from time to time, the seller sets prices low (sales) in order to serve segments with lower willingness to pay (see Conlisk *et al*., 1984).

Also, incomplete information can yield a search equilibrium where prices differ randomly between stores (cf. Burdett and Judd, 1983; Moen, 1999).[5] Lastly, Kaplan and Menzio (2015) show that price discrimination can also yield price dispersion beyond store effects.

III. Data

We use monthly micro data collected for the consumer price index (CPI).[6] The data cover monthly price observations of 766 different products and services from 4,297 stores from January 2000 to December 2004. In total, our sample consists of 2,774,494 price observations. The median number of observed prices in a store (in one month) is 46 with an interquartile

[5] Moen (1997) shows that price dispersion can also emerge with a directed search.

[6] See the Statistics Norway web site, http://www.ssb.no/en/priser-og-prisindekser/statistikker/kpi/, and Statistics Norway (2001).

range (IQR) between 19 and 187.[7] Online Appendix A reports some further descriptives on the size of the dataset.

The products represent all COICOP divisions, such as food, apparel, furnishing, transport, services, recreation, electronics, and fuels to name a few. Compared with Kaplan and Menzio (2015) and Kaplan *et al*. (2019), who analyze grocery prices (mostly products categorized by COICOP division 1), our data include a larger variety of products covering more than 70 percent of household expenditures. The panel is unbalanced, as some products and stores are replaced each year by Statistics Norway to ensure the representativeness of the consumption basket (Statistics Norway, 2001). Each product is defined with a varying degree of precision depending on its type. For example, “Coca-Cola, bottle, 0.33 liters” is more precise than “Dress, ladies, simple”. Thus, for some products, price dispersion might reflect differences in quality between stores. Thus, the products are a sample and not the population of the products in each store.

Each store is located in one of eight regions: (1) Oslo, (2) Bergen, (3) Trondheim, (4) Akershus county, (5) Eastern Norway, (6) Southern and Western Norway, (7) Central Norway, and (8) Northern Norway. The regions are heterogeneous, differing in geographical size and population density. Regions (1)–(3) are main cities, region (4) is a county just east of Oslo, while regions (5)–(8) are larger entities (see the map in Figure A1 in Online Appendix A).

Some prices might change more often than each month, so there might be some dispersion in prices that these data will miss. However, using the same data, Wulfsberg (2016, see Table B2 in that paper) reports that the mean duration of a price spell (i.e., the time between two price changes) varies between 5.8 months for food products (COICOP division 1) and 16.0 months for transport services (COICOP division 7).

IV. Stylized Facts of Price Dispersion

In this section, we present different measures of price dispersion in our sample, and we show how dispersion varies between products and over time. We denote P_{ist} as the price observation for product i in store s at month t. First, we construct a price distribution for each product–month sample, in total 40,567 distributions.[8] We drop product–month distributions

[7]The observed prices are a sample of all the goods sold by a store. Our underlying assumption is that the price of a basket of sampled goods is representative for the overall price level in that store. This seems reasonable given that the purpose of collecting price information is to measure the CPI.

[8]Unfortunately, there are not enough observations to construct price distributions at the product–region–month level. If we were to keep regions as one of the dimensions, we would need to

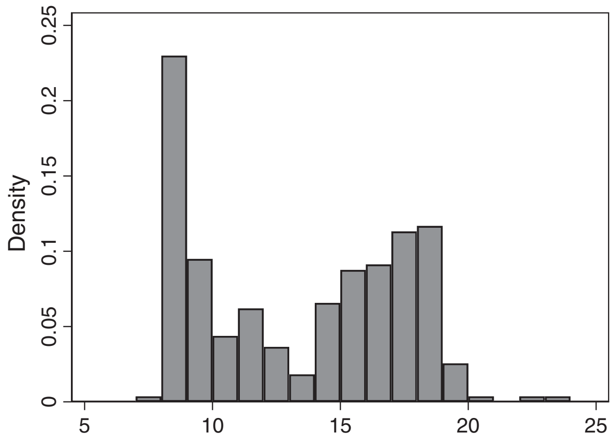

Fig. 1. Distribution of prices (in NOK) of Coca-Cola in January 2000

with fewer than 20 observations (stores) in order to reduce sampling errors. The median number of observations in a product–month distribution is 58 with an IQR of 39–87.

As an example, Figure 1 shows the distribution of observed prices P_{ist} for a bottle of Coca-Cola (in NOK) from 274 stores in January 2000. The lowest price is 7 NOK and the highest price is more than three times higher at 24 NOK. The third-quartile price is 89 percent higher than the first-quartile price, and the standard deviation is 3.70 NOK, which is 28 percent relative to the mean price. Interestingly, the price distribution is clearly bimodal. One possible interpretation of this feature is that stores are either "cheap" (e.g., supermarkets) or "expensive" (e.g., convenience stores).

Each product–month distribution of P_{ist} will obviously have different means μ_{it} and variances Σ_{it}, which depend on the the scale (i.e., the mean μ_{it}). In order to compare the dispersion of prices across products and months, we normalize all prices P_{ist} with respect to the mean price for each product in each month $\bar{P}_{it}$:

$$\tilde{P}_{ist} = \frac{P_{ist}}{\bar{P}_{it}}. \tag{1}$$

drop either the product or the time dimension (i.e., to analyze region–month or region–product distributions). As documented, price dispersion varies a lot between products, but it is also important to control for idiosyncratic changes in the mean product price over time. Furthermore, by analyzing product–time distributions, we can directly compare our results with previous studies. Hence, we analyze product–month distributions in our main analysis, which also allows us to compare our results with previous studies.

Table 1. *Descriptive statistics for relative price dispersion across products and over time*

Dispersion measure	Median	Q1–Q3 range
Standard deviation	0.327	0.193–0.504
IQR	0.319	0.166–0.562
P95–P5 range	0.945	0.548–1.480

Here, $\tilde{P}_{ist}$ has a mean of 1 and a variance σ_{it}^2, which we can compare across products and over time.[9]

As the product–month price distributions can be skewed or multimodal (e.g., the distribution for Coca-Cola in Figure 1), we measure price dispersion as the IQR and the 95/5 percentile range (the P95–P5 range) in addition to the variance and standard deviation. As the distributions are normalized, the standard deviation, the IQR, and the P95–P5 range are in percent of the mean.

The first column of Table 1 reports the median product–month estimates of these measures, and in Column 2 we illustrate the variation in each of these measures (across products and over time) by the range between the first and third quartiles (i.e., the Q1–Q3 range). Table 1 shows that price dispersion is significant with an estimated median standard deviation of 32.7 percent (corresponding to a variance of 0.107). The median IQR is 31.9 percent and the median P95–P5 range is 94.5 percent. However, there is a lot of variation in dispersion between the product–month distributions, as indicated by a Q1–Q3 range of the standard deviation between 19 and 50 percent. Similarly, there is a lot of variation in the other measures of price dispersion. The Q1–Q3 range for the IQR is between 16.6 and 55.8 percent, and between 54.8 and 149.0 percent for the P95–P5 range. This variation might reflect that some products are precisely defined while others are aggregates of products that are close substitutes.

Figure 2 shows a histogram of the 40,567 standard deviations illustrating the variation in price dispersion. The distribution is skewed right, which is why we report on the median.[10]

By grouping the products in COICOP categories, we show systematic differences in dispersion between types of products. Table 2 shows the median standard deviation and IQR for each COICOP division in the top panel and the degree of durability in the bottom panel. The groups are ranked according to their standard deviation. Clothing and footwear

[9]We exclude 4,010 outliers (i.e., 0.14 percent of the observations) with a relative price greater than 5 or less than 0.05, and then we renormalize. The outliers represent all regions and COICOP divisions non-systematically.

[10]The mean variance and standard deviation are 0.180 and 36.3 percent, respectively.

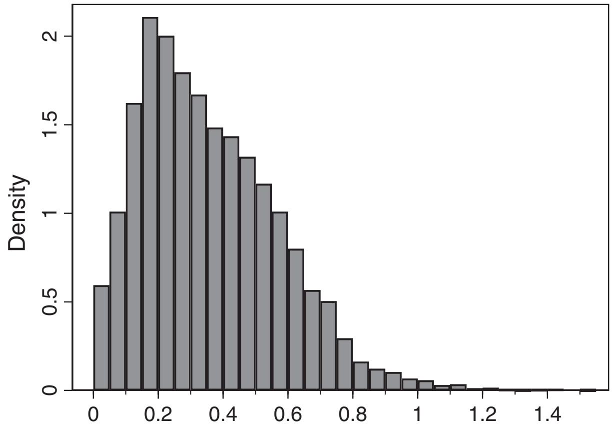

Fig. 2. Histogram of the standard deviation for all product–month distributions truncated at 2

Table 2. *Median dispersion in relative prices across COICOP categories, ranked by median standard deviation*

COICOP division	*N*	Standard deviation	
		Median	Q1–Q3 range
3 Clothing and footwear	102	0.551	0.425–0.666
5 Furnishings, household equipment and routine maintenance	120	0.424	0.254–0.584
11 Restaurants and hotels	38	0.408	0.307–0.488
8 Communications	7	0.401	0.343–0.523
7 Transport	37	0.351	0.161–0.515
9 Recreation and culture	81	0.325	0.199–0.488
12 Miscellaneous goods and services	60	0.325	0.219–0.536
1 Food and non-alcoholic beverages	255	0.253	0.170–0.367
4 Housing, water, electricity, gas and other fuels	15	0.250	0.134–0.385
2 Alcoholic beverages and tobacco	12	0.099	0.068–0.179
6 Health	39	0.089	0.035–0.223
Semi-durables	184	0.508	0.384–0.644
Services	72	0.388	0.248–0.497
Durables	86	0.372	0.213–0.553
Non-durables	424	0.243	0.154–0.376

Notes: N is the number of products in each category.

products have the highest price dispersion with a median standard deviation of 55.1 percent. The least-dispersed categories are Health products and Alcoholic beverages and tobacco, with a median standard deviation of less than 10 percent. In particular, the dispersion we find in normalized prices for food products in Norway is 25.3 percent (measured by the median standard deviation), which is similar to what Kaplan and Menzio (2015) and Kaplan *et al*. (2019) find for grocery products in the United States (19–36 percent).

The COICOP system also classifies products as durables, semi-durables, non-durables, and services (see Table A2 in the Online Appendix). For example, clothing and footwear products are classified as semi-durables. The bottom part of Table 2 shows the median product–month standard deviation within these categories. The standard deviation for semi-durable products is about twice as high as for non-durable products (50.8 versus 24.3 percent). The dispersion for services and durables are positioned in between these values at around 38 percent. The right column of Table 2 shows that there is a lot of heterogeneity within each consumption category, indicated by the Q1–Q3 range of the standard deviation.

Price dispersion is positively correlated with the price level with a correlation coefficient of 0.20 between the standard deviation and the log of the mean price. However, the relationship is actually hump-shaped (see Figure 3). This color figure (like Figure C2 in the Online Appendix) shows that non-durables dominate the low-price segment, durables dominate the high-price segment, while semi-durables are in between. The average prices of services are represented over the whole price range. These systematic differences shed light on the difference in price dispersion across COICOP divisions in Table 2. In particular, COICOP divisions

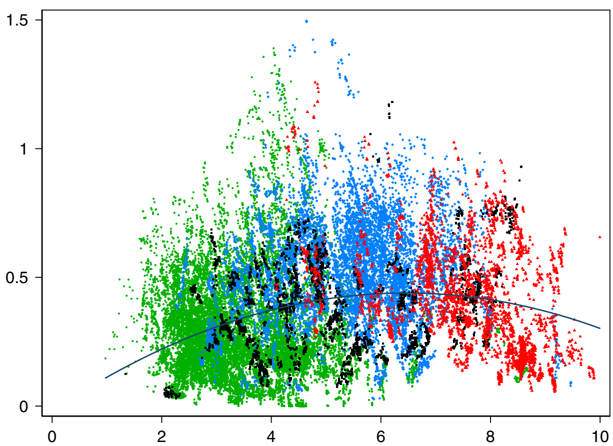

Fig. 3. Scatter plot of the standard deviation (vertical axis) and log of the mean price (horizontal axis) with a quadratic fitted line [Colour figure can be viewed at wileyonlinelibrary.com]

Notes: Non-durables are shown in green, semi-durables are blue, durables are red, and services are black.

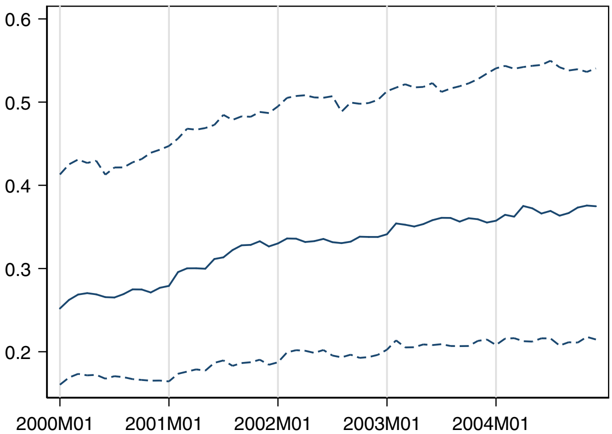

Fig. 4. The first, second (median), and third quartiles of the standard deviations over time

3 (Clothing and footwear) and 5 (Furnishings, household equipment and routine maintenance) are dominated by semi-durable products.

Figure 4 shows that there is a clear upward trend in the three quartiles (Q1, median, and Q3) of the standard deviation, indicating that dispersion increased over time. The median standard deviation increased from around 25 percent at the beginning of the sample to almost 40 percent at the end. This finding is consistent with that of Wulfsberg (2016), who reports that the mean size of non-zero price changes increased over the same period.

How much of the variation in price dispersion reported in Table 1 is accounted for by this trend? The decomposition of the variation in the standard deviation σ_{it} into cross-sectional variation between products ($\bar{\sigma}_i$) and time variation within products ($\sigma_{it} - \bar{\sigma}_i$) shows that the cross-sectional variation accounts for 84 percent while time variation only accounts for 16 percent of the overall variation in σ_{it}.

What can explain this trend? In an Ss-type menu cost model (see Caplin and Leahy, 2010), the size of price adjustments increases with inflation. Hence, we would expect more price dispersion when inflation is high. However, during this period, the 12-month annual inflation rate has varied between 4 and −2 percent with a slightly negative trend (if any). The trend increase in price dispersion might be related to a changing composition of products over time. To investigate this possibility, we have estimated price dispersion for the subset of products that is observed over the whole period. Table E1 in the Online Appendix shows that the estimates of price dispersion for the balanced panel is hardly changed. Figure E1 shows that the trend prevails, which indicates that the trend is not explained by changes to the composition of products.

Because the distributions are constructed at the product–month level, we cannot report estimates of price dispersion across regions based on these distributions. However, to get an idea of the regional variation on price dispersion, we have also constructed normalized price distributions for product and regions (omitting the time dimension); that is, $\tilde{P}_{isr}$, where subscript r is an index for region. Table D1 in the Online Appendix shows that price dispersion does not vary much between regions. Surprisingly, given the huge geographical size, Northern Norway is the region with the lowest price dispersion of 27.8 percent, and Southern and Western Norway has the highest price dispersion of 33.1 percent. Also, surprisingly perhaps, there is no less price dispersion in the three cities than in the geographically larger regions. Thus, regions do not seem to play an important role in price dispersion.

We noted above that a possible explanation for the bimodality of the price distribution for Coca-Cola in Figure 1 is that the product is sold by stores with different characteristics such as "cheap" stores (e.g., supermarkets) and "expensive" stores (e.g., convenience stores). Multimodal product-specific distributions of normalized prices seem to be prevalent. Instead of visual inspection of each product, we employ the Hartigan dip test (Hartigan and Hartigan, 1985), which rejects unimodality at a 1 percent level of significance for as many as 576 (75 percent) of the products.[11] Another explanation for multimodality is that there are regional differences in the price distributions. However, Figure D1 in the Online Appendix shows that the regional distribution of prices for Coca-Cola is still strongly bimodal.

While product-specific distributions are often multimodal, the pooled distribution of normalized prices is single peaked, kurtotic, and slightly left skewed, as seen from Figure B1 in the Online Appendix. The standard deviation of the pooled distribution is 39.9 percent, which is 7.2 percentage points higher than the median product–month standard deviation.

V. The Store Component

We assume that the relative price $\tilde{P}_{ist}$ can be decomposed into the mean (by definition equal to 1), a store component v_s, and a residual ε_{ist}:

$$\tilde{P}_{ist} = 1 + v_s + \varepsilon_{ist}. \tag{2}$$

The store component is assumed to be equal for all products sold in the same store s in all periods. The idea of the store component is illustrated in Figure 5 where the dots represent four observations of the relative price

[11] Cavallo and Rigobon (2014) use the dip test to inspect the distribution of price changes.

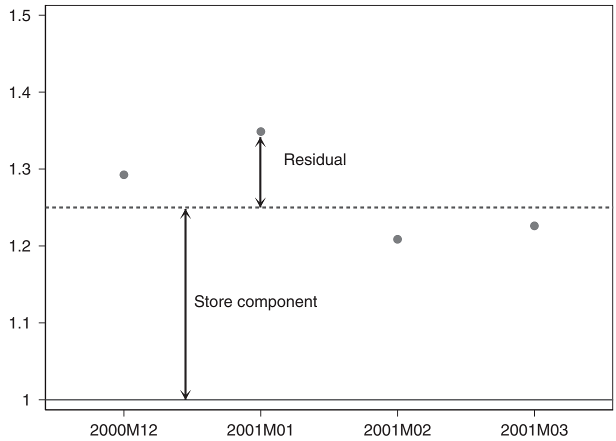

Fig. 5. Illustration of the store component
Notes: The vertical axis is the relative price and the horizontal axis is the month. The dots are observations from the same store.

for a bottle of Coca-Cola sold in one particular store s from December 2000 to April 2001. The distance from the mean relative price (equal to 1) to the dashed line represents the store component, v_s. Thus, the conditional mean of the relative price for all products sold by this particular store is $1 + v_s$. Because the store component is around 0.25, the store is on average 25 percent more expensive compared with other stores. However, the four price observations vary around the store mean $(1 + v_s)$, represented by the residual ε_{ist}.[12]

An intuitive estimate of the store component v_s is the mean normalized price for all products in all periods sold by the same store:

$$\hat{v}_s = \frac{1}{N_s} \sum_n (\tilde{P}_{ist} - 1). \tag{3}$$

Here, $n = 1, \ldots, N_s$ is the number of observations from store s over all products and months. Note that the estimation of the store effect is based on a large number of observations. As the median number of products per store is 46 and the median store is observed over 31 months, the median number of observations behind the estimate of the store component is 1,426 observations.

[12]While the store effect in equation (2) is fixed over time, Kaplan *et al.* (2019) estimate a time-varying store effect by decomposing the error terms further into a transitory and a persistence component. Their results indicate that 95 percent of the sample store effect is persistent.

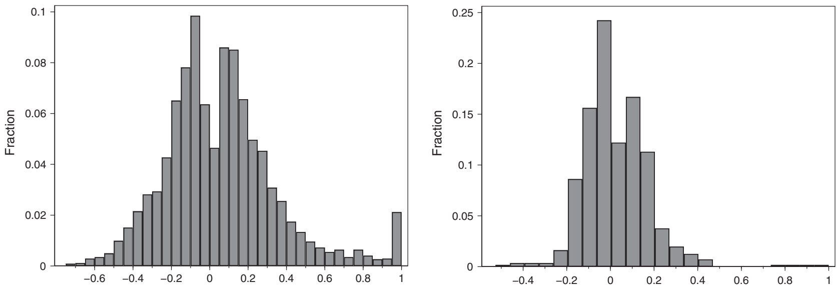

Fig. 6. Histograms of the significant store effects (significance at the 1 percent level) for all stores (left) and Coca-Cola stores (right)
Notes: The histograms are truncated at 1.

A t-test shows that one-fifth of the store effects are insignificant at the 1 percent level.[13] The significant (non-zero) store effects are plotted in Figure 6. Their sizes vary between −70 and 300 percent. The histogram is clearly bimodal with one mode below zero around −10 percent and one mode above zero around 15 percent. The mean store effect is 28.5 percent for expensive stores and −18.0 percent for cheap stores. (The mean absolute size of the store effects is 23.6 percent.) The histogram to the right is for stores selling Coca-Cola, which is also possibly bimodal with modes on each side of zero. In Figure F2 in the Online Appendix, the store effects are plotted by COICOP division. Bimodality is also reflected in these histograms with the exception of COICOP division 9 (Recreation and culture) and 12 (Miscellaneous goods and services). We note that the store effects seem particularly strong for division 3 (Clothing and footwear).

Interestingly, there is a clear tendency for the more expensive stores to have more variation in (normalized) prices within the store. Figure 7 plots the store effects versus the variation in prices between products sold in the same store. One possible explanation for this is that expensive stores selling specialized products also need to sell standardized products (as in our example above of a store selling cross-country skis and ski wax).

The residual component $\hat{\varepsilon}_{ist}$ is computed by subtracting the estimated store effects from each normalized price (i.e., $\hat{\varepsilon}_{ist} = \tilde{P}_{ist} - 1 - \hat{v}_s$, following equation (2)). The variance of ε_{ist} represents the price dispersion for products sold at equally expensive stores. Table 3 reports the same measures of residual price dispersion as for normalized prices in Table A1.

[13] 15 percent of the store effects are insignificant at the 5 percent level.

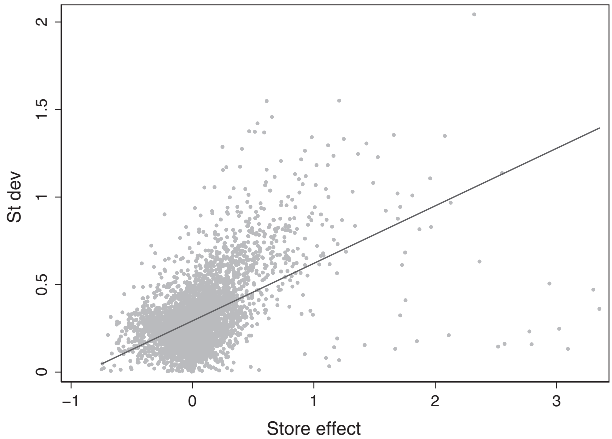

Fig. 7. Expensive stores have higher price variation

Table 3. *Descriptive statistics for residual price dispersion across products and over time*

Dispersion measure	Median	Q1–Q3 range
Standard deviation	0.274	0.174–0.392
IQR	0.277	0.157–0.416
P95–P5 range	0.806	0.511–1.185

Focusing on the dispersion of residual prices, Lach (2002) finds less dispersion (but only for four products). All three dispersion measures of the residual prices are around 85 percent of the corresponding measure for normalized prices. Thus, there is substantial variation in prices, even after controlling for store effects.

VI. The Importance of Store Heterogeneity

In this section, we explore how much of the variation in prices, which we documented in the previous section, can be attributed to store heterogeneity. We interpret the statistical model (2) as an error component model where the store effect v_s and the residual price ε_{ist} are stochastic, each with a zero mean and with time and product-specific variance σ^2_{vit} and $\sigma^2_{\varepsilon it}$. Note, in particular, that the variance of the store effect can vary between products and over time, even if the store effect v_s does not vary between products and over time. The reason for this is that the product range varies between stores over time. To illustrate this point, assume that there are, in total, three products A, B, and C. There are many stores, but each store sells only two products (A, B), (A, C), or (B, C). The store effect for a particular store is equal for both products sold in that store. However, the variance of the store effect for stores selling (A, B) products can be different from the variance of the store effect for stores selling the (A, C) products or the (B, C) products. Hence, the variance of the store effect for product A can differ from product B and product C. More generally, the store effect is drawn from a distribution, which depends on the set of products sold by the store and on the period.

Assuming that v_s and e_{ist} are independent, the variance of $\tilde{P}_{ist}$ is thus decomposed into

$$\sigma_{it}^2 = \sigma_{vit}^2 + \sigma_{\varepsilon it}^2. \tag{4}$$

The ratio of the variance of the store component σ_{vit}^2 to the total variance σ_{it}^2 measures the importance of store heterogeneity for price dispersion. Thus, our goal is to estimate σ_{vit}^2 and $\sigma_{\varepsilon it}^2$ and their implied share of the total variance σ_{it}^2.

We first estimate the variance components for each product–month distribution based on the estimates $\hat{v}_s$ and $\hat{\varepsilon}_{ist}$ according to

$$\hat{\sigma}_{vit}^2 = \sum_s \frac{(\hat{v}_s - \bar{v}_s)^2}{S_{it} - 1} \tag{5}$$

and

$$\hat{\sigma}_{\varepsilon it}^2 = \sum_s \frac{(\hat{\varepsilon}_{ist} - \bar{\varepsilon}_{ist})^2}{S_{it} - 1}, \tag{6}$$

where $s = 1, \ldots, S_{it}$ is an index for stores selling product i in month t.

In Column 1 of Table 4, we report the median variance of the store component and the median residual price variance, which turn out to be 0.032 and 0.075, respectively. This yields a total median variance of normalized prices of 0.107 (which is consistent with the estimates in Table 1). Thus, the store effect accounts for 30 percent of the total variance of the median product–month distribution. The ratio of the store variance varies obviously between product–month distributions. To illustrate this

Table 4. *Estimates of variance components, with share of total variance in parentheses*

Variance	Median		Q1–Q3 range	
Store $\hat{\sigma}_{vit}^2$	0.032	(30)	0.007–0.100	(19–39)
Residual $\hat{\sigma}_{eit}^2$	0.075	(70)	0.030–0.154	(81–61)
Total $\hat{\sigma}_{vit}^2 + \hat{\sigma}_{eit}^2$	0.107	(100)	0.037–0.254	(100)

Table 5. *Variance decomposition by COICOP division: mean estimates*

COICOP division	σ^2	σ_v^2	σ_v^2/σ^2
11 Restaurants and hotels	0.151	0.074	49%
3 Clothing and footwear	0.333	0.150	45%
12 Miscellaneous goods and services	0.181	0.057	32%
2 Alcoholic beverages and tobacco	0.055	0.017	32%
5 Furnishings, household equipment and routine maintenance	0.222	0.066	30%
4 Housing, water, electricity, gas and other fuels	0.098	0.026	26%
9 Recreation and culture	0.143	0.036	25%
6 Health	0.078	0.018	23%
1 Food and non-alcoholic beverages	0.109	0.025	23%
7 Transport	0.196	0.035	18%
8 Communications	0.189	0.022	12%
Services	0.159	0.060	38%
Semi-durables	0.293	0.106	36%
Durables	0.191	0.067	35%
Non-durables	0.112	0.027	24%

variation, we report the same statistics for the IQR (Q1–Q3 range) in the second column. We see that the store effect accounts for 19–39 percent of the total variance of product–month distributions measured by the Q1–Q3 range.

In Table 5, we report the variance decomposition by COICOP categories using the same approach as in Table 4. The store effect is particularly important for division 11 (Restaurants and hotels) in addition to division 3 (Clothing and footwear), at 49 and 45 percent, respectively. Typically, for services, we would expect variation in the store component to be an important part of the price dispersion. The store effect is least important for division 8 (Communications) with a ratio of 12 percent to the total variance. For food products (1 Food and non-alcoholic beverages), the store effect accounts for 23 percent of the total variance, which is similar to Kaplan and Menzio (2015).

It is likely that stores selling the same product(s) are less heterogeneous than stores in general. For example, food stores are probably less heterogeneous than food stores compared with hotels. Hence, for retail prices in general, the variance in the store component would be expected to be more important for price dispersion than for the median product–month sample. To investigate this possibility, we pool the sample and assume that $\sigma_{it}^2 = \sigma^2$, $\sigma_{vit}^2 = \sigma_v^2$, and $\sigma_{\varepsilon it}^2 = \sigma_\varepsilon^2 \quad \forall\, i, t$. In this exercise we estimate σ_v^2

Table 6. *Estimates of variance components: pooled distribution*

Variance	Pooled distribution		Mixed effects method	
Store $\hat{\sigma}_v^2$	0.104	(66)	0.098	(48)
Residual $\hat{\sigma}_e^2$	0.053	(34)	0.107	(52)
Total $\hat{\sigma}_v^2 + \hat{\sigma}_e^2$	0.157	(100)	0.205	(100)

Notes: Share of total variance in parentheses.

and σ_ε^2 by

$$\hat{\sigma}_v^2 = \sum_s \frac{(\hat{v}_s - \bar{v}_s)^2}{S - 1} \tag{7}$$

and

$$\hat{\sigma}_\varepsilon^2 = \sum_n \frac{(\hat{\varepsilon}_{ist} - \bar{\varepsilon}_{ist})^2}{N - 1}, \tag{8}$$

where $s = 1, \ldots, S$ is an index for all stores, and $n = 1, \ldots, N$ is an index for all observations in the sample. As reported in Column 1 of Table 6, this yields an estimate of the variance of the store component of 0.104, which is significantly larger than the estimate in Table 4, as expected. This accounts for 66 percent of the pooled total variance, leaving 34 percent for the residual variance, indicating a significant larger role for store effects when we look at the whole sample of products.

For robustness, we estimate σ_v^2 and σ_e^2 directly using the mixed effects method (for details, see Rabe-Hesketh and Skrondal, 2012). This method estimates $\hat{\sigma}_v^2$ and $\hat{\sigma}_e^2$ simultaneously using maximum likelihood without first estimating the store component v_s, but assuming normality and homoskedasticity (i.e., $v_s \sim N(0, \sigma_v^2)$). This approach yields an estimate of the store component variance σ_v^2 equal to 0.098, which is a share of 48 percent of the total variance (see Column 2 of Table 6). Note that the mixed effects estimate of the total variance is larger than the pooled variance because of the normality assumption, while the empirical (pooled) distribution is kurtotic (see Figure B1 in the Online Appendix).

Thus, the different approaches yield an estimated share of the store effects from 30 percent for the median product–month sample to around 50 percent for the pooled sample. So, we attribute a somewhat stronger importance to store heterogeneity than Kaplan *et al.* (2019), who attribute 10–36 percent of the observed price dispersion to store heterogeneity. The main reason for this difference is that we analyze prices for a wider product range and probably more heterogeneous stores.

VII. The Persistence of Store Heterogeneity

Store heterogeneity is an important component of the observed price dispersion, as documented above. In order to investigate the persistence of the store heterogeneity, we inspect the ranking of stores within the price distributions over time, following Lach (2002). Is a store's ranking in the price distribution persistent as indicated by the estimated store effects?

For each product–month, we partition each price distribution by the three quartiles (Q1, median, and Q3), and we assign each store into one of the four quartile bins: $QB1_{it}$, $QB2_{it}$, $QB3_{it}$, and $QB4_{it}$.[14] For how long does a store remain in the same quartile bin? Furthermore, how likely is it that a store that changes its nominal price will jump from one quartile bin to another or remain in the same part of the relative price distribution? If a store is systematically more expensive, consumers can learn this information and take advantage of price differences. If some consumers are informed about prices while others are not (Varian, 1980), then this shows that it is optimal for a store to randomize its price.

We denote the probability for store s to move from QBk_{it} to QBj_{it+1} as γ_{kji}. We estimate γ_{kji} by the fraction of stores selling product i, which moves from quarterly bin k in month t to quarterly bin j in the next month.[15] For each product i, the transition probabilities γ_{kji} give us all the elements of the one-month transition probability matrix. Table 7 reports the one-month transition probability matrix for the median product.[16] Each row represents the probability of either staying in the same bin or moving to another bin.[17] For example, the median probability of moving from the first quartile bin QB1 to the second quartile bin QB2 in the next month is $\gamma_{12} = 7.9$ percent.

We see that a store is most likely to stay in the same quartile bin in the next month as the probabilities along the diagonal are the largest varying between 83 and 93 percent. Given that a store moves from one quartile bin to another, it is most likely to move to an adjacent bin. The closest elements to the diagonal vary between 6.2 and 8.0 percent. The probability of jumping two quartile bins (e.g., from QB3 to QB1) ranges between 0.8 and 1.5 percent. A store is least likely to move from one tail to the other. The median probability of moving from the lowest to the highest bin is 0.6 percent while the probability of the reverse transition is 0.3 percent.

[14] However, random sampling errors will induce statistical errors in our ranking of stores.

[15] This estimate corresponds to the predicted probabilities of a probit model with the initial quarterly bin as a regressor.

[16] See Table G1 in the Online Appendix for the mean probabilities.

[17] Note that the rows do not sum to 1, as each element is the median value of each cell. However, the rows sum to 1 for each individual product.

Table 7. *One-month transition probability matrix, normalized prices* $\tilde{P}_{ist}$

Origin bin	Destination bin			
	$QB1_{t+1}$	$QB2_{t+1}$	$QB3_{t+1}$	$QB4_{t+1}$
$QB1_t$	0.885	0.079	0.015	0.006
	(0.096)	(0.061)	(0.030)	(0.053)
$QB2_t$	0.071	0.830	0.076	0.011
	(0.056)	(0.145)	(0.062)	(0.084)
$QB3_t$	0.010	0.080	0.831	0.069
	(0.023)	(0.059)	(0.152)	(0.109)
$QB4_t$	0.003	0.008	0.062	0.925
	(0.010)	(0.017)	(0.039)	(0.058)

Notes: Median estimates with standard errors in parentheses. All observations. The rows do not sum to 1, as each element is the median value. However, the rows sum to 1 for each individual product.

The matrix is quite symmetric, but the upper elements are somewhat larger than the corresponding lower elements. This indicates that the likelihood of moving down from, for example, the third quartile bin to the first quartile bin γ_{31} is smaller than moving up from the first quartile bin to the third quartile bin γ_{13}. The transition probability matrix varies across products, as indicated by the standard deviations.

We find the same pattern when we estimate the 12-month median transition probability matrix (see Table 8).[18] Even 12 months ahead, a store is most likely to remain in the same quartile bin than to move to any other bin. The median probability of being in the same quartile bin after 12 months varies between 51 and 73 percent compared with between 83 and 93 percent for the one-month ahead estimates in Table 7.

A change in a store's ranking within the relative price distribution can happen as a result of not only the store changing its own price, but also if other stores have changed their prices. It is interesting to know the transition probabilities conditional on the store changing its own nominal price. Table 9 reports the conditional one-month transition probability matrix. Still, the largest probabilities are found on the diagonal ranging from 60 to 79 percent. If a store does change its ranking following a nominal price change, it is most likely to move to an adjacent quartile, with probabilities ranging between 14.0 and 17.1 percent. These probabilities are roughly double compared with the corresponding unconditional probabilities. The probability of jumping two quartiles (e.g., from QB1 to QB3) now ranges between 3.8 and 5.6 percent, increasing by a factor of 4 compared with the unconditional probabilities. Finally, the median probabilities of moving

[18] See Table G2 in the Online Appendix for the mean probabilities. Table G3 reports the six-month transition probability matrices for the median, the mean, and the standard deviation.

Table 8. *Twelve-month transition probability matrix, normalized prices* $\tilde{P}_{ist}$

Origin bin	Destination bin			
	$QB1_{t+12}$	$QB2_{t+12}$	$QB3_{t+12}$	$QB4_{t+12}$
$QB1_t$	0.632	0.216	0.065	0.031
	(0.194)	(0.125)	(0.092)	(0.124)
$QB2_t$	0.174	0.506	0.206	0.056
	(0.098)	(0.186)	(0.134)	(0.109)
$QB3_t$	0.046	0.200	0.521	0.202
	(0.056)	(0.099)	(0.178)	(0.158)
$QB4_t$	0.019	0.047	0.183	0.734
	(0.036)	(0.068)	(0.082)	(0.129)

Notes: See Table 7.

Table 9. *One-month transition probability matrix conditional on nominal price changes, normalized prices* $\tilde{P}_{ist}$

Origin bin	Destination bin			
	$QB1_{t+1}$	$QB2_{t+1}$	$QB3_{t+1}$	$QB4_{t+1}$
$QB1_t$	0.722	0.162	0.056	0.025
	(0.156)	(0.086)	(0.070)	(0.116)
$QB2_t$	0.140	0.595	0.171	0.045
	(0.081)	(0.166)	(0.104)	(0.114)
$QB3_t$	0.043	0.158	0.602	0.166
	(0.056)	(0.075)	(0.164)	(0.127)
$QB4_t$	0.017	0.038	0.140	0.788
	(0.035)	(0.044)	(0.063)	(0.096)

Notes: See Table 7.

between the tail bins are 1.7 and 2.5 percent. The standard deviations are larger than in the unconditional estimation, so there is more variation in the transition probability matrices when we condition on a nominal price change.

Our results suggest that there are persistent patterns in the ranking of stores within a distribution, consistent with the finding that fixed store effects are an important component of variation in prices. Knowledge of the ranking of stores from a previous period might imply significant search cost savings for consumers as the previous ranking is a fair bet for the current ranking.

Fixed store effects are likely to be related to the persistence of relative prices. Thus, it is possible for consumers to learn what stores are cheaper on average. What is the relative price mobility of equally expensive stores? To answer this question, we control for store effects and we estimate transition

Table 10. *One-month transition probability matrix, residual prices $\hat{\varepsilon}_{ist}$*

Origin bin	Destination bin			
	$QB1_{t+1}$	$QB2_{t+1}$	$QB3_{t+1}$	$QB4_{t+1}$
$QB1_t$	0.907	0.069	0.011	0.007
	(0.076)	(0.043)	(0.020)	(0.013)
$QB2_t$	0.069	0.846	0.074	0.011
	(0.041)	(0.093)	(0.043)	(0.018)
$QB3_t$	0.011	0.077	0.844	0.064
	(0.021)	(0.044)	(0.093)	(0.039)
$QB4_t$	0.006	0.010	0.064	0.917
	(0.013)	(0.018)	(0.035)	(0.036)

Notes: Median estimates with standard errors in parentheses. All observations. The rows do not sum to 1 as each element is the median value. However, the rows sum to 1 for each individual product.

Table 11. *Twelve-month transition probability matrix, residual prices $\hat{\varepsilon}_{ist}$*

Origin bin	Destination bin			
	$QB1_{t+12}$	$QB2_{t+12}$	$QB3_{t+12}$	$QB4_{t+12}$
$QB1_t$	0.676	0.199	0.064	0.043
	(0.147)	(0.083)	(0.062)	(0.048)
$QB2_t$	0.185	0.529	0.218	0.062
	(0.079)	(0.132)	(0.086)	(0.057)
$QB3_t$	0.062	0.204	0.517	0.191
	(0.060)	(0.087)	(0.136)	(0.092)
$QB4_t$	0.039	0.058	0.183	0.697
	(0.052)	(0.005)	(0.075)	(0.133)

Notes: See Table 10.

probability matrices for the residual prices $\hat{\varepsilon}_{ist}$. Tables 10–12 report the unconditional transition probability matrices for the residual prices for one month, 12 months, and one month, respectively, conditional on a nominal price change.[19]

Comparing the diagonal elements of the residual price transition probability matrix in Tables 10 and 11 with the corresponding probabilities for relative prices in Tables 7 and 8, we see that they are even higher for QB1 and QB2, but a little lower for QB3 and QB4. Thus, relative prices for equally expensive stores are still very persistent. Lach (2002) finds more flexibility for residual prices than we do. However, he is analyzing only

[19]See Tables H1 and H2 in the Online Appendix for the mean probabilities. Table H3 reports the six-month transition probability matrices for the median, the mean, and the standard deviation.

Table 12. *One-month transition probability matrix conditional on nominal price changes, residual prices* $\hat{\varepsilon}_{ist}$

Origin bin	Destination bin			
	$QB1_{t+1}$	$QB2_{t+1}$	$QB3_{t+1}$	$QB4_{t+1}$
$QB1_t$	0.769	0.136	0.044	0.028
	(0.129)	(0.072)	(0.051)	(0.040)
$QB2_t$	0.130	0.661	0.146	0.043
	(0.072)	(0.133)	(0.078)	(0.048)
$QB3_t$	0.037	0.135	0.669	0.137
	(0.051)	(0.076)	(0.130)	(0.078)
$QB4_t$	0.024	0.035	0.121	0.809
	(0.036)	(0.043)	(0.068)	(0.104)

Notes: See Table 10.

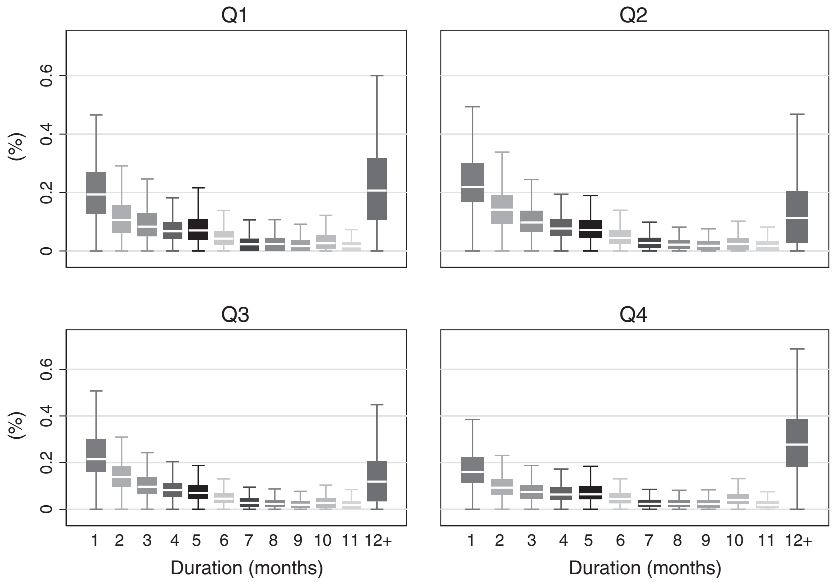

Fig. 8. Box plots of the monthly durations across the four quartiles
Notes: See Table I1 in the Online Appendix for the data used to make the box plots.

four products and with only one product for each store, which might lead to a biased store effect if prices are not perfectly correlated within each store.

We also measure the duration that stores are in a particular bin. Figure 8 presents box plots (across products) showing the fraction of different spells within each quartile bin. Most spells are typically between one and three months within a quartile bin. However, there is also a huge fraction of products where stores remain in the same quartile bin for 12 months or more, in particular for the lower quartile bin QB1 and the top quartile bin QB4.

We find the following relationship between the ranking spells and the transition probability matrix. The conditional probability of changing to a different quartile is the sum of the off-diagonal elements in the transition probability matrix. Taking the average across the four quartiles, we obtain the probability of changing a quartile one month ahead. This probability is equal to the probability of observing a one-month spell (Lach, 2002). These probabilities are very similar for each individual product in our estimations.

Based on the ranking spells and the transition probability matrix, stores in our sample are persistently cheap or expensive. Combined with the result from the variance decomposition, this result indicates that store heterogeneity is an important factor for price dispersion.

VIII. Conclusion

We document the empirical facts of price dispersion for a wider range of retail products and services than in earlier studies. The standard deviation for the median product is 33 percent. Dispersion varies between products and months, indicated by the IQR of the standard deviation from 19 to 50 percent. Prices appear more dispersed for clothing and footwear, and for other semi-durable goods, than for other products. Furthermore, price dispersion increased over time, as illustrated by an increase in the standard deviation for the median product from 25 to 40 percent over the sample period.

Our results suggest that store heterogeneity is an important component in price dispersion. By decomposing the variance in relative prices into a fixed store component and an idiosyncratic term, we find that 30 percent of the observed variance in relative prices for the median product–month can be accounted for by store heterogeneity. For the sample as a whole, store heterogeneity accounts for 50 percent of the variance in relative prices, which is a larger share than reported in previous studies.

The distribution of the store component is bimodal with a long right tail. The mean store effect for cheap stores is −18.0 percent while for expensive stores it is 28.5 percent.

The consistency of a store’s ranking within a distribution indicates that most stores are likely to be in the same part of the distribution one month ahead, and even 12 months ahead.

From a consumer point of view, it is possible to learn what stores are cheap by searching for prices.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Online Appendix

References

Basso, L. J., Clements, M. T., and Ross, T. W. (2009), Moral Hazard and Customer Loyalty Programs, *American Economic Journal: Microeconomics 1* (1), 101–123.

Berardi, N., Sevestre, P., and Thébault, J. (2017), The Determinants of Consumer Price Dispersion: Evidence from French Supermarkets, in L. Matyas (ed.), *The Econometrics of Multi-dimensional Panels: Theory and Applications*, Advanced Studies in Theoretical and Applied Econometrics, Vol. 50, Springer, Berlin, 427–449.

Burdett, K. and Judd, K. (1983), Equilibrium Price Dispersion, *Econometrica 51*, 955–969.

Caplin, A. and Leahy, J. (2010), Economic Theory and the World of Practice: A Celebration of the (S, s) Model, *Journal of Economic Perspectives 24* (1), 183–202.

Cavallo, A. and Rigobon, T. (2014), The Distribution of the Size of Price Changes, NBER Working Paper 16760.

Conlisk, J., Gerstner, E., and Sobel, J. (1984), Cyclic Pricing by a Durable Goods Monopolist, *Quarterly Journal of Economics 99*, 489–505.

Dixit, A. K. and Stiglitz, J. E. (1977), Monopolistic Competition and Optimum Product Diversity, *American Economic Review 67* (3), 297–308.

Grossman, S. J. (1981), The Informational Role of Warranties and Private Disclosure about Product Quality, *Journal of Law and Economics 24*, 461–483.

Hartigan, J. and Hartigan, P. (1985), The Dip Test of Unimodality, *Annals of Statistics 13*, 70–84.

Kaplan, G. and Menzio, G. (2015), The Morphology of Price Dispersion, *International Economic Review 56*, 1–41.

Kaplan, G., Menzio, G., Rudanko, L., and Trachter, N. (2019), Relative Price Dispersion: Evidence and Theory, *American Economic Journal: Microeconomics 11* (3), 68–124.

Lach, S. (2002), Existence and Persistence of Price Dispersion: An Empirical Analysis, *Review of Economics and Statistics 84*, 433–444.

Lal, R. and Matutes, C. (1989), Price Competition in Multimarket Duopolies, *RAND Journal of Economics 20*, 516–537.

Mackowiak, B. and Wiederholt, M. (2009), Optimal Sticky Prices under Rational Inattention, *American Economic Review 99* (3), 769–803.

Moen, E. R. (1997), Competitive Search Equilibrium, *Journal of Political Economy 105*, 385–411.

Moen, E. R. (1999), Education, Ranking, and Competition for Jobs, *Journal of Labor Economics 17*, 694–723.

Rabe-Hesketh, S. and Skrondal, A. (2012), *Multilevel and Longitudinal Modmodel Using Stata. Volume I: Continuous Responses*, Stata Press, College Station, TX.

Statistics Norway (2001), Consumer Price Index 1995–2000, Official Statistics of Norway.

Varian, H. (1980), A Model of Sales, *American Economic Review 70* (4), 651–659.
Weitzman, M. L. (1994), Monopolistic Competition with Endogenous Specialization, *Review of Economic Studies 61*, 45–56.
Wildenbeest, M. (2011), An Empirical Model of Search with Vertically Differentiated Products, *RAND Journal of Economics 42*, 729–757.
Wulfsberg, F. (2016), Inflation and Price Adjustments: Micro Evidence from Norwegian Consumer Prices 1975–2004, *American Economic Journal: Macroeconomics 8* (3), 175–194.

First version submitted August 2017;
final version received March 2019.