CHICAGO BOOTH | The Initiative on Global Markets

Working Paper No. 145

Booth Working Paper Series 17-18

Prices and Promotions in U.S. Retail Markets: Evidence from Big Data

Günter J. Hitsch
University of Chicago Booth School of Business

Ali Hortacsu
University of Chicago Department of Economics and NBER

Xiliang Lin
University of Chicago Booth School of Business

Initiative on Global Markets
The University of Chicago, Booth School of Business

"Providing thought leadership on international business, financial markets and public policy"

Prices and Promotions in U.S. Retail Markets: Evidence from Big Data*

Günter J. Hitsch
University of Chicago Booth School of Business

Ali Hortaçsu
University of Chicago and NBER

Xiliang Lin
University of Chicago Booth School of Business

May 2017
(First version: March 2015)

Abstract

We document the extent of price dispersion for identical products in the U.S. retail industry. Our analysis is based on "big data" that allow us to draw general conclusions based on the prices for close to 50,000 products (UPC's) in 17,184 stores that belong to 81 different retail chains. Both at the national and local market level we find a substantial degree of price dispersion for UPC's and brands at a given moment in time. We document that both persistent base price differences across stores and price promotions contribute to the overall price variance, and we provide a decomposition of the price variance into base price and promotion components. There is substantial heterogeneity in the degree of price dispersion across products. Some of this heterogeneity can be explained by the degree of product penetration (adoption by households) and the number of retail chains that carry a product at the market level. Prices and promotions are more homogenous at the retail chain than at the market level. In particular, within local markets prices and promotions are substantially more similar within stores that belong to the same chain than across stores that belong to different chains. Furthermore, the incidence of price promotions is strongly coordinated within retail chains, both at the local market level and nationally. We present evidence, based on store-level demand estimates for 2,000 brands, that price elasticities and promotion effects are also more similar within stores that belong to the same chain. Hence, the limited level of store-level price discrimination by retail chains reflects, in part, that their stores attract customers with similar demand.

*We thank Susan Athey, Pierre Dubois, and Steve Tadelis for their helpful comments and suggestions. Jacob Dorn, George Gui, and Jihong Song provided outstanding research assistance. Part of this research was funded by the Initiative on Global Markets (IGM) at the University of Chicago Booth School of Business and the Becker Friedman Institute at the University of Chicago. All correspondence may be addressed to the authors by e-mail at guenter.hitsch@chicagobooth.edu, hortacsu@uchicago.edu, or xiliang.lin@chicagobooth.edu.

1 Introduction

The goal of this paper is to document the extent of price dispersion and the similarities as well as differences in pricing and promotion strategies across stores in the U.S. retail industry. The analysis is based on a sample from the Nielsen RMS scanner data that comprehensively tracks weekly store-level prices and quantities sold for nearly 50,000 products in 17,184 stores that belong to 81 different retail chains, including grocery stores, drug stores, and mass merchandisers. This is a "big" data set with close to 25 billion observations. This comprehensive data set allows us to provide *general* insights that are not specific to a small number of products, categories, stores, or retailers as in most of the extant literature. Our focus is on products, defined either as UPC's (universal product code) or brands—UPC's that share the same brand name and are physically identical apart from the packaging or volume. This is different from an analysis of the overall price and promotion tactics across the whole assortment of products carried by a store, which is not easily possible without assumptions on how to aggregate products and how to account for differences in the assortments across stores. The pricing and promotion patterns documented in this paper are intended as a basis for future research by marketers, industrial organization economists, and macroeconomists.

We begin our analysis by documenting the basic facts of the dispersion of prices of identical or almost identical products across stores at a given moment (week) in time. We measure price dispersion for products defined as UPC's, which are identical across stores, and also for products defined as brands, which are only "almost identical" because they are frequently offered in different assortments of UPC's that differ in form or pack size, although the main product content is physically identical. We examine price dispersion at different geographic levels, and find that even at the most narrowly defined geographic level analyzed—3-digit ZIP codes—the law of one price is violated by an often substantial degree. For example, the ratio of the 95th to 5th percentile of prices is 1.294 for the median UPC and 1.433 for the median brand. Equally important is the heterogeneity in the degree of price dispersion across products, ranging from small to very large degrees of price dispersion.

The next step in the analysis is an examination of the sources of price dispersion at a given moment in time. In practice, the price of most products alternates between periods when the product is sold at the every-day shelf price or *base price*, which typically changes infrequently, and promotional or sales periods when the product is offered at a temporarily discounted price. Both persistent differences in the level of base prices across stores or temporary price promotions that are not perfectly coordinated across stores can be the source of the price dispersion at a given moment in time. To distinguish between these two sources of price dispersion we use a newly developed algorithm that classifies prices as base or promoted prices. We first document the degree of the dispersion of base prices across stores at a given moment in time. As expected, we find that the degree of base price dispersion is smaller than the degree of price dispersion, although the differences are not large. For example, the ratio of the 95th to 5th percentile of base prices at the 3-digit ZIP code level is 1.220 for the median UPC, compared to 1.294 for the corresponding ratio of prices (for brands, the difference in the dispersion between base prices and prices is even smaller). Furthermore, we document that the median product is promoted once in 6.8 weeks, that the median promotional discount is 19.5 percent on the every-day price, and that 28.7 percent of all product volume is sold on promotion for the median product.

We then provide a price variance decomposition to quantify the exact contribution of the different sources of price dispersion to the overall price variance during a year. Within markets, persistent base price differences across stores account for the largest share of the price variance (46.5 percent for UPC's and 70.9 percent for brands), whereas the within-store variance of base prices during a year accounts for only about 20 percent of the base price variance. The contribution of price promotions to the within-market price variance is 35.2 percent for UPC's and 10.0 percent for brands. The relatively small contribution of price promotions is due to a strong strong EDLP (every-day low price) vs. Hi-Lo pricing pattern, whereby stores with systematically high base prices offer deeper or more frequent price discounts than stores with systematically low base prices. Thus, this EDLP vs. Hi-Lo pattern compresses the overall price dispersion across stores. Indeed, for some products the contribution of price promotions to the overall price variance is negative due to particularly pronounced EDLP vs. Hi-Lo pricing.

We examine if the heterogeneity in the degree of price dispersion across products is related to observed product attributes. We find that price dispersion is positively related to the distribution of a product, especially to the number of retail chains that carry the product at the local market level, and to the level of product penetration (the percentage of all households that purchase a product within a year). Both factors, distribution and product penetration, may reflect the level of heterogeneity in product demand.

We also assess the extent by which the variation in prices and the variation in the promotion frequency and promotion depth across stores can be attributed to market versus retail chain-specific factors. We find that prices are significantly more similar within the 81 retail chains than within the 840 3-digit ZIP codes in our sample. Furthermore, at the local market level prices are relatively more homogenous within stores that belong to the same chain. We obtain almost identical results when we examine the degree to which market and chain-specific factors can explain the differences in the promotion frequency and promotion depth of a product across stores.

To gain a more detailed understanding of the similarity in pricing patterns across stores that belong to the same chain we first conduct a principal components analysis (PCA) of the store-level price vectors (the time-series of prices in the store). PCA is a dimensionality reduction method that allows us to visualize the price vectors, which are high-dimensional objects, in a low-dimensional space. We find that a small number of principal components explains a large percentage of the variance in the original price data, such that the original price vectors can be reconstructed using a small number of the principal components. When we visualize the store-level price vectors in a two-dimensional space it is apparent that the the price vectors are substantially more similar if the corresponding store belongs to the same retailer than if the store is in the same local market. We confirm this finding by regressing the projected prices on market, chain, and market/chain fixed effects and find that chain and market/chain factors explain a much larger percentage of the variance in projected prices than market factors.

We then study price promotions and document the extent to which price promotions are coordinated across stores in the same retail chain. An analysis of retail chain promotion percentages—the fraction of stores in a local retail chain that promote a product in the same week—reveals that the observed distribution of the promotion percentages has a significantly larger fraction of values that are close to 0 or 1 (the extremes of perfect promotion coordination) than a simulated distribution of promotion percentages assuming independence in the promotion incidence across stores.

We test for promotion coordination more formally using a linear probability model of store-level promotion incidence. The covariates included in the regression are the *inside percentage* of stores in the same chain (not including the focal store itself) contemporaneously promoting a product and the *outside percentage* of stores belonging to other chains that promote the product in the same week. The model is estimated separately for each product and market. The results provide strong evidence for promotion coordination within a retail chain. Most of the inside percentage coefficients are positive with a median close to 1, indicating that the probability of a promotion in a store increases in the percentage of other stores in the same chain promoting the product. Furthermore, promotions are not only coordinated within a retail chain at the market level but also at the national level. On the other hand, the estimates of the outside percentage coefficients are close to 0, indicating that store-level promotions are conditionally independent of promotions in other retail chains. However, we find evidence for unconditional promotion dependence—the store-level promotion incidence increases in the outside percentage if we do not condition on the inside percentage in the linear probability model—which may reflect promotional allowances that are common to multiple retailers or seasonality in demand.

Finally, we test if demand side factors are able to explain the similarity in prices and promotions among stores that belong to the same retail chain. We estimate demand models for the top 2,000 brands (based on total revenue) at the store-level , which results in a total of 27.2 million estimated price and promotion coefficients. The estimates are stable across different sets of controls for market-level time fixed effects (quarter vs. month vs. week), suggesting that endogeneity bias is unlikely to affect the estimated coefficients. We also compare the OLS estimates to the estimates obtained from a Bayesian hierarchical model that yield regularized estimates of the price and promotion effects. We then use the price and promotion estimates to test if demand is more similar in stores that belong to the same chain. We find that market factors explain 14.6 percent of the variation in price elasticities across stores, chain factors explain 17.2 percent, and market/chain factors explain 47.3 percent of the variation. The results for promotion effects reveal a similar pattern—market factors explain 19.3, chain factors explain 30.5, and market/chain factors explain 56.4 percent of the variation in the estimated promotion effects. Hence, demand is more similar at the chain than at the market level, and in particular demand is much more similar within the stores in a market that belong to the same retail chain. This similarity in demand within a retail chain is consistent with the similarity of prices and promotions within a retail chain that we documented, although the estimated degree of similarity in demand is less than the estimated degree of similarity in prices and promotions. Whether this reflects unexploited price discrimination opportunities across stores, or simply measurement error in the price and promotion effect estimates is beyond the scope of this paper but an important question to be asked in future research.

The paper is organized as follows. We discuss the related literature in Section 2, and then provide an overview of the data sources in Section 3. In Section 4 we present the basic facts of price dispersion 5, and we separately document the degree of base price dispersion and the prevalence of price promotions in Section 6. In Section 6 we decompose the overall price variance into base price and promotion factors. In Section 7 we examine if price dispersion is systematically related to product characteristics and to market or retail chain factors. A detailed discussion of price similarity and promotion coordination within chains is provided in Section 8. In Section 9 we examine if demand similarity is a source of price similarly within retail chains. Section 10 concludes.

2 Literature review

Most closely related to our study is the work by Kaplan and Menzio (2015), which studies price dispersion in the U.S. retail industry using the Nielsen Homescan household panel data set. Their analysis overlaps with one part of this paper, the discussion of the basic facts of price dispersion in Section 4. The sample of product prices employed in Section 4 differs substantially from the sample in Kaplan and Menzio (2015).

First, the prices recorded in the Nielsen Homescan data that are studied by Kaplan and Menzio (2015) only cover products that are *purchased* by a household. The sample of prices for products that households choose to buy is not a random sample of prices at which products are sold, and is systematically more likely to include low prices and omit high prices from the distribution of prices at which the products are sold. This problem affects our sample of prices to a much lesser degree, because we observe the price of a product in a given store and week whenever at least one unit of the product was sold (Section 3.1). Furthermore, we impute store-level prices in weeks when the product did not sell using the most recent base (non-promoted) price. We predict base prices based on an algorithm that classifies prices into promoted and non-promoted prices.

Second, using the purchase records of the approximately 50,000 Homescan households in Kaplan and Menzio (2015) it is only possible to systematically capture the prices of a small number of products at the market (Scantrack) level. The baseline results in Kaplan and Menzio (2015) are based on a sample of products with a minimum of 25 observations at the quarter/Scantrack-market level. Although not reported in Kaplan and Menzio (2015), we replicated their sample-selection approach and found that the median number of products (UPC's) at the quarter/Scantrack level in their sample was 147. In contrast, in our sample the median number of products (UPC's) with at least 25 observations at the quarter/Scantrack level is 32,416. The difference in product coverage implies a substantial difference in the generality of our findings compared to Kaplan and Menzio (2015).

Third, and related to the previous point, using the RMS data we can study price dispersion at the 3-digit ZIP code level. The 840 3-digit ZIP codes in our sample represent substantially smaller market areas compared to the 54 Scantrack markets in Kaplan and Menzio (2015).

Fourth, because our initial goal is to document the distribution of prices at a given moment in time, we summarize price dispersion at the week level, the shortest time period at which prices are reported in the Nielsen RMS data. We also decompose the yearly variance of prices into components including temporary price promotions and the variance of regular (base) price levels over time. In contrast, Kaplan and Menzio (2015) document the distribution of prices at the quarterly level and do not distinguish between regular and promoted prices.

In principle, the Homescan data provides better coverage of all retail chains than the RMS data provided by the Kilts Center for Marketing. For example, the Homescan data cover shopping trips at Walmart (although the exact retail chain identity is not revealed in the Kilts Nielsen data), whereas Wal-Mart is not among the retail chains included in the RMS data. To measure price dispersion across stores, the sample in Kaplan and Menzio (2015) only includes the approximately one third of all transactions for which a store-identifier (`store_code_uc`) is present in the Homescan data. However, the 2004-2009 Homescan data do not include a store-identifier for transactions at Wal-Mart (Wal-Mart stopped sharing its data with Nielsen in 2001 but resumed a sharing agreement in 2011). Therefore, de facto the sample in Kaplan and Menzio (2015) does not provide better coverage of retail chains than the RMS sample used in this paper.

In contemporaneous work that has a different focus than our study, Kaplan et al. (2016) analyze the extent to which product-level price dispersion is due to persistent price-level differences across *stores*, based on a sample of 1,000 UPC's from the Nielsen RMS data.

There are several other studies that document price dispersion for products sold in retail stores (supermarkets). These studies are based on data for a small number of products or categories (e.g. Lach 2002, Dubois and Perrone 2015, and Eizenberg et al. 2017) or have limited coverage of markets or retail chains (e.g. Eden 2014).

Our work is related to a literature that documents the frequency of price adjustments and price rigidity, which has important implications for macroeconomics (Nakamura and Steinsson 2008, Nakamura and Steinsson 2013). Our work is also related to research on assortments across stores. For a sample of products in four categories, Hwang et al. (2010) find that stores that belong to the same retail chain in a market (state) carry similar assortments. This finding mirrors our results on price and promotion similarities within retail chains.

3 Data description

Our analysis is primarily based on the Nielsen RMS (Retail Measurement Services) retail scanner data that is made available for academic research purposes through a partnership between the Nielsen Company and the James M. Kilts Center for Marketing at the University of Chicago Booth School of Business.[1] We also use the Nielsen Homescan consumer panel data to select the products in our sample.

[1] http://research.chicagobooth.edu/nielsen/

3.1 Nielsen RMS retail scanner data

The Nielsen RMS retail scanner data includes information on store-level sales units and prices at the UPC (universal product code) level. The data that are available from the Kilts Center for Marketing covers close to 40,000 stores across various channels, including grocery stores, mass merchandisers, drug stores, convenience stores, and gas stations. Although the data cover a broad range of stores and retailers, the subset of the RMS data available from the Kilts Center for Marketing is neither a census nor a randomly selected sample. Some big retail chains, most notably Walmart, are not included in the data. The data have broad geographic coverage and, on average, account for between 50 and 60 percent of all market-level spending in grocery and drug stores and for one third of all spending at mass merchandisers.[2] The data contain a retail-chain ID for each store, and hence we can identify all stores that belong to the same retailer. However, the exact identity (name) of a retail chain is concealed.

The RMS scanner data record sales units and prices at the week-level, separately for all stores and UPC's. Over time, a UPC can be reassigned to a different product. Therefore, the Kilts Center for Marketing also provides a version code (`upc_ver_uc`) such that the combination of the UPC and UPC version code uniquely identifies a product.[3] A change of the brand name (description) is one of the reasons why a new UPC version is created. Sometimes, a new UPC version reflects a different spelling or abbreviation of the brand name, for example "MOUNTAIN DEW R" versus "MTN DEW R." We attempt to identify and correct all such instances. From now on, we will refer only to UPC's, with the understanding that at the most disaggregated level a product is characterized by a unique combination of a UPC and UPC version code.

In the data, a week is a seven-day period that ends on a Saturday. If the shelf price of a UPC changes during this period the quantity-weighted average over the shelf prices is recorded.

The RMS data only contain records for weeks when at least one unit of the product was sold. For the top products that sell in most weeks this is of no concern, but for smaller products the incidence of weeks without a data record is more frequent. For such products the observed sample of prices will be more likely to include relatively low, promoted price levels than regular price levels, and hence the sample will not accurately reflect the true distribution of prices. To ameliorate this problem we impute the missing prices using on an algorithm that first classifies the observed prices as either base (regular) prices or promoted prices. The algorithm distinguishes between regular and promoted prices based on the frequently observed saw-tooth pattern in a store-level time series of prices whereby prices alternate between periods with (almost) constant regular price levels and shorter periods with temporarily reduced price levels. We perform this classification separately for each store. We assume that weeks without sales are non-promoted weeks (this assumption is justified by the frequently large sales spikes observed in promoted weeks), and hence we impute the missing prices using the predictions of the current regular (non-promoted) price levels based on the price classification algorithm. Two examples of UPC,

[2] See the *Retail Scanner Dataset Manual* provided by the Kilts Center for Marketing for the Scantrack market-level data indicating the coverage of spending for the three main retail channels.

[3] A new UPC version is created when one or more of the "core" UPC attributes change. The core attributes include the product module (category) code, brand code, pack size (volume), and a multi-pack variable indicating the number of product units bundled together.

store-level price series and the corresponding predicted base prices are given in the Appendix in Figures 21 and 22.

3.2 Nielsen Homescan household panel data

We use the Nielsen Homescan household panel data to select the product sample for our analysis (see Section 3.3 below). During the sample period in this paper, 2008-2010, the Homescan panel includes more than 60,000 households. Nielsen provides sampling weights (projection factors) to make summary statistics from the data, such as total spending in retail stores, representative of the U.S. population at large. The participating households scan all purchased items after each shopping trip, and thus Homescan provides us with comprehensive data on the UPC's purchased and the corresponding prices that the households paid.

3.3 Sample selection and size distribution

The number of products in the Nielsen RMS data is large. Between 2008 and 2010 the RMS data include information on almost one million (to be exact: 967,832) products (UPC's).[4] However, a large percentage of the total sales revenue is concentrated among a relatively small number of products. To illustrate, in Figure 1 we rank all products based on total revenue between 2008 and 2010 and plot the cumulative revenue of the top N products on the y-axis. For example, the top 1,000 products account for 20.7 percent, the top 10,000 products account for 56.5, and the top 50,000 products account for 89.3 percent of the total revenue in the 2008-2010 data, respectively.

We intend to base our analysis on a product sample that is as comprehensive as possible. However, including all products in our analysis poses some problems, in particular because the small (in terms of revenue) products rarely sell. As discussed in Section 3.1 we do not observe the product price in a week when store-level sales are zero, and using our price imputation algorithm is likely to yield noisy results if price observations are only available in a small fraction of all weeks. Therefore, we use only a subset of all products in our empirical analysis. To select these products, we choose all products (UPC's) that are observed in both the Nielsen RMS scanner data and the Homescan household panel data. We then choose the top 50,000 products based on total Homescan expenditure. These 50,000 products account for 73 percent of total Homescan expenditure and 79 percent of revenue in the RMS scanner data. We select products based on Homescan expenditure instead of RMS revenue for a better assessment of how general our sample of product is. As discussed above, the Homescan data provide a comprehensive view of household purchasing behavior across all retailers, while the RMS data contain a select sample of retailers and products that are not necessarily fully representative of the whole population distribution. The top 50,000 still represent a large number of products, and some of these products are only infrequently sold. As indicated above, this presents problems for the price imputation algorithm, and we hence exclude some of the infrequently sold products from the analysis.

[4] If we define a product as a combination of UPC and UPC version (the variable `upc_ver_uc`) the number is 967,863.

Ultimately, our sample includes 47,355 products. Table 1 provides summary statistics of the product sample that covers the years 2008-2010. In each year we observe more than 8 billion prices. In total, we have 24.75 billion price observations corresponding to 517 billion dollars in sales. Including the imputed prices the number of price observations is 37.46 billion . Overall, 34 percent of all the prices at the product/store/week level used in our analysis are imputed.

3.4 Products/UPC's versus brands

We often compare results at the product (UPC) level with results if products are aggregated to the brand level. Our basis of brand-level aggregation is the brand name of one or multiple products. For example, all products with the brand description "COCA-COLA CLASSIC R" or "COCA-COLA R" belong to the same brand, Coca-Cola, while products with the description "COCA-COLA DT" belong to a different brand, Diet Coke. We aggregate products based on equivalent units such as ounces or counts. The sales volume of a brand is measured in the number of corresponding equivalent units (e.g. 10,000 ounces) and the brand price is measured as the average price per equivalent unit (e.g. 12 cents per ounce). We calculate *weighted* average prices, using the total product-level revenue summed over all stores and weeks as weights. Thus, differences in brand prices are entirely due to differences in the underlying product prices, not due to differences in the aggregation weights.

We obtain 11,279 brands using the aggregation process. Summary statistics of the brand sample are contained in Table 1. The sample includes more than 2.3 billion brand-level price observations per year, and 7.1 billion price observations in total.

3.5 Private label products

The Nielsen data contain both national brand and private label products. However, the brand description of private label products is always "CTL BR" (control brand), and hence we do not know the brand name under which the product is sold. Also, we cannot infer the brand name based on the store where the product was sold because the name of the retail chain that the store belongs to is not revealed. However, we know the product (UPC) description of a product, such as "CTL BR RS BRAN RTE" for a private label Raisin Bran product in the ready-to-eat breakfast cereal category.

In our analysis we treat all private label UPC's as the same product if they share the same product description and contain the same volume. In particular, we treat such UPC's as the same product even if the UPC's are different. The UPC's are typically different because the product is sold by different retail chains, whereas the product itself of often physically identical because it is produced by the same manufacturer that supplies multiple retailers. Even if the product is identical the packaging and specific brand name (e.g. "Kroger Raisin Bran") will differ across retailers. Hence, treating different private label UPC's as the same product is not entirely innocuous, but it is the best we can do to compare the price dispersion of national brands to the price dispersion of private label products across retail chains.

Table 1 shows the percentage of observations accounted for by private label products. Private label products account for 16.2 percent of all price observations and 18.6 percent of total revenue.

3.6 Chain and store coverage

For some stores we observe a large incidence of weeks when the products in our sample do not sell. If the price of a product is not recorded for a large percentage of l weeks the price prediction based on our imputation algorithm becomes unreliable. Hence, to avoid a large degree of measurement error, we exclude such predominantly small stores, especially convenience stores and gas stations, from the sample used in our analysis. The final data include 17,184 stores that belong to 81 different retail chains, including grocery stores, drug stores, and mass merchandisers. These stores represent more than 90 percent of total revenue in the Kilts-Nielsen RMS data.

In Table 2 we summarize the observed number of chains and stores in our sample at the DMA level and the ZIP+3 level. Note that three-digit ZIP codes and counties are the smallest geographic areas available to researchers in the Kilts Center of Marketing release of the Nielsen RMS data. At the DMA level the median number of chains is 6 and the median number of stores is 32. At the ZIP+3 level the corresponding numbers are 4 and 10, respectively. Hence, even at the smallest geographic level that we will analyze there are typically multiple retailers and stores. This is an important fact, because the measured price dispersion in a specific geography would obviously be limited if there were only a small number of stores or retailers present.

Table 2 also provides chain-level summary statistics on the geographic coverage and the number of stores of retail chains. The median number of different DMA's where a chain is present is 5, and the corresponding median number of different ZIP+3 codes is 13. The median number of stores that belong to the same retail chain is 77 at the national level, 5 at the DMA level, and 2 at the ZIP+3 level.

3.7 Product assortments

The degree of price dispersion is limited by the extent to which a product is available at different stores or retail chains. Hence, we document the distribution of product and brand availability across stores and retail chains in our sample. We classify a product (brand) as available in a specific store or retail chain if it was sold in the store or chain at least once during 2010.

Figure 2 displays the distribution of store availability for brands (column one) and products (column two). The histograms are shown separately for the top 100 (based on total revenue), top 1,000, and top 10,000 brands and products, and—at the bottom of the figure—also for all brands and products included in the analysis. The median product in the top 100 group is sold in 12,771 stores, whereas the corresponding median brand is sold in 15,985 stores, representing 93 percent of all 17,184 stores. Hence, the top products and in particular the top brands are widely available. However, even the top 100 products and brands are not consistently available across all stores, indicating differences in store-level assortment choices. Also, the top brands are more consistently available across stores than products, implying assortment differences whereby stores that carry the same brand offer the brand in different pack sizes or forms (e.g. cans versus bottles).

Whereas the top 100 and also top 1,000 products and brands are widely available, the corresponding distributions for the top 10,000 and all products and brands indicate much less consistent availability across stores. For example, the median product among all products in the sample is available only at 3,854 stores, and the median brand is available at 5,281 stores, representing 31 percent of all stores. Both the product and brand availability distributions are right-skewed with a mode close to zero. Overall, we find that assortments across stores tend to be specialized, with the exception of a relatively small number of top-selling product and brands that are typically available at a vast majority of all stores and chains.

The distributions of brand and product availability across retail chains, shown in columns three and four of Figure 2, are similar to the corresponding distributions across stores. In particular, whereas the top-selling brand and products are widely available (for example, the median brand among the top 100 is sold in 76 out of 81 retail chains), availability is much more limited among the top 10,000 and among all brands and products in the sample. Compared to the store availability distributions, however, the differences across the top and bottom groups are less pronounced. For example, the median brand among all brands in the sample is still available in the majority of retail chains (45 out of 81). In particular, the brand availability distribution for all brands exhibits a pronounced bi-modal shape, indicating a mass of brands available at most retailers and a mass of brands available at a very limited number of retail chains.

4 Price dispersion: The basic facts

We start our analysis by presenting the basic facts—the price dispersion of identical or almost identical products across stores at any given moment in time. We present the results separately for the case when a product is defined as a UPC and the case when a product is defined as a brand. UPC's are identical across stores. In the case of brands, we calculate the store-level brand price per equivalent unit as a weighted average over the prices of the individual UPC's that share the same brand name (see Section 3.4). These UPC's typically differ along pack size (15 oz, 20 oz, etc.) or form factor (bottles, cans, etc.). Different stores may offer the same brand in different UPC's. Hence, products defined as brands may not be exactly identical across stores but may exhibit some degree of product differentiation due to differences in the size or form factor of the UPC's. However, apart from the packaging, the main product is *de facto* physically identical across UPC's, and consumers are also likely to *perceive* the main product as identical across UPC's that share the exact same brand name. Therefore, a comparison of the average price at which one (equivalent) unit of a brand can be purchased across stores is meaningful, and we will refer to brands as almost identical products.

4.1 Dispersion measures

For each product j in the sample we measure the dispersion of prices in week t using two statistics. Both statistics are based on the sample of store-level prices, $\mathcal{P}_{jt} = \{p_{jst} : s \in \mathcal{S}_{jt}\}$, where $\mathcal{S}_{jt}$ is the set of all stores that sell product j in week t.

The first statistic is the standard deviation of the log of prices from the overall mean,

$$\sigma_{jt} = \sqrt{\frac{1}{N_{jt}-1} \sum_{s \in \mathcal{S}_{jt}} \Big(\log(p_{jst}) - \overline{\log(p_{jt})}\Big)^2}.$$

σ_{jt} measures the dispersion of prices as percentage differences from the geometric mean of prices across stores. The second statistic, $r_{jt}(0.05)$, is the ratio of the 95th to the 5th percentile of the price observations $\mathcal{P}_{jt}$. We calculate the statistics for each week in 2010, and then take the mean over all weeks to report the average, representative dispersion statistics σ_j and $r_j(0.05)$.

4.2 Price dispersion: UPC's

Figure 3 and Table 3 summarize the distribution of the price dispersion statistics across all 47,355 products in our sample. To account for differences in the "importance" of each product we summarize the weighted distributions of the dispersion statistics using total RMS revenue (across all stores and weeks) for each product in 2010 as weights.

The panels in the top row of Figure 3 display the weighted distributions of the product price dispersion statistics at the national level. Overall, the degree of price dispersion for identical products across stores at any given moment in time is large. The log-price standard deviation for the median product (based on the revenue-weighted distribution of σ_j) is 0.161, which roughly indicates that 95 percent of prices vary over a range from 32 percent below up to 32 percent above the average national price of the median product. The ratio of the 95th to 5th percentile of prices is 1.646 for the median product, indicating a similarly large degree of price dispersion.

The large degree of price dispersion at the national level may simply reflect systematic differences in price levels across regions due to differences in regional wage levels and the cost of living. Our main focus, however, is on documenting the price dispersion of identical products in local markets, where consumers could at least in principle buy these products at any store. Hence, to account for systematic regional price differences, we first calculate the dispersion statistics separately for each market m based on the price observations $\mathcal{P}_{jmt} = \{p_{jts} : s \in \mathcal{S}_{jmt}\}$, where $\mathcal{S}_{jmt}$ is the set of all stores that sell product j in market m in week t. We then take the weighted average over all market-level dispersion statistics for product j using the number of observations in each market as weights. We use two separate market definitions: DMA's (designated market areas) and 3-digit ZIP codes. Our sample contains 205 DMA's and 840 3-digit ZIP codes with at least one store. Also, as already discussed in Section 3.6 there are 32 stores in the median DMA and 10 stores in the median 3-digit ZIP code. A small fraction of markets (2 DMA's and 45 3-digit ZIP codes) contain only one store. We exclude these markets from the analysis, and—more generally—we do not include markets where only one store carries product j in the average of regional dispersion statistics for product j.

We summarize the market-average price dispersion statistics in the middle and lower panels of Figure 3 and in Table 3. The standard deviation of log prices for the median product is 0.110 at the DMA level and 0.099 at the 3-digit ZIP code level. Hence, even at the 3-digit ZIP code level we find a large degree of price dispersion for identical products, although systematic price level differences across markets account for up to 39 percent of the price dispersion at the national level. The ratio of the 95th to 5th percentile of prices—1.360 for the median product at the DMA level and 1.294 at the 3-digit ZIP code level—confirms this finding.

While the *overall* degree of price dispersion of identical products at any given moment in time is large, Figure 3 and Table 3 also reveal another, equally important fact: There is substantial heterogeneity in the dispersion statistics across products. Focusing on the market-level (3-digit ZIP code) results, the standard deviation of log prices ranges from from 0.021 at the 5th percentile to 0.196 at the 95th percentile of the dispersion statistics. Similarly, the 95th to 5th percentile ratio of prices ranges from 1.045 to 1.713 when comparing the 5th and 95th percentile values.

4.3 Price dispersion: Brands

We now compare the degree of price dispersion for UPC's to the analogous brand-level statistics. As discussed, brand prices are measured as the weighted average across the prices of UPC's that share the same brand name, expressed in dollars per equivalent unit. The weights are based on the total revenue of a UPC across all observations and hence neither store nor time-specific. Therefore, differences in brand prices across stores with the same brand-level assortment of UPC's are entirely due to differences in the underlying UPC-level prices, not due to differences in how these prices are weighted. However, stores are often differentiated along their product assortments (Section 3.7), and therefore brand-level prices across stores may differ even if the prices of the common UPC's are identical.[5]

Measuring price dispersion using brands instead of UPC's as product definition has pros and cons. A disadvantage of comparing brand prices across stores is that we compare products that are not exactly identical but may differ in pack size or form. On the other hand, when comparing brand prices instead of UPC-level prices we directly focus on a price comparison of the main product (content) that is sold, irrespective of pack size or form differences across stores that may only be of secondary importance to consumers.

The distributions of the brand price dispersion statistics are summarized in Figure 4 and Table 3. We generally find a larger degree of price dispersion at the brand level compared to the UPC level. At the national level, the standard deviation of log brand prices is 0.175, compared to 0.161 when measured using UPC-level prices. At the 3-digit ZIP code level the corresponding numbers are 0.129 (brand-level) and 0.099 (UPC-level), respectively. Similarly, the 95th to 5th percentile ratio of brand prices at the 3-digit ZIP code level is 1.433, whereas the UPC price ratio is 1.294.

The large heterogeneity in the dispersion statistics across products that we documented for UPC's is also evident for products defined as brands. For example, at the 3-digit ZIP code level the standard deviation of log brand prices is 0.060 at the 5th percentile, compared to 0.239 at the 95th percentile of the dispersion statistics.

4.4 Comparison to Kaplan and Menzio (2015)

Our results are not directly comparable to Kaplan and Menzio (2015) because their work is based on different data and a substantially smaller number of products, as we already discussed in Section 2. Also, they measure price dispersion using the standard deviation of prices normalized relative to the market-average price level, $p_{jst}/\bar{p}_{jmt}$, whereas our main price dispersion measure

[5]Suppose store A carries 12 oz Cheerios at $3.13 and 18 oz Cheerios at $3.58, while store B carries 18 oz Cheerios at $3.58 and 21 oz Cheerios at $3.96. If all UPC's have equal weights, then the average brand-level price of the Cheerios in store A is $3.355 and the average price in store B is $3.77.

is the standard deviation of log-prices. However, as expected, the different dispersion statistics yield almost identical dispersion measures (see the sensitivity analysis in Appendix C.1) and hence they are not a source of differences in the results.

Kaplan and Menzio (2015) report that the standard deviation of normalized prices for the mean UPC at the Scantrack/quarter level is 0.19. In our data, the corresponding standard deviation is 0.10 for the median UPC at the 3-digit ZIP code/week level and 0.12 at the the Scantrack/week level. Hence, the price dispersion of identical products at a given moment in time is substantially smaller than the dispersion level that Kaplan and Menzio report at the quarter level for a small product sample. Comparable brand-level results are not reported in Kaplan and Menzio (2015).[6]

5 Base prices and promotions

For many products, prices alternate over time between periods when the product is sold at the base price (regular or every-day shelf price) and periods when the product is promoted and offered at a discount over the base price. Base prices change only infrequently over time. This pricing pattern is visible in a typical time series plot of prices, and it reflects the institutional practice by which prices are set in the retail industry. The Nielsen RMS data do not contain information that indicates if a price was intended as a promotion. Instead, as discussed in Section 3.1, we use an algorithm that allows us to classify prices as base prices or promoted prices.

In Section 4 above we documented a large degree of price dispersion for identical products at any given moment in time, even at the local market (DMA or 3-digit ZIP code) level. This price dispersion could be due either to differences in base prices across stores, reflecting a relatively persistent component in the dispersion of prices, or price promotions that are not coordinated across stores or retail chains. In this section we document the extent of price dispersion arising from cross-sectional base price differences and the promotion policies across the products in our sample. We also document the "importance" of price promotions, measured by the percentage of total product volume sold at a discounted price.

5.1 Base prices

To document the dispersion of base prices, $\mathcal{B}_{jt} = \{b_{jst} : s \in \mathcal{S}_{jt}\}$, we use the same approach as before in Section 4 and provide two dispersion statistics, the standard deviation of the log of base prices and the ratio of the 95th to the 5th percentile of base prices across stores. We show the results separately using UPC's as product definition in Figure 5 and brands as product definition in Figure 6. The key summary statistics and percentiles are also contained in Table 3.

Generally, the degree of base price dispersion is substantial. The standard deviation of the log of base prices is 0.133 for the median product at the national level and 0.078 at the 3-digit ZIP code level. The degree of base price dispersion is not much smaller than the degree of price dispersion. For example, the standard deviation of the log of prices at the 3-digit ZIP code level

[6] In Kaplan and Menzio (2015) a brand aggregate is obtained using a "set of products that share the same features and the same size, but may have different brands and different UPCs."

is 0.099, compared to 0.078 for the log of base prices. When using brands as the definition of a product the difference is even smaller: The standard deviation of the log of base prices is 0.118 for the median brand, compared to 0.129 for the median product. The 95th to 5th percentile statistics mirror these findings. Similar to the results in Section 4 we also find a substantial degree of heterogeneity in the base price dispersion statistics across products. For example, for brands at the 3-digit ZIP code level the standard deviation of log base prices ranges from 0.055 at the 5th percentile to 0.231 at the 95th percentile.

In summary, there is a substantial degree of dispersion in the every-day shelf prices of products, defined either as UPC's or brands, both at the national and at the local market level. In particular, the dispersion statistics for base prices are not much smaller than the corresponding statistics for prices, including promoted and every-day shelf prices. Hence, base price differences across stores are an important component in the overall dispersion of prices.

5.2 Price promotions

The Nielsen RMS data do not directly indicate if a product was promoted. Instead, we infer a price promotion from the difference between the imputed base price and the realized price. First, we define the percentage price discount, or promotion depth, as follows:

$$\delta_{jst} = \frac{b_{jst} - p_{jst}}{b_{jst}}.$$

Then we classify the product as promoted if the percentage price discount is at least as large as some threshold $\bar{\delta}$. The indicator variable $D_{jst} = \mathbb{I}\{\delta_{jst} \geq \bar{\delta}\}$ captures promotional events, such that product j is promoted if and only if $D_{jst} = 1$.

We conducted the price dispersion analysis in Section 4 using data from 2010. Here we extend the sample period to 2008-2010 to reduce measurement error, in particular in the promotion frequency statistic discussed below.

Choice of promotion threshold $\bar{\delta}$

Assuming that every event when the price of a product is strictly less than the base price is a price promotion, we could define the promotion indicator as $D_{jst} = \mathbb{I}\{\delta_{jst} > 0\}$. However, it is unlikely that any brand or category manager designs a price promotion that offers only a negligible price discount. Hence, to find a suitable threshold value $\bar{\delta}$ to use in our analysis we examine the distribution of the percentage price discounts, δ_{jst}, pooled across all products, stores, and weeks when $p_{jst} < b_{jst} \Leftrightarrow \delta_{jst} > 0$. This distribution is shown in the top right panel of Figure 7 and summarized in Table 4. The median percentage price discount across all events is 17.7 percent. There are instances of small percentage price discounts, but the overall incidence of such events is small. For example, in only slightly less than 10 percent of all events the price discount is less than 5 percent, $0 < \delta_{jst} < 0.05$. Again, it is implausible that these observations represent a planned price promotion. Rather, such observations are likely due to measurement error in either the price or base price. Such measurement error can arise due to differences between the promotional calendar in a store and the Nielsen RMS definition of a week. For example, suppose a product was offered at a 20 percent price discount during a two week period starting on a Monday and ending on Sunday, May 30. Because a week in the RMS data ends on a Saturday, the RMS week that begins on May 30 and ends on Saturday, June 4 will include one day when the product was offered at the 20 percent price discount and six days when the product was sold at the regular (base) price. The data report the average price over these seven days, which is an average over the promoted and non-promoted prices. The inferred percentage price discount, δ_{jst}, is likely to be small in this example, and it will not accurately represent the promotional price discount. In order to ameliorate measurement error we use a promotion threshold of $\bar{\delta} = 0.05$ in our analysis. Even then some measurement error will remain, and—if measurement error is predominantly due to the averaging problem that we discussed—the consequence will be that we understate both the true promotional discounts and promotion frequency.

Promotion frequency and promotion depth

We measure the promotion frequency of product j in store s as

$$\pi_{js} = \frac{1}{N_{js}} \sum_{t \in \mathcal{T}_{js}} D_{jst},$$

where $\mathcal{T}_{js}$ is the set of all observation periods (weeks) in 2008-2010 when product j was sold in store s, and N_{js} is the corresponding number of observations. Using π_{js} we then calculate a product-level promotion frequency statistic, π_j, by taking a weighted average over π_{js} across all stores using the number of store-level observations, N_{js}, as weights.[7] We also measure the heterogeneity of the promotion frequency across stores based on the difference between the 95th and the 5th percentile among all π_{js} observations (weighted by N_{js}) across stores.

Correspondingly, we measure the average promotional discount or promotion depth of product j in store s across all weeks when the product was promoted, $D_{jst} = 1$:

$$\delta_{js} = \frac{1}{N_{js}^D} \sum_{t \in \mathcal{T}_{js}, D_{jst}=1} \delta_{jst},$$

where N_{js}^D is the number of promotion events. We obtain a product-level promotion depth statistics, δ_j, based on a weighted average of δ_{js} across all stores using the number of observations, N_{js}^D, as weights. Also, as in the case of promotion frequency, we measure the heterogeneity in the promotion depth across stores using the difference between the 95th and the 5th percentile of δ_{js} across stores (the distribution of δ_{js} is weighted using N_{js}^D).

The left panel in the middle row of Figure 7 displays the weighted distribution (using total product revenue) of the promotion frequency, π_j, across products (see also Table 4 for some key summary statistics). The average promotion frequency for the median product is 0.147, which implies that the product is promoted about once in 6.8 weeks on average. The average promotion frequency varies strongly across products, ranging from 0.011 (once in 91 weeks) at

[7] This is equivalent to calculating π_j based on all D_{jst} observations, pooled across stores and weeks.

the 5th percentile to 0.370 (once in 2.7 weeks) at the 95th percentile level. The left panel in the middle row of Figure 7 shows the corresponding differences between the 95th and the 5th percentile of π_{js} across stores s. For the median product this difference is 0.314, compared to the average promotion frequency level of 0.147. Hence, the average promotion frequency masks large differences in the promotion frequencies across stores.

The bottom row of Figure 7 shows the distributions of the average promotion depth, δ_j, and the across-store heterogeneity in δ_{js}. Note that—unlike in the top left panel of Figure 7—the promotion depth statistics are calculated based on promotional events, defined as $D_{jst} = \mathbb{I}\{\delta_{jst} \geq 0.05\} = 1$. The average promotional discount for the median product is 19.5 percent on the base price, and the whole distribution across products ranges from 10.2 percent at the 5th percentile to 31.5 percent at the 95th percentile. Much as in the case of the promotion frequencies, the promotion depth differs strongly across stores. For example, for the median product the difference between the 95th and 5th percentiles of δ_{js} across stores is 18.0 points.

5.3 Volume sold on promotion

We end this section by documenting the "importance" of promotions, measured using the percentage of product volume that is sold during a promotional period:

$$\nu_{js} = \frac{\sum_{t \in \mathcal{T}_{js}, D_{jst}=1} q_{jst}}{\sum_{t \in \mathcal{T}_{js}} q_{jst}}.$$

Here, q_{jst} is the number of product j units sold in store s in week t. To calculate a corresponding product-level statistic, ν_j, we take a weighted average of ν_{js} over all stores s, with weights N_{js}, the number of observations for store s. We also document the ratio of the average product volume sold during a promotion relative to the average product volume when the product was not promoted:

$$L_{js} = \frac{\frac{\sum_{t \in \mathcal{T}_{js}, D_{jst}=1} q_{jst}}{N_{js}^D}}{\frac{\sum_{t \in \mathcal{T}_{js}, D_{jst}=0} q_{jst}}{N_{js} - N_{js}^D}}.$$

In the retail industry and in brand management L_{js} is called a *lift factor* or *promotion multiplier*.[8] To obtain a product-level lift factor L_j we aggregate over L_{js} using the same process that we used to aggregate the volume percentages above.

The top left panel in Figure 8 displays the weighted distribution of the percentage volume sold on promotion across products (as always, we use total product revenue weights), and Table 4 provide detailed summary statistics. The median percentage of volume sold on promotion is 28.7 percent, and ranges from 1.8 percent to 61.4 percent at the 5th and 95th percentiles. The bottom left panel displays the corresponding distribution of the promotion multipliers, with a median of 3.04 and a range from 1.34 to 9.22. Hence, as expected, the volume sold on promotion is disproportionately high (relative to the overall incidence of promotions), and units sales spike

[8] Alternatively, we could calculate L_{js} using the predicted volume in the absence of a promotion in the denominator, based on a demand model or a weighted average of the observed non-promoted volume.

relative to the non-promoted volume when a product is promoted. As can be seen in the right panels in Figure 8, the product-level averages mask a large degree of heterogeneity in ν_{js} and L_{js} across stores for most products.

6 Price variance decomposition

In Section 5 we documented a substantial degree of dispersion of the base prices of UPC's and brands. Furthermore, we found that many UPC's are frequently promoted, with an average promotional discount of 19.5 percent for the median product. In this section we quantify the relative contribution of these and other related sources to the overall level of price dispersion. In particular, we decompose the overall price variance of a product into components that capture (i) price differences across markets, (ii) persistent price or base price differences across stores at the market level, (iii) within-store price or base price variation over time, and (iv) price variation due to promotions.

We calculate the price variance decomposition separately for each product (UPC or brand) j, and we drop the j subscript to simplify the notation. $\mathcal{M}$ is the set of all markets, and $\mathcal{S}$ is the set of all stores in the whole sample. For each store s we observe prices in periods $t \in \mathcal{T}_s$. N_s is the number of observations for stores s, N_m is the total number of observations (across stores and time periods) in market m, and N is the total number of observations across all markets. $\bar{p}$ is defined as the overall (national) average price, $\bar{p}_m$ is the average price in market m, and $\bar{p}_s$ is the average price in stores s. We calculate the overall price variance of a product using the sample $\mathcal{P} = \{p_{st} : s \in \mathcal{S}, t \in \mathcal{T}_s\}$:

$$\text{var}(p_{st}) = \frac{1}{N} \sum_{s \in \mathcal{S}} \sum_{t \in \mathcal{T}_s} (p_{st} - \bar{p})^2.$$

All results below are derived in detail in Appendix A.

6.1 Basic decomposition

We first provide a decomposition that does not distinguish between base and promoted prices. $\text{var}(\bar{p}_m)$ is the variance of average market-level prices across markets, $\text{var}(\bar{p}_s|m)$ is the within-market variance of average store-level prices, and $\text{var}(p_{st}|s)$ is the within-store variance of prices over time. $\text{var}(p_m)$ and $\text{var}(\bar{p}_s|m)$ are calculated as weighted averages, using the number of observations in each market and the number of observations for each store as weights (see Appendix A). We can decompose the overall price variance into the variance of (average) prices across markets, the within-market variance of prices across stores, and the within-store variance of prices over time :

$$\begin{aligned} \text{var}(p_{st}) &= \text{var}(\bar{p}_m) && \text{(across-market)} \\ &+ \tfrac{1}{N} \textstyle\sum_{m \in \mathcal{M}} N_m \text{var}(\bar{p}_s|m) && \text{(across-store)} \\ &+ \tfrac{1}{N} \textstyle\sum_{s \in \mathcal{S}} N_s \text{var}(p_{st}|s). && \text{(within-store)} \end{aligned} \tag{1}$$

This decomposition is based on weighted averages of the within-market and within-store variances, using the number of observations as weights.

We report the mean of the variance components (weighted by total product revenue) in Table 5, calculated based on all price observations in 2010. The percentage of the price variance due to price-level differences across markets, defined as ZIP+3 areas, is 32.7 percent for UPC's and 29.7 percent for brands. Hence, at least about 68 percent of the national variance in prices is due to price variation at the local, ZIP+3 level. For UPC's, 27.0 percent of the overall price variance is due to price-level differences across stores, and 40.3 percent is due to within-store price variation during 2010. For brands, we find that the across-store price differences explain a substantially higher percentage of the overall price variation—42.3 percent—whereas 28.0 percent of the price variation is due to the within-store price variation.

6.2 Decomposition into base prices and promotions

We now provide a more detailed decomposition that distinguishes between the contribution of base price differences, both across stores and within stores over time, and the contribution of price promotions to the overall variance of prices:

$$\begin{aligned}
\text{var}(p_{st}) \quad & = \text{var}(\bar{p}_m) && \text{(across-market)} \\
& + \tfrac{1}{N} \textstyle\sum_{m \in \mathcal{M}} N_m \text{var}(\bar{b}_s | m) && \text{(across-store base price var.)} \\
& + \tfrac{1}{N} \textstyle\sum_{s \in \mathcal{S}} N_s \text{var}(b_{st} | s) && \text{(within-store base price var.)} \\
& + \tfrac{1}{N} \textstyle\sum_{m \in \mathcal{M}} N_m \text{var}(b_{st} - p_{st} | m) && \text{(promotional discount var.)} \\
& - 2 \tfrac{1}{N} \textstyle\sum_{m \in \mathcal{M}} N_m \text{cov}(b_{st} - p_{st}, b_{st} | m) && \text{(EDLP vs. Hi-Lo adjustment)}
\end{aligned} \quad (2)$$

The first component captures the variance of price levels across markets. The second component is the within-market variance of base prices, indicating persistent base price differences across stores, whereas the third component is the within-store variance in base prices over time.

The last two terms capture the contribution of price promotions to the overall price variance. The fourth term is the variance of promotional price discounts, $b_{st} - p_{st}$. This variance is zero if and only the promotional discounts are identical across all stores s and periods t, which is equivalent to $b_{st} = p_{st}$ for all observations because at least sometimes the product is sold at the base price. In other words, the variance of promotional price discounts can only be zero in the absence of any price promotions.

The last term in (2) is negative if the observation-weighted average of the market-level covariances between the promotional price discounts and the store-level base price is positive. Such a positive correlation indicates an EDLP (everyday low price) vs. Hi-Lo pricing pattern at the product level: Stores with above average base prices offer larger promotional price discounts than stores with below average base prices. Correspondingly, we call the last term in the decomposition (2) the "EDLP vs. Hi-Lo adjustment."

Price promotions increase the overall price variance in the absence of an EDLP vs. Hi-Lo pricing pattern. For example, in the special case when all base prices in a market are identical,

$b_{st} \equiv \bar{b}_m$, the EDLP vs. Hi-Lo adjustment is zero and the combined contribution of the price promotion terms in (2) to the price variance is positive. However, if there is EDLP vs. Hi-Lo pricing, then the adjustment term will be negative and the overall variance in prices will be reduced. Below we present an example to demonstrate that price promotions may even *decrease* the overall price variance if there is a pronounced EDLP vs. Hi-Lo pricing pattern.

Example: Price promotions may decrease the overall price variance For notational simplicity we use expectations instead of sample averages, and we focus only on price variation within one market and thus drop the subscript m. Assume that base prices in each store s are constant over time, $b_{st} \equiv \bar{b}_s$, and uniformly distributed around the mean base price $\bar{b}$ on the interval $[\bar{b} - \nu, \bar{b} + \nu]$. Suppose that only stores with above average base prices, $b_{st} = \bar{b}_s > \bar{b}$, promote the product, and that the promoted price is always $p_{st} = \bar{b}$. All stores with base prices $b_{st} = \bar{b}_s \leq \bar{b}$ always sell the product at the base price, $p_{st} = b_{st}$. The incidence of promotions is constant, $\pi \equiv \Pr\{D_{st} = 1 | b_{st} > \bar{b}\}$. There is a continuum of stores with mass 1, and promotions are independent across stores and across time periods. As shown in Appendix A.3, the EDLP vs. Hi-Lo adjustment factor is positive (if $\pi > 0$) and strictly increasing in π, and the variance of prices is strictly decreasing in the frequency of promotions, π:

$$\text{var}(p_{st}) = v^2 \left(\frac{1}{3} - \frac{\pi}{6} - \frac{\pi^2}{16} \right).$$

Results The results in Table 5 indicate that the within-market variance of mean base prices across stores accounts for 31.3 percent of the overall price variance for UPC's and for 49.9 percent of the overall price variance for brands. Hence, for UPC's the within-market variance of the average every-day shelf price is almost as large as the variance in average prices across markets, and for brands the within-market mean base price variance is even larger than the variance in mean prices across markets. In other words, at the local, 3-digit ZIP code level the persistent variation in base prices is comparable to or exceeds the price variation across 3-digit ZIP codes in the U.S. On the other hand, the within-store base price variance over the course of a year accounts for a comparably small percentage of the overall price variance—12.3 percent for UPC's and 13.4 percent for brands. Hence, we confirm that store-level base prices are highly persistent over the course of a year.

Of particular interest if the role of price promotions. The promotional price discount component in the variance decomposition (2) is large and positive—36.0 percent for UPC's and 29.9 percent for brands. However, the EDLP vs. Hi-Lo adjustment term is negative, -12.3 percent for UPC's and -22.9 percent for brands. Hence, there is strong evidence for EDLP vs. Hi-Lo pricing at the product-level, such that stores with above average base prices offer deeper promotional discounts than stores with low base prices. The EDLP vs. Hi-Lo pricing pattern reduces the contribution of price promotions to the overall price variance. Table 5 also shows the total contribution of promotions to the overall price variance, including the variance of the promotional price discounts and the EDLP vs. Hi-Lo component. The total contribution of promotions to the overall price variance is 23.7 percent for UPC's and 7.0 percent for brands.

The last two columns in Table 5 express the contributions of the across-store and within-store base price variances and the contribution of price promotions as a percentage of the *within-market* variance of prices. These results highlight the importance of persistent base price differences across stores (46.5 percent of the within-market variance for UPC's and 70.9 percent for brands) relative to the total contribution of price promotions (35.2 percent for UPC's and 10 percent for brands).

More detailed statistics, including the key percentiles of the price variance components, are provided in Table 12 in the Appendix. In particular, Table 12 reports the percentage of products for which price promotions *decrease* the overall price variance: 3.6 percent for UPC's and 30.8 percent for brands.

7 What factors explain price dispersion?

7.1 Product-level price dispersion and product characteristics

In Section 4 we showed that there is a substantial degree of price dispersion at the 3-digit ZIP code market level and we found much heterogeneity in the degree of price dispersion across products. Our goal is to understand if this across-product heterogeneity in price dispersion is related to observed product characteristics. In particular, we focus on two measures of product size or "importance": (i) The purchase volume of a product, measured in total revenue dollars in a year, and (ii) the product penetration rate, measured as the percentage of all households that buy a product at least once within a year. To capture these two "importance" measures we first rank all UPC's in the sample according to total product revenue or the product penetration rate in in 2010, and then we create dummy variables that indicate if a product is among the top 100, 101-250, 251-500, 501-1000, 1001-5000, 5001-10000, or 10001-20000 products. We also estimate if the level of price dispersion varies across national brands and private-label products, and we examine if the dispersion levels cluster, i.e. are more similar within a product category than across product categories. We measure price dispersion based on the standard deviation of the log of prices for a UPC at the market (3-digit ZIP code) level. Due to the large number of UPC/market combinations we base our analysis on a random sample of one million observations, which is sufficient to achieve a high degree of statistical power.

In Table 6, specification (1), we show the results from a regression of the price dispersion measure on the purchase volume rank indicators. The estimates indicate a larger degree of price dispersion for the top-selling products, and the differences in the degree of price dispersion across the product size groups are substantial. For example, the difference in the log-price standard deviation among the top thousand products compared to the top 10-20 thousand products is between 0.017 and 0.026 (recall that the standard deviation of log-prices for the median product is 0.099 at the 3-digit ZIP code level).

High-selling products are likely available at a larger number of stores and chains. Hence, to account for differences in the distribution of the products, we include the log of the number of stores and retail chains that carry a UPC at the local market level in specification (2). The monotonically decreasing pattern in the estimated degree of price dispersion associated with the rank indicators largely disappears. On the other hand, the number of stores and especially the number of retail chains carrying a product are positively related to the degree of price dispersion. This pattern is unchanged when we also control for the market size using the log of the local population size in specification (3). One interpretation is that controlling for market size, "large" products with inherently high demand are carried by a larger number of retail chains that employ heterogeneous pricing policies, possibly reflecting heterogeneous consumer demand. Note that the R^2 in regressions (2) and (3) is 0.372, compared to 0.030 in specification (1). Hence, the distribution variables—the number of stores and retail chains that carry a product in a local market—explain a substantial degree of the variation in price dispersion across products.[9]

We repeat this analysis using rank indicators based on the level of product penetration. The results are reported in Table 7. We find a strong, positive relationship between penetration and price dispersion (specification 1). Unlike in the results above that are based on revenue rank indicators, this pattern persists even controlling for the number of stores and chains that carry the product and the market size (specifications 2 and 3). Possibly, a larger degree of product penetration among households is associated with a larger degree of heterogeneity in demand, and this may be the source of a larger degree of local price dispersion.

In contrast to our findings, Kaplan and Menzio (2015) document a hump-shaped relationship between the log of total product expenditure and product-level price dispersion. Their analysis does not account for market size or distribution.

In Tables 6 and 7, specification (4), we add a dummy variable that indicates if a UPC is a private-label product. Generally, private-label products have a larger degree of price dispersion than national brands—the difference in the log-price standard deviation is between 0.012 and 0.014.

Finally, we examine if the degree of price dispersion systematically differs across product categories. We employ two definitions of product category: Product group (e.g. "DETERGENTS"), and product module (e.g. "DETERGENTS - HEAVY DUTY - LIQUID"), which represents more narrowly defined sub-categories of a product group. The results (specifications 5 and 6) reveal only a small increase in the R^2 of the regressions. Hence, the across-product heterogeneity in price dispersion cannot be explained by systematic differences in the degree of price dispersion across product categories.

7.2 Price dispersion across stores and market vs. retail chain factors

We now assess how much of the overall variance in prices, as well as the variation in the promotion frequency and promotion depth for a product across stores can be attributed to market versus retail chain-specific factors.

We know from the analysis in Sections 4 and 6 that there is a substantial degree of price variation at the local market (3-digit ZIP code) level. We confirm this finding by regressing the UPC prices, p_{jst}, on dummy variables for all 3-digit ZIP codes in our sample. We estimate

[9] If we control for the distribution factors using a completely flexible approach with a separate indicator variable for each value of the number of stores or chains that carry a product the R^2 increases modestly, from 0.372 to 0.405.

the regression models separately for each product j and week t in 2010, and then we take the average over all weeks to obtain a single R^2 value for each product. The revenue-weighted distribution of the corresponding product-level R^2 values (the percentage of the price variance explained by market-specific factors) is displayed in the top row of Figure 9, and Table 13 in the Appendix contains details of the results. For the median product, somewhat less than half—46.5 percent—of the price variance can be attributed local market factors.[10] When we regress p_{jst} on chain dummies we find that the distribution of R^2 values shifts to the right, with a median of 69.9 percent. Hence, prices are substantially more homogenous within the 81 different retail chains than within the 840 different 3-digit ZIP code areas in our sample. When we regress the prices on market/chain dummy variables the distribution shifts even further to the right, with a median of 88.1 percent—for half of all products, chain/market-specific factors explain at least 88.1 percent of the overall variance in prices. Indeed, for 90 percent of all products in the sample the R^2 values are above 26.6 percent for the market-level regressions and above 70.7 percent for the market/chain-level regressions,[11] indicating that retail chain-specific factors can explain a substantial fraction of the overall price variance at the local, 3-digit ZIP code level.

We perform a similar analysis for the store-level promotion frequency and promotion depth statistics, π_{js} and δ_{js}. The results are shown in the middle row (promotion frequency) and bottom row (promotion depth) of Figure 9. The results mirror the findings that we obtained for prices. Comparing the median R^2 statistics based on regressions of promotion frequency on the different indicator variables, the values increase from 36.1 percent for market dummies to 62.3 percent for chain dummies and to 80.0 percent for market/chain dummies. For promotion depth, the corresponding R^2 values are 38.0 percent (market), 58.9 percent (chain), and 80.7 percent (market/chain). Hence, both prices and promotions are more homogenous in the 81 retail chains compared to the 840 3-digit ZIP codes in our sample, and prices and promotions are also relatively more homogenous in stores belonging to the same retail chain at the local market level.

Adding to the results above in Section 7.1, we note that the across-product heterogeneity of the *within-chain* price dispersion at the 3-digit ZIP code level is largely unrelated to product "importance," measured either based on total product revenue or the penetration rate. However, products that are carried by a larger number of stores of the same retail chain exhibit a somewhat higher degree of price dispersion. Also, there is almost no difference in the degree of price dispersion between private-label and national brands.

8 Price similarity and promotion coordination within chains

We documented in Section 7.2 that chain identity explains a large fraction of the variance in prices, promotion frequency, and promotion depth across products. These results hold with and

[10] The R^2 values from the market-dummy regressions are comparable to the across-market price variance component in the variance decomposition (1). The values are not identical, however, because the variance decomposition in Section 6 was performed using all weeks in 2010, whereas the product-level R^2 values in this section are obtained by averaging over the R^2 values from week-level regressions.

[11] The details of the percentiles of the distributions of the R^2 values are shown in Table 13 in the Appendix.

without controlling for market fixed effects. In this section we provide a closer examination of the similarity in product pricing and promotions across stores that belong to the same retail chain.

8.1 Similarity in pricing patterns

We first examine the similarity of pricing patterns across stores based on the whole time series of prices observed in a store. For any product j, we observe the vectors of prices, $\boldsymbol{p}_s = (p_{s1}, \ldots, p_{sT})$ for each store $s \in \mathcal{S}$ (we suppress the index j for notational simplicity). Our sample of prices for product j then consists of $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_S$. Our goal is to visualize the price vectors $\boldsymbol{p}_s$. However, this is not possible because of the dimensionality of $\boldsymbol{p}_s$—each price vector has components corresponding to each sample period $t = 1, \ldots, T$, typically the whole period from 2008-2010. Instead, we conduct a principal components analysis (PCA) of the store-level price vectors. PCA is an unsupervised dimensionality reduction technique that allows us to represent each $\boldsymbol{p}_s$ in a low-dimensional space while maintaining as much of the original information (variance) contained in $\boldsymbol{p}_s$ as possible.[12]

For a given dimension $K < T$, the principal components algorithm finds the optimal linear reconstruction of the original price data,

$$\tilde{\boldsymbol{p}}_s \approx \hat{\boldsymbol{p}}_s = \boldsymbol{V}_K \boldsymbol{\lambda}_s.$$

Here, $\tilde{\boldsymbol{p}}_s = \boldsymbol{p}_s - \bar{\boldsymbol{p}}$ is the centered vector prices in store s, relative to the overall (national) average of prices. $\boldsymbol{V}_K$ is a $T \times K$ matrix with columns $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K$. These columns, called the principal components of the data, are orthogonal and have unit length, $\|\boldsymbol{v}_k\| = 1$. The principal components $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K$ form a basis for a K-dimensional linear subspace of $\mathbb{R}^T$. $\hat{\boldsymbol{p}}_s$ is the projection of $\tilde{\boldsymbol{p}}_s$ into this space, such that $\boldsymbol{\lambda}_s = (\boldsymbol{V}'_K \boldsymbol{V}_K)^{-1} \boldsymbol{V}'_K \tilde{\boldsymbol{p}}_s = \boldsymbol{V}'_K \tilde{\boldsymbol{p}}_s$. Expressed within the coordinate system defined by the basis $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K$, reconstruction of the price vector $\tilde{\boldsymbol{p}}_s$ in K dimensions is given by $\boldsymbol{\lambda}_s = (\lambda_{1s}, \ldots, \lambda_{Ks})$. $\boldsymbol{\lambda}_s$ is the desired low-dimensional representation of the original data.

The first principal component $\boldsymbol{v}_1$ has the property that the variance of the data projected onto the linear subspace spanned by $\boldsymbol{v}_1$, $\boldsymbol{v}'_1 \tilde{\boldsymbol{p}}_1, \ldots, \boldsymbol{v}'_1 \tilde{\boldsymbol{p}}_S$, has the largest variance among all possible directions $\boldsymbol{v}$. Furthermore, the variance of the data projected onto the space spanned by $\boldsymbol{v}_2$ has the largest variance among all possible directions that are orthogonal to $\boldsymbol{v}_1$, etc. Thus, the linear reconstruction of the data provided by the principal components algorithm is optimal in the sense that the highest amount of the variance in the original data is maintained.

We perform a PCA for the top 1,000 products (UPC's) in our sample, based on total revenue rank. We only choose these top products because we need to be able to consistently observe the weekly prices, p_{st}, across stores for the analysis to be feasible. For smaller products there is a larger incidence of missing values.

The top panel in Figure 10 displays box plots of the percentage of the price variance that is explained by the first twenty principal components. Each box plot shows the distribution

[12] See, for example, Hastie et al. (2009) for a thorough introduction to principal components analysis.

(weighted by total product revenue) of these percentages across the products in our sample. The first principal component explains 20 percent of the price variance for the median product, and all the first five principal components explain at least 5 percent of the price variance. In the bottom panel of Figure 10 we display box plots of the cumulative percentage of the price variance explained by the top principal components. The top five principal components explain 53 percent and the top ten principal components explain 68 percent of the price variance for the median product. Hence, a large percentage of the information in the original store-level price vectors over the 2008-2010 period can be explained by a small number of principal components. A representation of the original, high-dimensional price data in a low-dimensional space is therefore meaningful.

In Figure 11 we present a two-dimensional representation of the store-level price vectors using the projections onto the first two principal components for the case of Tide HE Liquid Laundry Detergent (100 oz). The graph is split into six panels that contain identical gray dots representing all projected store-level price vectors. In each of the panels some of the dots are colored according to the retail chain that the corresponding store-level prices belong to. All stores that belong to the same retail chain appear in exactly one panel. However, the color labels are not mutually exclusive across the panels. For example, red dots in two different panels represent the projected prices for stores that belong to two different retail chains. The graph allows us to visually examine the similarity in prices within a chain. Figure 11 reveals a strong degree of similarity—the projected price vectors that belong to the same retail chain cluster and exhibit much less variance compared to the overall variance in projected prices. In Figure 12 we color the projected store-level price vectors according to the market (DMA) that a store belongs to. We use DMA's, not 3-digit ZIP codes, as the market definition because the large number of 3-digit ZIP codes is hard to visualize on one page. We see some clustering of prices also at the DMA level, but the similarity of prices within a DMA appears to be much smaller than the similarity of prices within a retail chain.

We present additional examples in Figure 13, including Prilosec (42 count), Pepsi (12 oz cans 12 pack), and private-label milk (2 percent, 1 gallon). In this graph we only include the store-level price vectors for a subset of all retail chains. For Prilosec and Pepsi we find a pattern that is similar to the case of Tide laundry detergent—a large degree of price similarity within retail chains and a significantly smaller degree of price similarity at the market level. The case of private-label milk is very different: there is much heterogeneity in prices both at the chain and the market level.

To summarize the price similarity patterns across all products in our sample we project the price vectors onto the first principal components and then regress (for each product separately) the projections on DMA, chain, and chain-DMA fixed effects. Figure 14 presents histograms of the percentage of the projected price variance (R^2) explained by these three factors. The results are shown separately for each of the top six principal components. The results indicate that there is a consistently higher degree of price similarity at the chain than at the DMA level, and a moderately higher degree of price similarity at the chain-DMA level compared to the chain level. For example, for the first principal component we find a median R^2 of 0.291 at the DMA level, compared to an R^2 at the chain level of 0.844 and an R^2 at the chain-DMA level of 0.913. The R^2 levels are generally smaller for the lower-ranked principal components, but the general pattern persists.

8.2 Promotion coordination

We now focus on price promotions in particular, and examine if promotions are coordinated, in the sense that the same product is systematically promoted at the same time among the stores in a retail chain or in a market. We measure promotion incidence using the indicator $D_{jst} \in \{0,1\}$, such that $D_{jst} = 1$ if product j is promoted in stores s in week t.

Overview of promotion coordination

To provide an overview of promotion coordination we first summarize the percentage of all stores in a retail chain that promote a specific product during week t. Let $\mathcal{S}_{jcmt}$ be the set of stores that belong to the retail chain c in market m and carry product j in week t. We then calculate the chain/market level *promotion percentage* for product j:

$$\phi_{jcmt} = \frac{\sum_{s \in \mathcal{S}_{jcmt}} D_{jrs}}{|\mathcal{S}_{jcmt}|}.$$

The graphs in the top row of Figure 15 display histograms of the promotion percentages ϕ_{jcmt}, pooled over all products, chains, markets, and time periods between 2008 and 2010 (see Table 8 for the detailed summary statistics). The distributions are weighted using total product revenue weights. We display the promotion percentage distributions conditional on $\phi_{jcmt} > 0$, i.e. weeks when at least one store in chain $c$and market m promotes the product, to avoid that the histograms are dominated by large mass points at 0. The percentage of observations when none of the stores promoted the product, $\phi_{jcmt} = 0$, is indicated separately at the bottom of each graph. We define markets as DMA's to ensure that the retail chains have a larger number of stores in a market compared to 3-digit ZIP codes as market definition (see Table 2 for summary statistics on the number of stores per retail chain at the DMA and ZIP+3 level). With a larger number of stores the promotion percentage ϕ_{jcmt} can take a larger number of values compared to a small number of stores. In the extreme case where a chain has only one local store, the promotion percentage is always 0 or 1, indicating perfect promotion coordination. For the same reason, we summarize the promotion percentage distributions only for observations when the retail chain c carries the product in at least five stores in the local market.

The top row in Figure 15 shows the promotion percentage distributions for all products (left panel) and the top 1,000 products as measured by annual, national product revenue (right panel). Both histograms reveal a large mass point at 1, indicating perfect promotion coordination. The promotion percentages, conditional on $\phi_{jcmt} > 0$, are overall larger among the top 1,000 products, with a median of 0.548 compared to a median of 0.41 for all products. This indicates a larger degree of promotion coordination among the top-selling products, although the percentage of $\phi_{jcmt} = 0$ observations is smaller for the top 1,000 products: 55.06 percent versus 62.18 percent among all products. However, the latter finding may simply reflect an overall higher promotion frequency among the top 1,000 products, not a smaller degree of coordination on weeks when none of the stores in the chain promote the product.

Although suggestive of promotion coordination, in particular due to the mass of promotion percentages at or close to 1, the overall extent of promotion coordination conveyed by Figure 15 is difficult to assess without a comparison to a baseline where promotions are not coordinated across stores. To provide such a baseline we simulate a data set assuming that promotions are chosen independently at the store level. For each product j and stores s we calculate the promotion frequency π_{js} using the 2008-2010 data, as in Section 5.2. For each store s and week t we then draw a promotion indicator $\tilde{D}_{jst}$ from a Bernoulli distribution with success probability π_{js}. We show the distributions of promotion percentages for the simulated data in the second row of Figure 15 and in Table 8.[13] The histograms indicate a much smaller degree of promotion coordination compared to the observed promotion percentages. The median promotion percentage among all products in the simulated data is 0.200, compared to 0.41 in the actual data, and the corresponding percentages among the top 1,000 products are 0.279 in the simulated data versus 0.548 in the observed data. Indeed, the displayed distributions understate the difference between the simulated and the original data because they are conditional on $\phi_{jcmt} > 0$, and hence do not reveal the large difference in observations where none of the stores in a chain promotes a product. In the observed data, $\phi_{jcmt} = 0$ in 62.18 percent of all observations, compared to 28.27 percent of all observations in the simulated data.

The distributions the top rows of Figure 15 are based on observations at the chain/market level in a given week, conditional on at least one store in the chain promoting product j. This leads to an asymmetry between observations with highly coordinated promotions and observations with a small number of stores promoting a product. For example, suppose that all promotions were perfectly coordinated within a retail chain such that $\phi_{jcmt} = 1$ in the week when the promotion is held. Suppose there were some small differences in the timing of the end date of the promotion across stores in the chain, such that a small number of stores would still offer the promotion for a few days in week $t + 1$. Then each perfectly coordinated promotion observation, $\phi_{jcmt} = 1$, would have an associated observation with a small, positive $\phi_{jcm,t+1}$ value, suggesting that perfectly coordinated promotion events were as frequent as almost completely uncoordinated promotion events. Hence, as an alternative summary of promotion coordination we associate each store-level promotion event, $D_{jst} = 1$, with the corresponding promotion percentage, ϕ_{jcmt}, and display the distribution of the promotion percentages based on all store level observations such that $D_{jst} = 1$. This is equivalent to displaying the distribution of ϕ_{jcmt} weighted by the number of stores that promote product j in chain c and market m in week t. The results are shown in the third row of Figure 15 (see Table 8 for the detailed numbers), and strongly indicate that store-level promotions are coordinated at the chain-market level. Furthermore, the differences between the actual and simulated data, shown in the bottom row of Figure 15, are large.

[13] To make the observed and simulated distributions comparable we only display the simulated promotion percentages corresponding to observations where a chain sells product j in at least five stores in the local market.

Estimating the dependence in promotion incidence

We now test more formally if the incidence of promotions is dependent across stores. For any store s in market (DMA) m we define the *inside* promotion percentage of all stores in market m that belong to the same chain as store s and promote the product in period t:

$$I_{jst} = \frac{1}{|\mathcal{S}_{sm}| - 1} \sum_{r \in \mathcal{S}_{sm},\, r \neq s} D_{jrt}.$$

Here, $\mathcal{S}_{sm}$ is the set of all stores in market m that belong to the same retail chain as store s. Note that the inside percentage is calculated based on the promotion incidence of all other stores in the local retail chain, not including store s. Vice versa, the *outside* promotion percentage of stores that belong to any other chain in market m and promote the product in week t is

$$O_{jst} = \frac{1}{|\bar{\mathcal{S}}_{sm}|} \sum_{r \in \bar{\mathcal{S}}_{sm}} D_{jrt},$$

where $\bar{\mathcal{S}}_{sm}$ is the set of all stores in market m that belong to a retail chain other than the chain that store s belongs to.

In our main regression specification we estimate the relationship between the promotion indicator D_{jst} and the inside and outside promotion percentages, $x_{jst} = (I_{jst}, O_{jst})$, separately for each product j and store s.

$$\mathbb{E}(D_{jst}|x_{jst}) = \Pr\{D_{jst}|x_{jst}\} = \alpha_{js} + \beta_{js} I_{jst} + \gamma_{js} O_{jst}. \quad (3)$$

If the promotions in store s are set independently of the same-chain or other-chain promotions in market m, then $\beta_{js} = \gamma_{js} = 0$.

Figure 16 displays histograms of the estimates of α_{js}, β_{js}, and γ_{js}, pooled across products and stores, and Table 9 contains corresponding summary statistics and percentiles. As always, we present weighted distributions using the total product revenue for product j as weights. For now we focus on the market-level results in the left column of Figure 16. The median across all inside percentage coefficients, β_{js}, is 1.005, and 97.7 percent of all estimates are positive. Furthermore, we reject the null hypothesis that the inside percentage coefficient is not positive, $\beta_{js} \leq 0$, for 95.8 percent of all estimates, and 52.8 percent of the estimates are not statistically different from 1 at a 5 percent level. These results provide clear evidence that the price promotions for most products are dependent (coordinated) across stores within the same retail chain. On the other hand, the median of the outside percentage coefficients, γ_{js}, is -0.001, and 89.7 percent of the γ_{js} estimates are not statistically different from 0. Hence, conditional on on the inside promotion percentage, I_{jst}, information on the contemporaneous promotion incidence in other retail chains in the local market is typically not predictive of D_{jst}. Also, the estimates of the intercept are small and mostly not distinguishable from 0, indicating that the probability of a promotion in a store is zero if none of the other stores in the chain promote the product. This is further evidence that promotions are coordinated within a retail chain. Ignoring the heterogeneity in the estimates (see Table 9), these results can be summarized using a stylized statistical model of the promotion data-generating process, whereby in any week a retail chain chooses a promotion intensity $\pi \in [0, 1]$ for a product, and the product is promoted in any store within the chain with probability π. The probability of a promotion in store s is conditionally independent (given I_{jst}) of the promotion intensities in the other retail chains in the local market.

To investigate if promotions are also unconditionally independent of promotions in other retail chains, we estimate a restricted version of (3) where D_{jst} is regressed on the outside percentage, O_{jst}, only. The corresponding results are presented in the bottom panel of the left column in Figure 16 and in Table 9. The median of the outside percentage coefficients, γ_{js}, is 0.082, and the distribution of the estimates is skewed to the right. Hence, for some products there is evidence that the promotion incidence in store s is unconditionally dependent on the promotion incidence in the other retail chains at the market level. This dependence is likely due to promotional allowances—trade deals that are offered by the product manufacturers to multiple or all retail chains. Another explanation is seasonality in demand, although seasonality is unlikely to account for the large degree in promotion coordination that we documented.

We also test if promotions are coordinated nationally, across markets. We thus define the national inside promotion percentage based on all stores that belong to the same chain as store s but are not in the same market:

$$I'_{jst} = \frac{1}{|\mathcal{S}_{s,-m}|} \sum_{r \in \mathcal{S}_{s,-m},\, r \neq s} D_{jrt}.$$

Here, $\mathcal{S}_{s,-m}$ includes all stores in the same chain at the national level, excluding the market that store s belongs to. We similarly define the national outside promotion percentage, O'_{jst}, using only promotion observations in other chains outside the market. The right column in Figure 16 shows the corresponding estimates (see also Table 9). The estimates reveal that promotions are strongly coordinated also at the national level. The median of the national inside promotion percentage estimates is 1.017. 97.2 percent of the estimates are positive, and for 93.3 percent we reject the null hypothesis that the estimate is non-positive. Furthermore, mirroring the market-level results, promotions in store s are conditionally independent of the promotions in other retails chains outside the market. However, in the restricted regression where we regress D_{jst} on the outside percentage only, the median of the national outside percentage coefficients is 0.363. Thus, the national-level estimates provide stronger evidence for unconditional promotion dependence across retail chains than the market-level results.

8.3 Discussion

The strong similarity in pricing and promotion strategies among stores that belong to the same retail chain may contradict the documented heterogeneity in pricing strategies within the same retail chain in Ellickson and Misra (2008) (see, for example, the results in Tables 2 and 3 in their paper). Ellickson and Misra (2008) use data from the 1998 *Trade Dimensions Supermarkets Plus Database*, which provides information on store-level pricing strategies based on surveys of retail chain managers. Hence, the data are not directly comparable and cover a different time period than our work. In particular, the managers surveyed in the *Trade Dimensions* data classify store-level pricing policies as EDLP (everyday low price), promotional/Hi-Lo, or as a hybrid of EDLP and Hi-Lo. These qualitative responses may be consistent with the residual variation in pricing and promotion policies after accounting for market/chain dummies as shown in Figure 9. We leave a more detailed comparison of our results with the results in Ellickson and Misra (2008) for future research.

9 Can similarity in demand explain the similarity in prices within chains?

We have extensively documented that prices and promotions are much more similar at the chain than at the local market level. There are different reasons that can explain this fact. One demand side explanation is that retail chains target customers who are similar in their shopping behavior and product preferences. This targeting can be achieved by the choice of store locations—even at the ZIP+3 market level there is often a substantial degree of heterogeneity among the local customers—or by sorting of customers into retail chains. Other explanations include supply-side or institutional factors. For example, retail chains may be unwilling to engage in micro-marketing because they advertise prices and promotions uniformly at the market level, and they may differ in their technical abilities to analyze data or in their data sources, which explains why they choose different price and promotion strategies.

A systematic test of all these potential theories is beyond the scope of this paper. Here, we will focus on a demand side explanation only, and investigate if product demand is more similar at the chain than at the local market level. To this end, we first estimate product (brand) demand models for a large number of products in our sample.

Providing a general overview of price elasticities and promotion effects is interesting not just for the specific purpose of explaining the similarity patterns in prices and promotions that we documented. Given the extremely comprehensive nature of our sample, the estimated distribution of price elasticities and promotion effects provides a general overview of the magnitude of these effects for the kind of products that are contained in our data. We also discuss the usual endogeneity problem and provide some evidence that the estimated price elasticities and promotion effects can be interpreted as causal.

9.1 Demand model

We estimate demand at the brand level separately for each store. Alternatively we could estimate demand at the UPC level, but this is challenging because of the large number of UPC's in most product categories and because even stores that belong to the same retail chain often carry a brand in different pack sizes or forms. Correspondingly, the majority of studies on product demand in the Industrial Organization and Marketing literature (e.g. Hoch et al. 1995 and Nevo 2001) use brands a product definition.

Our sample includes observations on brand j in store s in week t. Let $\mathcal{J}_{js}$ be the set of all products in store s that are in the same category that brand j belongs to, including brand j

itself. We estimate a log-linear demand model for brand j in store s:

$$\log(1 + q_{jst}) = \alpha_{js} + \sum_{k \in \mathcal{J}_{js}} \beta_{jks} \log(p_{kst}) + \sum_{k \in \mathcal{J}_{js}} \gamma_{jks} D_{kst} + \tau_j(s,t) + \epsilon_{jst}. \tag{4}$$

α_{js} is a brand-store fixed effect. q_{jst} denotes unit sales of brand j, p_{kst} is the price of brand k, and D_{kst} is a promotion indicator. Note that (4) is not strictly speaking a log-linear demand model because we add 1 to q_{jst} before taking the log to be able to incorporate observations $q_{jst} = 0$. Despite using $\log(1 + q_{jst})$ instead of $\log(q_{jst})$ as the dependent variable, β_{jjs} is the own-price elasticity and β_{jks} $(k \neq j)$ is the cross-price elasticity with respect to product k.

$\tau_j(s,t)$ is a time fixed effect to account for demand shocks that may be observable to retail chains hence potentially correlated with store-level prices and promotions. The fixed effects are constant across stores in the same local market, defined as a 3-digit ZIP code, and they are also constant within a time period, defined either as a quarter, month, or week in any given year. For example, consider the case of year/month fixed effects at the 3-digit ZIP code market level. If s and s' are two stores in the same 3-digit ZIP code, and if t and t' are two weeks in the same year and month, then $\tau_j(s,t) = \tau_j(s',t')$.

We estimate demand models for the top 2,000 brands during the 2008-2010 sample period, ranked according to total revenue. We choose this subset to avoid brands with many missing price observations (when $q_{jst} = 0$) in the raw data and the corresponding potential for measurement error if a large percentage of prices are imputed based on our price algorithm. Similarly, for each chosen brand we only include stores when prices are observed in at least 80 percent of weeks. In total, we estimate 27.2 million brand-store demand models.

In many categories there is sizable number of competing brands. Including the prices and promotions for all these brands in our demand model is infeasible, and hence we limit the number of brands included to either the brands that account for at least 80 percent of the category revenue or a maximum of five.

Because the time fixed effects, $\tau_j(s,t)$, are common across all stores in a local market we estimate the regressions separately for each brand/market combination.

9.2 Estimation results

The top panel in Figure 17 displays the distribution of the estimated own-price elasticities, pooled across all brand-store estimates, and Table 10 contains the corresponding summary statistics of the distribution. All estimates are weighted based on total brand revenue.[14] We color estimates that are not statistically significant (in the sense that we cannot reject the null hypothesis $\beta_{jjs} = 0$ at a 5 percent level) in gray and all other estimates in blue.

Figure 17 shows the estimated elasticities based on our main specification that accounts for local demand shocks using 3-digit ZIP code/month fixed effects. The median price elasticity across all brand-store estimates is -1.93. There is a large degree of heterogeneity in the estimates, ranging from -6.647 at the 5th percentile to 2.025 at the 95th percentile of the distribution (see

[14] The weights are brand, not store-specific.

Table 10). The percentage of own-price elasticity estimates that is negative is 85.6 percent, although only a small percentage of the non-negative estimates is statistically significant, as is evident from Figure 17. Exactly 4.1 percent of all own-price elasticity estimates are positive and statistically different from 0. Note that all demand model parameters are brand/store-specific, and the number of observations to estimate these parameters is at most 156 weeks (the 2008-2010 period). Hence, it should be expected that not all elasticities are precisely estimated or have the expected sign. Finally, 70.6 percent of the estimated elasticities indicate elastic demand, $\beta_{jjs} < -1$.

Endogeneity

So far we have referred to the estimates, $\hat{\beta}_{jjs}$, as own-price elasticities, and thus we have given the estimates a causal interpretation corresponding to the demand model (4). However, in the presence of price endogeneity the estimates will be biased and will not correspond to the true elasticities. Before we delve into this problem more deeply we note that even if the estimates were biased it would still be valid to examine if the biased elasticity estimates were more similar at the chain level than at the market level, at least if the bias was approximately constant across all estimates for a given brand. Nonetheless, a more careful discussion of price endogeneity is warranted.

Price endogeneity arises if the retail chains are able to set prices or promotions conditional on demand shocks that are observed to them, but not to us, the researchers. A large literature, starting with Berry (1994) and Berry et al. (1995), highlights the endogeneity problem and provides a solution based on instrumental variables. However, in practice good instruments are often hard to find, a point that many researchers have recognized and that has recently received more attention in the literature. Rossi (2014) argues that many instruments that have been used in the literature are either not valid (such as lagged prices) or weak. We agree with these points, especially in our context where the goal of providing a general overview of price elasticities for a larger number of brands makes it impossible to find good instruments that will generally be suitable for all brands and categories.

Our main strategy to avoid endogeneity bias, in addition to the time-invariant store fixed effects α_{js}, is to include the market/time fixed effects $\tau_j(s,t)$ in the estimated models to account for demand shocks and brand-specific trends in demand at a narrowly defined geographic level. We can interpret the estimates $\hat{\beta}_{jks}$ as store-specific price elasticities if (i) $\tau_j(s,t)$ captures all time-varying demand components that may be correlated with the prices p_{kst}, (ii) there is variation in the price *changes* over time across stores, and (iii) the difference in price changes across stores reflects store or chain-specific changes in costs, wholesale prices, markups, or other factors that affect prices but not directly demand. These assumptions are not directly testable, but we can perform an analysis to indicate if our estimates are sensitive to the inclusion and exact specification of the fixed effects. Thus, we first estimate demand without fixed effects, and then including $\tau_j(s,t)$ defined at the 3-digit ZIP code/quarter, 3-digit ZIP code/month, and 3-digit ZIP code/week levels.

The estimated distributions of the price effects are shown in Figure 18 and summarized in

Table 10. The median elasticity estimate is -1.767 in the model without fixed effects, -1.924 in the model with 3-digit ZIP code/quarter fixed effects, -1.93 with 3-digit ZIP code/month fixed effects, and -1.859 with 3-digit ZIP code/week fixed effects. Hence, controlling for time fixed effects at the local market level moderately changes the distribution of the estimates, and the direction of this change is consistent with price endogeneity if positive demand shocks are correlated with higher prices. However, the elasticity estimates are not particularly sensitive to the exact choice of fixed effects, and the direction of the change in the estimated elasticities when we use year-month or year-week fixed effects instead of year-quarter fixed effects is not indicative of a price endogeneity problem. Indeed, for a serious price endogeneity problem to exist it would have to be true that there are high-frequency demand shocks that occur at level that is more local than a 3-digit ZIP code area, and that the store or chain managers are able to predict these shocks and correspondingly change prices. To us, this seems a priori implausible, and in particular such localized price-setting is inconsistent with the strong similarity in price and promotion patterns at the retail chain level that we documented.

Figure 19 displays the distributions of the estimated cross-price elasticities with respect to the two largest (measured using total revenue) competitors in the product category for each brand, and the estimated distribution of the own-promotion effects γ_{jjs}. Although the medians of the cross-price elasticities are positive, a large fraction of the estimates is negative (42.8 percent for the largest and 44.7 percent for the second largest competitor), and the majority of the estimates is not statistically different from zero. Hence, estimating cross-price effects at the brand/store level with weekly data during a three-year period is daunting. To improve on these estimates we would need to impose parameter restrictions or estimate a demand model that relies on a smaller number of parameters, such as a logit market share model (Berry 1994). Among the own-promotion effect estimates a larger percentage have the expected sign: 77.5 percent of all estimates and 94.8 percent of the statistically significant estimates are positive.

Sensitivity analysis: Bayesian hierarchical demand model

We compare the OLS estimates of the parameters in the log-linear demand model (4) to estimates (the posterior means) obtained using a Bayesian hierarchical model. For brand j, let θ_{js} be a parameter vector that includes the store-specific intercept, α_{js}, the own and cross-price elasticities, β_{jks}, and the promotion parameters, γ_{jks}. In the Bayesian hierarchical model specification θ_{js} is drawn from a prior distribution $p(\theta)$. We assume that the prior is normal, $N(\bar{\theta}_j, V_j)$, although more flexible priors, such as a mixture of normals distribution, have been used in the literature (see Rossi et al. 2005). We first project the data on 3-digit ZIP code/month fixed effects, $\tau_j(s,t)$, and then use the residualized data to estimate the brand/store-level demand parameters. The posterior distribution of the model parameters is obtained using MCMC sampling. We use diffuse prior settings (the default values in the `bayesm` package). See Appendix B for a more detailed summary of the model specification.

There are two reasons that motivate us to provide these additional estimates. First, the posterior means in the Bayesian hierarchical model are shrinkage estimators, whereby store-level estimates that are imprecise, in particular due to insufficient variation in the independent variables, are shrunk to the population mean. This shrinkage property provides a form of regularization to guard against unreasonable parameter estimates. This is particularly important given our goal to obtain a large number of brand/store-level demand estimates. Second, Bayesian hierarchical models are widely used in the industry by companies that provide demand estimates as part of their consulting services, especially for brand manufacturers and possibly also for retail chains.

The bottom panel in Figure 17 displays the distribution of the own-price elasticity estimates from a Bayesian hierarchical model, and Table 10 provides the corresponding detailed summary statistics. The median of all own-price elasticity estimates is almost identical for the OLS and Bayesian hierarchical model estimates. However, from Figure 17 it is evident that the distribution of the Bayesian hierarchical model estimates has thinner tails than the distribution of the OLS estimates, which is expected due to the shrinkage property of the Bayesian hierarchical model. The detailed percentiles in Table 10 confirm this result. Also, comparing the Bayesian hierarchical model estimates to the OLS estimates we find that the percentage of negative elasticities is larger (90.3 versus 85.6 percent), and there is also a higher incidence of estimates indication elastic demand (74.9 versus 70.6 percent). In this sense, the Bayesian hierarchical model estimates conform more to prior expectations, although the overall difference with respect to the OLS estimates is only moderate.

9.3 Similarity of price elasticities and promotion effects: Market vs. retail chain factors

We now conduct an analysis that is similar to the analysis in Section 7.2 where we measured the percentage of the variance in prices, promotion frequency, and promotion depth that can be attributed to market versus retail chain-specific factors. Here, we regress the estimated store-level own-price elasticities, $\hat{\beta}_{jjs}$, and promotion coefficients, $\hat{\gamma}_{jjs}$, on market (3-digit ZIP code) dummies, retail chain dummies, and market/chain dummies.

The distribution of the revenue-weighted R^2 values is displayed in Figure 20 (see Table 14 in the Appendix for the details of the results). The results in the top row are based on the own-price elasticity estimates from the main model specification that includes 3-digit ZIP code/month fixed effects. Local market factors explain 14.6 percent (the median across products) of the overall variance in elasticities across stores, whereas chain factors explain a slightly larger fraction, 17.2 percent. In contrast, market/chain factors explain 47.3 percent, i.e. close to half of the overall variation in elasticities across stores. Hence, while market factors account for some of the differences in the own-price elasticity of demand, the elasticities are much more similar within the stores that belong to the same retail chain at the local market level.

Even at the local chain level, however, the price elasticities are not identical—slightly more than half of the variation is not captured by the market/chain fixed effects. This remaining variation in elasticities might represent an unexploited opportunity to price discriminate across stores, but it may also simply reflect that our analysis is based on *estimates* of the true own-price elasticities that are affected by some measurement error. Hence, we compare the distribution of the R^2 values based on the OLS estimates to the results in the second row of Figure 20 that are based on the posterior means of the store-level own price elasticities from the Bayesian hierarchical demand model (recall that we account for 3-digit ZIP code/months fixed effects by projecting the data on the local time dummy variables before obtaining the posterior distribution of the main model parameters using MCMC sampling). The shrinkage properties of the Bayesian estimation method may ameliorate the potential for measurement error in the price elasticities, which is consistent with the results. The R^2 values are now uniformly higher, and we find that local market factors explain 20.7 percent of the overall variance in elasticities, whereas chain factors explain 23.5 percent and market/chain factors explain more than half of the variation in own-price elasticities—52.3 percent. Still, a substantial percentage of the variation in elasticities is not captured by market/chain factors.

In the bottom two rows of Figure 20 we perform an identical analysis for the estimated own-promotion coefficients. Here, chain factors explain a substantially larger percentage of the variance in the estimated coefficients, and the percentage of the variance explained by market/chain factors is somewhat higher than in the case of own-price elasticities. For example, based on the estimates using the Bayesian hierarchical model, market factors explain 19.3, chain factors explain 30.5, and market/chain factors explain 56.4 percent of the overall variance in the estimated promotion effects.

10 Conclusions

The central fact documented in this work is the substantial degree of price dispersion at a given moment in time for identical products (UPC's) and almost identical products (brands) across U.S. retail stores. We observe this deviation from the law of one price even at a fairly narrow geographic level (3-digit ZIP codes) in a large sample of products representing almost 80 percent of retail revenue. This dispersion in prices is due both to persistent differences in base prices and temporary price promotions. However, although the majority of products is frequently promoted, the overall contribution of price promotions to the variance of prices is diminished due to a pervasive EDLP vs. Hi-Lo pricing pattern that is observed for most products.

The degree of price dispersion varies strongly across products, which in part can by explained by differences in the distribution of a product across chains and the level of product "importance" as measured by the product penetration rate. The degree of distribution and product penetration may reflect heterogeneity in product demand. More importantly, the price dispersion for a given product across stores follows a systematic pattern. Prices are overall more homogenous within a local market area (3-digit ZIP code), and in particular prices are substantially more homogenous among the local stores that belong to the same retail chain. Similarly, both the frequency and depth of price promotions is relatively more homogenous within stores in the same retail chain at the local market level. Furthermore, price promotions are strongly coordinated within retail chains, both in local markets and at the national level.

We are able to provide some evidence on the potential causes of the chain-level homogeneity in prices and promotions using store-level demand estimates for 2,000 brands. Both for the estimates of store-level price elasticities and for store-level promotion effects we find that market/chain factors explain a substantially larger fraction of the overall variance in the demand parameters than either market or chain factors alone. Hence, in a local market demand is significantly more homogenous within a retail chain than across stores that belong to different retail chains. Therefore, the local chain-level homogeneity in prices and promotions may simply reflect similarity in demand.

In this work, we are unable to address if the similarity in demand in a retail chain reflects location choices or a a form of sorting whereby chains attract similar consumers. We also do not address if the unexplained portion of the variance in demand parameters simply reflects measurement error or hints at unexploited gains of store-level price discrimination, possibly due to institutional constraints on how category management is performed. Related, we do not take a stance on whether the observed prices and promotions are optimal given the estimated demand systems. Indeed, such an exercise is infeasible without either additional data (wholesale prices and market development funds provided to the retailers) or auxiliary assumptions.

References

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–890.

Berry, S. T. (1994): "Estimating Discrete-Choice Models of Product Differentiation," *Rand Journal of Economics*, 25, 242–262.

Dubois, P. and H. Perrone (2015): "Price Dispersion and Informational Frictions: Evidence from Supermarket Purchases," *manuscript*.

Eden, B. (2014): "Price Dispersion and Demand Uncertainty: Evidence from US Scanner Data," *manuscript*.

Eizenberg, A., S. Lach, and M. Yiftach (2017): "Retail Prices in a City," *manuscript*.

Ellickson, P. B. and S. Misra (2008): "Supermarket Pricing Strategies," *Marketing Science*, 27, 811–828.

Hastie, T., R. Tibshirani, and J. Friedman (2009): *The Elements of Statistical Learning*, Springer, second ed.

Hoch, S. J., B.-D. Kim, A. L. Montgomery, and P. E. Rossi (1995): "Determinants of Store-Level Price Elasticity," *Journal of Marketing Research*, 32, 17–29.

Hwang, M., B. J. Bronnenberg, and R. Thomadsen (2010): "An Empirical Analysis of Assortment Similarities Across U.S. Supermarkets," *Marketing Science*, 29, 858–879.

Kaplan, G. and G. Menzio (2015): "The Morphology of Price Dispersion," *International Economic Review*, 56, 1165–1205.

Kaplan, G., G. Menzio, L. Rudanko, and N. Trachter (2016): "Relative Price Dispersion: Theory and Evidence," *NBER Working Paper 2193*.

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

Nakamura, E. and J. Steinsson (2008): "Five Facts About Prices: A Reevaluation of Menu Cost Models," *Quarterly Journal of Economics*, 123, 1415–1464.

——— (2013): "Price Rigidity: Microeconomic Evidence and Macroeconomic Implications," *Annual Review of Economics*, 5, 133–163.

Nevo, A. (2001): "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 69, 307–342.

Rossi, P., G. Allenby, and R. McCulloch (2005): *Bayesian Statistics and Marketing*, John Wiley & Sons.

Rossi, P. E. (2014): "Even the Rich Can Make Themselves Poor: The Dangers of IV and Related Methods in Marketing Applications," *Marketing Science*, 33, 655–672.

Table 1: Product sample descriptive statistics

Year	Observed prices (million)	Observed and imputed prices (million)	Revenue ($ million)
	Panel A: Products (UPC's)		
2008	8072.1	12131.0	167339.0
2009	8243.6	12573.7	173423.0
2010	8438.3	12756.4	176472.0
All	24754.0	37461.1	517234.0
	Percent imputed prices 33.92		
	Panel B: Brands		
2008	2302.9	2958.2	140157.0
2009	2401.8	3131.4	146830.0
2010	2409.5	3146.7	146608.0
All	7114.2	9236.3	433595.0
	Percent imputed prices 22.97		
	Panel C: Percent private label		
Products	16.2	14.6	18.6
Brands	8.6	7.6	15.4

Note: The first column indicates the number of price observations obtained from the RMS data. The second column also includes the imputed prices for weeks with zero sales. The observation numbers are expressed in millions, and the revenue data are expressed in millions of dollars. The table also indicates the percentage of imputed prices among all observed and imputed prices, and the percentage of private label observations among all product or brand price observations and revenue.

Table 2: Retail chain and store descriptive statistics

	Mean	SD	Min	Percentiles					Max
				10	25	50	75	90	
Panel A: At National level									
No. of retail chains	81								
No. of stores	17184								
Panel B: At DMA level									
No. of retail chains	6.0	2.6	1	3	4	6	7	9	18
No. of stores	83.8	143.7	1	6	13	32	82	224	1061
Panel C: At ZIP+3 level									
No. of retail chains	4.4	2.2	1	2	3	4	6	8	11
No. of stores	20.5	24.9	1	2	5	10	27	54	170
Panel D: Markets covered by retail chains									
No. of DMA's	15.1	34.8	1	1	2	5	9	31	192
No. of ZIP+3	45.9	105.7	1	4	6	13	32	112	572
Panel E: Stores per retail chain									
National	212.1	443.8	2	10	25	77	171	505	3007
DMA	14.0	26.7	1	1	2	5	13	36	320
ZIP+3	4.6	6.6	1	1	1	2	5	11	109

Table 3: Price and base price dispersion statistics

		Median	Mean	Percentiles							
				0.01	**0.05**	**0.1**	**0.25**	**0.75**	**0.9**	**0.95**	**0.99**
Prices											
Product definition: UPC											
Log-price SD	National	0.161	0.163	0.015	0.059	0.090	0.124	0.200	0.238	0.268	0.318
	DMA	0.110	0.114	0.006	0.032	0.050	0.078	0.147	0.185	0.208	0.249
	ZIP+3	0.099	0.103	0.002	0.021	0.039	0.066	0.136	0.174	0.196	0.236
95/5 percentile ratio	National	1.646	1.690	1.018	1.171	1.300	1.460	1.869	2.103	2.331	2.836
	DMA	1.360	1.398	1.011	1.075	1.129	1.232	1.524	1.721	1.831	2.071
	ZIP+3	1.294	1.329	1.004	1.045	1.094	1.182	1.438	1.614	1.713	1.946
Product definition: Brand											
Log-price SD	National	0.175	0.185	0.074	0.098	0.114	0.141	0.216	0.260	0.320	0.427
	DMA	0.138	0.146	0.045	0.067	0.084	0.105	0.174	0.214	0.250	0.356
	ZIP+3	0.129	0.137	0.035	0.060	0.072	0.097	0.165	0.202	0.239	0.341
95/5 percentile ratio	National	1.728	1.841	1.232	1.340	1.414	1.537	1.981	2.289	2.753	4.002
	DMA	1.508	1.583	1.113	1.208	1.264	1.367	1.697	1.929	2.146	3.042
	ZIP+3	1.433	1.492	1.081	1.170	1.211	1.303	1.593	1.786	1.978	2.666
Base prices											
Product definition: UPC											
Log-price SD	National	0.133	0.138	0.006	0.046	0.073	0.103	0.168	0.205	0.240	0.320
	DMA	0.088	0.093	0.002	0.023	0.039	0.062	0.118	0.151	0.177	0.239
	ZIP+3	0.078	0.083	0.001	0.014	0.029	0.052	0.107	0.140	0.166	0.225
95/5 percentile ratio	National	1.516	1.564	1.001	1.130	1.230	1.373	1.702	1.933	2.131	2.710
	DMA	1.276	1.310	1.003	1.050	1.096	1.177	1.399	1.556	1.680	1.983
	ZIP+3	1.220	1.250	1.001	1.028	1.066	1.134	1.328	1.464	1.569	1.832
Product definition: Brand											
Log-price SD	National	0.162	0.174	0.062	0.091	0.101	0.127	0.206	0.253	0.310	0.425
	DMA	0.126	0.137	0.034	0.062	0.071	0.093	0.166	0.209	0.244	0.357
	ZIP+3	0.118	0.127	0.025	0.055	0.064	0.086	0.157	0.197	0.231	0.345
95/5 percentile ratio	National	1.654	1.780	1.187	1.297	1.350	1.478	1.912	2.229	2.671	3.857
	DMA	1.460	1.541	1.081	1.192	1.225	1.315	1.659	1.893	2.133	2.918
	ZIP+3	1.389	1.453	1.054	1.154	1.185	1.257	1.564	1.759	1.953	2.645

Table 4: Price promotion statistics

	Median	**Mean**	**Percentiles**							
			1%	**5%**	**10%**	**25%**	**75%**	**90%**	**95%**	**99%**
Pooled across products, stores, and weeks										
Percentage price discount	0.177	0.207	0.013	0.032	0.051	0.101	0.287	0.402	0.486	0.595
Product-level statistics										
Promotion frequency	0.147	0.163	0.001	0.011	0.023	0.074	0.231	0.317	0.370	0.502
Promotion depth	0.195	0.201	0.083	0.102	0.119	0.149	0.246	0.290	0.315	0.390
Percentage volume on promotion	0.287	0.298	0.002	0.018	0.046	0.157	0.426	0.544	0.614	0.755
Promotion multiplier	3.043	3.783	1.054	1.335	1.610	2.169	4.458	6.790	9.220	23.297
Product-level statistics: Differences between 95th and 5th percentiles across stores										
Promotion frequency difference	0.314	0.316	0.007	0.046	0.083	0.184	0.436	0.545	0.591	0.667
Promotion depth difference	0.180	0.190	0.023	0.070	0.096	0.135	0.233	0.292	0.342	0.444
Percentage promoted volume difference	0.541	0.510	0.013	0.085	0.163	0.381	0.658	0.758	0.821	0.933
Promotion multiplier difference	5.531	7.854	0.603	1.316	1.897	3.225	9.096	16.289	25.518	65.013

Table 5: Price variance decomposition

			Within-Market	
	UPC	Brand	UPC	Brand
Basic decomposition				
Across-market	32.7	29.7		
Across-store	27.0	42.3	40.2	60.1
Within-store	40.3	28.0	59.8	39.9
Decomposition into base prices and promotions				
Across-market	32.7	29.7		
Across-store mean base price variance	31.3	49.9	46.5	70.9
Within-store base price variance	12.3	13.4	18.3	19.0
Total contribution of promotions	23.7	7.0	35.2	10.0
Promotional price discounts	36.0	29.9	53.5	42.6
EDLP vs. Hi-Lo adjustment	-12.3	-22.9	-18.3	-32.6

Table 6: Relationship between price dispersion and product characteristics (product rank indicators based on revenue)

	Dependent variable: log-price standard deviation					
	(1)	(2)	(3)	(4)	(5)	(6)
Revenue rank, 1-100	0.043***	0.004***	0.004***	0.002**	0.012***	0.017***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Revenue rank, 101-250	0.034***	−0.001	−0.001	−0.002**	0.005***	0.011***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Revenue rank, 251-500	0.041***	0.007***	0.007***	0.006***	0.011***	0.015***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Revenue rank, 501-1000	0.040***	0.007***	0.007***	0.007***	0.011***	0.016***
	(0.001)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
Revenue rank, 1001-5000	0.030***	0.002***	0.002***	0.002***	0.005***	0.009***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Revenue rank, 5001-10000	0.025***	0.002***	0.002***	0.002***	0.004***	0.006***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Revenue rank, 10001-20000	0.017***	0.002***	0.002***	0.002***	0.003***	0.004***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Log(no. stores)		0.006***	0.007***	0.006***	0.006***	0.006***
		(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Log(no. chains)		0.064***	0.064***	0.067***	0.066***	0.065***
		(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0002)
Log(population)			−0.0005***	−0.001***	−0.0002**	0.00003
			(0.0001)	(0.0001)	(0.0001)	(0.0001)
Private-Label				0.014***	0.013***	0.013***
				(0.0002)	(0.0002)	(0.0002)
Constant	0.062***	0.020***	0.019***	0.017***		
	(0.0001)	(0.0001)	(0.0002)	(0.0002)		
Product group dummy	No	No	No	No	Yes	No
Product module dummy	No	No	No	No	No	Yes
Observations	1,000,000	1,000,000	1,000,000	1,000,000	1,000,000	1,000,000
R^2	0.030	0.372	0.372	0.376	0.399	0.421
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 7: Relationship between price dispersion and product characteristics (product rank indicators based on product penetration)

	Dependent variable: log-price standard deviation					
	(1)	(2)	(3)	(4)	(5)	(6)
Penetration rank, 1-100	0.073***	0.045***	0.045***	0.043***	0.047***	0.050***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Penetration rank, 101-250	0.066***	0.036***	0.037***	0.034***	0.037***	0.039***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Penetration rank, 251-500	0.063***	0.035***	0.035***	0.034***	0.037***	0.040***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Penetration rank, 501-1000	0.052***	0.027***	0.028***	0.027***	0.029***	0.030***
	(0.001)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Penetration rank, 1001-5000	0.037***	0.019***	0.020***	0.019***	0.022***	0.023***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Penetration rank, 5001-10000	0.023***	0.012***	0.012***	0.012***	0.014***	0.015***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Penetration rank, 10001-20000	0.015***	0.008***	0.008***	0.008***	0.009***	0.009***
	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0002)
Log(no. stores)		0.006***	0.006***	0.005***	0.006***	0.006***
		(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Log(no. chains)		0.063***	0.063***	0.066***	0.064***	0.064***
		(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0001)
Log(population)			0.001***	0.001***	0.001***	0.001***
			(0.0001)	(0.0001)	(0.0001)	(0.0001)
Private-Label				0.013***	0.012***	0.013***
				(0.0002)	(0.0002)	(0.0002)
Constant	0.060***	0.015***	0.017***	0.015***		
	(0.0001)	(0.0001)	(0.0002)	(0.0002)		
Product group dummy	No	No	No	No	Yes	No
Product module dummy	No	No	No	No	No	Yes
Observations	1,000,000	1,000,000	1,000,000	1,000,000	1,000,000	1,000,000
R^2	0.045	0.385	0.385	0.388	0.409	0.430
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 8: Promotion coordination: Summary statistics

		Median	Mean	% = 0	Percentiles							
					1%	5%	10%	25%	75%	90%	95%	99%
DMA/chain promotion percentages												
Top 1,000	Data	0.548	0.541	55.060	0.011	0.026	0.048	0.143	0.991	1.000	1.000	1.000
	Simulated	0.279	0.313	22.451	0.025	0.060	0.091	0.167	0.429	0.600	0.667	0.833
All	Data	0.410	0.484	62.184	0.010	0.023	0.039	0.111	0.886	1.000	1.000	1.000
	Simulated	0.200	0.255	28.274	0.021	0.048	0.071	0.125	0.353	0.500	0.600	0.800
DMA/chain promotion percentages, store/promotion-weighted												
Top 1,000	Data	0.957	0.822		0.047	0.188	0.379	0.745	1.000	1.000	1.000	1.000
	Simulated	0.416	0.419		0.060	0.125	0.174	0.274	0.556	0.667	0.732	0.833
All	Data	0.901	0.779		0.038	0.158	0.323	0.650	0.991	1.000	1.000	1.000
	Simulated	0.333	0.355		0.045	0.094	0.132	0.212	0.482	0.606	0.681	0.800

Note: Results are based observations when a product is carried in at least 5 stores by the retailer in the DMA.

Table 9: Promotion coordination: Promotion dependence regressions

		Median	Mean	NR = 0	NR = 1	R ≥ 0	Est. > 0	Percentiles							
								0.01	**0.05**	**0.1**	**0.25**	**0.75**	**0.9**	**0.95**	**0.99**
DMA															
(1)	Inside Percentage	1.005	0.975		0.528	0.958	0.977	-0.147	0.297	0.553	0.861	1.086	1.233	1.401	2.247
	Outside Percentage	-0.001	0.008	0.897		0.036	0.490	-0.768	-0.272	-0.168	-0.064	0.064	0.183	0.309	0.916
	Intercept	0.000	0.008	0.872			0.497	-0.125	-0.056	-0.037	-0.015	0.021	0.060	0.097	0.235
(2)	Inside Percentage	1.006	0.977		0.525	0.961	0.978	-0.131	0.311	0.563	0.866	1.085	1.232	1.399	2.243
	Intercept	0.000	0.009	0.838			0.493	-0.093	-0.042	-0.027	-0.011	0.018	0.053	0.088	0.218
(3)	Outside Percentage	0.082	0.173	0.777		0.252	0.631	-1.036	-0.363	-0.222	-0.067	0.308	0.646	0.957	2.106
	Intercept	0.147	0.186	0.289			0.956	-0.044	0.003	0.017	0.060	0.272	0.419	0.513	0.686
National															
(1)	Inside Percentage	1.017	1.036		0.503	0.933	0.972	-0.294	0.183	0.458	0.825	1.140	1.369	1.629	3.296
	Outside Percentage	0.000	0.027	0.876		0.044	0.498	-1.374	-0.560	-0.360	-0.144	0.164	0.429	0.685	1.666
	Intercept	-0.002	0.006	0.873			0.476	-0.267	-0.115	-0.074	-0.029	0.030	0.091	0.154	0.383
(2)	Inside Percentage	1.018	1.033		0.497	0.938	0.973	-0.254	0.204	0.475	0.834	1.138	1.366	1.624	3.288
	Intercept	-0.002	0.010	0.799			0.463	-0.153	-0.064	-0.041	-0.017	0.021	0.070	0.125	0.338
(3)	Outside Percentage	0.363	0.510	0.649		0.394	0.754	-1.299	-0.499	-0.268	0.006	0.875	1.495	1.961	3.553
	Intercept	0.077	0.114	0.635			0.794	-0.255	-0.101	-0.047	0.010	0.190	0.338	0.449	0.693

Note: NR stands for “Not Rejected", R stands for “Rejected", and “Est." stands for “Estimate"

Table 10: Demand estimates

	Median	Mean	% > 0	% < 0	% < -1	% significant	Percentiles							
							1%	5%	10%	25%	75%	90%	95%	99%
OLS														
Own-price, no FE's	-1.767	-1.941		0.839	0.670	0.564	-15.182	-6.803	-4.907	-3.078	-0.585	0.689	2.093	9.095
Year-quarter/ZIP+3	-1.924	-2.046		0.857	0.704	0.595	-13.849	-6.608	-4.906	-3.204	-0.753	0.534	1.958	8.851
Year-month/ZIP+3	-1.930	-2.050		0.856	0.706	0.601	-13.948	-6.647	-4.929	-3.216	-0.763	0.557	2.025	9.116
Year-week/ZIP+3	-1.859	-1.972		0.847	0.692	0.592	-14.153	-6.658	-4.878	-3.137	-0.691	0.669	2.198	9.568
Cross-price 1	0.153	0.157	0.572			0.203	-8.767	-3.251	-1.855	-0.538	0.925	2.144	3.424	8.338
Cross-price 2	0.067	0.080	0.553			0.180	-6.733	-2.508	-1.401	-0.410	0.611	1.578	2.587	6.482
Cross-price 3	0.048	0.049	0.540			0.163	-6.827	-2.499	-1.395	-0.411	0.547	1.474	2.518	6.657
Own-promotion effect	0.178	0.373	0.775			0.347	-0.610	-0.218	-0.112	0.016	0.494	1.216	1.913	3.579
Bayesian hierarchical														
Own-price	-1.970	-2.026		0.903	0.749	0.708	-8.215	-5.116	-4.163	-2.989	-0.994	-0.033	0.754	4.070
Cross-price 1	0.159	0.193	0.626			0.191	-3.511	-1.364	-0.768	-0.195	0.615	1.205	1.734	3.899
Cross-price 2	0.089	0.084	0.592			0.177	-3.253	-1.226	-0.676	-0.196	0.405	0.871	1.300	2.751
Cross-price 3	0.070	0.063	0.584			0.148	-2.392	-1.029	-0.614	-0.184	0.336	0.718	1.080	2.389
Own-promotion effect	0.184	0.299	0.844			0.452	-0.271	-0.104	-0.042	0.052	0.408	0.835	1.247	2.353

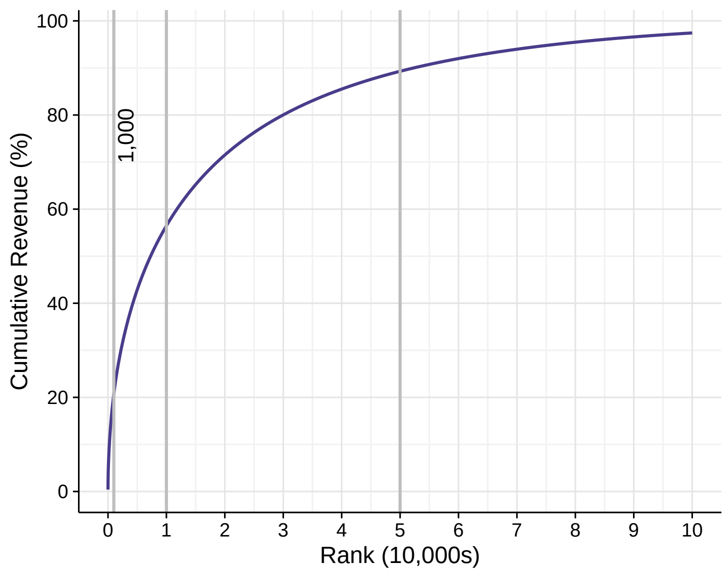

Figure 1: Cumulative revenue for top 100,000 products

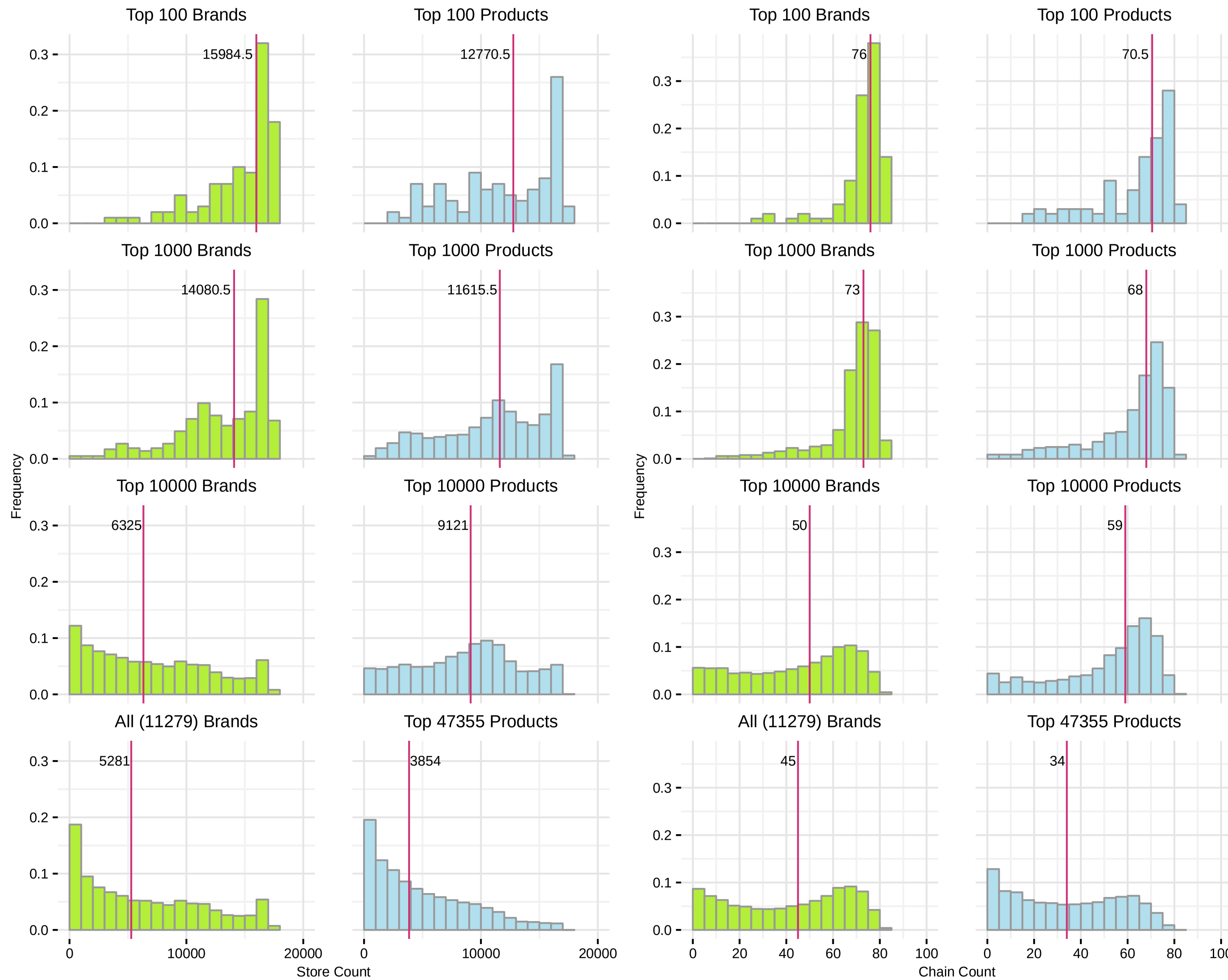

Figure 2: Product and brand availability across stores and retail chains

Note: The graph displays the distribution of the number of stores (the two columns to the left) and chains (the two columns to the right) at which a product or brand is available. The median of each distribution is indicated using a vertical line. A product (brand) is classified as available if it was sold at least once in a store or chain in 2010. The sample includes 17,184 stores and 81 retail chains.

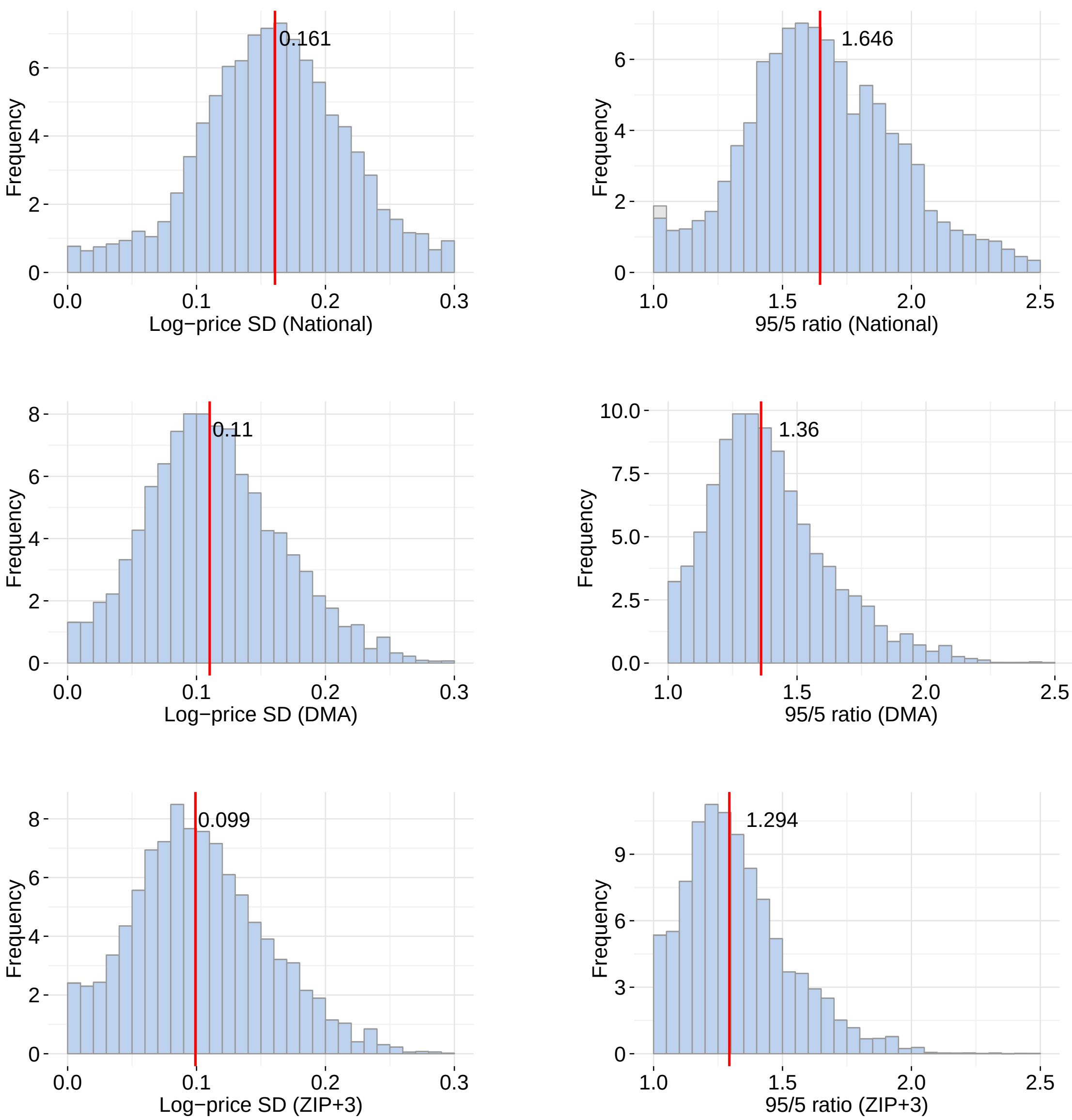

Figure 3: Price dispersion statistics: UPC prices

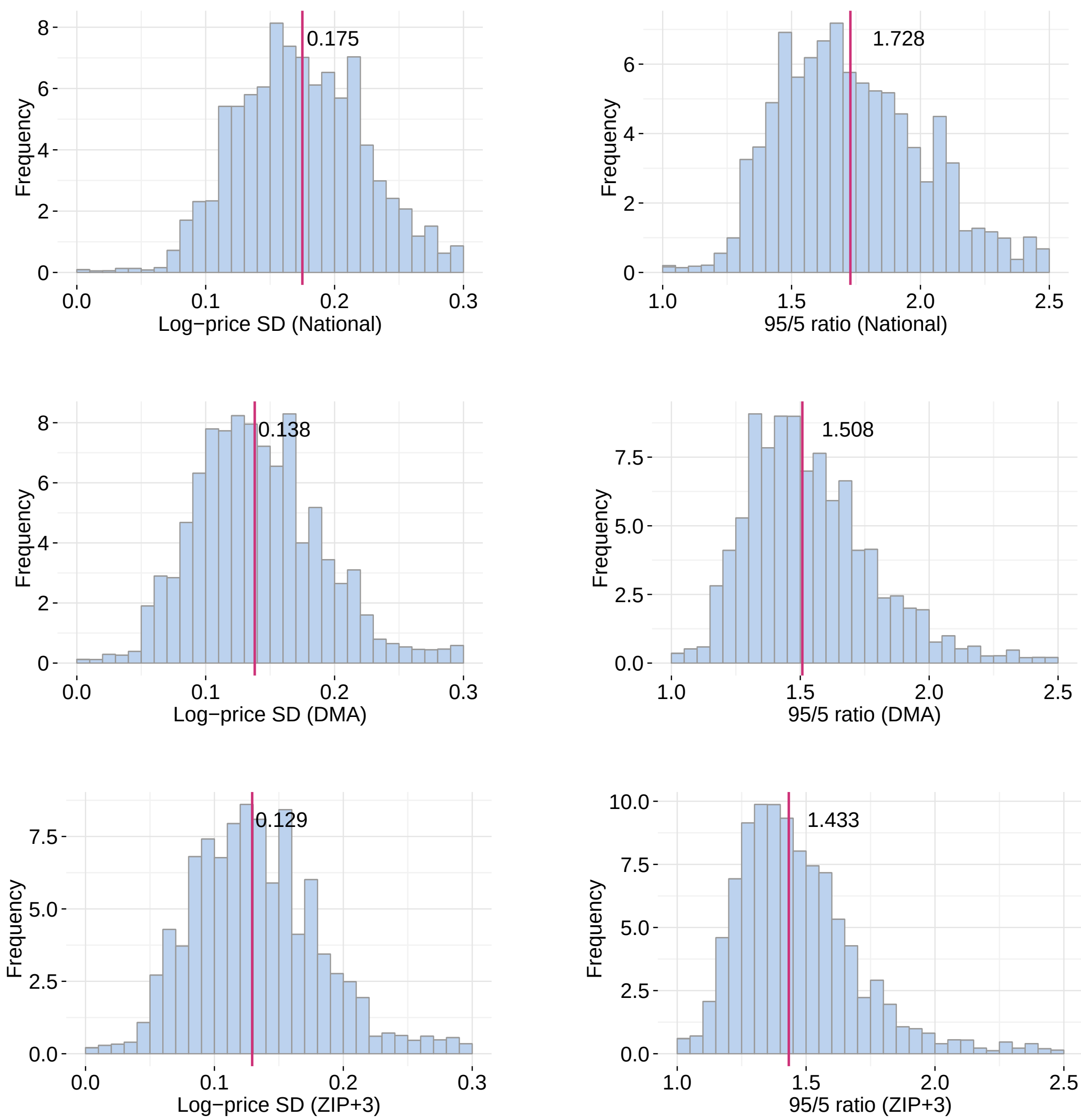

Figure 4: Price dispersion statistics: Brand-level prices

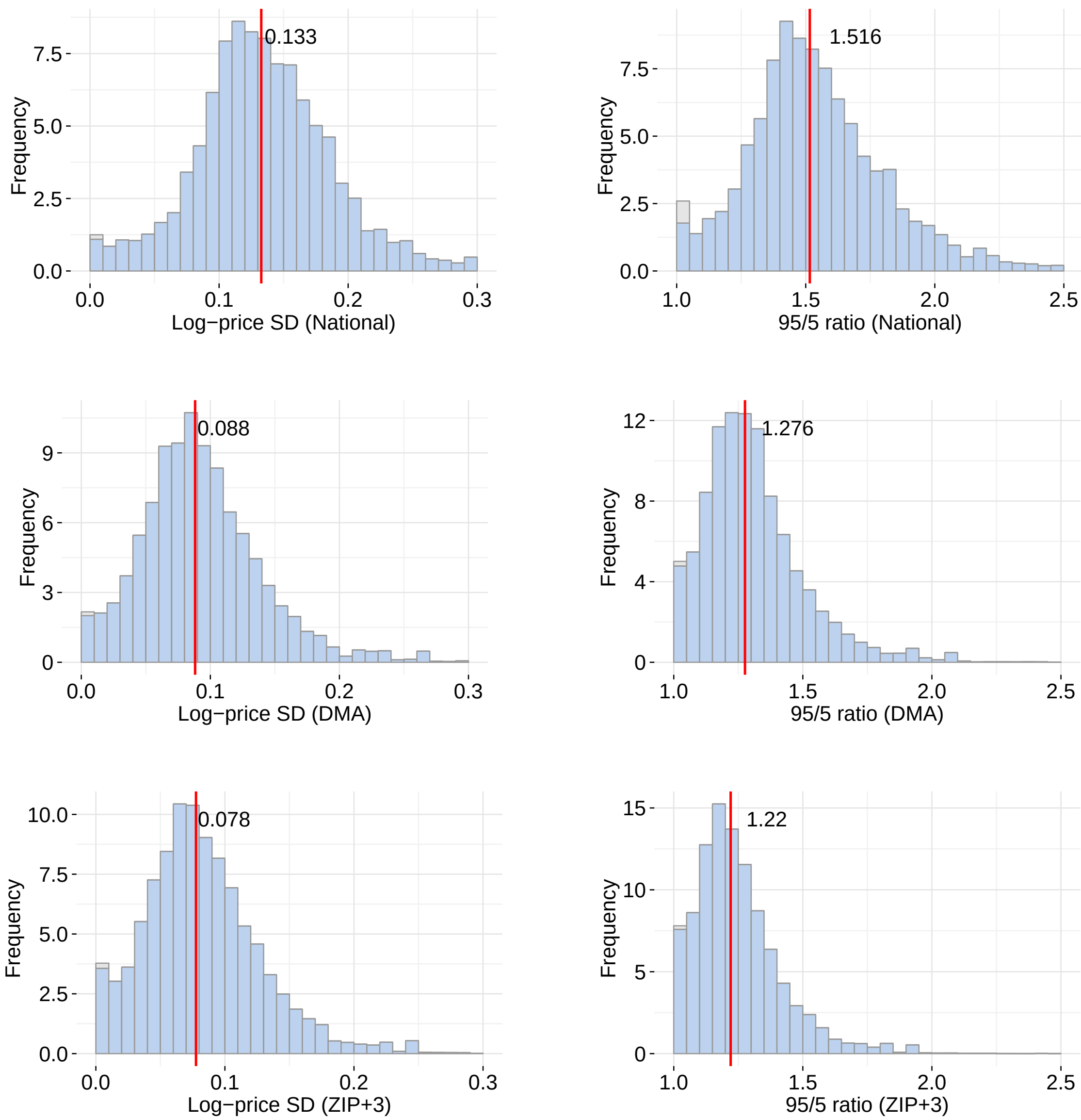

Figure 5: Base prices dispersion statistics: UPC base prices

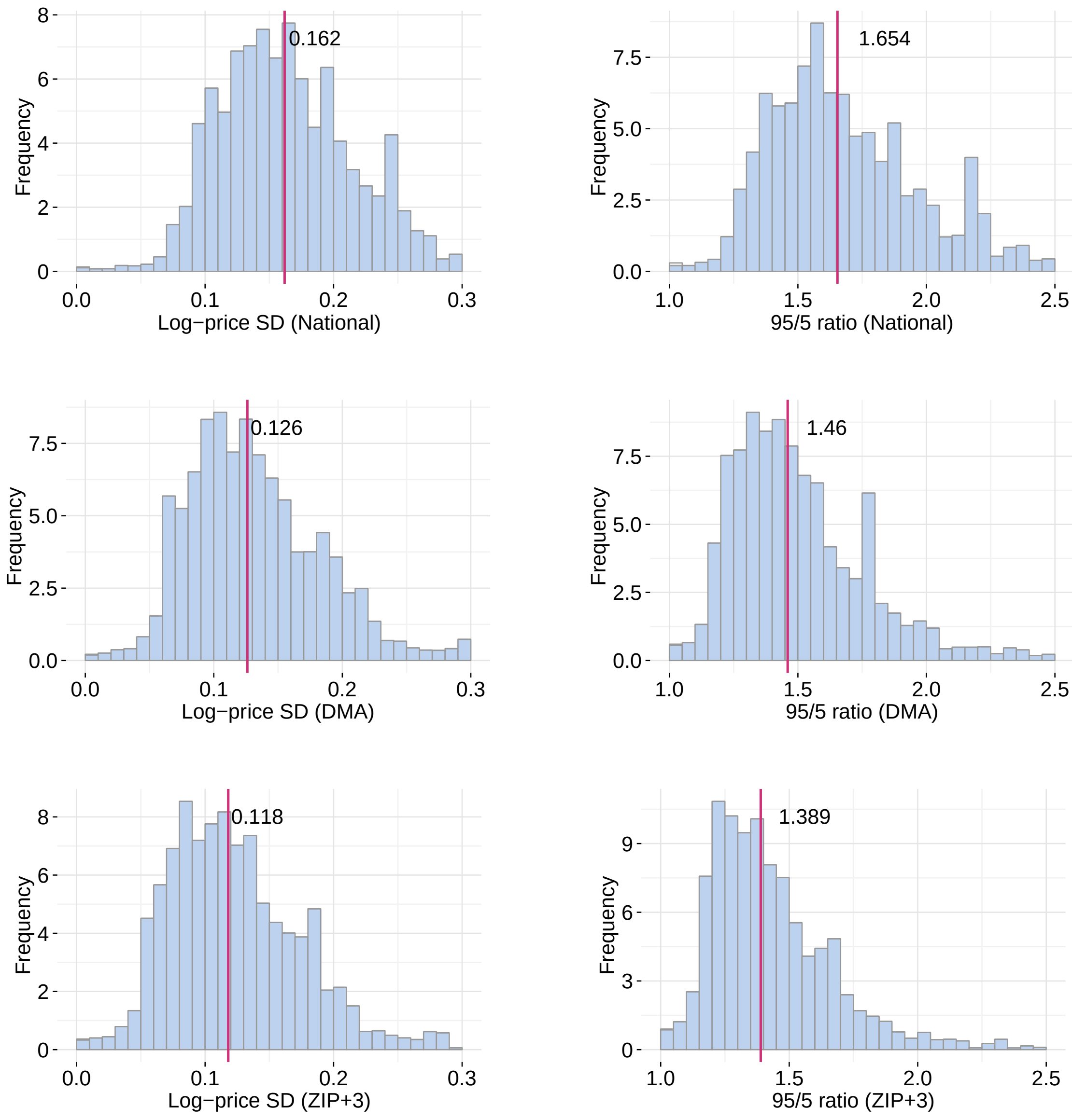

Figure 6: Base prices dispersion statistics: Brand-level base prices

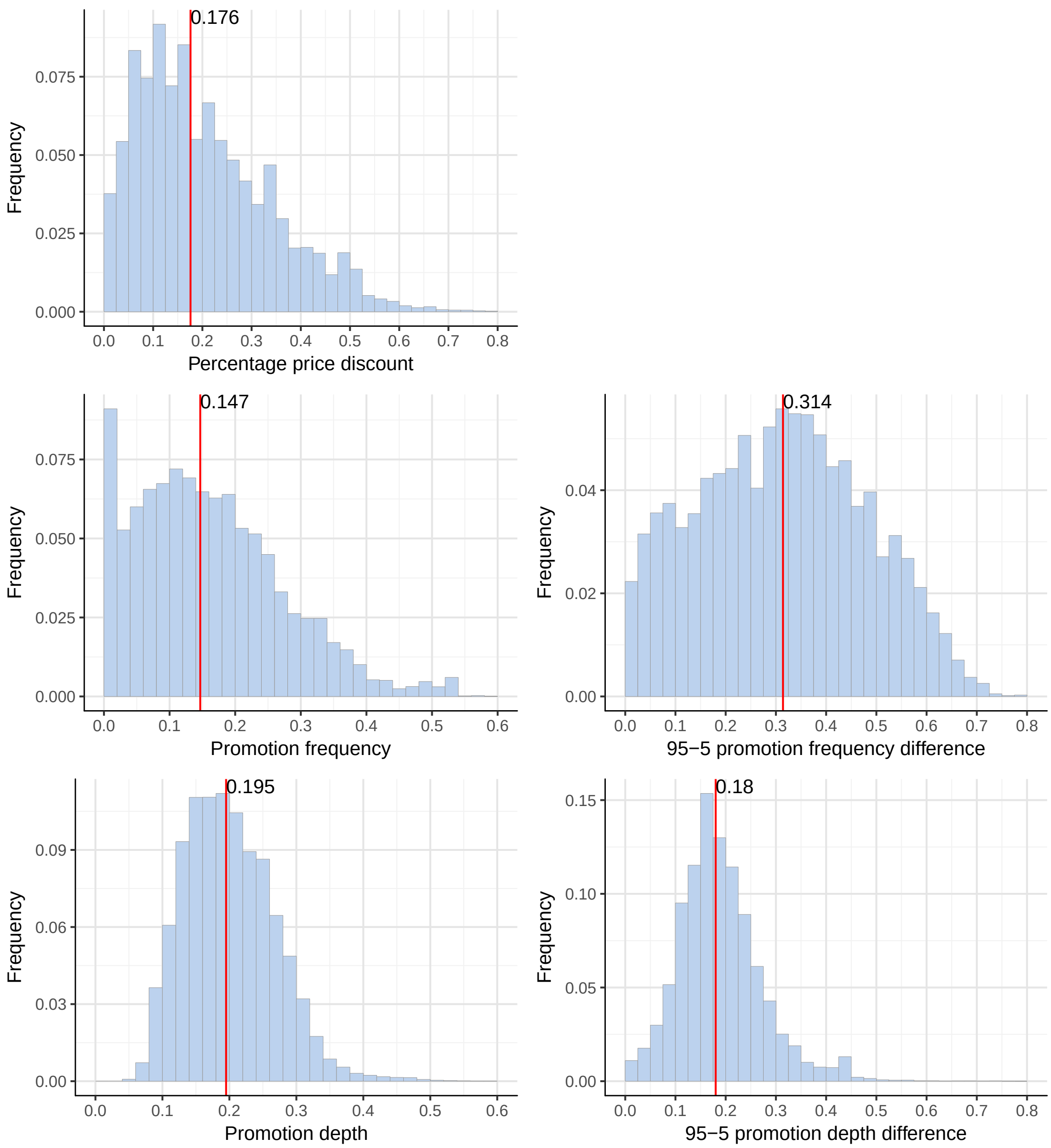

Figure 7: Promotion depth, frequency, and differences in promotion depth and frequency across stores

Note: The top-left panel displays the distribution of the percentage price discounts, δ_{jst}, pooled across all products j, stores s, and weeks t when the price is strictly less than the base price, $p_{jst} < b_{jst}$. The middle and bottom-left panels display the weighted distribution of promotion frequency and promotion depth across products, j. Here, promotion depth is measured conditional on the product being promoted at a 5 percent promotion threshold. The middle and bottom-right panels summarize across-store differences in promotion frequency and promotion depth for all products. In particular, for each product j the differences are based on the 95th and 5th percentile of promotion frequency and promotion depth across all stores where the product is sold.

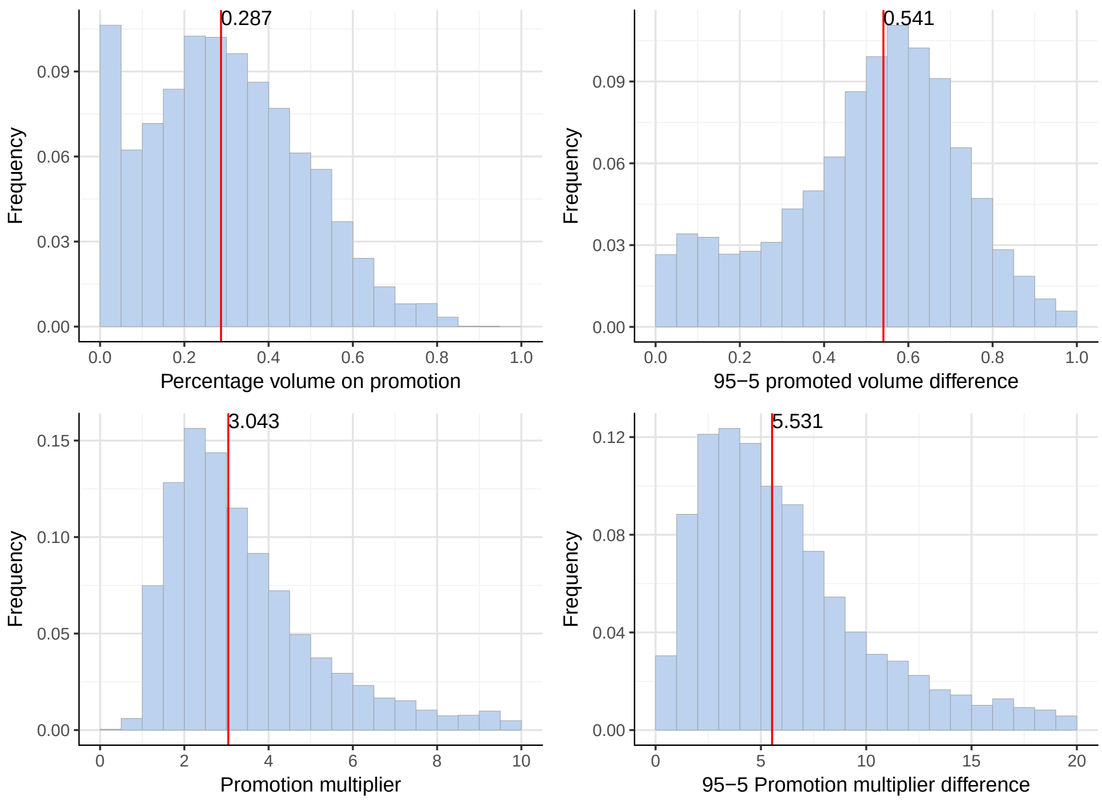

Figure 8: Promotion volume and multiplier, and differences in volume and multipliers across stores

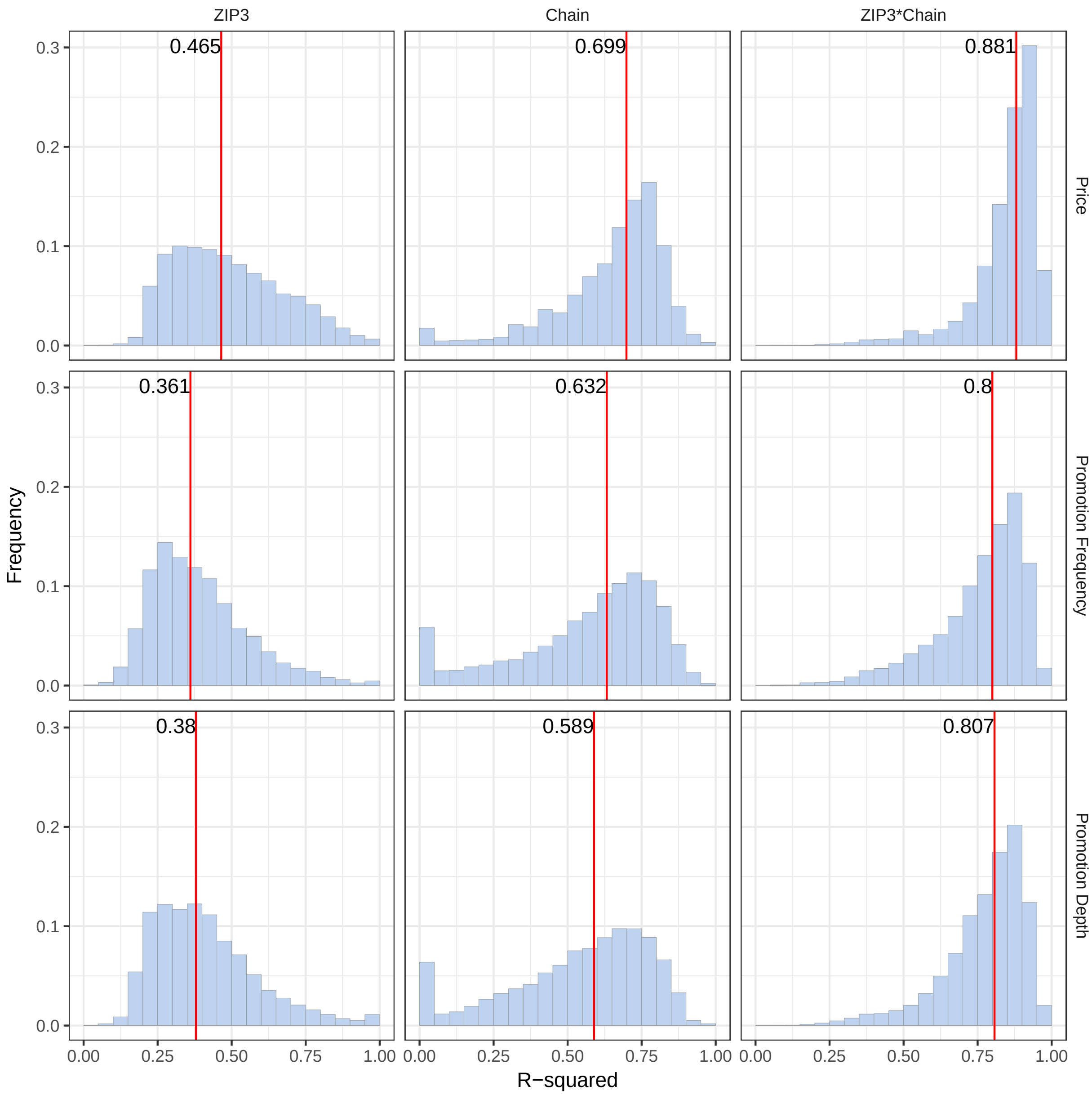

Figure 9: Percentage of variance of prices, promotion frequency, and promotion depth explained by market and chain factors

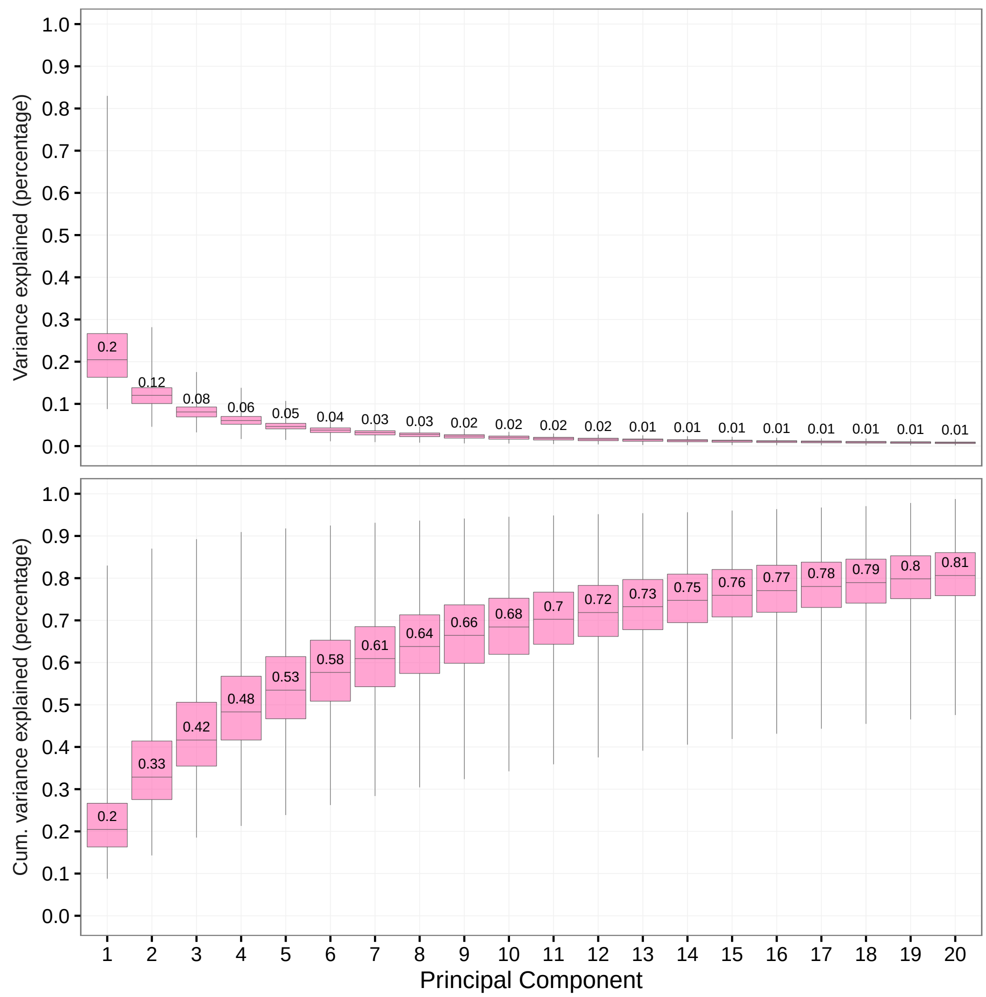

Figure 10: Percentage and cumulate percentage of price variance explained by principal component

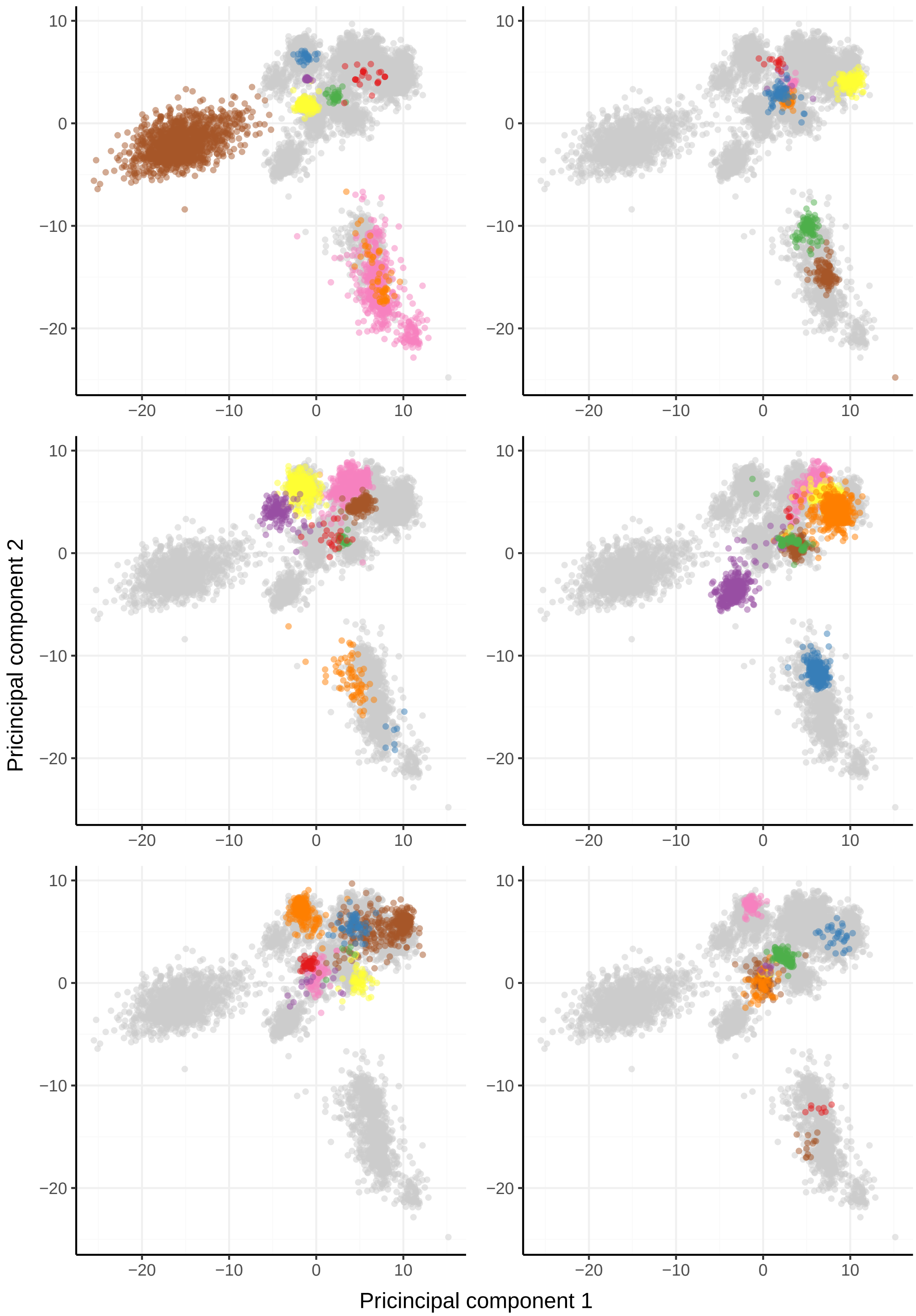

Figure 11: Projected store-level price vectors colored by retail chain

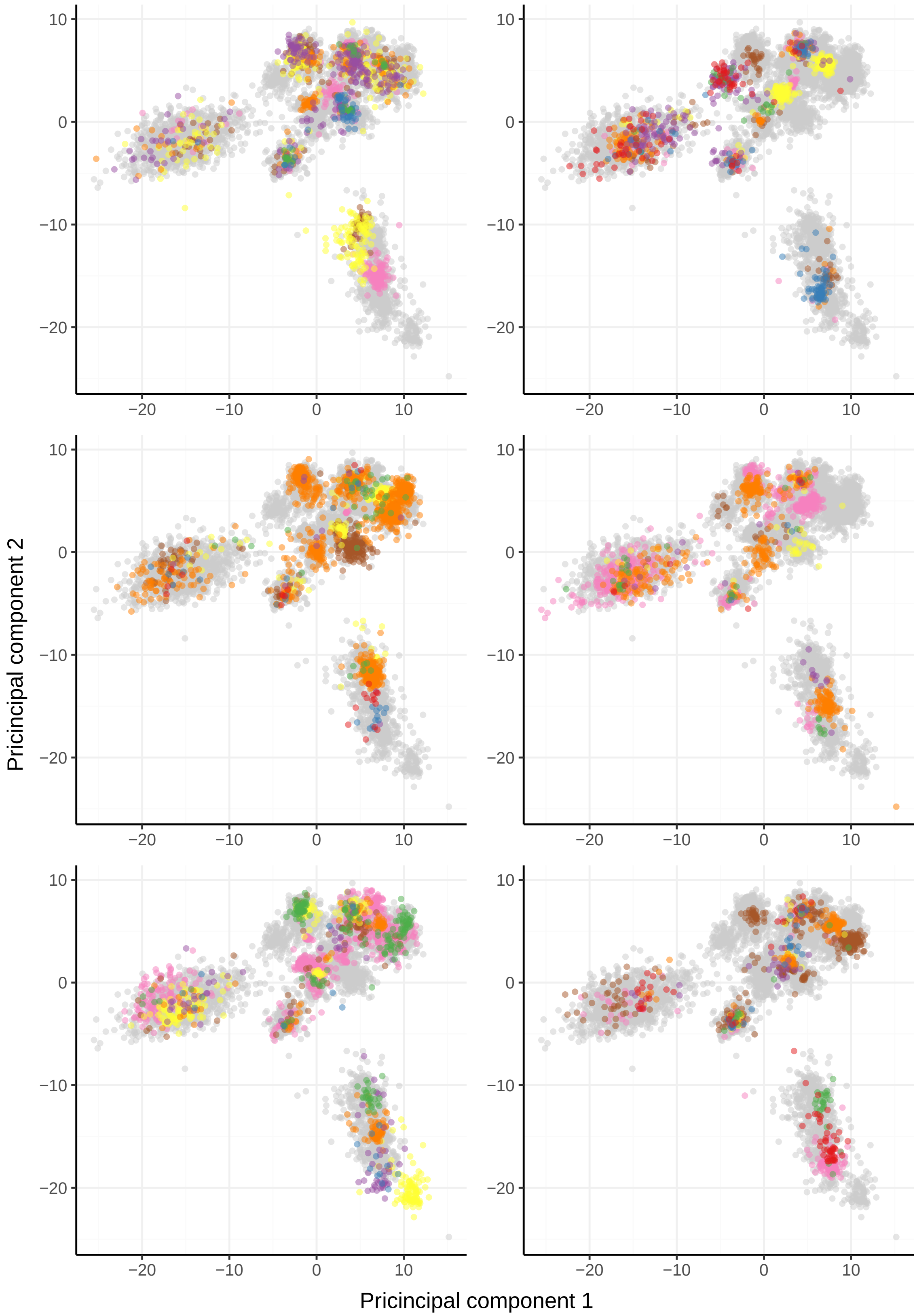

Figure 12: Projected store-level price vectors colored by DMA

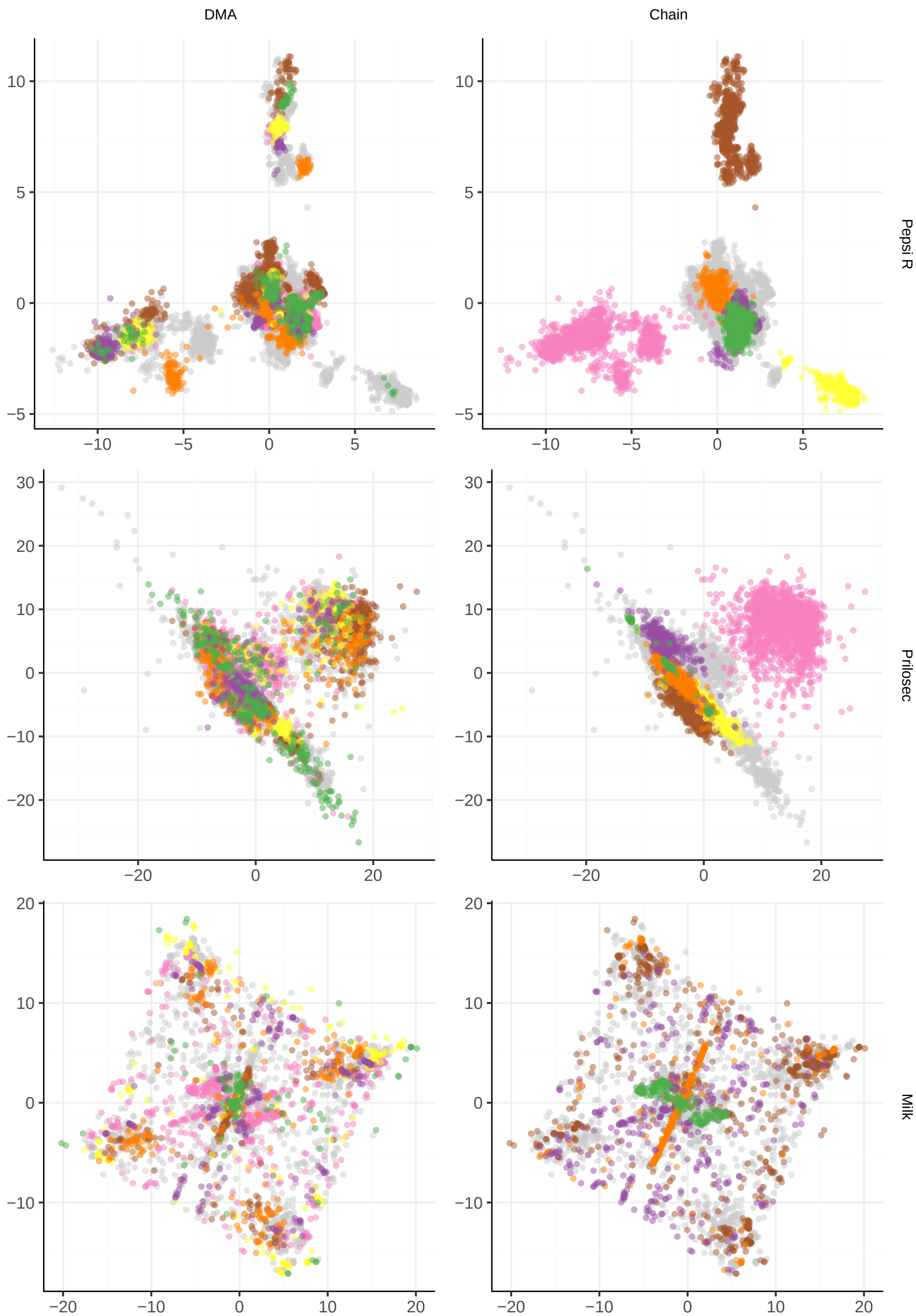

Figure 13: Projected store-level price vectors colored by retail chain and DMA

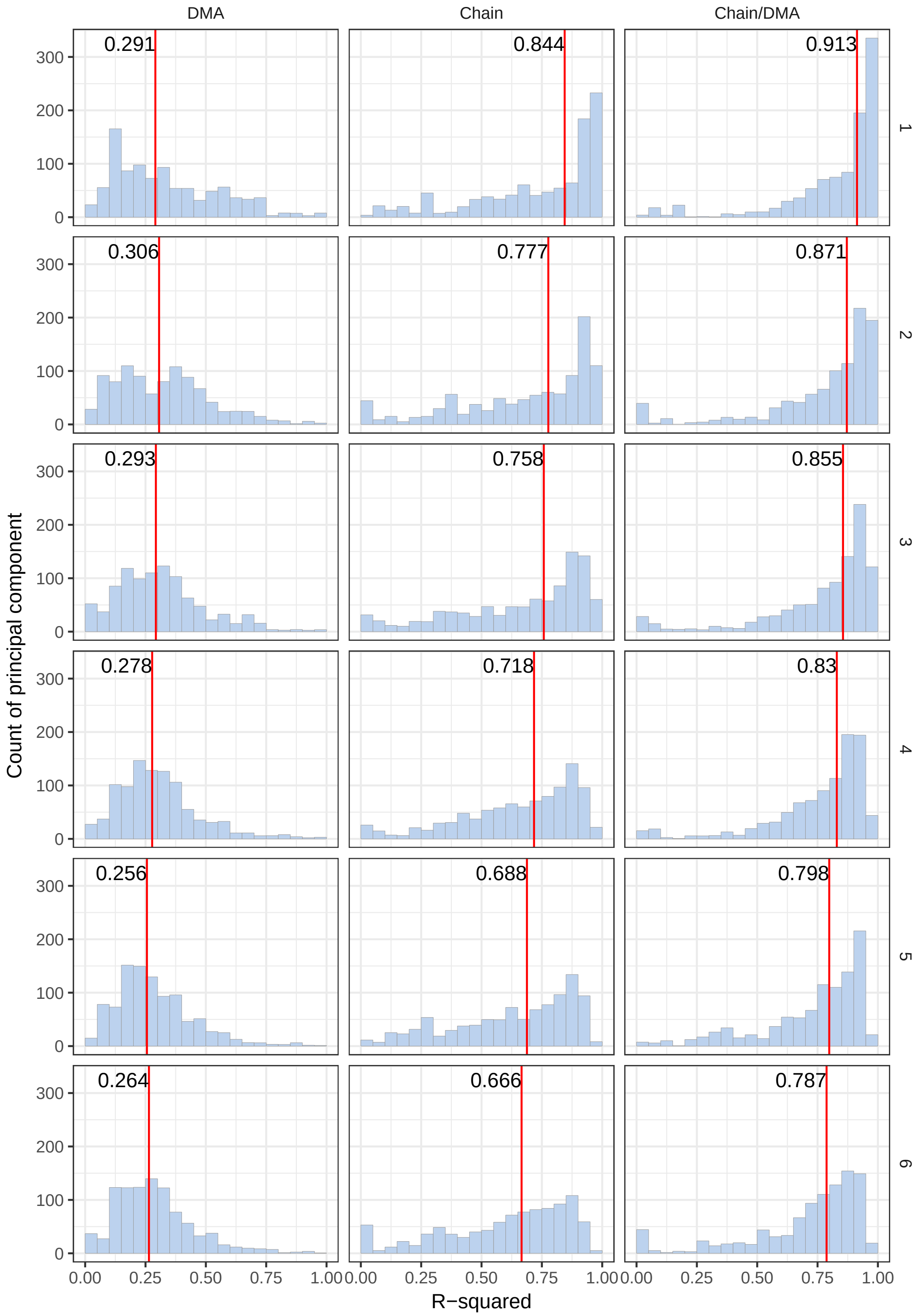

Figure 14: Percentage of projected price variance (R^2) explained by DMA, chain, and chain-DMA fixed effects

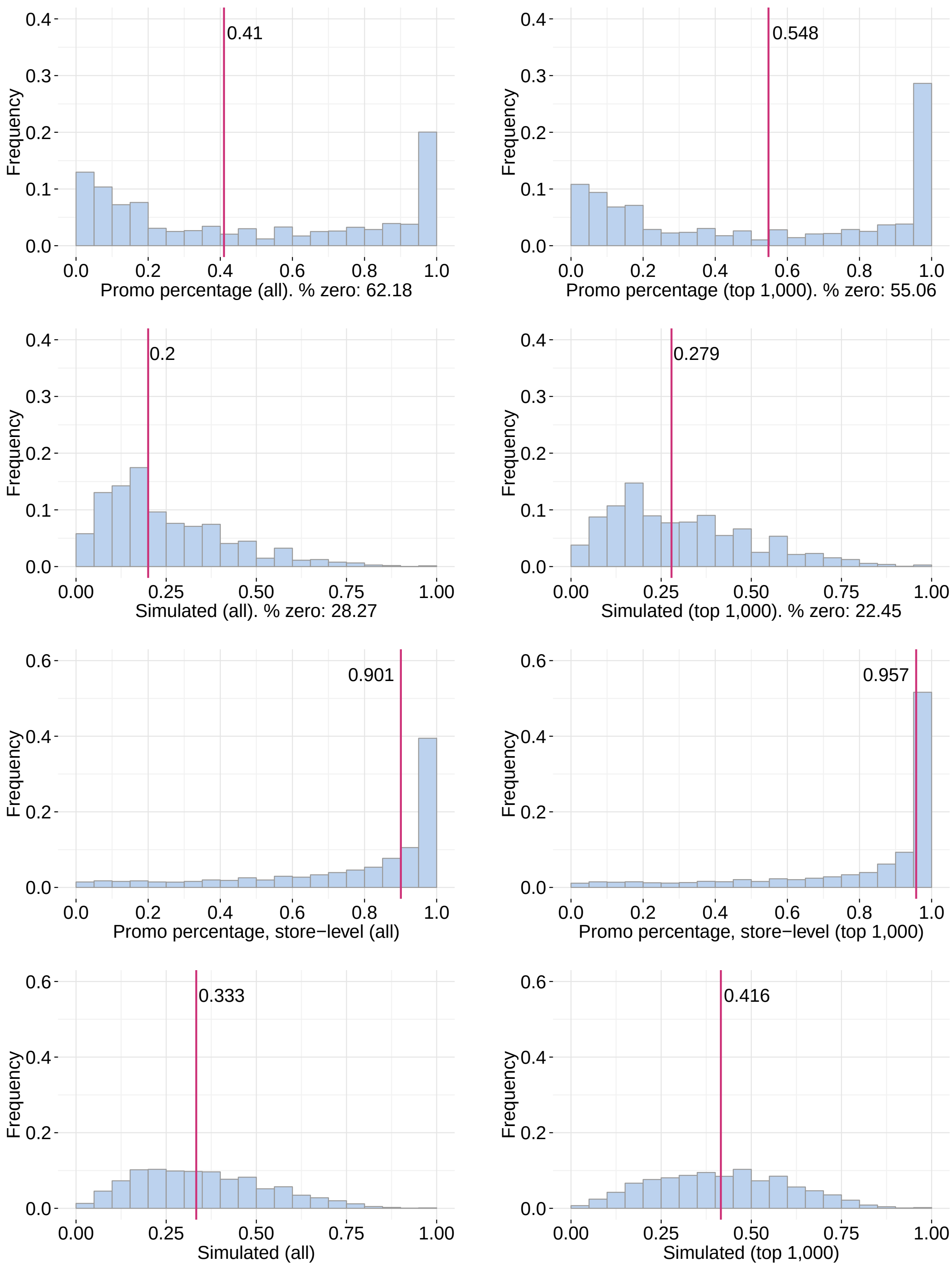

Figure 15: Distribution of chain/DMA promotion percentages

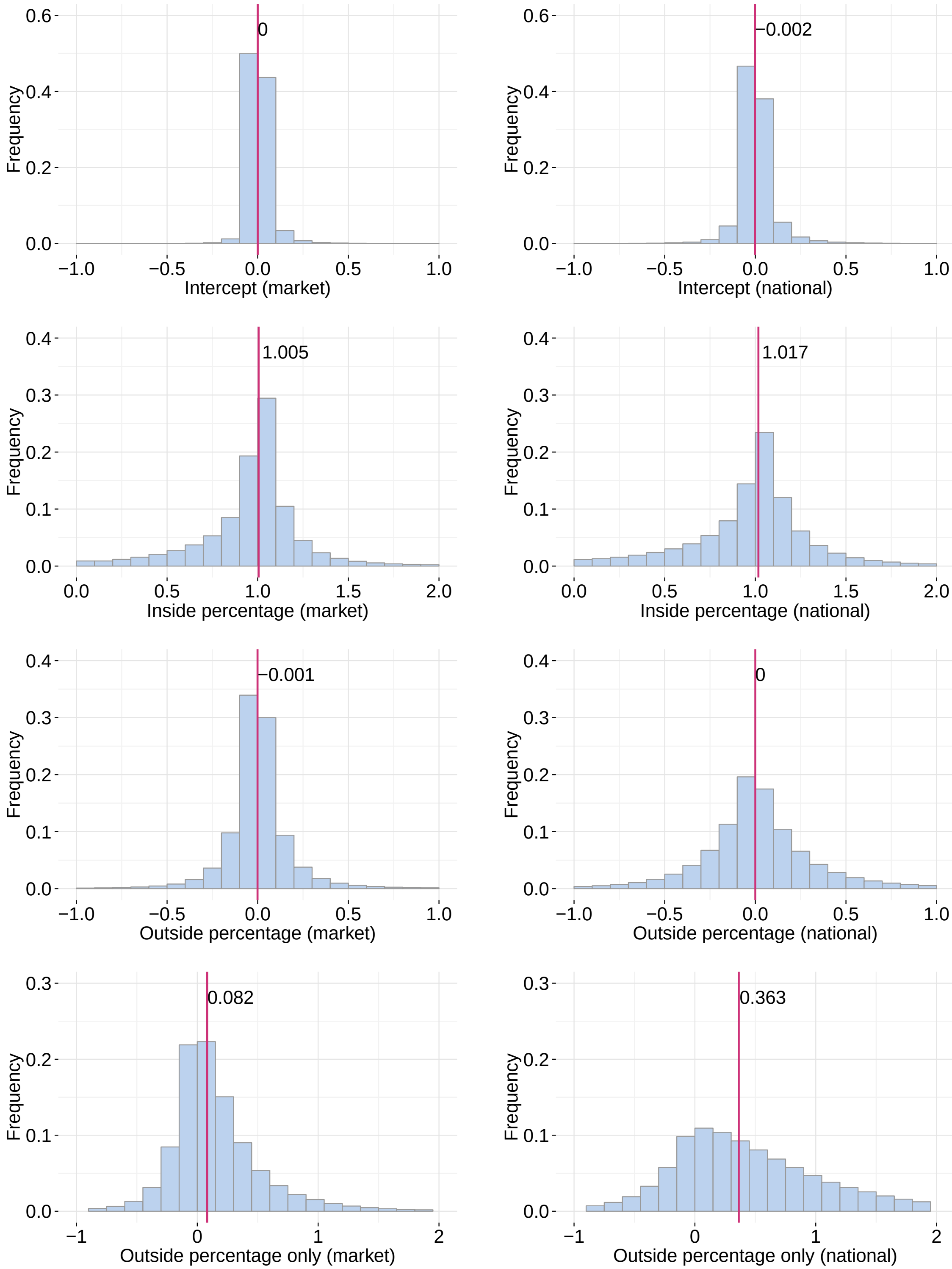

Figure 16: Regression results: Promotion incidence and inside and outside promotion percentages

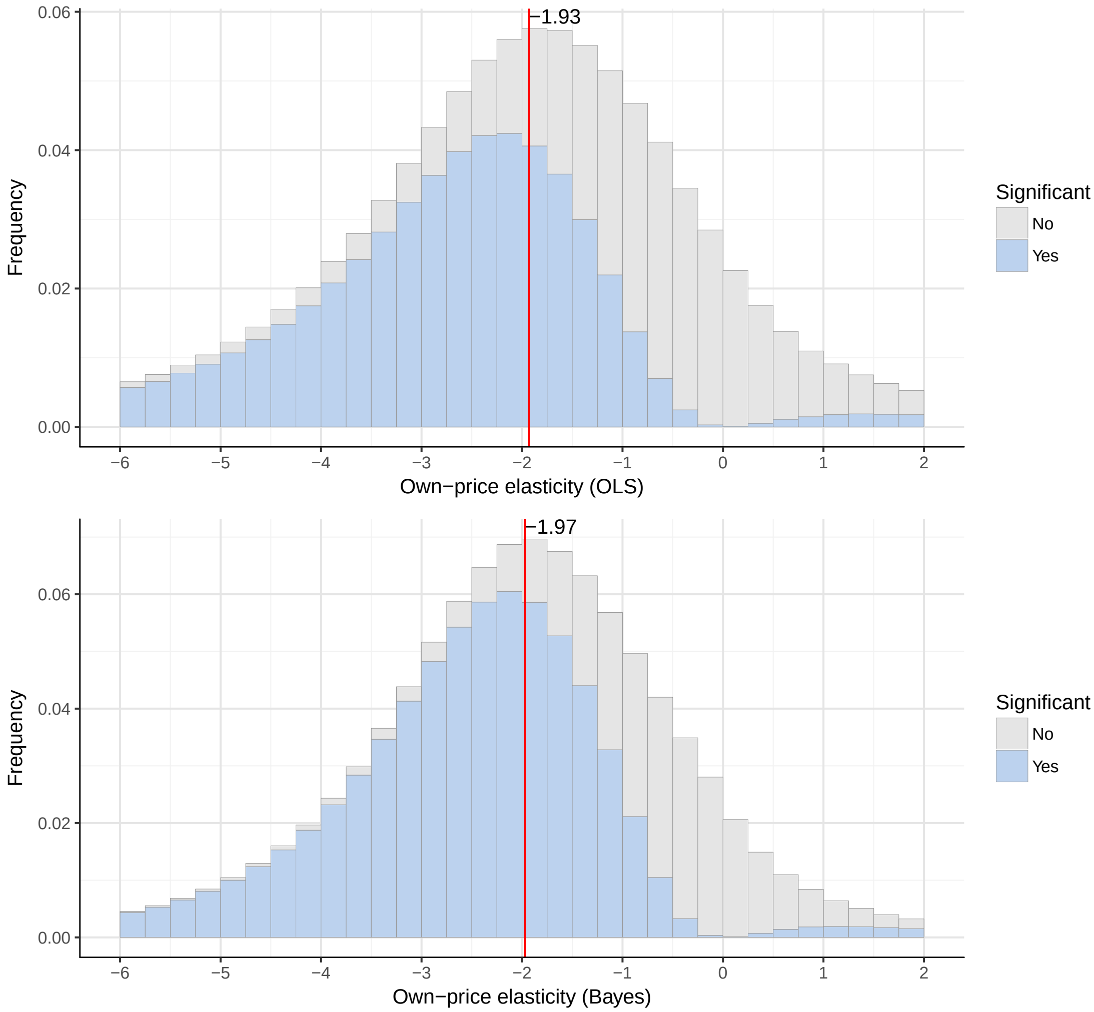

Figure 17: Own-price elasticity estimates

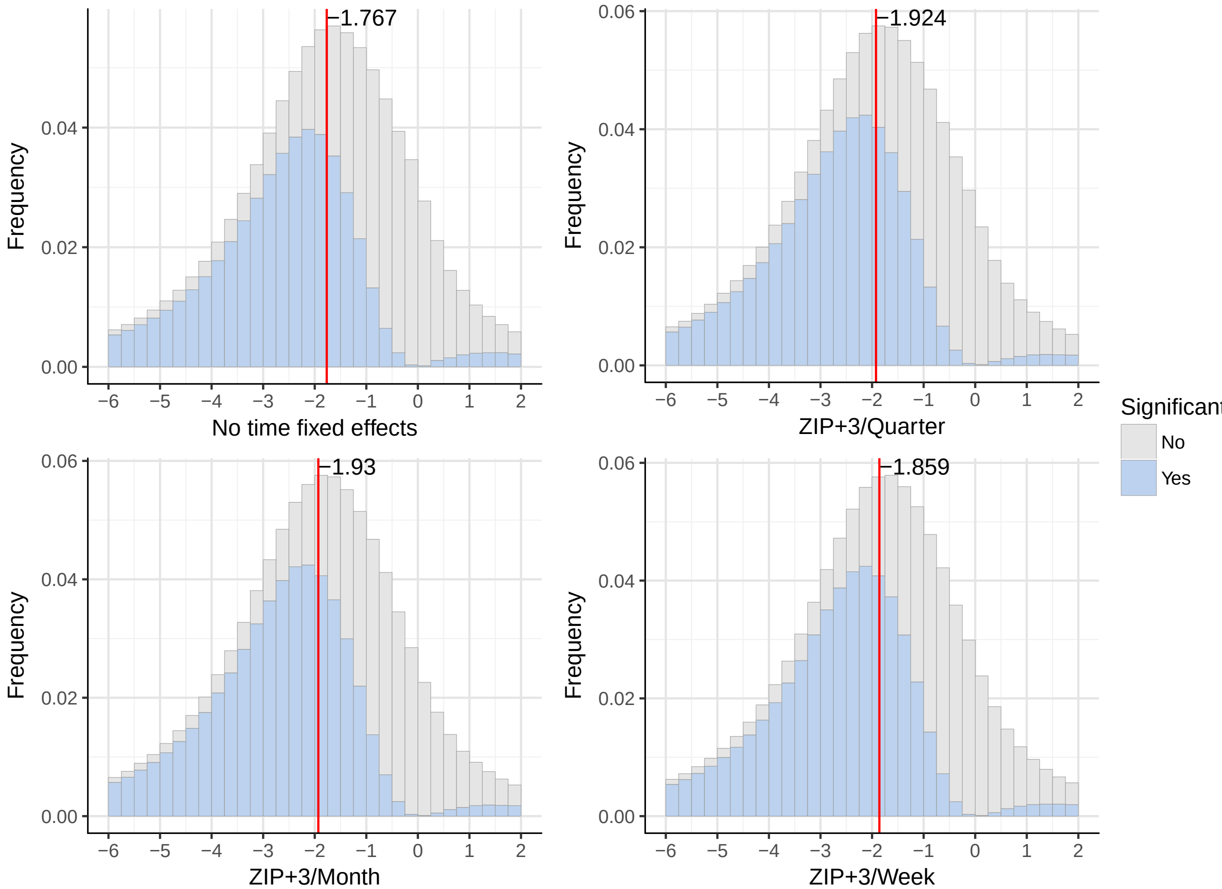

Figure 18: Own-price elasticity estimates for different fixed effects definitions

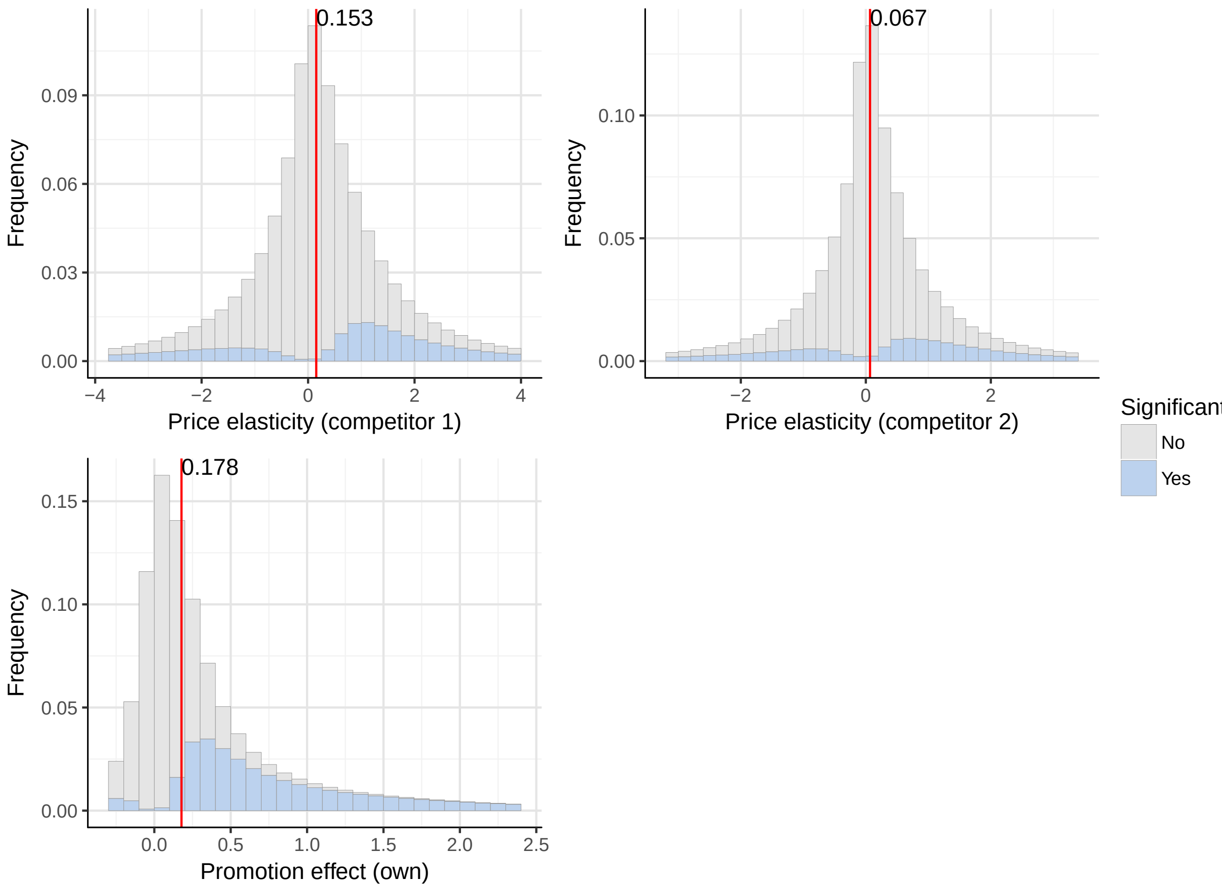

Figure 19: Demand model parameter estimates

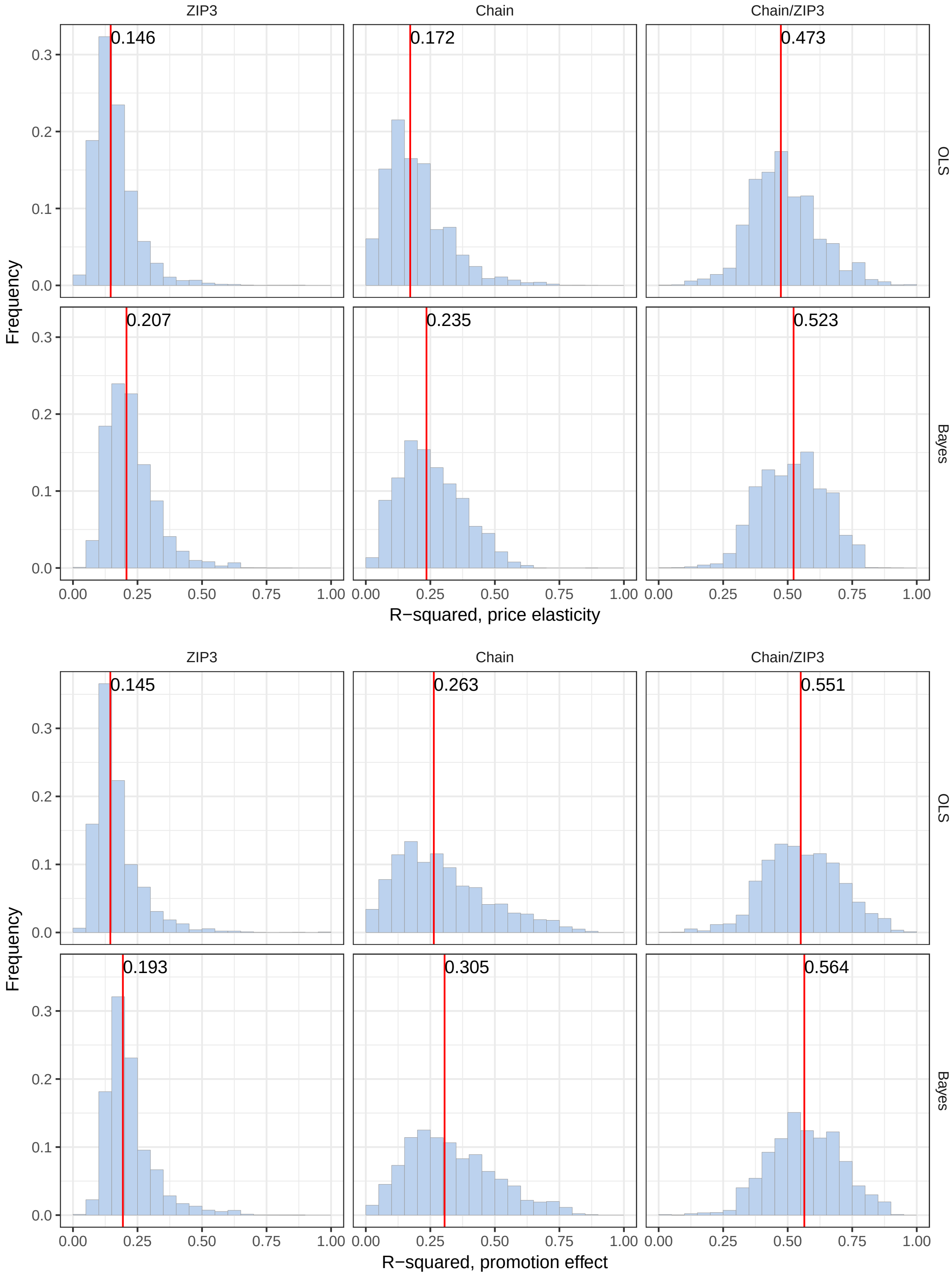

Figure 20: Percentage of variance of own-price elasticities and own-promotion effects explained by market and chain factors

Appendix

A Derivation of price variance decompositions

All decompositions are performed at the product level, hence we drop the subscript j. $\mathcal{M}$ is the set of all markets, $\mathcal{S}_m$ is the set of all stores in market $m \in \mathcal{M}$, and $\mathcal{S} = \cup_{m\in\mathcal{M}}\mathcal{S}_m$ is the set of all stores. For each store s we observe prices in periods $t \in \mathcal{T}_s$. Correspondingly, S_m is the number of stores in market m, $S = \sum_{m\in\mathcal{M}} S_m$ is the number of all stores, and N_s is the number of observations for stores s. Then the total number of observations is $N = \sum_{s\in\mathcal{S}} N_s$, and the number of observations in market m is $N_m = \sum_{s\in\mathcal{S}_m} N_s$.

Define the overall (national) average price, the average price in market m, and the average price in stores s:

$$\bar{p} = \frac{1}{N}\sum_{s\in\mathcal{S}}\sum_{t\in\mathcal{T}_s} p_{st},$$

$$\bar{p}_m = \frac{1}{N_m}\sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s} p_{st},$$

$$\bar{p}_s = \frac{1}{N_s}\sum_{t\in\mathcal{T}_s} p_{st}.$$

Similarly, define the average base price in market m and the average base price in store s :

$$\bar{b}_m = \frac{1}{N_m}\sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s} b_{st},$$

$$\bar{b}_s = \frac{1}{N_s}\sum_{t\in\mathcal{T}_s} b_{st}.$$

Our goal is to provide a decomposition for the overall variance of prices,

$$\text{var}(p_{st}) = \frac{1}{N}\sum_{m\in\mathcal{M}}\sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s}(p_{st}-\bar{p})^2.$$

A.1 Basic decomposition

Define

$$\text{var}(\bar{p}_m) = \frac{1}{N}\sum_{m\in\mathcal{M}} N_m(\bar{p}_m-\bar{p})^2,$$

$$\text{var}(\bar{p}_s|m) = \frac{1}{N_m}\sum_{s\in\mathcal{S}_m} N_s(\bar{p}_s-\bar{p}_m)^2,$$

$$\text{var}(p_{st}|s) = \frac{1}{N_s}\sum_{t\in\mathcal{T}_s}(p_{st}-\bar{p}_s)^2.$$

$\text{var}(p_m)$ is the variance of average market-level prices across markets, $\text{var}(\bar{p}_s|m)$ is the within-market variance of average store-level prices, and $\text{var}(p_{st}|s)$ is the within-store variance of prices over time. Note that $\text{var}(p_m)$ and $\text{var}(\bar{p}_s|m)$ are calculated as weighted averages, using the number of observations in each market and the number of observations for each store as weights.

We first decompose the overall variance of prices, $\text{var}(p_{st})$, into an across-market and a within-market term:

$$\begin{aligned}
\text{var}(p_{st}) &= \frac{1}{N}\sum_{m\in\mathcal{M}}\sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s}(p_{st}-\bar{p})^2 \\
&= \frac{1}{N}\sum_{m\in\mathcal{M}}\sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s}(p_{st}-\bar{p}_m+\bar{p}_m-\bar{p})^2 \\
&= \frac{1}{N}\sum_{m\in\mathcal{M}}\sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s}(p_{st}-\bar{p}_m)^2+\frac{1}{N}\sum_{m\in\mathcal{M}}\sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s}(\bar{p}_m-\bar{p})^2 \\
&= \text{var}(\bar{p}_m)+\frac{1}{N}\sum_{m\in\mathcal{M}}\sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s}(p_{st}-\bar{p}_m)^2. \qquad (5)
\end{aligned}$$

Note that the third line in this formula follows because

$$\sum_{m\in\mathcal{M}}\sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s}(p_{st}-\bar{p}_m)(\bar{p}_m-\bar{p}) = \sum_{m\in\mathcal{M}}(\bar{p}_m-\bar{p})\sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s}(p_{st}-\bar{p}_m)=0.$$

To further decompose the within-market term, note that

$$\begin{aligned}
\sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s}(p_{st}-\bar{p}_m)^2 &= \sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s}(p_{st}-\bar{p}_s+\bar{p}_s-\bar{p}_m)^2 \\
&= \sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s}(p_{st}-\bar{p}_s)^2+\sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s}(\bar{p}_s-\bar{p}_m)^2 \\
&= \sum_{s\in\mathcal{S}_m}N_s\text{var}(p_{st}|s)+N_m\text{var}(\bar{p}_s|m). \qquad (6)
\end{aligned}$$

Here, to derive the second line we used

$$\sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s}(p_{st}-\bar{p}_s)(\bar{p}_s-\bar{p}_m) = \sum_{s\in\mathcal{S}_m}(\bar{p}_s-\bar{p}_m)\sum_{t\in\mathcal{T}_s}(p_{st}-\bar{p}_s)=0.$$

Substituting equation (6) in (5), we obtain the desired decomposition of the overall price variance into the variance of average market-level prices, the weighted average of the within-market variances of average store-level prices, and the weighted average of the within-store variances of prices:

$$\begin{aligned}
\text{var}(p_{st}) &= \text{var}(\bar{p}_m) + \frac{1}{N}\sum_{m\in\mathcal{M}}\sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s}(p_{st}-\bar{p}_m)^2 \\
&= \text{var}(\bar{p}_m) + \frac{1}{N}\sum_{m\in\mathcal{M}} N_m\text{var}(\bar{p}_s|m) + \frac{1}{N}\sum_{m\in\mathcal{M}}\sum_{s\in\mathcal{S}_m} N_s\text{var}(p_{st}|s) \\
&= \text{var}(\bar{p}_m) + \frac{1}{N}\sum_{m\in\mathcal{M}} N_m\text{var}(\bar{p}_s|m) + \frac{1}{N}\sum_{s\in\mathcal{S}} N_s\text{var}(p_{st}|s). \qquad (7)
\end{aligned}$$

A.2 Decomposition into base price and promotion components

We start with an alternative decomposition of the within-market term (6):

$$\begin{aligned}
\sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s}(p_{st}-\bar{p}_m)^2 &= \sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s}(p_{st}-b_{st}+b_{st}-\bar{p}_m)^2 \\
&= \sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s}(p_{st}-b_{st})^2 + \sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s}(b_{st}-\bar{p}_m)^2 + 2\sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s}(p_{st}-b_{st})(b_{st}-\bar{p}_m). \qquad (8)
\end{aligned}$$

Note that

$$\begin{aligned}
\sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s}(b_{st}-\bar{p}_m)^2 &= \sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s}(b_{st}-\bar{b}_m+\bar{b}_m-\bar{p}_m)^2 \\
&= \sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s}(b_{st}-\bar{b}_m)^2 + N_m(\bar{b}_m-\bar{p}_m)^2 \\
&= \sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s}(b_{st}-\bar{b}_s+\bar{b}_s-\bar{b}_m)^2 + N_m(\bar{b}_m-\bar{p}_m)^2 \\
&= \sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s}(b_{st}-\bar{b}_s)^2 + \sum_{s\in\mathcal{S}_m} N_s(\bar{b}_s-\bar{b}_m)^2 + N_m(\bar{b}_m-\bar{p}_m)^2 \\
&= \sum_{s\in\mathcal{S}_m} N_s\text{var}(b_{st}|s) + N_m\text{var}(\bar{b}_s|m) + N_m(\bar{b}_m-\bar{p}_m)^2. \qquad (9)
\end{aligned}$$

Substituting equation (9) in (8) and rearranging terms, we obtain

$$\begin{aligned}
\sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s}(p_{st}-\bar{p}_m)^2 = N_m\text{var}(\bar{b}_s|m) &+ \sum_{s\in\mathcal{S}_m} N_s\text{var}(b_{st}|s) \\
&+ N_m(\bar{b}_m-\bar{p}_m)^2 + \sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s}(b_{st}-p_{st})^2 \\
&- 2\sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s}(b_{st}-p_{st})(b_{st}-\bar{p}_m). \qquad (10)
\end{aligned}$$

Define

$$\begin{aligned}\text{var}(b_{st}-p_{st}|m) &= \frac{1}{N_m}\sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s}\big((b_{st}-p_{st})-(\bar{b}_m-\bar{p}_m)\big)^2\\ &= \frac{1}{N_m}\sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s}(b_{st}-p_{st})^2-(\bar{b}_m-\bar{p}_m)^2. \qquad (11)\end{aligned}$$

Rearranging (11) and substituting in equation (10), we obtain

$$\begin{aligned}\sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s}(p_{st}-\bar{p}_m)^2 = {} & N_m\text{var}(\bar{b}_s|m)+\sum_{s\in\mathcal{S}_m}N_s\text{var}(b_{st}|s)\\ &+N_m\text{var}(b_{st}-p_{st}|m)+2N_m(\bar{b}_m-\bar{p}_m)^2\\ &-2\sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s}(b_{st}-p_{st})(b_{st}-\bar{p}_m). \qquad (12)\end{aligned}$$

Define the within-market covariance between the promotional price discounts, $b_{st}-p_{st}$, and the difference between the store-level base price and the average market price, $b_{st}-\bar{p}$:

$$\begin{aligned}\text{cov}(b_{st}-p_{st},b_{st}-\bar{p}_m|m) &= \frac{1}{N_m}\sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s}\big((b_{st}-p_{st})-(\bar{b}_m-\bar{p}_m)\big)\big((b_{st}-\bar{p}_m)-(\bar{b}_m-\bar{p}_m)\big)\\ &= \frac{1}{N_m}\sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s}(b_{st}-p_{st})(b_{st}-\bar{p}_m)-(\bar{b}_m-\bar{p}_m)^2. \qquad (13)\end{aligned}$$

Rearranging and substituting (13) in equation (12), we then obtain

$$\begin{aligned}\sum_{s\in\mathcal{S}_m}\sum_{t\in\mathcal{T}_s}(p_{st}-\bar{p}_m)^2 = {} & N_m\text{var}(\bar{b}_s|m)+\sum_{s\in\mathcal{S}_m}N_s\text{var}(b_{st}|s)\\ &+N_m\text{var}(b_{st}-p_{st}|m)+2N_m(\bar{b}_m-\bar{p}_m)^2\\ &-2N_m\text{cov}(b_{st}-p_{st},b_{st}-\bar{p}_m|m)-2N_m(\bar{b}_m-\bar{p}_m)^2\\ = {} & N_m\text{var}(\bar{b}_s|m)+\sum_{s\in\mathcal{S}_m}N_s\text{var}(b_{st}|s)\\ &+N_m\text{var}(b_{st}-p_{st}|m)-2N_m\text{cov}(b_{st}-p_{st},b_{st}-\bar{p}_m|m). \qquad (14)\end{aligned}$$

Finally, we substitute (14) in (5) and note that $\text{cov}(b_{st}-p_{st},b_{st}-\bar{p}_m|m)=\text{cov}(b_{st}-p_{st},b_{st}|m)$ to obtain the variance decomposition:

$$\begin{aligned}\text{var}(p_{st}) = {} & \text{var}(\bar{p}_m)\\ &+\frac{1}{N}\sum_{m\in\mathcal{M}}N_m\text{var}(\bar{b}_s|m)+\frac{1}{N}\sum_{s\in\mathcal{S}}N_s\text{var}(b_{st}|s)\\ &+\frac{1}{N}\sum_{m\in\mathcal{M}}N_m\text{var}(b_{st}-p_{st}|m)-2\frac{1}{N}\sum_{m\in\mathcal{M}}N_m\text{cov}(b_{st}-p_{st},b_{st}|m). \qquad (15)\end{aligned}$$

A.3 Example: Price promotions may decrease the overall price variance

Here we provide the derivation of the formula describing the price variance in the example in Section 6.2. Define the promotion indicator $D_{st} = \mathbb{I}\{p_{st} < b_{st}\}$ (this corresponds to the promotion definition in Section 5.2 with a threshold $\bar{\delta} = 0$).

The mean price is given by:

$$\begin{aligned}
\mathbb{E}[p_{st}] &= \frac{1}{2}\mathbb{E}[p_{st}|b_{st} \leq \bar{b}] + \frac{1}{2}\left((1-\pi)\mathbb{E}[p_{st}|D_{st}=0, b_{st} > \bar{b}] + \pi\mathbb{E}[p_{st}|D_{st}=1, b_{st} > \bar{b}]\right) \\
&= \frac{1}{2}\left(\bar{b} - \frac{\nu}{2}\right) + \frac{1}{2}\left((1-\pi)\left(\bar{b} + \frac{\nu}{2}\right) + \pi\bar{b}\right) \\
&= \bar{b} - \frac{\pi\nu}{4}.
\end{aligned}$$

The across-store variance of base prices is given by the variance of a uniform distribution,

$$\text{var}(\bar{b}_s) = \frac{\nu^2}{3}.$$

The derive the variance of the promotional price discounts we first calculate

$$\begin{aligned}
\mathbb{E}\left[(b_{st} - p_{st})^2\right] &= \frac{1}{2}\pi\mathbb{E}\left[(b_{st} - p_{st})^2|D_{st}=1, b_{st} > \bar{b}\right] \\
&= \frac{1}{2}\pi\mathbb{E}\left[(b_{st} - \bar{b})^2|D_{st}=1, b_{st} > \bar{b}\right] \\
&= \frac{1}{2}\pi\int_{\bar{b}}^{\bar{b}+\nu}(x - \bar{b})^2\frac{1}{\nu}dx \\
&= \frac{\pi\nu^2}{6}.
\end{aligned}$$

Similarly, to derive the covariance between the promotional price discounts and the base price we use the expression for $\mathbb{E}\left[(b_{st} - p_{st})^2\right]$ above to obtain

$$\begin{aligned}
\mathbb{E}\left[(b_{st} - p_{st})(b_{st} - \bar{b})\right] &= \frac{1}{2}\pi\mathbb{E}\left[(b_{st} - p_{st})(b_{st} - \bar{b})|D_{st}=1, b_{st} > \bar{b}\right] \\
&= \frac{1}{2}\pi\mathbb{E}\left[(b_{st} - \bar{b})^2|D_{st}=1, b_{st} > \bar{b}\right] \\
&= \frac{\pi\nu^2}{6}.
\end{aligned}$$

Also, the squared difference between the mean base price and shelf price is:

$$(\bar{b} - \bar{p})^2 = \frac{\pi^2\nu^2}{16}.$$

Hence,

$$\begin{aligned}
\text{var}(b_{st} - p_{st}) &= \mathbb{E}\left[(b_{st} - p_{st})^2\right] - (\bar{b} - \bar{p})^2 \\
&= \frac{\pi\nu^2}{6} - \frac{\pi^2\nu^2}{16}.
\end{aligned}$$

Also,

$$\begin{aligned}\operatorname{cov}(b_{st}-p_{st},b_{st}) &= \operatorname{cov}(b_{st}-p_{st},b_{st}-\bar{b}) \\ &= \mathbb{E}\left[(b_{st}-p_{st})(b_{st}-\bar{b})\right] \\ &= \frac{\pi\nu^2}{6}.\end{aligned}$$

Combining the three components we obtain the variance of prices,

$$\begin{aligned}\operatorname{var}(p_{st}) &= \operatorname{var}(\bar{b}_s)+\operatorname{var}(b_{st}-p_{st})-2\operatorname{cov}(b_{st}-p_{st},b_{st}) \\ &= \frac{\nu^2}{3}+\frac{\pi\nu^2}{6}-\frac{\pi^2\nu^2}{16}-2\frac{\pi\nu^2}{6} \\ &= \nu^2\left(\frac{1}{3}-\frac{\pi}{6}-\frac{\pi^2}{16}\right).\end{aligned}$$

B Bayesian hierarchical demand model

The approach to estimate a Bayesian hierarchical linear regression model that we use is standard in the literature. See Rossi et al. (2005), Chapter 3.7, for a detailed exposition.

We do not attempt to estimate the large number of 3-digit ZIP code/month fixed effects, $\tau_j(s,t)$. Instead, we first project all variables that enter the demand model onto the fixed effects and then use the residuals from this projection to estimate the store-level demand parameters in the model:

$$\begin{aligned}\log(1+q_{st}) &= \alpha_s + \sum_{k\in\mathcal{J}_s}\beta_{ks}\log(p_{kst}) + \sum_{k\in\mathcal{J}_s}\gamma_{ks}D_{kst} + \epsilon_{st},\\ \epsilon_{st} &\sim N(0,\sigma_s^2).\end{aligned}$$

ϵ_{st} is i.i.d. across stores and time. Note that we drop the brand index j and that we do not distinguish between the original and the residualized data to simplify the notation. The parameter vector θ_s includes the store-specific intercept, α_s, the own and cross-price elasticities, β_{jks}, and the promotion parameters, γ_{jks}.

The store-level demand parameters, θ_s, are drawn from a common first-stage prior distribution:

$$\theta_s \sim N(\mu, V_\theta).$$

These draws are independent, conditional on μ and V_θ. We further specify the second-stage prior distribution of V_θ and μ:

$$\begin{aligned}V_\theta &\sim \mathrm{IW}(\nu, V),\\ \mu|V_\theta &\sim N(\bar{\mu}, V_\theta\otimes A^{-1}).\end{aligned}$$

IW denotes an inverse Wishart distribution. The error variances, σ_s^2, are independent draws from an inverse chi-squared distribution,

$$\sigma_s^2 \sim \frac{\nu_\epsilon r_s^2}{\chi^2_{\nu_\epsilon}}.$$

Here, ν_ϵ denotes the degrees of freedom and r_s^2 is a scale parameter.

The MCMC algorithm to obtain the posterior distribution of the model parameters is performed using Peter Rossi's `bayesm` package[15] in R. We run the algorithm using the default settings

[15] `https://cran.r-project.org/web/packages/bayesm/index.html`

for the hyper-parameters in the `bayesm` package:

$$\begin{aligned} v &= 3 + n, \\ V &= \nu I_n, \\ \bar{\mu} &= 0, \\ A &= 0.01, \\ \nu_\epsilon &= 3, \\ r_s^2 &= \text{var}(\log(1 + q_{st})). \end{aligned}$$

Here, n is the dimension of the parameter vector θ_s.

We choose a chain length of 20,000 (after 2,000 initial burn-in draws) and keep every 10th draw to calculate the posterior means and the 95 percent credible intervals of the parameters. A visual inspection of the trace plots for a large number of randomly selected parameters (across brands and stores) indicates convergence of the chain.

C Additional results

C.1 Price dispersion: Sensitivity analysis

We calculate two alternative dispersion statistics that are related to the standard deviation of log-prices. First, the distribution of percentage price differences can be measured using the standard deviation of prices normalized relative to the mean price (nationally or at the market level), $p_{jst}/\bar{p}_{jmt}$, which is the approach used in Kaplan and Menzio (2015). Second, we can report the square root of the variance of log-prices calculated using the following approach:

$$\begin{aligned} \text{var}(\log(p_{jst})|m) &= \frac{1}{N_{jmt}} \sum_{s \in \mathcal{S}_{jmt}} (\log(p_{jst}) - \overline{\log(p_{jmt})})^2, \\ \text{var}(\log(p_{jst})) &= \frac{1}{N_{jt}} \sum_{m \in \mathcal{M}} N_{jmt} \text{var}(\log(p_{jst})|m). \end{aligned} \tag{16}$$

Note that we do not use Bessel's correction in these two variance formulas. This approach is equivalent to demeaning each $\log(p_{jst})$ observation with respect to the average log price in market m, and then calculating the variance over all observations. We include this approach because it is more closely related to the variance decomposition in Section 6.

Summary statistics for these two alternative approach are shown in Table 11, separately for products defined as UPC's and brands. As expected the difference between the dispersion statistics based on the standard deviation of log prices and the standard deviation of normalized prices is negligible. On the other hand, the standard deviation calculated as the square root of (16) is slightly larger at the DMA and 3-digit ZIP code level compared to the standard deviation of the log of prices. Overall, our main conclusions are unchanged using these two alternative dispersion statistics.

Table 11: Additional price and base price dispersion statistics

		Median	Mean	Percentiles							
				0.01	**0.05**	**0.1**	**0.25**	**0.75**	**0.9**	**0.95**	**0.99**
Prices											
Product definition: UPC											
Normalized price SD	National	0.157	0.160	0.014	0.058	0.088	0.122	0.197	0.233	0.267	0.327
	DMA	0.107	0.112	0.006	0.031	0.049	0.076	0.142	0.181	0.205	0.258
	ZIP+3	0.097	0.101	0.002	0.021	0.038	0.066	0.132	0.170	0.192	0.247
Demeaned log-price SD	National	0.161	0.163	0.014	0.058	0.090	0.124	0.200	0.238	0.268	0.318
	DMA	0.121	0.125	0.010	0.040	0.060	0.089	0.158	0.195	0.218	0.265
	ZIP+3	0.113	0.116	0.002	0.026	0.048	0.080	0.151	0.187	0.209	0.257
Product definition: Brand											
Normalized price SD	National	0.174	0.187	0.073	0.097	0.115	0.140	0.217	0.263	0.330	0.450
	DMA	0.135	0.145	0.045	0.066	0.082	0.104	0.173	0.216	0.251	0.388
	ZIP+3	0.127	0.135	0.035	0.059	0.071	0.096	0.162	0.203	0.237	0.362
Demeaned log-price SD	National	0.175	0.185	0.074	0.098	0.114	0.141	0.216	0.260	0.319	0.427
	DMA	0.146	0.154	0.055	0.073	0.092	0.114	0.181	0.223	0.261	0.364
	ZIP+3	0.139	0.145	0.044	0.067	0.081	0.107	0.172	0.209	0.244	0.332

Table 12: Details of price variance decompositions

	Median	**Mean**	**% > 0**	**Percentiles**							
				1%	**5%**	**10%**	**25%**	**75%**	**90%**	**95%**	**99%**
UPC's											
Basic decomposition											
Across-market	0.269	0.327		0.009	0.068	0.100	0.164	0.444	0.662	0.774	0.980
Across-store	0.251	0.270		0.000	0.028	0.076	0.158	0.373	0.486	0.557	0.670
Within-store	0.396	0.403		0.000	0.033	0.075	0.214	0.578	0.719	0.802	0.977
Decomposition into base prices and promotions											
Across-market	0.269	0.327		0.009	0.068	0.100	0.164	0.444	0.662	0.774	0.980
Across-store mean base price variance	0.303	0.313		0.000	0.028	0.084	0.182	0.432	0.545	0.612	0.760
Within-store base price variance	0.096	0.123		0.000	0.008	0.022	0.051	0.160	0.244	0.323	0.622
Total contribution of promotions	0.188	0.237	0.964	-0.096	0.000	0.007	0.070	0.359	0.537	0.653	0.938
Promotional price discounts	0.329	0.360		0.000	0.005	0.029	0.143	0.546	0.735	0.828	0.975
EDLP vs. Hi-Lo adjustment	-0.084	-0.123	0.103	-0.772	-0.375	-0.286	-0.183	-0.017	0.000	0.004	0.040
Covariance between price discounts and store base price level											
	0.042	0.062	0.920	-0.020	-0.002	0.000	0.009	0.091	0.143	0.188	0.386
Brands											
Basic decomposition											
Across-market	0.242	0.297		0.066	0.107	0.133	0.175	0.372	0.564	0.674	0.794
Across-store	0.420	0.423		0.052	0.157	0.201	0.291	0.555	0.652	0.710	0.793
Within-store	0.251	0.280		0.020	0.043	0.062	0.139	0.394	0.528	0.616	0.776
Decomposition into base prices and promotions											
Across-market	0.242	0.297		0.066	0.107	0.133	0.175	0.372	0.564	0.674	0.794
Across-store mean base price variance	0.513	0.499		0.053	0.167	0.234	0.351	0.642	0.752	0.812	0.881
Within-store base price variance	0.094	0.134		0.008	0.023	0.033	0.053	0.159	0.272	0.328	0.755
Total contribution of promotions	0.059	0.070	0.692	-0.613	-0.255	-0.101	-0.012	0.174	0.312	0.411	0.635
Promotional price discounts	0.242	0.299		0.006	0.027	0.053	0.126	0.426	0.621	0.722	0.990
EDLP vs. Hi-Lo adjustment	-0.151	-0.229	0.028	-1.466	-0.785	-0.457	-0.260	-0.067	-0.021	-0.008	0.017
Covariance between price discounts and store base price level											
	0.076	0.115	0.974	-0.008	0.004	0.010	0.033	0.130	0.229	0.393	0.733

Table 13: Percentage of variance of prices, promotion frequency, and promotion depth explained by market and chain factors

	Median	**Mean**	**Percentiles**							
			0.01	**0.05**	**0.1**	**0.25**	**0.75**	**0.9**	**0.95**	**0.99**
Price										
Market (ZIP+3)	0.465	0.490	0.198	0.241	0.266	0.341	0.620	0.757	0.820	0.932
Chain	0.699	0.652	0.001	0.299	0.410	0.573	0.774	0.826	0.855	0.917
Market/chain	0.881	0.845	0.374	0.589	0.707	0.812	0.921	0.946	0.957	0.978
Promotion frequency										
Market (ZIP+3)	0.361	0.389	0.125	0.182	0.211	0.270	0.477	0.611	0.708	0.879
Chain	0.632	0.587	0.000	0.019	0.177	0.444	0.747	0.822	0.856	0.910
Market/chain	0.800	0.759	0.287	0.443	0.541	0.688	0.873	0.910	0.929	0.963
Promotion depth										
Market (ZIP+3)	0.380	0.407	0.148	0.191	0.215	0.279	0.505	0.649	0.753	0.965
Chain	0.589	0.562	0.000	0.005	0.177	0.402	0.724	0.804	0.840	0.894
Market/chain	0.807	0.772	0.305	0.483	0.589	0.711	0.875	0.913	0.930	0.980

Table 14: Percentage of variance in estimated own-price elasticities and own-promotion effects explained by market and chain factors

	Median	Mean	Percentiles							
			0.01	**0.05**	**0.1**	**0.25**	**0.75**	**0.9**	**0.95**	**0.99**
OLS estimates										
Own-price elasticity										
Market (ZIP+3)	0.146	0.163	0.047	0.072	0.082	0.108	0.197	0.263	0.311	0.476
Chain	0.172	0.195	0.016	0.042	0.068	0.110	0.250	0.353	0.412	0.601
Market/chain	0.473	0.486	0.186	0.292	0.341	0.394	0.563	0.668	0.732	0.827
Own-promotion effect										
Market (ZIP+3)	0.145	0.170	0.057	0.077	0.089	0.111	0.198	0.283	0.346	0.519
Chain	0.263	0.297	0.013	0.067	0.087	0.158	0.409	0.564	0.653	0.772
Market/chain	0.551	0.558	0.209	0.338	0.381	0.455	0.659	0.749	0.807	0.876
Bayesian hierarchical model estimates										
Own-price elasticity										
Market (ZIP+3)	0.207	0.224	0.082	0.106	0.124	0.158	0.269	0.340	0.400	0.563
Chain	0.235	0.252	0.039	0.072	0.099	0.161	0.334	0.421	0.479	0.563
Market/chain	0.523	0.520	0.239	0.323	0.356	0.424	0.608	0.681	0.724	0.793
Own-promotion effect										
Market (ZIP+3)	0.193	0.217	0.087	0.109	0.126	0.157	0.247	0.328	0.407	0.597
Chain	0.305	0.332	0.044	0.087	0.130	0.202	0.439	0.568	0.657	0.773
Market/chain	0.564	0.567	0.250	0.347	0.392	0.471	0.667	0.744	0.801	0.867

Figure 21: Predicted base prices, Tide liquid laundry detergent (70 oz)

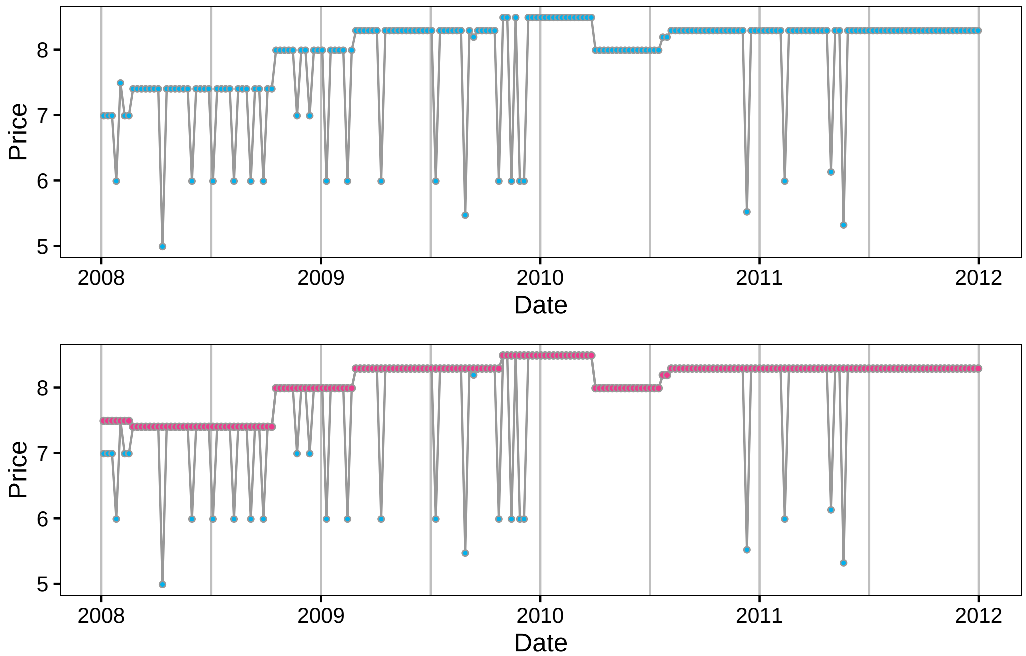

Figure 22: Predicted base prices, Kellogg's Raisin Bran (20 oz)