Multiproduct Search and Retail Pricing: Some Empirical Results

Hyunchul Kim*

and

Jungwon Yeo†

April 30, 2017

Abstract

The existence of fixed costs of shopping (or search) and consumers' tendency to shop for multiple products induces multiproduct retailers to price otherwise independent products with consideration for one another. Theoretical papers have provided insight into how multiproduct retailers price their products in such an environment. We herein empirically test theories on multiproduct pricing, focusing on Lar and Matutes (1994), Chen and Ray (2012) and Rhodes (2015). Using supermarket scanner data, we provide evidence supporting the exploitive cross-subsidization proposed by Chen and Ray (2012). We also find the effect of economies of scale in search on firms' pricing behavior characterized by Rhodes (2015).

*Department of Economics, Sungkyunkwan University; hchkim@skku.edu

†Corresponding author. School of Economics, Singapore Management University; jwyeo@smu.edu.sg

1 Introduction

In this paper, we empirically examine some predictions made by theoretical work on multiproduct firms' pricing and advertising behavior when they face consumers who pay fixed costs to visit them. When consumers are uninformed about prices or product-consumer match valuations of the multiple products they wish to purchase, such costs are called search costs. Otherwise, they can be understood as time or transaction costs for making a trip to a store and waiting in line for check-out. Regardless of the level of information consumers possess, the existence of fixed costs creates a hold-up problem: once consumers are in a store, their reservation surplus is reduced. This hold-up problem motivates multiproduct firms to set the price of each product in consideration of its effects on the sales of other products.

We focus on three theoretical papers: Lal and Matutes (1994), Chen and Rey (2012), and Rhodes (2015). They all assume that consumers face fixed costs of shopping (or search), but they differ in whether consumers are informed about prices and heterogeneous in their valuations for multiple products or shopping costs, and whether stores are heterogeneous.

Lal and Matutes (1994) (hereafter, L-M) consider a multiproduct firm's pricing when faced with consumers who are imperfectly informed about prices and advertising is costly. This setting is shared by Rhodes (2015) (hereafter, Rhodes), but the latter assumes that consumers are heterogeneous in their valuations, whereas in the former, consumers are heterogeneous in their proximity to each store, as captured by the Hotelling's linear city set-up.[1] Both L-M and Rhodes allow firms to engage in costly advertising to pass price information onto consumers. In Chen and Rey (2012) (hereafter, C-R), both consumers, who are perfectly informed about prices, and firms are heterogeneous: consumers are heterogeneous in their shopping costs, and firms are heterogeneous in their product offerings. C-R focus on loss-leader pricing, which can be viewed as an extreme case of cross-subsidization pricing.

These differences in settings lead the studies to generate distinct predictions that pivot on each setting. In L-M, two symmetric retailers advertise the same product and set the price of the advertised product below the common valuation H, but the unadvertised product exactly at H. In Rhodes, a local monopolist sets an advertised price below the single product monopoly price and unadvertised prices strictly higher than that, while the relationship between unadvertised prices and advertised prices is positive. C-R predict an exploitative use of loss-leader pricing: the entry of a hard-discount store with limited product offerings engenders multistop shopping. Faced with a mix of one-stop and multistop shoppers, a full-line

[1] Rhodes (2015) extends consumer heterogeneity to loyal and non-loyal consumers when introducing competition to his model.

store raises the price of the product it sells to both multistop and one-stop shoppers (monopolized product) while reducing the price of the product that only one-stop shoppers purchase (competitive product) . Thus, the full-line store's pricing exploits multistop shoppers—those who visit both stores to cherry-pick—while keeping the price of a one-stop shopper's shopping basket constant (Proposition 1 in C-R).

We begin our investigation with a prediction from C-R; that is, we examine whether and how the presence of a small store with limited product offerings affects the pricing of a large store that offers a broader range of products. More specifically, we see whether the large store increases the differentials between the prices of products that both one-stop and multistop shoppers are likely to purchase and those that only one-stop shoppers are likely to purchase. Regardless of the result, this test will unveil the basic premise behind multiproduct retailing pricing: Unless there are fixed costs for visiting a store, or natural complementarities among products, the entry of a limited-line store should not affect the full-line store's pricing of products whose stand-alone availabilities are unaffected by the entry.

We then extend our investigation to the relationship between a store's product range and its pricing and advertising policies, based on Rhodes. In his model, a monopoly firm with a broad product range has little incentive to advertise (Proposition 4 in Rhodes) because consumers can infer, from the broadness of the product range, that the store sets low prices (Proposition 2). Rhodes notes, however, that when there is competition, a large product range makes it more profitable for the firm to steal business from its rivals, and therefore advertising might increase with an increased product range.

Finally, we examine the relationship between advertised prices and unadvertised prices. L-M and Rhodes make distinct predictions in this regard. According to L-M, the two prices should not show any systematic relationship (Proposition 2 in L-M) because the store sets the prices of unadvertised products at levels that fully extract consumer surplus regardless of the advertised prices. Rhodes predicts a positive relationship (Lemma 2 and Proposition 5 in Rhodes): a lower advertised price on a product will shift the pool of shoppers towards lower valuations, and thus the store will find it optimal to lower the prices of unadvertised products as well.

C-R is silent about this question, and rightly so given that in their setting consumers are perfectly informed about prices. One way to think about advertised prices, however, is that they are simply promotional prices that reflect temporary price reductions. Under this interpretation, one can deduce a promotional pricing pattern that can arise under cross-subsidization. The (temporarily) reduced prices on some products, especially competitive products, would enable the store to raise the prices on other products, possibly monopolized products. The underlying assumption is that such an event occurs for exogenous and isolated reasons, such as demand or cost shocks confined to the products whose prices are temporarily reduced.

Before we summarize our findings, we should note that the three papers do not make many opposing predictions regarding multiproduct retail pricing. Instead, they ask different questions and have different focuses. Therefore, our goal is not to compare them against each other but to evaluate each theory's unique prediction on its own merits against our empirical findings, and provide further empirical guidance to interested theorists.

Our regression analyses find evidence for cross-subsidization pricing that arises from store heterogeneity as in C-R. Specifically, when we examine the prices of large, full-line stores, we find that 1) there are price differentials among products depending on how competitively they are supplied—i.e., lower prices for products that are more widely available—and 2) the price differentials increase when there is a small, limited-line store in the market, compared with markets without a small store. This finding is consistent with C-R's prediction both in the direction of cross-subsidization (rarely-found products subsidize widely available products) and in the environment that likely induces cross-subsidization (co-existence of one-stop and multistop shoppers).

Additionally, when we examine week-to-week price deviations from the regular prices of products sold by large stores in markets with a small store, we find that the weekly deviations for competitive products are negatively related to those for monopolized products. This pricing pattern indicates that when stores offer price promotions for competitive products, they raise their prices of monopolized products, thereby keeping the price of one-stop shoppers' shopping baskets more or less constant. This finding corroborates C-R's cross-subsidization pricing by demonstrating a pattern of promotional pricing that can arise under the cross-subsidization pricing policy.

Next, we find that stores' product ranges are positively related to advertising activity, which is at odds with Rhodes's prediction drawn under monopoly but is consistent with his conjecture for a case of (imperfect) competition. As for the relationship between product ranges and prices, we find it negative, as Rhodes proposes. However, it becomes positive when chain- and product-specific factors are controlled for through fixed effects. This motivates us to investigate the relationship at the chain level, and we find it negative. This result, together with the previous results, suggests that both product ranges and pricing and advertising policies might be determined at the chain level. Although our empirical findings are largely in accordance with Rhodes's prediction, we cannot corroborate that the mechanism that generates the negative relationship is the low-price image effect of a broad range of products as in Rhodes. Other explanations can be offered. For example, cost factors—e.g., economies of scale from operating a mega-store with many products—are also likely to generate a negative relationship, especially at the chain level. Walmart, for instance, has been known to use its high bargaining power to pressure its suppliers to cut costs.

Finally, when we relate the weekly price deviations of advertised products to those of unadvertised products, we find a negative relationship. However, we argue that the mechanism through which advertised prices are positively related to unadvertised prices—which we call the *signaling effect* of advertised prices—might not be reflected in week-to-week price deviations if consumers update their beliefs about prices infrequently or with a time lag. To unveil the mechanism, we then look at the sign of the relationship at each store and relate it to the store's promotional and regular pricing pattern. The basis for this analysis is that stores that adopt a policy that regularly charges low prices—the so-called everyday-low-price (EDLP) policy—are less subject to the signaling effect of advertised prices, whereas stores that adopt a policy that keeps regular prices high but offers deep, frequent price cuts—the so-called PROMO (also known as Hi-Lo) policy—are likely more susceptible to that effect.

We indeed find that stores characterized by a positive relationship between advertised and unadvertised prices exhibit the pricing characteristics of the PROMO policy, whereas stores with a negative relationship exhibit the pricing characteristics of the EDLP policy. We take this as evidence for the existence of the mechanism that draws the positive relationship prediction in Rhodes.

Our work makes contributions to several strands of the economic and marketing literature. First, we add to the literature on retail pricing, especially multiproduct retail pricing. Besides the aforementioned three papers and several related papers that we discuss in detail in the next section, there is a large body of theoretical work on this subject. For example, McAfee (1995) shows the existence of equilibria in the multiproduct search environment that are qualitatively different from the unique equilibrium in the single product case. A more recent example is Hosken and Reiffen (2007); they show that the multiproduct search environment engenders the use of intertemporal price changes for price discrimination. Shelegia (2012) characterizes the relationship between the prices of two goods under the assumption that consumer valuations for the two goods are related—as opposed to the assumption that fixed shopping costs give rise to the interdependence between two prices.

We contribute to the literature by providing empirical evidence for some theoretical predictions. Compared to the vast theoretical literature on retail pricing, only a few attempts have been made to put those theories under empirical scrutiny. Among the few, Chevalier et al. (2003) show that retail prices are counter-cyclical and that the loss-leader pricing that L-M propose is the most probable explanation for that phenomenon. In contrast, Nevo and Hatzitaskos (2006) find another explanation more plausible—a change in price sensitivity and brand preferences. Hosken and Reiffen (2004) show that the typical grocery product has a "regular" price with temporary downward price deviations and use the finding to reject some price distributions drawn under a single-product environment. Our paper differs from these papers in that we focus on theoretical predictions for which multiproduct shopping/search is the critical building block.

There has been relatively more empirical work documenting shopping and pricing patterns in relation to multiproduct shopping and search. For example, Smith and Thomassen (2012) document empirical features of grocery shopping and find that the primary shopping trip—in terms of expenditure—is more likely done in full-line stores, whereas secondary shopping trips are made at various types of grocery stores. Mulhern and Padgett (1995) find that consumers who visit a store to purchase a product on sale also purchase regular-price products. Several empirical papers estimate shopping/search costs based on a structural approach. Florez-Acosta and Herrera-Araujo (2014) estimate consumer shopping costs using French scanner data. They document that many consumers engage in multistop shopping. Richards et al. (2016) estimate consumer search costs in U.K. online grocery stores while highlighting the importance of accounting for multiproduct search.

We also contribute to the literature on supermarket pricing by documenting pricing patterns in the supermarket industry and providing interpretations based on theories on multiproduct retail pricing. Marketing scientists and economists have made many efforts to unveil supermarket pricing mechanisms. For example, Besanko et al. (2005) examine a supermarket chain's pass-through behavior to find cross-brand pass-through effects, and Pesendorfer (2002) finds evidence for intertemporal effects in demand for ketchup using supermarket sales and pricing data. Our work is distinguished by our effort to establish evidence for pricing practices that are directly related to the multiproduct feature of supermarket pricing.

A few papers in this line of literature provide guidance for some of the approaches we take in this paper. Bell and Lattin (1998) find that large-basket and infrequent shoppers tend to prefer EDLP stores to PROMO stores. Lal and Rao (1997) develop a duopoly model in which stores make pricing and advertising decisions while competing for both time-constrained and cherry-picking consumers; they show that the EDLP policy attracts the former and the

PROMO policy attracts the latter. Ellickson and Misra (2008) estimate a discrete choice model of a store's decision as to whether to adopt EDLP or PROMO, and find that the choice is made in consideration of local consumer demographics.

The remainder of the paper proceeds as follows. In the next section, we describe the three papers in more detail, especially in relation to the literature on multiproduct retail pricing. In section 3, we describe the IRI scanner data used for the tests. In section 4, we state the propositions to be tested, formulate them into regression equations, and present the test results one by one. Section 5 concludes.

2 Theories on Multiproduct Pricing

Interdependence in the demand for various products makes multiproduct firms' pricing different from single-product firms' pricing. Traditionally, natural complementarities and substitutability have been identified and studied as a source of interdependence. Early work on multiproduct retail pricing goes back to Cournot (1838), who showed that when goods are complements. a multiproduct monopolist sets a price for each product lower than the price a single-product monopolist would set. Bliss (1988) shows that the margins a multiproduct retail store sets reflect cross-price elasticities among the products it carries.

One strand of literature has considered multiproduct firms' pricing in the context of bundle pricing. Those researchers pose various questions regarding bundling and bundle pricing, such as when bundling—pure versus mixed—is optimal under different market structures. Some papers in this strand consider fixed search costs as the source of demand interdependence instead of natural complementarities or substitutability. A recent example is Armstrong and Vickers (2010), who show that the bundling discount (or one-stop shopping discount) is positive under imperfect competition, in contrast to the monopoly case, and in some cases, the discount becomes smaller as the search cost increases.

L-M consider multiproduct pricing under the Hotelling's linear city setting—that is, consumers are heterogeneous in their proximity to each store. Prices are unknown, and advertising is costly. Because consumers are uninformed about prices unless advertised, but rationally expect the unadvertised prices to be set to extract all of the consumer surplus, stores need to guarantee a positive surplus from a visit. Under the assumption that stores, consumers, and goods are symmetric, only one-stop shoppers arise in equilibrium. Faced with high enough fixed costs for advertising, stores advertise only one good and the same good in equilibrium. The price of the advertised good is strictly below the price the firm would set in a single-product environment; the price of the unadvertised good is equal to the common consumer valuation, and the positive profit on the unadvertised good can drive the advertised price below cost.

Rhodes makes a prediction different from L-M regarding the relationship between advertised and unadvertised prices. He studies a monopoly multiproduct retailer's pricing when consumers pay a fixed cost to learn prices. Unlike in L-M, however, consumers are heterogeneous in their valuations for n products but homogeneous in their fixed cost of search. This is a critical departure because, for example, a lower (expected or advertised) price on a product not only increases the number of searchers but also lowers the average valuation of the searchers, which is the key mechanism behind the positive relationship between advertised and unadvertised prices that we seek to examine.[2]

The mechanism also generates two other predictions we seek to test: the breadth of a store's product range is negatively related to its price levels and advertising efforts. Under Rhodes's setting, one more product offering induces additional consumers to embark search. Because they have lower valuations than the consumers who are unaffected by the additional offering, the average consumer valuation is now lower, and so the store finds it optimal to charge lower prices. It also follows that a local monopolist with many product offerings does not have much incentive to engage in costly advertising because its large product range is another signal for low prices. He notes, however, that the negative relationship between a store's product offerings and its advertising effort is likely to be reversed when the store faces competition in which more product offerings imply a higher benefit from stealing business from rivals.

Zhou (2014) draws a conclusion similar to that of Rhodes. He considers an environment in which prices are unknown to consumers, but unlike L-M and Rhodes, consumers must visit a store to learn prices and match valuations. Given economies of scale in search, a store that lowers the prices of some products can induce more consumers who have visited the store to stop their search and buy other products there also. This effect, which he terms the *joint search effect*, leads multiproduct retailers to charge less than single-product retailers in a duopoly market.[3]

Ellison (2005) considers a setting similar to that in L-M to study multiproduct firms' pricing in the context of add-on pricing. Consumers are heterogeneous in terms of marginal

[2] McAfee (1995) shows the existence of a class of equilibria in which profits for each of the multiple products are invariant to price and prices for multiple products are negatively correlated.

[3] Zhou (2014) also shows that if the joint search effect dominates the negative effect of low margins, higher search costs would lead to lower prices, in contrast to the single product search case.

utility of income and face a constant and common cost of visiting a store. He shows that there arise equilibria in which all consumers are one-stop shoppers, but only high type consumers buy the add-on.[4]

Chen and Rey (2012, 2016) differ from the above papers in that multistop shoppers arise in equilibrium and are profitable to multiproduct retailers, despite the fact that they are cherry-pickers. Consumers are perfectly informed about prices and heterogeneous in their fixed shopping/travel costs. Stores are also heterogeneous. In Chen and Rey (2012), stores are heterogeneous in their product offerings: a large store (store L) sells products A and B, and a hard-discount, specialty store (store S) sells only product B. In Chen and Rey (2016), two stores offer the same product range, but they have comparative advantages (in net consumer value, which is the consumer valuation minus the cost of production) in different products. In both studies, the key ingredients are the store and consumer heterogeneity, which lead both one-stop and multistop shoppers to exist in equilibrium.

Under the former setting, multistop shoppers buy the monopolized product (product A) in store L and the competitive product (product B) in store S. Consumers with high shopping costs are unaffected by the presence of the hard-discount, specialty store; they remain one-stop shoppers and are charged the monopoly margin for the bundle of both products. The presence of the small store gives the full-line store an opportunity to have a mix of shoppers, as opposed to all one-stop shoppers, through a loss-leader pricing scheme (an extreme form of cross-subsidization pricing). The large store's duopoly profit could increase over its monopoly profit if the advantage from its broad product range is sufficiently large.

3 Data

We use supermarket scanner data collected by IRI from scanning the barcodes of purchased products. The data encompass the weekly sales of 1,503 grocery and drug stores across 50 U.S. cities in 2006. The vast majority of the stores are grocery stores, and all stores in the data are part of grocery chains (as opposed to being independent), such as Target Corp. and Stop & Shop, Inc. The sample stores belong to 111 chains, but the chain names of the stores are disguised by IRI to protect confidentiality.

The data include the weekly prices and quantity sold for each product at the universal

[4] Ellison (2005)'s set-up on consumer heterogeneity—that low type consumers are not only unwilling to pay high prices for high quality add-ons but also are willing to travel far to pay less—makes unobservable add-on pricing a tool for softening competition by creating adverse selection faced by a store wishing to lower prices.

product code (UPC) level. A UPC is specific to an item, manufacturer, and packaging. For example, a 12-ounce can of Coca Cola has a different UPC than a 16-ounce bottle of Coca Cola. Our data span more than 49,000 UPCs in 30 categories, 17 of which are food and beverages, such as carbonated beverages and frozen pizza; the remaining categories are non-food household products such as shampoo and diapers.[5]

The data also provide information on marketing/advertising activities at the product level, mainly under two variables: promotion and feature. The promotion variable indicates whether a temporary price reduction (TPR) of 5% or larger was placed on each product. IRI determines this based on the regular price computed using its proprietary algorithm. The feature variable indicates whether the product is featured/advertised (via newspaper inserts or weekly circular) during the week. IRI classifies the levels of advertising into four categories: retailer coupon or rebate (A+), large ad (A), medium-size ad (B), and small ad (C)—usually one line of text. Table 1 summarizes the promotion and feature variables, which are highly correlated. The average probability that a TPR of 5% or larger is placed on a product in any given week is 0.25, and when such a TPR is placed on a product, the probability of it being advertised is 0.34, most likely in a medium-sized ad. The probability of a product being advertised without a TPR of 5% or larger is only 0.03.

[Table 1 about here.]

IRI provides rather crude information about the location of each store in the sample with five-digit ZIP codes. Using this information, we compute the distance between any two stores as the shortest driving distance between two centroids of the ZIP codes. As a result, stores in the same ZIP code are recorded as having zero distance between them. We use the Google Maps API for the computation.

The data also include information about demographic characteristics within a 2 mile radius of the actual store locations. The demographic information includes population size, population of households with children, and population by age, marital status, education, race, and income level. Table 2 presents the summary statistics for the demographic characteristics.

[Table 2 about here.]

[5] For more details on the categories and the database in general, please refer to Bronnenberg et al. (2008).

3.1 Market definition

How to define markets is important in our study because it determines the boundaries of the strategic interaction in the pricing and advertising game we conceptualize. This is particularly relevant to our test of C-R's hypothesis because we exploit the variation across markets in the store type distribution that could foster multistop shopping behavior, more specifically, whether a store that can be classified as a "hard-discount, specialty store" exists in each market.

A multiproduct search environment inherently involves the decision of which store to shop at, and the decision is likely governed by geographical considerations rather than administrative boundaries. As such, we define markets based on the distance between stores.[6] Specifically, we consider two stores to be in the same market when they are located no farther than a certain driving time apart. We set the baseline cut-off at 20 minutes, which is approximately the average driving time for grocery shoppers according to the 2003-2007 American Time Use Survey collected by the Bureau of Labor Statistics. The definition inevitably means some stores belong to more than one market. In those cases, we include such a store in a market such that the average distance between the store and other stores in the market is the smallest. When we strictly apply this criterion, many stores are left to be local monopolists. If among the monopolists a store has a neighbor within a 40 minute driving distance, then we assign it to that market.[7]

This definition yields 333 markets, including 47 isolated markets. A city is divided into 6.7 markets on average; there are cities with only one market (e.g., Omaha) and as many as 25 markets (New York City).[8] Table 3 summarizes the distribution of markets and stores in our baseline market definition. Each market includes 4.5 stores on average, but several markets include more than 10 stores. The average driving distance between any two stores in a single market is 10.2 minutes.

[Table 3 about here.]

[6] This is in contrast to other studies that mainly focus on consumer demand for goods, whereby city- or county-level geographical market definition is conventionally used.

[7] This raised cut-off was applied to less than 10% of stores in the sample. When there is no store in the 40 minute-driving vicinity of a store, we consider it to be a local monopolist; there are 47 such stores.

[8] As expected, the number of markets in a city is positively correlated with the size of city, in terms of both population and area.

3.2 Key variables

In this section, we illustrate the construction of several key variables used in the hypothesis tests. Consider a market with J stores and N products. The most straightforward way of formulating the distribution of products and stores in the market is to construct an $N \times J$ matrix where the (i, j)-th argument is 1 if store j carries product i and 0 otherwise. Many theoretical papers on multiproduct pricing, including L-M and C-R, consider settings in which $N = 2$ and $J = 2$ (Rhodes is an exception; he lets N take any number when $J = 1$), which makes working with such matrices simple and easy. Unfortunately, in reality N easily exceeds tens of thousands. We therefore construct a few variables that capture store heterogeneity (in the product range) and product heterogeneity (in how competitively they are supplied) based on the matrix representation.

■ *StrRPR*

First, let StrNPR_j denote simply the number of products store j carries. Because a product sold in a store is not observed unless it is scanned, it is also the number of distinct products that were sold at all throughout 2006.[9] Similarly to *StrNPR*, *StrRPR* measures the breadth of a store's product range but against the market it belongs to. It is the ratio of the products carried by each store to all the products offered in its market:

$$\mathit{StrRPR}_{jm} = \frac{\mathit{StrNPR}_{jm}}{\#\text{ of products ever sold in market } m}.$$

Because it is defined relative to the market's product range, the values of this variable are not directly comparable across markets.

■ *PRCOMP*

PRCOMP is a continuous variable that captures the overall competitiveness of a product in a market. It is defined as the ratio of the number of stores carrying a particular UPC to the total number of stores in the market:

$$\mathit{PRCOMP}_{im} = \frac{\#\text{ of stores carrying product } i \text{ in market } m}{\#\text{ of stores in market } m}.$$

A small value of PRCOMP_{im} means product i is carried by only a small fraction of stores in the market, which might confer some market power to those stores. We also construct an

[9] Although a store could carry products that were not sold in the whole one year period, omitting them will not bias our analysis in a meaningful way because having products on the shelves no one buys is unlikely to affect shoppers' decisions on where to shop.

indicator of whether a product is monopolistically sold by one store in the market (I_{Mono}) and an indicator of whether a product is offered by all stores (I_{Comn}).

We construct another store-level variable that captures store heterogeneity in product offerings: *StrPRC* is obtained by taking the average of *PRCOMP* over all products the store carries. A small *StrPRC* value implies that a store carries products that are not widely carried by other stores in the market.

[Table 4 about here.]

Table 4 shows how the key variables are distributed in the whole sample and within each non-monopoly market. There are considerable variations in product offerings and the overall competitiveness of the products stores carry. The bottom panel highlights store heterogeneity in these measures within and across markets. While the average difference in *PRCOMP* between the store with the broadest product range and the store with the narrowest one is 0.13, the difference in one market is 0.38, implying that stores in that market are considerably heterogeneous in their product offerings. Similar variation is also found in terms of the average competitiveness of the products stores sell (*StrRPR*).

Additionally, *StrRPR* and *StrPRC* are weakly positively correlated, with a correlation coefficient of 0.062. This positive correlation, albeit very weak, implies stores that carry a relatively broad range of products tend to carry products supplied by many other stores.

4 Tests and Results

In this section, we state each hypothesis to be tested as a regression equation and report the test results. For the ease of presentation, we divide the section into three subsections devoted to the three statements we test. In the last subsection, we further examine C-R's cross-subsidization pricing under a different light.

4.1 Exploitive cross-subsidization

In this section, we examine C-R's main proposition. Consider a market in which a monopoly store sells two products. When a store that supplies only one product, but at a lower price, enters the market, the used-to-be monopoly store will raise the price (margin) of the monopolized product and lower the price (margin) of the competitive product while keeping the sum of the two prices constant. One way to think about this pricing strategy is to imagine a seesaw of price margins. Tilting the seesaw will reduce the price (margin) of one product while raising the price (margin) of the other product by the same amount—although this analogy holds only for certain parameter values. The entry of a small store represents a circumstance in which both one-stop and multistop shoppers exist, which enables the large store to exploit cherry-pickers (multistop shoppers).

An ideal setting to test this proposition would be one in which we observed the same full-line stores with and without a rival specialty store, but such events are rarely present in the data.[10] We instead explore cross-market variation in the competition structure. More specifically, we examine whether large stores in markets with small stores price differently from large stores in markets without small stores. The underlying premise is that the presence of small stores is not endogenous to unobserved market characteristics.

Our approach calls for classification of store types. We determine whether a store is small (S type) or large (L type) based on its product range and the competitiveness of the products it carries. Recall that *StrRPR* measures the breadth of a store's product range and *StrPRC* measures a store's overall product competitiveness. Our baseline criteria classify a store as type L if its *StrRPR* is no more than 25% below the highest *StrRPR* in the market, i.e., if its product range is the broadest in the market or not too far behind. By construction, every market has at least one L type store.

A store is classified as type S if its product range is narrow and it carries highly competitive, rather than rare and exclusive, products. Specifically, we mark a store as type S if i) its *StrRPR* is smaller than the highest *StrRPR* in the same market by at least 50%, and ii) its *StrPRC* is no more than 2% below the highest *StrPRC* of the market. These requirements leave some markets without any type S store.

These criteria were chosen in consideration of the fact that the small store in C-R not only carries a limited range of products but also offers lower prices for the products it sells.[11] Note that our criteria leave some stores unclassified—neither S nor L. Unclassified stores include stores carrying relatively a narrow range of products whose overall product competitiveness is high enough to be type S, possibly because it supplies some products monopolistically.

Table 5 presents the differences between type S and L stores. Of the 1,503 stores in our data, 192 are type S stores and 879 stores are type L. Among 286 markets (excluding the 47 isolated or monopoly markets), 134 (47%) markets contain at least one small store. The bottom panel of the table shows that there is a minimal difference in the number of large

[10] Note that our data do not allow us to track the entry and exit of stores because sample stores might have been included/excluded from the data depending on their contracts with IRI.

[11] To compare the prices between the two types of stores, we regress stores' yearly or monthly average prices (with the UPC-market fixed effects) on the type dummy variable. The regression results indicate that the stores classified as type S offer significantly lower prices than the type L stores.

stores between the markets with a small store and the markets without one. Also, neither the number of product offerings nor sales differ much between large stores in markets with a small store and those in markets without a small store.

[Table 5 about here.]

In the following, our discussion focuses on the indicator variable $I_{S,m}$, which is 1 when there is at least one type S store in market m and 0 otherwise. The presence of a type S store ($I_{S,m} = 1$) represents an environment in which multistop shoppers (cherry-pickers) can abound, providing an avenue for a large store to adopt a cross-subsidization pricing policy to exploit them.

We explore several specifications. In all specifications, the question is whether type L stores adopt (exploitive) cross-subsidization pricing—which we picture as tilting the margin-seesaw—when they face a type S store as a competitor. We do not observe margins, so we use prices instead. The premise is that the presence of a small store is not correlated with unobservable cost factors specific to the market.

■ Model 1

We first run a regression of the following price equation for products sold only by the large stores in each market:

$$\begin{aligned} Price_{ijmc} \;=\; & (\beta_1 + \beta_2 x_{im} + \beta_3 StrRPR_{jm} + \beta_4 StrPRC_{jm}) \cdot I_{S,m} \\ & + \alpha_1 x_{im} + \alpha_2 StrRPR_{jm} + \alpha_3 StrPRC_{jm} + \alpha_4 NStr_m \\ & + Z_{jm}\Gamma + \eta_{ic} + \varepsilon_{ijm}, \end{aligned} \tag{1}$$

where $Price_{ijmc}$ is the yearly average unit price of product i at store j in market m of city c and $I_{S,m}$ is an indicator that takes 1 if there is a small store in the market. $NStr_m$ is the number of stores in the market and Z_{jm} is the vector of the demographic characteristics of store j (within a 2-mile radius), whose list is given in Table 2. η_{ic} is the fixed effects of UPC-city pairs where $c = c(j)$ is the city in which store j is located. The variables $StrRPR$ and $StrPRC$ are as described above.

The UPC-city fixed effects capture unobserved factors that affect prices at the UPC and city level, possibly through marketing campaigns that affect city-specific consumer preferences, or supply factors. Note that they allow the same product to be differently priced across cities in a regular fashion; for example, a 12 pack of 12 ounce Diet Pepsi could be consistently priced higher in Pittsfield than in Detroit.

x_{im} denotes a product-level characteristic that not only affects the store's pricing for the product but also could induce a different pricing response to the entry of a small store—in particular, whether the product is cross-subsidized (its margin is tilted downward) or subsidizing (its margin is tilted upward). Recall that in C-R's world, the latter product (subsidizing product) is the one that multistop shoppers purchase—either because it is not carried by the specialty store (Chen and Rey, 2012), or because it is priced lower than at the rival store because the store has a comparative advantage in that product (Chen and Rey, 2016). Because we do not observe wholesale prices, we base our test on the former case, so use the product-level competitiveness as a proxy for x_{im}. In one specification, we use the continuous measure of product competitiveness: $PRCOMP_{im}$, and in the other, we take a dichotomous approach and divide products into two groups: those that are monopolistically supplied by the store ($I_{Mono,ijm} = 1$) and those supplied by multiple stores ($I_{Mono,ijm} = 0$).

The variable of interest is the interaction term of $I_{S,m}$ and x_{im}. Consider the specification in which we use $PRCOMP$ as a proxy for the product-level heterogeneity that determines whether a product is being subsidized or subsidizing. Whereas β_1 captures the overall competitive pressure from the presence of a hard-discount, specialty store (a shift of the margin-seesaw), β_2 captures the differential effect of how competitively the product is supplied in the market (a tilt of the margin-seesaw). If the large store tends to set high prices for less available products and low prices for widely available products, α_1 will be negative (positive when x=I_{Mono}), and if the gap widens with the presence of a small store, β_2 will be negative (positive when x=I_{Mono}). The negative (positive when x=I_{Mono}) sign of β_2 would support C-R's cross-subsidization pricing, implying that large stores tilt up the margins of the products multistop shoppers would purchase—products that cannot be easily found in a small store.

The coefficients α_2 and α_3 capture pricing strategies that vary at the store level depending on the store's product range and the average competitiveness of the products it sells, respectively, and the coefficients β_3 and β_4 capture how those effects differ when there is a small store in the market.

[Table 6 about here.]

Table 6 presents the results, with each column corresponding to a different specification. In specification (3), we further divide products that are competitively supplied into two groups. The indicator $I_{byS,ijm}$ is 1 if product i (of store j) is supplied by any of the type S stores in the market and 0 otherwise. Notice that the interaction term of $I_{byS,ijm}$ and $I_{S,m}$ is omitted because $I_{byS,ijm}$ is always zero if $I_{S,m}$ is 0.

The negative coefficient of $I_{S,m}$ implies that the presence of a small store in the market puts additional competitive pressure on all the products that the large store sells. [12]

The negative coefficient for $I_S \times PRCOMP$ and the positive coefficient for $I_S \times I_{Mono}$, along with the negative coefficient of $PRCOMP$ and the positive coefficient of I_{Mono}, are consistent with C-R's cross-subsidization pricing prediction. They together indicate that the presence of a small store leads large stores to widen the price differentials between products easily found in other stores and products rarely found elsewhere, which implies that the former are "subsidized" products while the latter are "subsidizing."

The signs of the other coefficient estimates are sensible and more or less in line with C-R's reasoning. It is interesting that the presence of a small store appears to weaken the negative relationship between a large store's product range and its prices, as shown by the negative coefficient of $StrRPR$ and the positive coefficient of $I_S \times StrRPR$, as well as the negative relationship between the average competitiveness of the products it sells and its prices. It could indicate that a large store's broader product range, even though it consists of products on average more commonly found, gives the store a competitive edge when it is directly compared to a rival store that offers a narrow product range that might include some rare products.[13]

■ Model 2

The previous model provides a useful benchmark by exploring variation in the market-wide indicator variable I_S to capture full-line stores' pricing responses to competition with limited-line stores. However, it assumes the same extent of pricing response among large stores regardless of their competitive relations to the small stores. In particular, it is possible that competition occurs at a more local level than our definition of markets indicates. To address the concern, we next consider a model in which a large store's cross-subsidization pricing arises in response to competition with its nearest competitor. Specifically, we replace $I_{S,m}$ in the previous regression with an indicator variable $I^{clst}_{S,jm}$, which takes 1 when store j's closest (driving distance-wise) rival store is type S. Accordingly, $I^{clst}_{byS,ijm}$, which is 1 when the

[12] Although the coefficient should be zero in ta world that perfectly resembles C-R's setting—as the sum of the two prices remain the same after the entry of the small store—this is somewhat expected: The entry of a specialty store is likely to generate one-stop shoppers who visit only the specialty store as well as multistop shoppers, which would then translate into store-wide competitive pressure for the full-line store.

[13] Note that we do not control for chain-specific factors with, for example, chain fixed effects. Some coefficient estimates are no longer statistically significant when including chain (or UPC-chain) fixed effects. We feel that the variables of interest, such as $PRCOMP$ and I_{Mono}, are highly correlated among stores within a chain, and thus most of the variations in the variables are absorbed by the chain fixed effects.

nearest small store carries product i (of store j), replaces $I_{byS,ijm}$. We present the estimation results for this model in Table 7.

[Table 7 about here.]

Clearly, the main results are robust to confining the competition with a small store to a more local level. Having a small, specialty store as the closest rival store appears to make a large, full-line store further widen the price differentials between products that are not easily found in other stores and products supplied competitively. It is noteworthy that the coefficient for whether the product is also sold by the closest small store is negative and statistically significant; it was also negative but insignificant when the competition against a small store was specified at the market level.

In the Appendix, we further vary market definitions and classification criteria for store types to check whether our findings are robust to the variations. The robustness checks results largely agree with the baseline results presented in this section.

4.2 Effect of product range on advertising and pricing

The main hypothesis we test in this subsection is derived from Rhodes (2015). In his model, a store with more product offerings advertises less and charge lower prices—possibly following the EDLP pricing strategy—compared to a store with limited product offerings. The key driving-force for this result lies in (continuous) consumer heterogeneity in their valuations for multiple products. A store with a broader product range attracts consumers who on average have lower valuations and so finds it optimal to charge lower prices. Thus a store's broad product range serves as a signal to consumers that overall prices are low, and so the store need not engage in costly advertising. We call this effect of the product range the *low-price image effect*. Rhodes adds that, however, the result might not hold when the market is competitive. In that case, a broader product range would make it more profitable to steal business from rival stores, and therefore a store with a broader product range might advertise more intensively. We call this effect of the product range the *business-stealing incentive effect*.

As for the negative relationship between product ranges and prices, a similar conclusion emerges in Zhou (2014) under the term *joint search effect*. It describes the effect of a reduced price for one product on another product's sales. In his model, the joint search effect leads multiproduct retailers to charge less than single-product retailers in a duopoly market. It is noteworthy that the negative relationship between product ranges and prices is drawn by two different mechanisms—one by heterogeneous consumer valuations and the other by competition. Thus, Rhodes's prediction of a negative relationship between a store's product range and its prices that is drawn under little- or no-competition is likely to be reinforced by competition.

■ Relationship between product range and advertising

To examine the effect of a store's product range on advertising, we run the following regression:

$$f_{ijm}^{Ad} = \beta_1 StrNPR_{jm} + [StrNPR_{jm} \times X_{jm}]\beta_2 + X_{jm}\alpha_1 \qquad (2)$$
$$+Z_{1,ijm}\gamma_1 + Z_{2,jm}\gamma_2 + \eta_{1m} + \eta_{2ic} + \varepsilon_{ijm},$$

where f_{ijm}^{Ad} denotes a measure of the advertising effort store j (in market m) makes for product i. $StrNPR_{jm}$ is our measure of store j's product range: the number of products (UPCs) store j (in market m) offers.

We measure advertising effort using the information contained in two variables in the data: promotion and feature, for which summary statistics are provided in Table 1. Our baseline criterion for counting (low-price) adverts requires a product be advertised (inclusive of retailer coupons or rebates) with a temporary price reduction of 5% or larger. Given this criterion, we compute how frequently a product is advertised—that is, the number of weeks in which the product is advertised divided by the total number of weeks—to obtain the advertising frequency in percentage: f^{Ad}.[14] The top panel of Table 8 provides some summary statistics for our measure of advertising intensity.

[Table 8 about here.]

X_{jm} denotes a set of variables that might alter the marginal effect of the breadth of store j's product range on its advertising effort, such as the degree of market competition in relation to business-stealing opportunities. Rhodes models it as the fraction of non-loyal consumers in the market—those who are indifferent about where to shop and so choose randomly. We approximate it using the number of stores in the market ($NStr_m$), the distance to the closest rival store ($Dist_{jm}$), and the closest rival store's product range ($RvlNPR_{jm}$). The coefficient

[14] We have also computed the advertising frequency varying the criteria for counting (low-price) adverts—for example, removing the requirement that advertising be accompanied by a price reduction of 5% or larger and differentiating the level of advertising under the assumption that A, B, and C correspond to 3, 2, and 1 lines of advertising, respectively. The different criteria led to results that were only quantitatively different. We therefore report the results from the baseline measure for advertising frequency.

for each proxy's interaction with *StrRPR* will be positive if the variable increases with the business-stealing opportunity and negative otherwise. Ideally, the inclusion of proxies for the degree of competition would factor out the business-stealing incentive effect, and isolate the low-price image effect of a broad product range.

$Z_{1,ijm}$ denotes the variables that capture product-level competitiveness—$PRCOMP$, I_{Mono}, and I_{Comn}—and $Z_{2,jm}$ includes store-specific controls other than *StrNPR*, such as the consumer demographic variables as in section 4.1 as well and the average availability of the products the store carries (*StrPRC*).

We control for various sources of spurious relationships between the dependent variable (f^{Ad} in the current regression and regular prices in the next regression) and the regressors of interest using a fixed effect approach. η_{1m} denotes the market fixed effects, which are expected to capture market specific factors that affect the store's advertising effort and pricing policy, such as advertising costs, effectiveness of existing advertising outlets, and level of competitive pressure. η_{2ic} denotes the product-level fixed effects: UPC-city and UPC-chain fixed effects are considered. UPC-city fixed effects account for product-specific idiosyncratic demand at the city level, whereas UPC-chain fixed effects account for the adverting frequency and pricing rule related to trade deals between the chain and the manufacturer as well as chain-wide advertising and pricing policies.

Table 9 presents the estimation results. It shows that advertising frequencies are positively related to a store's product range in all specifications, which is at odds with Rhodes's proposition drawn for a local monopolist. On the other hand, this result is in accordance with Rhodes's assertion that the business-stealing incentive effect of the product range will dominate the low-price image effect when stores face competition. The proxies for the degree of competition do not appear to factor out the business-stealing incentive effect, leaving the coefficient on *StrNPR* to capture the mix of this effect and the low-price image effect.

[Table 9 about here.]

The coefficient of *StrNPR*×*NStr* is negative, whereas that of *StrNPR*× *Dist* is positive in specifications (3) and (4), implying that a store's tendency to advertise more as its product range increases is elevated when fewer stores are in the market and the closest rival store is farther away. This seems inconsistent with the business-stealing incentive effect because such an environment likely represents a low degree of competition and the presence of many consumers "loyal" to the store. Also, the coefficient of *StrNPR*× *RvlNPR* is negative, which implies that as a rival store's product range becomes broader, the store's proclivity to advertise more with more product offerings decreases.

Among the unreported coefficient estimates, the effect of a store's carrying more widely available products on its advertising effort appears negative when UPC-city fixed effects are considered but positive otherwise. The reversal of the coefficient signs is also observed in the other product-level variables such as $PRCOMP$, I_{Mono}, and I_{Comn}, which suggests a possible correlation between a store's product range or assortment and its advertising policy at the chain level.

The regression analysis seems to largely support Rhodes's conjecture on the positive effect of a store's product range on its advertising in the presence of business-stealing opportunities. However, it fails to disentangle the low-price image effects from the business-stealing incentive effects of product ranges. In addition, other factors (besides the business-stealing incentive effect) could contribute to the positive relationship between product ranges and advertising activities. For example, a store's product range might be endogenous, especially with respect to overhead costs including the costs of advertising. Under this scenario, a store or chain of stores that can advertise more cost-efficiently might choose to carry more products.

■ Relationship between product range and prices

We proceed to the remaining part of the prediction: as the product range becomes broader, prices become lower. Because the result is drawn with respect to "unadvertised" prices, which Rhodes also refers to as "regular" prices, we begin by defining the regular price for each product. We follow the approach proposed by Hendel and Nevo (2006) and set the regular price at the modal price over the entire sample period. Because retail prices for consumer packaged goods tend to be stable over time with infrequent temporary price reductions (see Hosken and Reiffen, 2004), the use of modal prices is conformable to Rhodes's regular prices. The bottom panel of Table 8 provides some summary statistics for this variable.

The regression equation shares independent variables with equation (2), whereas the dependent variable is now the regular price. Note that the regression is similar to the price regressions in section 4.1. The differences are first, the dependent variable is the regular price in the current regression, whereas it was the average price previously, and second, the current regression includes all the stores and products in our sample unless they were dropped due to missing information or collinearity, whereas the previous one considers only large stores and their products.

[Table 10 about here.]

Table 10 presents the estimation results. The sign of the coefficient of interest differs depending on whether the product-level control is at the UPC-city level or at the UPC-chain level: it is negative in the former and positive in the latter, and this pattern survives the inclusion of the proxies for the degree of competition. The coefficients for the proxies for the degree of competition do not have the expected signs consistently across the specifications except for *StrNPR*$\times$ *Dist.* The positive sign for that variable implies that the farther away the closest rival is, the less the store faces the joint-search effect.

The aforementioned pattern suggests the possibility that the relationship between product offerings and pricing is determined at the chain level. Indeed, a piece of empirical evidence that Rhodes himself refers to is based on chain-level variations: Ellickson and Misra (2008) find that stores that operate on a larger scale also tend to offer lower prices through the EDLP policy—as opposed to the PROMO policy, and adopting such pricing policies tend to be enacted at the chain level.

To verify the supposition and explore variation across chains, we average both the regular prices for each product and stores' product ranges within each chain and run a regression with UPC fixed effects. The result is given in Table 11: the relationship is negative at the chain-level.

[Table 11 about here.]

Overall, the results seem to support the negative relationship between product ranges and (regular) prices proposed by Rhodes and by Zhou (2014). However, they do not provide evidence for the machinery behind that negative relationship. Especially, they raise the question of whether the negative relationship is drawn by cost factors such as economies of scale. In the supermarket industry, more product offerings are likely to translate into large operations with many check-out lanes and many employees, such as at Walmart stores.

4.3 Relationship between advertised and unadvertised prices

In regard to the relationship between the prices of advertised products and the prices of unadvertised products, the three papers we consider all draw distinct predictions (in some stretched sense). However, the most relevant is Rhodes. C-R's set-up—that consumers are perfectly informed about prices—makes it difficult to introduce advertising, and it is unclear how advertised prices are communicated to consumers who are heterogeneous in their fixed shopping costs. L-M's setting lacks the machinery to derive a meaningful relationship between the two. Thus we base our empirical models on Rhodes.

Rhodes considers a setting in which consumers search for n products, and the local monopolist firm can choose to advertise the price of one of them. In his model, consumers use the advertised price as a signal for the unadvertised prices. When the advertised prices decline for some exogenous reason, such as an attractive trade deal offered by the manufacturer, additional consumers would find it worthwhile to visit the store. Because those new consumers have lower valuations than those who are unaffected by the advert, the average consumer valuation decreases, and so the store optimally reduces the unadvertised prices as well. We call this effect of the advertising prices the *consumer-valuation shift effect* or the *signaling effect.*

To examine the relationship, we consider the weekly movements of prices of advertised and unadvertised products. Specifically, we consider temporary price deviations (TPDs) from the *regular prices* defined in the previous section.

While we are interested in the relationship between the movements of a TPD index for the sets of advertised products (TPD^{Ad}) and that for the unadvertised ($TPD^{Non\text{-}Ad}$), the products that compose each set change week by week. This makes it difficult to incorporate sales-based weights in calculating the weekly TPD^{Ad} and $TPD^{Non\text{-}Ad}$. Because the two sets are complements, having products with large weights in one set will accompany a decrease in the weights in the other set and vice versa, falsely generating a negative relationship between the two variables. We thus compute the percentage TPD for each product in the advertised and unadvertised sets, respectively, and then take a simple arithmetic average to obtain the week's TPD^{Ad} and $TPD^{Non\text{-}Ad}$.[15]

There could be other factors, irrelevant to the force that creates a positive relationship in Rhodes, that affect both sides—TPD^{Ad} and $TPD^{Non\text{-}Ad}$—in a systematic fashion. Seasonality in demand for consumer packaged goods is a clear example. It would add to the positive relationship, if any, between the two variables. There might also be some store-specific (or chain-specific) pricing policies that could compound the relationship in question. For example, a store with the EDLP policy might not face a decline in consumer valuations (because it has already established its low price image) as acutely as a store with the PROMO policy that uses price promotions extensively to lure shoppers into the store when they lower advertised prices (Ortmeyer et al., 1991; Lal and Rao, 1997; Ellickson and Misra, 2008; Bell and Lattin, 1998). Because the stores in our data probably represent a mix of stores with different policies along this dimension, it might average out the signaling effect of advertised

[15] The potential concern in using the simple average as opposed to the sales-weighted average would arise when products with small sales weights are associated with substantially large TPD value—whereby the simple TPD average can misrepresent the overall price deviations. However, when we examine the TPDs for the ten decile groups (in terms of sales), the TPD of the first decile is relatively lower on average (-11% and -0.7% for the advertised and unadvertised sets, respectively) but the $TPDs$ of the other nine decile groups are all similar to each other for both sets (the average TPD^{Ad} ranges between -19% and -24% and the average $TPD^{Non\text{-}Ad}$ between -1.5% and -1.7%). Therefore, we do not expect our approach to be of great concern.

prices that Rhodes proposes.

To factor out the seasonal effect as well as the effect from store-specific pricing policies, we first regress TPD^{Ad} and $TPD^{Non\text{-}Ad}$ on the week dummies and store dummies to obtain the residuals. We then run a single variable regression. In this way, we can focus on the within-store variation in deseasonalized price deviations. The first-stage regression equation is:

$$TPD_{jt}^{k} = \eta_t + \nu_j + \gamma_1 W_{jt}^{k} + \sum_{c\in\mathcal{C}} \gamma_c SW_{cjt}^{k} + \varsigma_{jt},\ k = Ad,\ Non\text{-}Ad. \tag{3}$$

where η_t is the week fixed effects and ν_j the store fixed effects. We vary the first-stage specification by adding two other variables to account for the fact that changes in TPD_{jt}^{k} ($k = Ad$, $Non\text{-}Ad$) are partly driven by changes in the products that compose the advertised and unadvertised sets. W_{jt}^{k} is the sum of the sales-weights (the product's annual sales divided by the store's total annual sales) of the products in set k. It controls for any correlation between products' *TPDs* and their sales-weights—e.g., products with large sales-weights might be prone to deeper discounts. SW_{cjt}^{k} is the revenue share of category c in set k. It accounts for possible correlations between the category of products and the size of their *TPD*s. If carbonated drinks, for example, tend to be more deeply discounted than other categories, their uneven inclusion in the advertised and unadvertised sets would steer the relationship in question to the negative side.

[Table 12 about here.]

Table 12 presents the estimation results from the second-stage single variable regression. The first two columns report the estimates from controlling for only the week and store fixed effects in the first-stage regression, whereas the next two columns are from the specification that also includes the two basket-specific variables.

Given that retail stores set the prices of both advertised and unadvertised products simultaneously to maximize store-wide profits, the second-stage regression is subject to the endogeneity problem. To address the potential endogeneity in the advertised prices on the right hand side, we instrument $TPD^{Ad} - \widehat{TPD^{Ad}}$ with the average price deviation of the same products in other markets. If the manufacturers charge similar wholesale prices or offer similar deals across stores, the price deviations in other markets for the same products will be correlated with the price deviations in the given market, but they are independent of the store-specific error term in the regression.

The results indicate that, consistently across the specifications and the use of an IV approach, the advertised and unadvertised prices are negatively correlated. The IV estimates imply that the estimated coefficient for the residuals of TPD^{Ad} is biased toward zero due to the simultaneity problem. Including the basket specific controls slightly increases the goodness of fit in the first-stage regression, but it does not qualitatively affect the relationship between the two price deviations. In fact, if stores randomize which products they advertise, as Rhodes suggests, the varying composition of the products in the advertised and unadvertised sets should not affect the results.[16]

Our finding of a negative relationship between advertised and unadvertised prices appears to reject Rhodes's prediction, but we argue that such a conclusion is not warranted. Recall that the driving force behind the positive relationship between advertised and unadvertised prices is the signaling effect of the advertised prices: uninformed consumers take lower advertised prices as a signal that all prices are lower (in accordance with stores' equilibrium pricing strategies). If consumers update their beliefs about prices infrequently or with a time lag, our empirical strategy to explore the weekly variations in price deviations might not be suitable to invoke the mechanism that gives rise to the positive relationship between advertised and unadvertised prices.

It could follow that we should define the advertised and unadvertised price deviations at a lower frequency level—for example, by averaging them over a year—and then compare the two price deviations at the store level. This approach, however, would be subject to the concern that how the two variables are related could differ across stores—especially between stores that adopt the EDLP policy and stores that adopt the PROMO policy—and that pooling stores together might average out any relationship between the two variables. Therefore, we take a rather indirect approach to unraveling the signaling effect of advertised prices.

We first estimate the relationship of the two price deviations store-by-store and then relate it to the pricing policy of each store: EDLP versus PROMO. Because our scanner data do not contain the information about which pricing policy is used in each store, instead we use pricing characteristics related to the two pricing policies. That is, we ask whether stores exhibiting a positive relationship between advertised and unadvertised prices are more likely to display pricing patterns characterizing a PROMO policy and whether stores with a negative relationship are characterized by an EDLP policy.[17]

[16] Rhodes (2015) shows that each store under duopoly randomizes its choice of a product to advertise as well as the size of the price cut in equilibrium so that it cannot second-guess the competitor's advertising strategy.

[17] Ellickson and Misra (2008) use store manager survey data which contain store managers' answers to the question of which pricing policy is adopted in the stores they oversee.

To estimate the store-by-store relationship between advertised and unadvertised prices, we first obtain the residuals of the weekly price deviations with a first-stage specification similar to equation (3), except that we now include only the week dummies to remove the seasonality in TPD^{Ad} and $TPD^{Non\text{-}Ad}$. We then estimate the Pearson's correlation coefficient between the residuals of TPD^{Ad} and $TPD^{Non\text{-}Ad}$ for each store separately. Table 13 summarizes the estimated correlation coefficients obtained store-by-store. Although the mean coefficient over the entire sample is negative, advertised and unadvertised prices are positively related for about 45% of stores. We observe a similar pattern at the chain level: the fraction of the chains with a positive relationship is about 46% from among the 111 different chains in the data.

[Table 13 about here.]

We relate the estimated correlation coefficients to three variables that manifest distinct features of the EDLP and PROMO policies. EDLP pricing is often described as offering relatively low prices that are stable over time, whereas PROMO pricing is associated with deep and frequent price cuts. The final step is to estimate the following equation:

$$Y_{ij}^{k} = \beta_k I_{+,j} + \nu_i^k + \eta_m^k + \varepsilon_{ij}^k, \quad k = 1, 2, 3 \tag{4}$$

where $Y_{ij}^1 = RP_{ij}$ is the regular price of product i in store j; $Y_{ij}^2 = TPR_{ij}$ is the average temporary price reduction (limited to 5% or larger) placed on product i in store j over the sample period; and $Y_{ij}^3 = f_{ij}^{Ad}$ is store j's advertising frequency for product i, as defined in section 4.2. $I_{+,j}$ is the indicator for the store being characterized with a positive correlation coefficient in the previous step. The UPC and market (where the store is located) fixed effects are denoted by ν_i and η_m, respectively.

[Table 14 about here.]

Table 14 presents the estimates from the regressions. The positive coefficient of $I_{+,j}$ in the first column, the negative coefficient in the second, and the positive in the third imply that stores characterized by a positive relationship between advertised and unadvertised prices tend to charge relatively higher regular prices and be prone to deeper and frequent price cuts.[18] Based on the previous argument that stores that charge higher prices regularly but offer deep price discounts are more likely subject to the signaling effect of advertised prices,

[18]The estimates do not qualitatively change when we use an alternative measure for the regular price—the highest price in past four weeks, following Hendel and Nevo (2006)—to compute TPDs.

we conclude that our results confirm the mechanism underlying the positive relationship between advertised and unadvertised prices that Rhodes proposes.

4.4 Promotional pricing under cross-subsidization

The regression analyses in section 4.3 show that weekly price deviations for advertised products are negatively related to those for unadvertised products. It is not straightforward to draw an implication from this result for C-R's cross-subsidization pricing because there is no room for advertising in C-R's setting—consumers are perfectly informed about prices. However, if we view advertised prices simply as temporary price reductions caused by favorable cost shocks to the advertised products, we can pose a question of how a store that adopts cross-subsidization pricing would price other products unaffected by the cost shocks. The store would raise its (unadvertised) prices for those products so that it can extract more from multistop shoppers while keeping the price of the one-stop shopper's basket constant. As such, the negative relationship we find might be a promotional pricing pattern that arises under the cross-subsidization policy. This observation motivates our next test on C-R's cross-subsidization pricing.

We conduct a regression analysis to examine the relationship between the basket prices for two types of products—those that are likely cross-subsidized and those that are likely cross-subsidizing. Following C-R, we let a basket of competitive products represent the former and a basket of monopolized products represent the latter. Additionally, we consider only stores that are classified as large in section 4.1 and belong to markets that contain at least one small store because those stores are the most likely to adopt the cross-subsidization pricing strategy. A product sold at a large store is called competitive if it is also sold by a small store in the market, and it is called monopolized if the large store is the sole supplier of the product.

The actual estimation procedure is very similar to the weekly price deviation regression in section 4.3. First, we fit the weekly price deviations with the week and store fixed effects. Then, the residuals from this regression for the competitive products ($TPD^C - \widehat{TPD^C}$) are regressed on the residuals of the monopolized products ($TPD^M - \widehat{TPD^M}$). Notice a subtle but important difference between the current approach and the one in section 4.3. We no longer consider whether the price, or price deviation, is advertised or unadvertised. We simply look at the weekly price deviations of two sets of products that differ in how likely they are to be included in multistop shoppers' shopping baskets. Because consumers are perfectly informed about prices in C-R's set-up, there is no need for price announcements through advertising to lure (multistop) shoppers.

We consider two models. In the first model, presented in the first two columns of Table 15, the competitive products in the large store are restricted to those sold by the closest small store. In the second model, the competitive products are those sold by any of the small stores in the market. As the price deviations of the monopolized products are likely endogenous, we construct an instrument variable for it in a fashion similar to that in section 4.3.

The price deviations of the competitive products are negatively correlated to those of the monopolized products. We take this result as another evidence for C-R's cross-subsidization on the basis that it can generate such a promotional pricing pattern.

[Table 15 about here.]

5 Conclusion

In many markets, retailers are predominantly multiproduct sellers, and shopping trips generally involve some fixed costs. When consumers are uninformed about the prices or product-consumer match valuations of the multiple products they wish to purchase, such costs are called search costs. Otherwise, they can be understood as time or transaction costs for making a trip to a store. The presence of fixed costs induces a multiproduct seller to set the price of each product with consideration of its effects on the sales of other products. In this paper, we empirically examine some of the predictions made by theoretical work on multiproduct firms' pricing and advertising behavior in such an environment.

Many theoretical papers have drawn interesting predictions in this domain. Among those, we focus on Lal and Matutes (1994), Chen and Rey (2012) and Rhodes (2015). The three papers particularly tally with the goal and scope of our work in that they all produce original predictions with simple but interesting machineries, and some of their predictions are relatively straightforward to test empirically.

Specifically, we examine three propositions. First, C-R proposes that a full-line store adopts a cross-subsidization pricing strategy—products that both one-stop and multistop shoppers buy cross-subsidize products that only one-stop shoppers buy—in the presence of a limited-line specialty store, which provides an environment for the rise of multistop shopping behavior. Second, Rhodes proposes a negative effect of a store's product range on advertising and prices. He also predicts that if a store lowers its advertised prices, it should lower its unadvertised prices as well. In contrast, in L-M's linear city world in which both consumers and firms are symmetric, advertised and unadvertised prices are unrelated. Rhodes and L-M together motivate our third question: are unadvertised prices positively related or unrelated to advertised prices?

We use scanner data collected by IRI that include information on weekly advertising and promotional activities and the sales of more than 1,000 grocery and drug stores in the US. The grocery store industry is suitable for testing the empirical relevance of the aforementioned propositions because grocery shopping is generally multiproduct shopping, and competition is more local than found in other industries. On the other hand, it is known that consumers in this industry are not as uninformed about prices as the theories—especially Rhodes and L-M—assume. This feature both limits and guides our approaches to formulating and interpreting some of our tests.

Our regression analyses find evidence for C-R's cross-subsidization pricing. Specifically, we find that full-line stores increase the price differentials between more and less competitive products when there is a limited-line store in the market. We submit this as evidence for the cross-subsidization pricing that arises when stores are faced with both one-stop and multistop shoppers. Additionally, our investigation in weekly temporary price deviations reveals that the (deseasonalized) temporary price deviations of monopolized products and competitive products move in the opposite direction. This pattern of promotional pricing is also consistent with cross-subsidization.

Our findings are also largely in line with Rhodes's equilibrium pricing and advertising characterization. First, we observe that a store's product range is negatively related to its non-advertised, regular prices (when variation across chains are explored). Because a chain is likely to set uniform policies regarding both product offerings and pricing—for example, whether to adopt EDLP or PROMO pricing policies, we take the chain-level regression result as evidence in support of Rhodes. However, we note that our analyses do not corroborate the mechanism that draws the negative relationship in Rhodes: the low price image effect of a broad product range. Second, we find the relationship between product ranges and advertising efforts to be positive. This result is in disagreement with Rhodes's prediction drawn for a local monopolist, but is consistent with his conjecture for a competitive environment.

Finally, we find that stores characterized by a positive relationship between advertised and unadvertised prices appear to exhibit the pricing characteristics of the PROMO policy, whereas stores with a negative relationship exhibit those of the EDLP policy. Because the signaling effect of advertised prices—the driving force of the positive relationship between the two prices in Rhodes—is likely stronger for stores that adopt PROMO policies than stores with EDLP policies, we find this result supportive of Rhodes's prediction.

Accordingly, we find against L-M's prediction that advertised and unadvertised prices are independent of one another. However, we note that L-M's model lacks an ingredient, such as consumer or store heterogeneity, necessary to draw a meaningful relationship between the two variables.

Although our empirical findings provide some evidence for the theories considered, it remains unanswered which theory is the most relevant and prevalent in the supermarket industry. The lack of a unified theory—in which all the different forces generating the propositions examined in this paper interplay—makes this question difficult to answer. We hope our empirical findings guide interested theorists to advance the theory of multiproduct pricing.

References

Armstrong, M. and J. Vickers (2010). Competitive Non-Linear Pricing and Bundling. *Review of Economic Studies 77*, 30–60.

Bell, D. and J. Lattin (1998). Shopping Behavior and Consumer Preference for Store Price Format: Why "Large Basket" Shoppers Prefer EDLP. *Marketing Science 17*(1), 66–88.

Besanko, D., J.-P. Dubé, and S. Gupta (2005). Own-Brand and Cross-Brand Retail Pass-Through. *Marketing Science 24*(1), 123–137.

Bliss, C. (1988). A Theory of Retail Pricing. *Journal of Industrial Economics 36*(4), 375–391.

Bronnenberg, B., M. Kruger, and C. Mela (2008). Database Paper-The IRI Marketing Data Set. *Marketing Science 27*(4), 745–748.

Chen, Z. and P. Rey (2012). Loss Leading as an Exploitative Practice. *American Economic Review 102*(7), 3462–3482.

Chen, Z. and P. Rey (2016). Competitive Cross-Subsidization. *Unpublished manuscript*.

Chevalier, J., A. Kashyap, and P. Rossi (2003). Why Don't Prices Rise During Periods of Peak Demand? Evidence from Scanner Data. *American economic review 93*(1), 15–37.

Cournot, A. (1838). *Researches into the Mathematical Principles of the Theory of Wealth* (1960 ed.). reprented by A. M. Kelly.

Ellickson, P. and S. Misra (2008). Supermarket Pricing Strategies. *Marketing Science 27*(5), 811–828.

Ellison, G. (2005). A Model of Add-on Pricing. *The Quarterly Journal of Economics 120*(2), 585–637.

Florez-Acosta, J. and D. Herrera-Araujo (2014). Multiproduct retailing and consumer shopping patterns: The role of shopping costs. *Unpublished manuscript*.

Hendel, I. and A. Nevo (2006). Sales and Consumer Inventory. *Rand Journal of Economics 37*(3), 543–561.

Hosken, D. and D. Reiffen (2004). Patterns of retail price variation. *Rand Journal of Economics 35*(1), 128–146.

Hosken, D. and D. Reiffen (2007). Pricing Behavior of Multiproduct Retailers. *The B.E. Journal of Theoretical Economics 7*(1), 1–43.

Lal, R. and C. Matutes (1994). Retail Pricing and Advertising Strategies. *Journal of Business 67*(3), 345–370.

Lal, R. and R. Rao (1997). Supermarket Competition: The Case of Every Day Low Pricing. *Marketing Science 16*(1), 60–80.

McAfee, P. (1995). Multiproduct Equilibrium Price Dispersion. *Journal of Economic Theory 67*(1), 83–105.

Mulhern, F. J. and D. T. Padgett (1995). The Relationship between Retail Price Promotions and Regular Price Purchases. *Journal of Marketing 59*(4), 83–90.

Nevo, A. and K. Hatzitaskos (2006). Why Does the Average Price Paid Fall During High Demand Periods? *Unpublished manuscript*.

Ortmeyer, G., J. Quelch, and W. Salmon (1991). Restoring Credibility to Retail Pricing. *Sloan Management Review 33*(1), 55–66.

Pesendorfer, M. (2002). Retail sales: A study of pricing behavior in supermarkets. *The Journal of Business 75*(1), 33–66.

Rhodes, A. (2015). Multiproduct Retailing. *Review of Economic Studies 82*(1), 360–390.

Richards, T., S. Hamilton, and W. Allender (2016). Search and price dispersion in online grocery markets. *International Journal of Industrial Organization 47*, 255–281.

Shelegia, S. (2012). Multiproduct pricing in oligopoly. *International Journal of Industrial Organization 30*(2), 231–242.

Smith, T. and O. Thomassen (2012). Multi-category demand and supermarket pricing. *International Journal of Industrial Organization 30*(3), 309–314.

Zhou, J. (2014). Multiproduct Search and the Joint Search Effect. *American Economic Review 104*(9), 2918–2939.

A Robustness checks

We examine whether our test results are robust to the definition of markets and the classification of stores. To this end, we vary the boundaries of markets and the classification criteria for small and large stores. Note that when we change the definition of markets, the sets of stores classified as small and large also change because our classification criteria are set based on a store's relative placement within its market in two variables: each store's product range and the average availability of its products.

We report the results from three different considerations. Rob-1 reduces the market boundaries by setting the maximum driving time between any two stores in the same market at 15 minutes, instead of 20 minutes in the baseline market definition. In addition, we set the raised cut-off applied to remote stores at 30 minutes instead of the baseline 40 minutes. Rob-1 produces 583 markets, including 322 isolated monopoly markets, whereas the baseline definition led to 333 markets with 47 monopoly markets. Rob-2 expands the market boundaries by setting the usual cut-off at 25 minutes and the raised cut-off at 35 minutes. Rob-2 generates 264 markets with 57 isolated markets.

Rob-3 use stricter rules for both small and large stores. Whereas the baseline classification defines a large store as one having a product range (*StrPRR*) no more than 25% below the highest in the market and a small store as one having at least 50% below the highest *StrPRR* in the market, Rob-3 sets the former at 15% and the latter at 60%. As a results, the numbers of large and small stores fall from 879 and 192 to 673 and 137, respectively.

Table 16 presents the robustness checks results for the cross-subsidization pricing test in section 4.1. The presented results are for Model 2, in which the closest rival store being a small store represents the event motivating large stores to undertake cross-subsidization pricing (in contrast to Model 1 in which the presence of any small store in the market represents the motivator). We believe Model 2 better captures the incentive for cross-subsidization pricing because competition is likely to occur at a level more local than whole markets. Recall that specification (1) takes a continuous approach and (3) a dichotomous approach in measuring the availability of a product, which is related to the likelihood that the product is cross-subsidizing or being subsidized.

The robustness checks results largely agree with the baseline results. With the dichotomous specification, it is shown that large stores widen the price gaps between monopolized products and others upon having a small store in the market or as the closest rival store. This effect is not apparent under Rob-3 (last column). When taking the continuous approach, large stores lower the price of more competitive products (products with high *PRCOMP*) more than they do the prices of less competitive products when the closest rival store is small (as shown by the negative coefficients on $I_S \times PRCOMP$). Unlike the baseline results, large stores do not appert to price their products differently depending on availability whenthey are not faced with a small store as the closest rival store (as shown by the insignificant coefficients of $PRCOMP$). Thus, large stores appear to (newly) create price differentials between more and less competitive products upon having a small store as their closest rival, as opposed to widening differentials that already exist even when the closest rival is not small. As for Model 1 (not reported), we find the cross-subsidization pricing pattern for both specifications (1) and (3) under Rob-3, but under Rob-1 and Rob-2, the pattern is observed only for specification (3).

[Table 16 about here.]

Table 17 presents the robustness checks results for the relationship between a store's product range and its advertising effort and prices (baseline results are presented in section 4.2). We only report the results for the specification that uses market and UPC-city fixed effects. The positive relationship between the product range and the advertising frequency shown in Table 9 is robust to the two different definitions of markets. The negative relationship between the product range and regular prices when UPC-city fixed effects are used also holds true. So does the tendency of the relationship to become positive when UPC-chain fixed effects are used instead (although all these results are unreported). The bottom panel confirms the baseline findings in Table 11: the relationship between a store's product range and its regular prices is negative at the chain level under the two different market definitions.

[Table 17 about here.]

Table 18 presents the robustness checks results regarding the consumer-valuation shift effect of advertised prices. We repeat the exercise described in section 4.3 for Rob-1 and Rob-2. We still find evidence for the consumer-valuation shift effect of advertised prices in stores that keep regular prices high but offer deep and frequent price discounts—stores that probably use the PROMO policy rather than the EDLP policy.

[Table 18 about here.]

Finally, we find, although unreported, that the pricing pattern shown in Table 15 remains in all robustness checks specifications. To reiterate our finding: large stores tend to raise the prices of their monopolized products when they cut the prices of their competitive products. We find this negative relation between the two price movements regardless of how we define markets or classify store types.

B Tables and Figures

Table 1: Promotions and Advertising

	Mean	Std	Min	Max
Promo	0.25	0.43	0	1
Feature	*Promo*= 1		*Promo*= 0	
large-size ad (A)	13.1 %		1.3 %	
medium-size ad (B)	18.9 %		1.8 %	
small-size ad (C)	0.8 %		0.1 %	
coupon/rebate (A+)	0.8 %		0.2 %	
None	66.4 %		96.7 %	

Table 2: Demographic Characteristics of Store Location

	Mean	Std	Min	Max
Age of household head (%)				
≤ 25	13.3	2.6	4.3	34.8
$25-34$	14.3	4.9	4.2	53.2
$35-44$	13.4	2.6	4.2	23.0
$45-54$	14.3	2.0	4.1	23.27
$55-64$	13.7	2.1	3.7	24.6
$65-$	16.4	4.4	3.3	61.8
Married	51.5	9.9	19.0	82.7
BA degree	42.0	14.3	6.7	83.8
Per capita income	29,959	11,664	3,451	86,604
Population	47,124	54,437	61	705,334
Hispanic population (%)	15.3	16.2	0.4	91.8
Asian population (%)	6.1	7.7	0.0	64.1
Black population (%)	12.4	15.8	0.0	96.1

Table 3: Distribution of Markets

Characteristics	Mean	Std	Min	Max
Number of stores	4.5	3.5	1	21
Number of products	8090.9	2358.8	159[a)]	13971
Avg. driving minutes to nearest store	10.2	8.3	0	38.7

[a)] There are three monopoly markets that were reported to have sold fewer than 1,000 different products (UPCs) only.

Table 4: Key Variables: *PRCOMP*, *StrPRC*, *StrNPR*, and *StrRPR*

	No. of obs	Mean	Std	Min	Max
Variation in the whole sample					
PRCOMP	2,694,043	0.53	0.29	0.05	1.00
StrPRC	1,053	0.66	0.10	0.29	0.98
StrNPR	1,053	4565	1638	70	7625
StrRPR	1,053	0.49	0.19	0.007	0.98
Variation across markets					
Range of *StrPRC* in a market	289	0.13	0.08	0.00	0.38
Range of *StrRPR* in a market	289	0.36	0.18	0.002	0.92

Table 5: Baseline classification of stores

	mean	std	median	min	max
type S (192 stores)					
StrRPR	0.23	0.08	0.23	0.012	0.401
StrPRC	0.70	0.07	0.69	0.449	0.877
type L (879 stores)					
StrRPR	0.62	0.12	0.61	0.382	0.984
StrPRC	0.65	0.11	0.64	0.310	0.982

	No. of L stores	No. of products	Sales (dollar)
Markets with S (134 markets)	2.8	5835.6	2,184,040
Markets w/o S (152 markets)	3.3	5498.4	1,843,904

Each column in the bottom panel displays the average among the large stores only.

Table 6: Cross-Subsidization Pricing (Model 1)

	x=*PRCOMP*	$x=I_{Mono}$	
Variables	(1)	(2)	(3)
I_S	-0.146***	-0.191***	-0.191***
	(0.006)	(0.007)	(0.007)
$I_S \times PRCOMP$	-0.028***		
	(0.005)		
PRCOMP	-0.006**		
	(0.003)		
$I_S \times I_{Mono}$		0.011***	0.011***
		(0.002)	(0.002)
I_{Mono}		0.001	0.001
		(0.004)	(0.004)
I_{byS}			-0.001
			(0.003)
StrRPR	-0.198***	-0.188***	-0.188***
	(0.007)	(0.006)	(0.006)
StrPRC	-0.013	-0.046***	-0.046***
	(0.012)	(0.010)	(0.010)
$I_S \times StrRPR$	0.072***	0.044***	0.044***
	(0.012)	(0.013)	(0.013)
$I_S \times StrPRC$	0.002***	0.248***	0.248***
	(0.0001)	(0.009)	(0.009)
NStr	-0.0001	-6.31e-06	3.44e-07
	(0.0004)	(0.0004)	(0.0003)
No. of observations	4,965,882	4,965,882	4,965,882
Adj. R^2	0.959	0.959	0.959

Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Table 7: Cross-subsidization Pricing (Model 2)

	x=*PRCOMP*	x=I_{Mono}	
Variables	(1)	(2)	(3)
I_S^{clst}	-0.051***	-0.069***	-0.071***
	(0.006)	(0.009)	(0.009)
$I_S^{clst} \times PRCOMP$	-0.015***		
	(0.004)		
$PRCOMP$	-0.001		
	(0.003)		
$I_S^{clst} \times I_{Mono}$		0.010*	0.012**
		(0.005)	(0.005)
I_{Mono}		0.0005	0.003
		(0.004)	(0.005)
I_{byS}^{clst}			-0.009***
			(0.002)
$StrRPR$	-0.191***	-0.191***	-0.193***
	(0.006)	(0.008)	(0.008)
$StrPRC$	0.043***	0.042***	0.033***
	(0.009)	(0.006)	(0.006)
$I_S^{clst} \times StrRPR$	0.036***	0.014*	0.018***
	(0.007)	(0.007)	(0.007)
$I_S^{clst} \times StrPRC$	0.001***	0.101***	0.111***
	(7.13e-05)	(0.012)	(0.012)
$NStr$	-0.001	-0.0005	-0.0004
	(0.0004)	(0.0003)	(0.0003)
No. of observations	4,965,882	4,965,882	4,965,882
Adj. R^2	0.959	0.959	0.959

Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Table 8: Advertising frequencies

						Coefficient of variation	
	No. of obs	Mean	Std	Min	Max	across stores	within chain
f^{Ad}	6,839,199[a)]	4.34	8.05	0	100	1.35	0.36
Regular price	6,861,335	4.68	5.74	0.01	140.0	0.15	0.05

[a)] Excludes 21 stores that report to have no advertising activity over the sample period. The coefficient of variation across stores is calculated for each UPC, and then averaged over all the UPCs.

Table 9: Relationship between product range and advertising

	Specification			
Independent Variables	(1)	(2)	(3)	(4)
StrNPR	.361***	.629***	.482***	.669***
	(.007)	(.008)	(0.011)	(.008)
StrNPR× *NStr*			-.003***	-.0005
			(.0007)	(.0003)
StrNPR× *Dist*			1.002***	.001***
			(.0004)	(.0002)
Dist			-.018***	-.004***
			(.002)	(.001)
StrNPR× *RvlNPR*			-.002***	-.008***
			(.0001)	(.0002)
RvlNPR			.093***	.040***
			(.008)	(.004)
PRCOMP	-.001***	.003***	-.001***	.003***
	(.0002)	(.0001)	(.0002)	(.0001)
I_{Comn}	.021*	-.012*	.025**	-.011**
	(.010)	(.007)	(.010)	(.007)
I_{Mono}	-.088***	.058***	-.095***	.057***
	(.008)	(.005)	(.008)	(.005)
Market FE	Y	Y	Y	Y
UPC-city FE	Y		Y	
UPC-chain FE		Y		Y
No. of observations	6,638,213	6,638,213	6,638,213	6,638,213
Adj. R^2	0.635	0.887	0.636	0.887

Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01. *StrNPR* and *RvlNPR* are in thousands.

Table 10: Relationship between product range and prices

Independent Variables	Specification (1)	(2)	(3)	(4)
StrNPR	-.015***	.043***	-.017***	.032***
	(.0007)	(.001)	(.002)	(.002)
StrNPR× *NStr*			.001***	.0004***
			(8e-05)	(.0001)
StrNPR× *Dist*			.0003***	.0004***
			(5e-05)	(6e-05)
Dist			-.001***	-.001***
			(.0003)	(.0003)
StrNPR× *RvlNPR*			-.003***	.001***
			(.0002)	(.0003)
RvlNPR			.015***	-.007***
			(.001)	(.002)
PRCOMP	-.016***	-.166***	-.012**	-.165***
	(.005)	(.008)	(.005)	(.008)
I_{Comn}	-.004*	.010***	-.007***	.009***
	(.002)	(.002)	(.002)	(.002)
I_{Mono}	-.002	-.005**	-.0003	-.004**
	(.002)	(.002)	(.002)	(.002)
Market FE	Y	Y	Y	Y
UPC-city FE	Y		Y	
UPC-chain FE		Y		Y
No. of observations	6,659,766	6,659,766	6,659,766	6,659,766
Adj. R^2	0.983	0.982	0.983	0.982

Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01. *StrNPR* and *RvlNPR* are in thousands.

Table 11: Relationship between product range and prices at the chain-level

Dependent variable: $\overline{RP}_{i,c}$	
$\overline{StrNPR}_c$	-.00006***
	(3×10^{-6})
UPC FE	Y
No. of observations	765,595
Adj. R^2	0.95

Table 12: Relationship between advertised and unadvertised prices

Dependent variable $=TPD^{Non\text{-}Ad} - \widehat{TPD}^{\,Non\text{-}Ad}$	(1) OLS	(1) IV	(2) OLS	(2) IV
$TPD^{Ad} - \widehat{TPD}^{Ad}$	-0.031***	-0.483***	-0.051***	-0.417***
	(0.011)	(0.023)	(0.017)	(0.026)
Constant	10.653***	5.658***	0.648***	0.413***
	(0.125)	(0.259)	(0.021)	(0.031)
No. of observations	76,574	76,574	76,574	76,574
Adj. R^2	0.02	–	0.03	–

Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Table 13: Relationship between advertised and unadvertised prices (store-by-store)

	Mean	Std	Min	Max	No. of entities with $I_{+,j} = 1$ /total No. of entities	Average of positive coefficients†
Store level	-0.021	0.206	-0.619	0.628	665/1503	0.163
Chain level	-0.023	0.163	-0.435	0.336	51/111	0.117

†: The average is taken across the positive correlation coefficients only.

Table 14: Pricing patterns and signaling-effect of advertised prices

Dependent variable	RP	TPR	f^{Ad}
$I_{+,j}$	0.071***	-1.300***	0.354***
	(0.001)	(0.076)	(0.006)
Constant	4.796***	-19.935***	4.111
	(0.010)	(0.562)	(3144)
No. of observations	6,861,335	6,861,335	6,839,199
Adj. R^2	0.943	0.02	0.46

Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Table 15: Large stores' promotional pricing

Dependent variable	Sold by the closest store S		Sold by any store S	
$=TPD^C - \widehat{TPD^C}$	OLS	IV	OLS	IV
$TPD^M - \widehat{TPD^M}$	-0.054*	-0.182***	-0.074***	-0.218***
	(0.030)	(0.020)	(0.012)	(0.013)
Constant	0.712***	0.615***	0.804***	0.678***
	(0.032)	(0.023)	(0.017)	(0.016)
No. of observations	8,306	8,306	19,662	19,662
Adj. R^2	0.010	–	0.014	–

Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Table 16: Robustness Checks: cross-subsidization pricing

	Specification					
	Rob-1	Rob-2	Rob-3	Rob-1	Rob-2	Rob-3
Variables	(1)	(1)	(1)	(3)	(3)	(3)
I_S	-0.132***	-0.063***	-0.213***	-0.123***	-0.049***	-0.187***
	(0.008)	(0.007)	(0.009)	(0.008)	(0.007)	(0.009)
$I_S \times PRCOMP$	-0.0231***	-0.015***	-0.035***			
	(0.005)	(0.005)	(0.004)			
$PRCOMP$	-0.005	0.002	-0.005			
	(0.005)	(0.003)	(0.003)			
$I_S \times I_{Mono}$				0.014***	0.032***	0.004
				(0.003)	(0.004)	(0.003)
I_{Mono}				0.008***	0.009***	0.017***
				(0.002)	(0.003)	(0.003)
I_{byS}				0.014***	0.009***	-0.028***
				(0.002)	(0.001)	(0.002)
$StrRPR$	-0.083***	-0.177***	-0.219***	-0.079***	-0.180***	-0.220***
	(0.008)	(0.005)	(0.007)	(0.008)	(0.005)	(0.008)
$StrPRC$	-0.008	-0.018***	0.006	0.002	-0.008	-0.003
	(0.006)	(0.006)	(0.008)	(0.007)	(0.009)	(0.006)
$I_S \times StrRPR$	-0.071***	-0.004***	0.024***	-0.048***	0.106***	-0.0006
	(0.009)	(0.010)	(0.007)	(0.015)	(0.011)	(0.008)
$I_S \times StrPRC$	0.148***	0.135***	0.349***	0.236***	0.097***	0.328***
	(0.011)	(0.012)	(0.015)	(0.017)	(0.014)	(0.015)
$NStr$	-0.002***	-0.002***	-0.002***	-0.001***	-0.002***	-0.002***
	(0.0001)	(0.0001)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
No. of obs.	4,016,994	4,846,140	3,932,986	4,016,994	4,846,140	3,932,986
Adj. R^2	0.959	0.967	0.960	0.959	0.967	0.960

Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Table 17: Robustness checks: product range and advertising/prices

Dependent Variable	Specification			
	Advertising frequency		*Regular price*	
Independent Variables	Rob-1 (1)	Rob-2 (2)	Rob-1 (3)	Rob-2 (4)
StrNPR	.497***	.343***	-.054***	-.025***
	(.015)	(.012)	(.002)	(.002)
StrNPR× *NStr*	.010***	.001***	.003***	.001***
	(.001)	(.0005)	(.0002)	(7×10^{-5})
StrNPR× *Dist*	.004***	.006***	.0006***	.0007***
	(.0004)	(.0003)	(7×10^{-5})	(5×10^{-5})
Dist	.030***	.026***	-.003***	-.003***
	(.002)	(.002)	(.0004)	(.0003)
StrNPR× *RvlNPR*	.041***	-.005***	.002***	.0007***
	(.002)	(.001)	(.0003)	(.0002)
RvlNPR	.241***	.018***	-.006***	-0.006***
	(.010)	(.008)	(.002)	(.002)
PRCOMP	-.017	-.149***	-.015**	-.010***
	(.028)	(.023)	(.006)	(.006)
I_{Comn}	-.016	.037***	-.010***	-.003
	(.012)	(.011)	(.002)	(.002)
I_{Mono}	-.118***	-.160***	.004***	.003
	(.010)	(.009)	(.002)	(.002)
No. of observations	5,215,829	6,485,274	5,234,760	6,485,274
Adj. R^2	0.643	0.637	0.983	0.983

Dependent Variable	*Average regular price in the chain*[a)]	
Market definition	Rob-1	Rob-2
Average *StrNPR* at chain	-0.059***	-0.058***
	(0.002)	(0.002)
No. of observations	752,876	768,367
Adj. R^2	0.951	0.951

Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01. [a)] Both specifications include UPC fixed effects. *StrNPR* and *RvlNPR* are in thousands.

Table 18: Robustness checks: signaling-effect of advertised prices

	Rob-1			Rob-2		
Dependent variable	RP	TPR	f^{Ad}	RP	TPR	f^{Ad}
$I_{+,j}$	0.063***	-1.236***	0.072***	0.072**	-1.429***	0.318***
	(0.002)	(0.086)	(0.001)	(0.001)	(0.079)	(0.006)
Constant	4.429***	-1.798	4.804***	4.804***	-22.22***	3.458***
	(0.028)	(1.488)	(.026)	(0.026)	(1.432)	(0.110)
No. of observations	6,515,567	4,515,993	6,807,087	6,807,087	4,714,066	6,784,951
Adj. R^2	0.943	0.020	0.943	0.943	0.018	0.468

Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01. All include market fixed effects.