



Selected Poster/Paper prepared for presentation at the Agricultural & Applied Economics Association's 2017 AAEA Annual Meeting, Chicago, Illinois, July 30-August 1, 2017

Copyright 2017 by [authors]. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

Retail Market Power in a Shopping Basket Model of Supermarket Competition

March 21, 2017

Abstract

Supermarket consumers typically purchase more than one item at a time. Retail prices, in turn, are likely to depend on demand relationships between multiple categories of goods in consumers' shopping baskets. In this paper, we develop a model of retail price competition that explicitly models the effect of complementary demand relationships between products that appear in consumer shopping baskets. We derive inferences for retail market power when shopping baskets contain products from complementary categories and compare outcomes with the predictions derived from conventional models that assume consumers make discrete choices among independent product categories. Our findings reveal that cross-category product complementarity in consumer shopping baskets facilitates substantially greater retail market power relative to the benchmark case of discrete choice over independent goods.

Keywords: complementarity, retailing, pricing, supermarkets, oligopolies.

JEL Codes: D43, L13, L81, M31.

1 Introduction

Consumers typically purchase more than one item at a time from supermarkets. Such behavior suggests that demand relationships among products frequently purchased together are likely to play an essential role in determining how multi-product retailers set prices on individual products comprising consumers' shopping baskets. Yet, despite empirical regularities in consumer purchasing behavior at the basket level, the literature to date examining supermarket pricing behavior tends to rely on data from only one product category, or else implicitly assumes independence across product categories by applying an identical, single-category demand model for all categories in consumers' shopping baskets at once. In this paper, we develop and test a model that allows us to jointly estimate equilibrium price-setting behavior among competing retailers across multiple categories of products that have interrelated demands in consumer shopping baskets.

We examine demand relationships among typical items in consumers' shopping baskets using a structural model of equilibrium retail prices that is conditioned on a multi-product demand framework. Specifically, we adopt a flexible demand structure that encompasses the entire range of demand relationships from perfect substitutes to perfect complements, a structure that nests by construction the case of shopping baskets comprised of independent products. Our nested empirical structure allows us to develop counterfactual experiments on the impact of substitution effects among items in consumers' shopping baskets relative to the limiting case of "category independence" commonly specified in models of supermarket pricing.

Empirical models of competition among retailers typically assume consumers make discrete choices among products within a category, discrete choices among categories within a retail store, and discrete choices among different retail stores (Bell and Lattin 1998; Briesch, Chintagunta and Fox 2009). Here, we maintain the assumption that consumers make discrete choices among different retail stores, an outcome that Smith and Thomassen (2012) demonstrate is approximately true, but depart from the conventional approach by considering a

shopping-basket model of demand in which consumer choices involve purchasing decisions over multiple product categories at a time. Relaxing the assumption in retail pricing models that consumers purchase multiple items with category-independent demands on each shopping occasion (Ainslie and Rossi 1998; Manchanda, Ansari, and Gupta 1999; Russell and Petersen 2000; Chib, Seetharaman, and Strijnev 2002; Kwak, Duvvuri, and Russell 2015) allows us to derive novel inferences on the effect of product complementarity among typical items in consumers’ shopping baskets on retail pricing behavior.

It is well-recognized that demand relationships among items in a shopping basket are essential to understanding many retail pricing strategies. For example, a “loss-leader” pricing strategy (Lal and Matutes 1994; Chevalier, Kashyap and Rossi 2003), in which a product is sold below its marginal cost, makes sense for a retailer only when such pricing behavior stimulates commensurate sales of complementary goods or services. Our analysis explicitly models the effect of complementary demand relationships among items in a shopping basket, which allows us to consider retail pricing strategies at the basket level, rather than at the individual product level of supermarket behavior.

We consider a structural model of consumer demand for items in a representative shopping basket at two, competing retail stores in a small US city. Our empirical model consists of a nested multi-variate logit (MVL) specification that encompasses a rich set of cross relationships between shopping basket items and between stores (Russell and Petersen 2000; Niraj, et al. 2008; Moon and Russell 2008; Kwak, Duvvuri, and Russell 2015). Conditioned on this MVL demand structure, we analytically derive the Bertrand-Nash solution to a retail pricing game over multiple products. Given that retailer rents are determined by store-level sales rather than brand-level sales, we condition equilibrium pricing decisions among competing supermarkets according to how consumers respond to these multi-product price changes.

Understanding how consumers respond to retail price changes of individual products requires internalizing demand relationships between the various goods in a shopping basket.

Because a typical shopping basket consists of many different categories of products, the net effect of internalizing substitute and complement relationships between products on retail prices is largely an empirical question. We derive implications for the role of cross-product demand effects on supermarket pricing behavior by relying on the structure of the nested MVL model to compare the equilibrium prices generated by our unrestricted MVL model to those generated by traditional multinomial logit (MNL) models that restrict the demand structure between goods.

We examine the effect of product complementarity on retail prices using panel data from supermarket retailers in the Eau Claire, Wisconsin market. Specifically, we construct shopping baskets for households in our sample and examine the effect of shopping basket composition on retail prices for four categories of goods: Milk, breakfast cereal, soft drinks, and snacks. Our empirical results provide strong support for the hypothesis that selling complementary goods softens retail price competition. That is, relative to the case of independent goods, we find evidence of higher overall retail prices for shopping baskets composed of complementary products.

Our findings suggest that price competition between oligopoly supermarkets is significantly less intense when retailers sell complementary products. We provide intuition for this outcome by developing a model of supermarket behavior that provides a clear decomposition between the *intra*-retailer margin and *inter*-retailer margin of supermarket behavior. On the *intra*-retailer margin, supermarkets act as monopolists for consumers that enter the store, fully internalizing cross-effects in demand when setting prices for items in the shopping basket (Hamilton and Richards 2009). Lower prices facilitate complementary purchases within the store, providing retailers with an incentive on the *intra*-retailer margin to set lower prices when supermarket products are complementary goods. On the *inter*-retailer margin, supermarkets compete with rivals to acquire store traffic. Because oligopoly retailers ignore the effect of price changes on the profit of rival retailers, introducing demand relationships between products in consumers' shopping baskets conveys an additional pricing externality on

the inter-retailer margin through product composition effects in the multi-product demand structure.

For complementary products, a selective price discount in one product category raises cross-category sales of complementary products, thereby increasing the value of a typical shopping basket. Complementary effects on demand are fully internalized on the intra-retailer margin; however, because retailers fail to recognize the effect of prices on sales at rival retailers, both direct pricing effects and cross-product pricing effects on rival retailers are ignored when setting prices. It is well known that oligopoly retailers set lower prices than monopoly prices due to the “business-stealing effect”, which arises for each independent product through a failure to internalize the positive effect of a price increase on rivals’ profits. Here, we show that product complementarity in multi-product pricing decisions introduces second-order effects on rival sales when retailers steal business at the basket level, which are also ignored when retailers set prices, and these cross-effects between typical items in consumers’ shopping baskets reinforce business-stealing effects on retail prices in the case of substitutes and offset them in the case of complements.

Our analysis reveals that the effect of product complementarity on retail margins depends on the intensity of inter-retailer competition. In markets with relatively weak competition between retailers, for instance when transportation costs between retailers is “high”, retailers set lower prices for shopping baskets containing complementary products than in the case of independent goods; however, when competition between retailers is relatively intense, retailers set higher prices for shopping baskets containing complementary products.¹ Thus, our findings suggest that the effect of product complementarity on raising retail margins is accentuated as retail markets become more saturated.

Our paper also contributes to the empirical literature on shopping-basket demand, and

¹Our argument is a variation on the harvest-invest story of Dube, Hitsch, and Rossi (2009) and Pavlidis and Ellickson (2012) in that retailers compete relatively more intensively in order to earn high-margin customers. On the other hand, when categories are complements within the store – when retailers sell only store-brands, which are complementary due to “umbrella branding” (Erdem and Chang 2012), for example – then price competition is less intense, and market conduct is less competitive.

retail pricing. While others have used the MVL model to examine consumer demand at the shopping-basket level (Russell and Petersen 2000; Niraj, et al. 2008; Moon and Russell 2008; Kwak, Duvvuri, and Russell 2015), the link between shopping basket composition, consumer store choice, and pricing among competing retail stores has remained unexplored. While understanding shopping-basket-demand patterns is critical for retail marketing, we know little about how consumers’ basket purchases are manifest in observed retail pricing strategies. For retailing practitioners, our findings provide new insight into how strategic pricing generalizes to the shopping-basket level. That is, prices that may be optimal at the product or category level may not, in fact, be ideal when interactions among items in multiple categories are taken into account. While traditional economic thinking maintains that complementarity enhances competitiveness (Smith and Thomassen 2012; Rhodes 2015) we show the opposite – that shopping-basket demand is a source of market power.

In the next section, we derive a theoretical model of retail pricing that accounts for shopping-basket purchases and show that equilibrium prices can rise when products in the basket are complements in demand. We test this theory using the empirical model of store-choice and shopping-basket demand described in Section 3. Section 4 describes our data source and provide some stylized facts that support the use of a MVL model to estimate demand inter-relationships among grocery categories. Section 5 summarizes our empirical findings and offers some implications for the conduct of retailing more generally, while we conclude and offer some suggestions for future research in Section 6.

2 Basket Composition and Pricing

In this section we present a simple theoretical model that isolates retail pricing incentives on the *intra*-retailer margin and *inter*-retailer margin of a supermarket. Consumers in the model engage in one-stop-shopping, acquiring a basket of goods comprised of multiple products on each shopping occasion. Retail pricing at the basket level, in turn, is driven by two opposing incentives: (*i*) on the inter-retailer margin, retailers wish to lower retail prices on all retail

products to steal business from rivals; whereas (ii) on the intra-retailer margin, retailers maintain an optimal mix of prices, fully internalizing externalities between goods in the representative shopping basket by setting “Ramsey” prices. Customers that visit a given supermarket purchase multiple products on a single shopping trip, which stimulates retailers to internalize complementary brand relationships in the multi-product demand system among consumers entering the retail store.

Consider duopoly supermarkets that stock products in multiple categories. The retailers differ in their spatial proximity to consumers in the Hotelling (1929) sense and stock product categories that contain complementary goods, for instance breakfast cereal and milk. Because our focus is on how equilibrium prices change with the degree of complementarity between products, we accordingly simplify the model by considering retailers that stock an exogenous number of products.²

Each retailer is located at the end of a unit line segment and consumers are distributed uniformly along the line segment between retailers so that no one retail location is inherently superior to any other retail location. Consumers incur increasing transportation costs of τ per unit of distance to visit retailers.³ The decision to shop with a given retailer consequently depends on the transportation cost required to visit the retailer relative to the consumption opportunity afforded by that retailer’s product assortment and prices.

Consumer preferences over retail products are represented by the utility derived from one-stop shopping. Specifically, given her choice of retailer and consumption bundle (x_1, x_2, \dots, x_n) , utility for the representative consumer is given by

$$u(x_1, x_2, \dots, x_n) = \sum_i p_{j,i} x_i.$$

Solving this problem for the optimal consumption bundle selected at retailer j yields the

²See Anderson and dePalma (1992, 2006) and Hamilton and Richards (2009) for analysis of product variety choices among multi-product retailers.

³Transportation costs for visiting retailers is assumed to be sufficiently high that consumers purchase multiple products on each shopping occasion and compare between supermarkets at the basket level rather than the individual product level.

indirect utility function

$$v^*(\mathbf{p}_j) = \max_{x_1, x_2, \dots, x_n} u(x_1, x_2, \dots, x_n) - \sum_i p_{j,i} x_i,$$

where $\mathbf{p}_j = (p_{j,1}, p_{j,2}, \dots, p_{j,n})$ is the vector of prices selected by retailer j .

Aggregate demand facing each retailer depends on the decisions made by consumers at all points on the line segment regarding where to shop. Given consumer transportation costs of τ per unit distance, a consumer at a distance of $\theta \in (0, 1)$ from retailer j could achieve surplus of $v^*(\mathbf{p}_j) - \theta\tau$ by purchasing from that retailer. Letting θ^* denote the location of the consumer who is indifferent between the alternative of shopping with either retailer, θ^* solves $v^*(\mathbf{p}_1) - \theta\tau = v^*(\mathbf{p}_2) - \tau(1 - \theta)$, which yields

$$\theta^*(\mathbf{p}_1; \mathbf{p}_2) = \frac{1}{2} + \frac{1}{2\tau} [v(\mathbf{p}_1) - v(\mathbf{p}_2)]. \quad (1)$$

All consumers located at a distance of $\theta \leq \theta^*$ prefer to shop with retailer 1 and all consumers located at a distance of $\theta^* \leq \theta$ prefer to shop with retailer 2. The demand for retail product i , at retailer 1 accordingly, is $X_i(\mathbf{p}_1; \mathbf{p}_2) = \theta^*(\mathbf{p}_1; \mathbf{p}_2)x_i(\mathbf{p}_1)$ and total store demand for retailer 1 is defined accordingly by aggregating products in the representative consumer's basket: $X(\mathbf{p}_1) = \theta^*(\mathbf{p}_1; \mathbf{p}_2) \sum_i x_i(\mathbf{p}_1)$.

Now consider the problem of retailer 1. Suppose each retailer pays a fixed set-up cost, F , and a constant unit cost of c to stock an individual product. Denoting *per-customer* profit for retailer 1 as

$$\pi(\mathbf{p}_1) = \sum_i (p_{1,i} - c)x_i(\mathbf{p}_1), \quad (2)$$

total retailer profit for retailer 1 is given by

$$\Pi(\mathbf{p}_1; \mathbf{p}_2) = \theta^*(\mathbf{p}_1; \mathbf{p}_2)\pi(\mathbf{p}_1) - F. \quad (3)$$

Differentiating (3) with respect to $p_{1,i}$ gives the first-order necessary condition

$$\theta^*(\mathbf{p}_1; \mathbf{p}_2) \frac{\partial \pi(\mathbf{p}_1)}{\partial p_{1,i}} + \frac{\pi(\mathbf{p}_1)}{2\tau} \frac{\partial v(\mathbf{p}_1)}{\partial p_{1,i}} = 0, \quad i = 1, 2, \dots, n, \quad (4)$$

where $\partial v(\mathbf{p}_1)/\partial p_{1,i} = -x_i(\mathbf{p}_1) < 0$ holds by Roy's identity. Notice that condition (4) decomposes the effect of a price change into an *inter*-retailer margin and an *intra*-retailer margin of profit. The first term on the left-hand side of equation (4) is the effect of a price change on the *intra*-retailer margin. For a given amount of store traffic (θ^* fixed), the retailer sets relative prices like a monopolist, selecting "Ramsey" prices that fully internalize demand relationships among products. The second term on the left-hand side of equation (4) defines the effect of a price change on the *inter*-retailer margin. A small decrease in price of $dp_{1,i}$ units shifts $(x_i/\tau)dp_{1,i}$ customers towards retailer i and away from his rival through the *business-stealing effect*. Because each customer purchases multiple products from the retailer, each additional customer results in retail profit of $\pi(\mathbf{p}_1)$, resulting in rents of $(x_i/\tau)\pi(\mathbf{p}_1)dp_{1,i}$ from a decrease in the price of good i .

Condition (4) implies oligopoly prices are proportionately lower than monopoly prices. A monopoly retailer not disciplined by competition on the inter-retailer margin would choose category prices such that $\partial\pi(\mathbf{p}_1)/\partial p_i = 0$; however, an oligopoly retailer selects prices below the monopoly price level, that is $\partial\pi(\mathbf{p}_1)/\partial p_i > 0$ in equation (4), because the business-stealing effect of a price increase (the second term on the left-hand side of (4) is negative).

The retail pricing implications of a wide variety of product configurations is completely characterized by equation (4). To understand how demand complementarity affects super-market prices, it is helpful to examine how the demand structure of items in a shopping basket alters pricing incentives on the intra- and inter-retailer margins. To simplify this comparison, consider the retail market equilibrium with n symmetric product categories, each of which contains one product. Dropping arguments for notational convenience, condition (4) can be written in the symmetric case as

$$\frac{p - c}{p} = \frac{1}{\varepsilon_{ii} - (n - 1)\varepsilon_{ij} + S/\tau}, \quad (5)$$

where $\varepsilon_{ii} = -\frac{\partial x_i p_i}{\partial p_i x_i} > 0$ is the own-price elasticity of demand, $\varepsilon_{ij} = \frac{\partial x_j p_i}{\partial p_i x_j}$ is the cross-price elasticity of demand between product categories, and $S = np_x$ is total retail sales. The first two terms in the denominator determine the intra-retailer margin. Absent competition from

rival retailers, retail margins are given by $\frac{p-c}{p} = \frac{1}{\varepsilon_{ii} - (n-1)\varepsilon_{ij}}$, which is the Ramsey pricing condition for a multi-product monopolist.⁴ On the intra-retailer margin, complementarity between retail categories, $\varepsilon_{ij} < 0$, is associated with narrower margins than in the case of independent goods, as retailers internalize the cross-product effect of a price discount on facilitating sales of complementary goods among consumers shopping within their retail store.

The third term in the denominator on the right-hand side of equation (5) captures the retail pricing incentive on the inter-retailer margin. This term is positive and larger in magnitude when consumer transportation cost (τ) is “small”; thus, the intensity of retail competition is essential for determining the effect of basket composition on retail market power. For larger values of τ , equilibrium retail prices are driven predominantly by incentives on the intra-retailer margin, resulting in decreased retail market power for categories with more complementary demand; however, for smaller values of τ , pricing incentives on the intra-retailer margin become more pronounced.

A complementary demand structure among products in consumer shopping baskets results in upward pressure on equilibrium retail prices on the inter-retailer margin. The intuition for this is the business-stealing effect: Oligopoly retailers set prices below the monopoly level due to a failure to internalize the positive effect of higher prices on rivals’ sales. When retailers set prices simultaneously for multiple products, the business-stealing effect involves losing an entire shopping basket of sales for each consumer that switches stores. The composition of products in consumers’ shopping baskets therefore introduces a second externality on the inter-retailer margin, as retailers fail to consider the cross-price effects among products comprising the shopping basket on rival’s sales. When consumers’ shopping baskets contain complementary products, higher retail prices on one category of products reduces sales of other products in the shopping basket, dampening business-stealing effects at the basket level, thereby softening price competition on the inter-retailer margin relative to the

⁴This equation reduces to the single-product monopoly markup in the case of independent goods, $\varepsilon_{ij} = 0$.

case of independent goods.

To confirm this intuition, we simulate price changes in response to changes in demand complementarity using two demand structures: (i) quadratic utility; and (ii) constant elasticity of substitution (CES). The outcome is qualitatively similar in each case, with a range of τ emerging in which retail provision of complementary product categories results in higher retail prices than in the case of independent (or substitute) goods.

Consider the quadratic utility structure

$$u(x_1, x_2, \dots, x_n) = \alpha \sum_i x_i - \frac{1}{2}\beta \sum_i x_i^2 - \lambda \sum_i \sum_{j \in n} x_i x_j,$$

where $\alpha, \beta, \lambda > 0$ and $\beta > \lambda$. This utility structure leads to demand of the form

$$x_i = a - bp_i + \delta \sum_{j \neq i} p_j, \text{ for all } i,$$

where $a = \alpha/(\beta + \lambda)$, $b = \beta/(\beta + \lambda)^2$, and $\delta = \lambda/(\beta + \lambda)^2$. Products are complements when $\delta < 0$.

Table 1 shows the results of numerical simulation for the case where $a = 1$, $b = 1$, $c = \frac{1}{2}$ and $n = 2$ for variations in transportation cost (τ) and demand complementarity (δ). Notice that retail prices decrease monotonically as the categories become more complementary when transportation costs are “high” ($\tau = 1$), decreasing from $p = \$0.73$ for the case of independent categories ($\delta = 0$) to $p = \$0.53$ for the case of “strong” complements ($\delta = -0.8$). The reason is that pricing outcomes are predominantly driven by the intra-retailer margin when consumers face high transportation costs.

[table 1 in here]

For the case when transportation costs are “low” ($\tau = 0.01$), retail prices follow a non-monotonic pattern with changes in demand complementarity. The retail price peaks at $\delta = -0.7$, rising with the degree of complementarity between product categories for $\delta \in (-0.7, 0)$, and then falling at still higher degrees of complementarity. The reason is that retail prices converge towards marginal cost on the intra-retailer margin for high levels of product complementarity, dampening business-stealing incentives that result from ignoring

product composition externalities on the inter-retailer margin. In all cases, retail prices are higher when retailers stock complementary product categories than in the case of independent product categories.

Consumer transportation costs are unobservable to the econometrician in scanner data. Therefore, it is unclear *a priori* whether the effect of complementarity in the first column or the second column is a better description of retail pricing in practice. In the next section we describe an empirical approach designed to test the effect of complementarity on equilibrium retail pricing.

3 Empirical Model of Retail Pricing and Demand Complementarity

Overview

We estimate a structural model of retail demand, and retailer pricing. Our primary objective is to determine the effect of demand structure in consumers' shopping baskets on retail market power. We begin by specifying a hierarchical empirical model in which consumers first choose stores, and choose baskets conditional on store choice, and then derive expressions for retail prices that are consistent with equilibrium in a Bertrand-Nash environment, conditional on the structure of demand between stores and within the shopping basket. We conclude this section by describing how both the demand and pricing elements of the model are estimated.

Model of Retail Demand

Consumers $h = 1, 2, 3, \dots, H$ in our model select items from among $i = 1, 2, 3, \dots, N$ categories, c_{iht} , in assembling a shopping basket, $\mathbf{b}_{ht} = (c_{1ht}, c_{2ht}, c_{3ht}, \dots, c_{Nht})$ on each trip, t , conditional on their choice of store, r . Define the set of all possible baskets in r as $\mathbf{b}_{ht}^r \in \mathbf{B}^r$. Utility is assumed to be separable between the component arising from the store (W_{rht}) and the basket of items (V_{bht}) so we write utility from the combined choice as

$$U_{bht} = W_{rht} + V_{bht} + \varepsilon_{bht}, \tag{6}$$

where the unobservable component of the basket-and-store choice utility, ε_{iht} , is distributed Generalized Extreme Value (GEV, Train 2003). In this section, we first develop the basket-choice component of utility, and then the store-choice part, before bringing them together in a nested model of basket-and-store choice.

We first focus on the probability of choosing a particular basket of items on a given shopping occasion, and model demand by assuming consumers purchase one item per category across multiple categories. Consumers choose among categories to maximize basket-utility, V_{bht}^r , which we further index by r to emphasize that the choice of basket at this point is conditional on the choice of store. We follow Song and Chintagunta (2006) and Kwak, Duvvuri, and Russel (2015) in writing basket-utility in terms of a discrete, second-order Taylor series approximation,

$$V_{bht}^r(\mathbf{b}_{ht}|r) = \sum_{i \in \mathbf{b}} \pi_{iht}^r c_{iht}^r + \sum_{i,j \in \mathbf{b}} \theta_{ijh}^r c_{iht}^r c_{jht}^r, \quad (7)$$

where π_{iht}^r is the baseline utility for category i earned by household h on shopping trip t in store r , c_{iht}^r is a discrete indicator that equals 1 when category i is purchased in store r , and 0 otherwise, and θ_{ijh}^r is a household-specific parameter that captures the degree of interdependence in demand between categories i and j in store r . Specifically, $\theta_{ijh}^r < 0$ if the categories are substitutes, $\theta_{ijh}^r > 0$ if the categories are complements, and $\theta_{ijh}^r = 0$ if the categories are independent in demand. To ensure identification, we restrict all $\theta_{ii}^r = 0$ and impose symmetry on the matrix of cross-purchase effects, $\theta_{ijh}^r = \theta_{jih}^r, \forall i, j \in r$ (Besag 1974, Cressie 1993, Russell and Petersen 2000).

The probability that a household purchases a product from a given category on a given shopping occasion depends on both perceived need, and marketing activities from the brands in the category (Bucklin and Lattin 1992, Russell and Petersen 2000). Therefore, we write baseline utility for each category as dependent on a set of category (\mathbf{X}_i) and household (\mathbf{Z}_h) specific factors such that

$$\pi_{iht}^r = \alpha_i^r + \beta_h \mathbf{X}_{it}^r + \gamma \mathbf{Z}_h, \quad (8)$$

where perceived need is determined by household factors such as the size of the household

(HH_h), the age of the household head (AGE_h), income (INC_h), and educational attainment (ED_h). Marketing mix elements at the category level include a price index of the individual items in each category (PR_{it}^r), the proportion of items featured during the purchase occasion (FT_{it}^r), the percentage on display (DP_{it}^r), and the share on temporary price reduction (TPR_{it}^r).⁵ While there are likely other factors that influence category choice, this set covers those typically used in the category-choice literature (Manchanda, Ansari, and Gupta 1999) and exhaust those available in our data.

Each of the variables entering (8) represent sources of observed heterogeneity, whether at the category (\mathbf{X}_{it}^r) or household (\mathbf{Z}_h) levels. However, Gentzkow (2007) argues that identifying complementarity, separate from unobserved heterogeneity, is difficult without either exploiting the panel structure of the data, or through a valid instrumentation strategy. Because one of the primary goals in estimating a MVL is to capture any complementarity that may be present in the data, we control for unobserved heterogeneity in household preferences, and in the interaction in preferences among categories, that may affect category incidence. To do so, we specify a subset of the estimated parameters to be randomly distributed in the most general form of the model in order to capture unobserved heterogeneity in marketing mix responsiveness, and the interaction terms that are key to the MVL model, respectively. In the most general form of the model, therefore, we specify:

$$\begin{aligned}\beta_{ph} &= \beta_{ph0} + \beta_{ph1}\nu_1, \quad \nu_1 \sim N(0, \sigma_1), \\ \theta_{ilh}^r &= \theta_{il0}^r + \theta_{il1}^r\nu_2, \quad \nu_2 \sim N(0, \sigma_2),\end{aligned}\tag{9}$$

for the price-element of the marketing-mix matrix, and the interaction between categories i and l for the inter-dependence parameters. By controlling for unobserved heterogeneity at the household-level, any remaining complementarity, either through price-effects or interaction effects, will be identified by our econometric model.

With the error assumption in equation (6), the conditional probability of purchasing in

⁵The price index is calculated as a time-varying, volume-share weighted index of prices for individual items offered in each category.

each category assumes a relatively simple logit form. Following Kwak, Duvvuri, and Russell (2015), we simplify the expression for the conditional incidence probability by writing the cross-category purchase effect in matrix form, suppressing the store index on the individual elements, where: $\Theta_h^r = [\Theta_{1h}, \Theta_{2h}, \dots, \Theta_{Nh}]$ and each Θ_{ih} represents a column vector of the $N \times N$ cross-effect Θ_h^r matrix defined as

$$\Theta_h^r = \begin{bmatrix} 0 & \theta_{12h} & \theta_{13h} & \dots & \theta_{1Nh} \\ \theta_{21h} & 0 & \theta_{23h} & \dots & \theta_{2Nh} \\ \theta_{31h} & \theta_{32h} & 0 & \dots & \theta_{3Nh} \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \theta_{N1h} & \theta_{N2h} & \theta_{N3h} & \dots & 0 \end{bmatrix}. \quad (10)$$

making use of this notation, we express the conditional utility of purchasing any specific category i within basket \mathbf{b}_{ht}^r as

$$V_{iht}^r(c_{iht}^r | c_{jht}^r) = \pi_{ht}^{r'} \mathbf{b}_{ht}^r + \Theta_{ih}^{r'} \mathbf{b}_{ht}^r, \quad (11)$$

for the items in the basket vector \mathbf{b}_{ht}^r .

Conditional utility functions of this type potentially convey important information, and are more empirically tractable than the full probability distribution of all potential assortments (Moon and Russell 2008). However, a limitation of this structure is that the conditional utility functions cannot describe the entire matrix of substitute relationships in a consistent way, and are not econometrically efficient in that they fail to exploit the cross-equation relationships implied by the utility maximization problem. To see this more clearly, we derive the estimating equation implied by the GEV error-distribution assumption, conditional on the purchases made in all other categories, c_{jht}^r . With this conditional assumption, the probability of purchasing in category $i = 1$ is written as

$$\Pr(c_{1ht}^r = 1 | c_{jht}^r, r) = \frac{[\exp(\pi_{1ht}^r + \Theta_{1h}^{r'} \mathbf{b}_{ht}^r)]^{c_{1ht}^r}}{1 + \exp(\pi_{1ht}^r + \Theta_{1h}^{r'} \mathbf{b}_{ht}^r)}, \quad (12)$$

and \mathbf{b}_{ht}^r represents the basket vector. Estimating each N of these equations separately is one option; however, Besag (1974) describes how the full distribution of \mathbf{b}_{ht}^r choices are

consistently, and efficiently, estimated together as a single system that uses all the available information to identify any cross-category relationships that may be present in the data.

Assuming the Θ_h^r matrix is fully symmetric, and the main diagonal consists entirely of zeros, then Besag (1974) shows that the probability of choosing each of the category-combinations that comprise the basket-vector \mathbf{b}_{ht}^r is written as

$$\Pr(\mathbf{b}_{ht}^r|r) = \frac{\exp(\boldsymbol{\pi}_{ht}^{r'} \mathbf{b}_{ht}^r + \frac{1}{2} \mathbf{b}_{ht}^{r'} \Theta_h^r \mathbf{b}_{ht}^r)}{\sum_{\mathbf{b}_{ht}^r \in \mathbf{B}^r} [\exp(\boldsymbol{\pi}_{ht}^{r'} \mathbf{b}_{ht}^r + \frac{1}{2} \mathbf{b}_{ht}^{r'} \Theta_h^r \mathbf{b}_{ht}^r)]}, \quad (13)$$

where $\Pr(\mathbf{b}_{ht}^r|r)$ is interpreted as the joint probability of choosing the observed combination of categories from among the 2^N potentially available from N categories, still conditional on the choice of store r .⁶ Assuming the elements of the main diagonal of Θ^r is 0 necessary for identification, while the symmetry assumption is required to ensure that (??) truly represents a joint distribution, a multi-variate logistic (MVL, Cox 1972) distribution, of the category-purchase events. Essentially, the model in (??) represents the probability of observing the simultaneous occurrence of N discrete events – a shopping basket – at one point in time. Due to the iid assumption of the GEV errors associated with each basket choice, the model in (??) implicitly assumes that the baskets are subject to the independence of irrelevant alternatives (IAA), but the categories within each basket are allowed to assume a more general correlation structure (Kwak, Duvvuri, and Russell 2015).

The model to this point describes the probability of basket choice, conditional on the choice of store. Because we are interested in the unconditional probability of choosing a particular basket of groceries, we next describe a model of store choice that allows us to derive the joint probability of choosing a basket and a store such that: $\Pr(\mathbf{b}_{ht}, r) = \Pr(\mathbf{b}_{ht}^r|r) \Pr(r)$, where $\Pr(r)$ is the probability of choosing retailer r . Estimating the model as a nested variant

⁶The practical limitations of describing 2^N choices are somewhat obvious. Recently, others have developed ways to either reduce the dimensionality of the \mathbf{b}_{ht} vector, or of estimating it more efficiently. Kwak, Duvvuri, and Russell (2015) focus on "clusters" of items within conventional category definitions, while Moon and Russel (2008) project the \mathbf{b}_{ht} vector into household-attribute space, so only 2 parameters are estimated. Kamakura and Kwak (2012) use the random-sampling approach of McFadden (1978) to reduce the estimation burden while leaving the size of the problem intact. Because our problem is well-described with only a small number of categories (5), we estimate the MVL model in its native form.

of the MVL also allows us to test the hypotheses that follow from our theoretical model as a complementary demand structure among products in a typical shopping basket will have implications for equilibrium retail prices on both the inter- and intra-retailer margin.⁷ Nesting category-and-store choices within a conventional logit-assumption is a well-accepted technique for accomplishing this purpose (Bucklin and Lattin 1992; Bell and Lattin 1998).

Store choice depends upon attributes of the store, and both observable and unobservable attributes of the consumer that may mean that one store is more desirable than another. Bell, Ho, and Tang (1998) differentiate between fixed costs of visiting a particular store – those that do not vary with the amount purchased – and the variable costs, or those that depend on how much the consumer buys. Fixed costs include the distance between the household and each store, as well as measures of state dependence that capture opportunity costs of departing from habits or familiar patterns of behavior. Similar measures are also used by Bell and Lattin (1998) and Briesch, Chintagunta, and Fox (2009) to describe store choice. Variable costs include measures of the relative cost of shopping at each store, typically captured by a price index (Bell, Ho, and Tang 1998), the assortment available at each store (Briesch, Chintagunta, and Fox 2009), and the relative attractiveness of purchasing a basket of items from each store (Bell and Lattin 1998).

We capture each of these elements of store choice in writing the indirect utility from choosing store r as

$$W_{rht} = \delta \mathbf{Y}_{rh} + \lambda IV_{rh} = \delta_{0r} + \delta_1 DIST_{rh} + \delta_2 LOY_{rh} + \delta_3 VAR_r + \delta_4 PRI_r + \lambda IV_{rh}, \quad (14)$$

for a set of household-and-store variables \mathbf{Y}_{rh} , where $DIST_{rh}$ is the distance, measured in Euclidean terms, between household h and store r , LOY_{rh} is a measure of state-dependence, defined as loyalty to a particular store, which we operationalize as the percentage of visits to store r in the 6 months prior to the estimation period (Bell and Lattin 1998), VAR_r is a measure of the depth of assortment across all categories in store, calculated as the total number

⁷Although consumers do not necessarily choose one store exclusively in practice, we model their choice of *primary* store, which typically accounts for over 80% of their grocery expenditure (Smith and Thomassen 2012).

of universal product codes (UPCs) offered each week in store r , PRI_r is a price index calculated as a sales-weighted average across all categories in the model, and the inclusive value (IV_{rh}) for store r and household h , which captures the relative attractiveness of the store as measured by the expected utility from a basket of groceries estimated in the conditional basket choice model, written as: $IV_{rh} = \log \left(\sum_{\mathbf{b}_{ht}^r \in \mathbf{B}^r} [\exp((\boldsymbol{\pi}_{ht}^{r'} \mathbf{b}_{ht}^r + \frac{1}{2} \mathbf{b}_{ht}^{r'} \Theta_h^r \mathbf{b}_{ht}^r)^{1/\lambda})] \right)$.⁸ The nesting parameter, λ , also referred to as the substitution parameter, measures the extent to which groups, stores in this case, substitute for each other. If $\lambda = 1$, our nested MVL model collapses to a non-nested MVL model of choice among baskets and stores as the GEV distribution simply becomes a product of logit distributions (Train 2003).

To complete the nested MVL model structure, the GEV error assumption in the combined basket-and-store choice utility model implies that the choice among stores takes the familiar logit form,

$$\Pr(r) = \frac{\exp(\delta \mathbf{Y}_{hr} + \lambda IV_{hr})}{\sum_{s \in R} \exp(\delta \mathbf{Y}_{hs} + \lambda IV_{hs})}, \quad (15)$$

with the same hierarchical structure as a more traditional nested logit, for each household, h . With estimates of the store choice model, we then express the joint probability of each basket-and-store choice as

$$\Pr(\mathbf{b}_{ht}, r) = \Pr(\mathbf{b}_{ht}^r | r) \Pr(r) = \sum_{\mathbf{c}_{jht}^r \in \mathbf{b}_{ht}^r} \left(\frac{\exp((\boldsymbol{\pi}_{ht}^{r'} \mathbf{b}_{ht}^r + \frac{1}{2} \mathbf{b}_{ht}^{r'} \Theta_h^r \mathbf{b}_{ht}^r)^{1/\lambda})}{\sum_{\mathbf{b}_{ht}^r \in \mathbf{B}^r} [\exp((\boldsymbol{\pi}_{ht}^{r'} \mathbf{b}_{ht}^r + \frac{1}{2} \mathbf{b}_{ht}^{r'} \Theta_h^r \mathbf{b}_{ht}^r)^{1/\lambda})]} \right) \left(\frac{\exp(\delta \mathbf{Y}_{hr} + \lambda IV_{hr})}{\sum_{s \in R} \exp(\delta \mathbf{Y}_{hr} + \lambda IV_{hr})} \right) \quad (16)$$

With this closed-form expression for the unconditional choice of each basket / store combination, we then derive the effect of a price change within one basket on equilibrium prices within each store on the intra-retailer margin, and across rival stores on the inter-retailer margin.

Demand Elasticities

⁸We include measures of price at both the category- and store-choice levels because relative prices are expected to affect category-choice and store-choice in different ways.

Intra- and inter-retailer price effects are measured by price elasticities of demand. We derive the individual-item elasticities implied by (16) and (15) in this section, and apply them to the analysis of competitive price-and-assortment response in the subsequent section. To foreshadow our results, we find that the elasticity expressions look very similar to the usual logit price-elasticity expressions, but with one critical difference: Because each category can appear in several bundles, the derivative must sum over the marginal effect of a change in price on the probability of observing each bundle that contains the category. Formally, the household-level marginal effect of a change in a same-category, same-store price is

$$\frac{\partial \Pr(c_{jht}^r, r)}{\partial PR_j} = \sum_{c_{jht}^r \in \mathbf{b}_{ht}^r} \left(\frac{\beta_{ph} \Pr(c_{jht}^r, r)}{\lambda} (1 - (1 - \lambda) \Pr(c_{jht}^r | r) - \lambda \Pr(c_{jht}^r, r)) \right), \quad (17)$$

where β_{ph} is the household-specific marginal utility of income, and $\Pr(c_{jht}, r)$ includes all baskets that contain the category j in store r . Similarly, the marginal effect of a change in the price index for a different category (i) in the same store on the probability of purchasing category j , when the categories are in the same baskets is given by

$$\frac{\partial \Pr(c_{jht}^r, r)}{\partial PR_i} = \sum_{c_{jht}^r \in \mathbf{b}_{ht}^r} \left(\frac{-\beta_{ph} \Pr(c_{jht}^r, r)}{\lambda} ((1 - \lambda) \Pr(c_{iht}^r | r) + \lambda \Pr(c_{iht}^r, r)) \right), \quad (18)$$

and the marginal effect of change in the price of a category that is not in the same bundle as j is given by

$$\frac{\partial \Pr(c_{jht}^r, r)}{\partial PR_k} = \sum_{c_{jht}^r \in \mathbf{b}_{ht}^r} \left(\frac{-\beta_{ph} (\Delta \Pr)}{\lambda} ((1 - \lambda) \Delta \Pr + \lambda \Delta \Pr) \right), \quad (19)$$

where the argument $\Delta \Pr = \Pr((c_{jht}^r, r), (c_{kht}^r, r)) - \Pr(c_{jht}^r, r) \Pr(c_{kht}^r, r)$ is interpreted as the difference between the joint probability of observing categories j and k purchased together in the same basket and store, less the product of the marginal probabilities of observing each category purchase, again in the same store.⁹ Finally, we derive the cross-price elasticity for a pair of products that are not in the same bundle, and are not purchased in the same store. Logically, the demand relationship between pairs of products purchased in different stores is

⁹Note that these probabilities are marginal with respect to categories, but still joint with respect to stores.

only manifest through the store-choice component of the model, so the cross-price elasticity becomes

$$\frac{\partial \Pr(c_{jht}^r, r)}{\partial PR_l} = \sum_{c_{jht}^r \in \mathbf{b}_{ht}^r} (-\beta_{ph} \lambda \Pr(c_{jht}^r, r)), \quad (20)$$

for all products l not in the same store as j .

With these expressions, we can estimate an entire matrix of price responses, for all categories with respect to all other categories, accounting for the fact that they may or may not be purchased in the same shopping basket.

Price Response

In this section, we complete the structural model of price response by deriving the optimal retailer response to choices made by rivals. Retailers maximize profit by choosing category-prices in Bertrand-Nash rivalry, conditioned on market demand aggregated over the household behavior described in (16). Retailers are assumed to purchase items sold in each category from manufacturers, and pass through input cost increases to consumers.¹⁰ Dropping time subscripts for clarity, the profit equation for retailer r is written as

$$\Pi^r = M \sum_{i \in I_r} S_i^r (PR_i^r - C_i^r) - F^r, \quad (21)$$

where Π^r is the profit for retailer r , M is the size of the aggregate market for all products, I_r is the set of all categories in retailer r , S_i^r is the share of retailer r in category i , C_i^r is the marginal cost of producing, and selling, category i in retailer r , and F^r reflects the retailer's fixed cost of selling items in all categories.

Marginal costs in equation (21), which reflect both the costs of producing and retailing items in each category, are specified as a linear function of input prices (Villas-Boas 2007). This results in the following expression for retailing costs:

$$C_i^r(\mathbf{v}_r) = \sum_{l \in L} \eta_{wl} v_{rl} + \epsilon_{ir}, \quad (22)$$

¹⁰To the extent that retailers have some purchasing power over manufacturers, this assumption may be an oversimplification. Our assumption simply implies that what we refer to as retailer margins below, may in fact be a combination of manufacturer and retailer margins. However, if complementarity does not influence upstream market power, which it should not, our qualitative conclusions will not be affected by this assumption.

where \mathbf{v}_r is a vector of L input prices, η_{wl} are estimates of the contribution of each input price to unit costs, and ϵ_{ir} is an iid error term. Retailing costs are estimated after substituting equation (22) into the first-order conditions derived below. Conditional on the structure of demand, retailer r 's first order condition for the price of category i is given by

$$\frac{\partial \Pi^r}{\partial PR_i^r} = MS_i^r + M \sum_{i \in I} (PR_i^r - C_i^r) \frac{\partial S_i^r}{\partial PR_j^r} = 0, \quad \forall i \in I, r \in R, \quad (23)$$

where $\partial S_i^r / \partial PR_j^r$ is one element of a matrix of share derivatives with respect to price for all categories, i and j . Notice that equation (23) implies that each retailer internalizes all cross-sectional pricing externalities across categories in the store, but does not take into account the effect of his category pricing on the sales of other retailers. Stacking the first-order conditions across retailers we define the ownership matrix as Ω , which has element $\omega_{ir} = 1$ if category i is sold by retailer r (and zero otherwise). Making use of this notation, we write the first-order condition as

$$\mathbf{PR} = \mathbf{C} - \phi(\Omega \mathbf{S}_p)^{-1} \mathbf{S} + \boldsymbol{\epsilon}, \quad (24)$$

where bold notation indicates a vector (or matrix), and \mathbf{S}_p is the matrix of share-derivatives with element $\partial S_i^r / \partial PR_j^r$.¹¹ We include the parameter ϕ in this model, the "conduct parameter" in order to measure the extent of deviation from the maintained form of the pricing game. If the estimate of $\phi = 1$, then retailers do indeed compete as Bertrand-Nash rivals and any markup is due entirely to the extent of product differentiation (category differentiation) reflected in the matrix of share derivatives. If, however, the estimate of $\phi = 0$, then margins are zero and retail prices are consistent with perfect competition. How complementarity affects margins, and pricing power, therefore is reflected in estimates of ϕ and the implied equilibrium prices that result.

We estimate equation (24) to recover the parameters of the retail cost function using information from the demand side and the structure of the game. Based on these estimates,

¹¹The specific form of these derivatives for the random-coefficient MVL model are provided in the technical appendix.

we next conduct a set of counter-factual simulations to investigate the importance of complementarity on equilibrium retail prices. In another estimation-and-simulation exercise, we compare fitted prices from our maintained model to alternative estimates that assume prices are conditioned on a more usual, logit model of retail demand. In this way, we achieve our dual objectives of both describing a model of retail pricing at the shopping basket level and testing our theoretical model of how complementarity affects retail prices.

Estimation and Identification

Category-level price indices are likely to be endogenous as retailers take into account many important demand-shifters (e.g., shelf-space, advertising, manufacturer-promotions, etc.) that are unobservable in our data. Our identification strategy is well-accepted in the literature, and relies on the control function approach of Petrin and Train (2010). Essentially, the control function approach consists of using the residuals from a first-stage instrumental variables regression as additional variables on the right-side of the demand model. Because the residuals from the instrumental variables regression contain information on the part of the endogenous price variable that is not explained by the instruments, they have the effect of removing the correlated part from the demand equation. Because input prices are expected to be correlated with retail prices, and yet independent of demand, our first-stage control function regression uses the set of input price variables as instruments. We also include category and retailer fixed-effects in order to account for any endogenous effects that are unique to each group of products. Although these variables should represent effective instruments, whether they are weak in the sense of Staiger and Stock (1997) is evaluated on the basis of the F-test that results from the first-stage instrumental variables regression. In this case, the F-statistic is 164.2, which is much larger than the threshold of 10.0 suggested by Staiger and Stock (1997). Therefore, we conclude that our instruments are not weak.

In the absence of unobserved heterogeneity, the MVL model is estimated using maximum likelihood in a relatively standard way. However, because we allow several parameters to vary randomly across panel observations, the likelihood function no longer has a closed form.

Therefore, we estimate the model using simulated maximum likelihood (Train 2003), using $g = 1, 2, 3, \dots, G$ simulations. For clarity of the likelihood function, we index the possible baskets ($15 = 2^N - 1$, excluding the null basket) by k , and define a set of indicator variables z_{rk} that assume a value of 1 if basket k is chosen in store r and 0 otherwise. We then estimate the likelihood function as a panel over h cross-sections and t shopping occasions per household so that the simulated likelihood function is written (Kwak, Duvvuri, and Russel 2015) as follows:

$$\mathcal{L}(\mathbf{b}_{ht}, r) = \frac{1}{G} \sum_{g=1}^G \prod_t \prod_k (\Pr(\mathbf{b}_{ht} = \mathbf{b}_{ht}^k, r)^{z_{rk}}), \quad (25)$$

where the joint distribution function for all possible baskets is given in (16). We then take the log of (25), sum over all households, and maximize with respect to all parameters: $LLF(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\lambda}, \boldsymbol{\theta}) = \sum_{h=1}^H \log \mathcal{L}_h(\mathbf{b}_{ht}, r)$. To increase the efficiency of the SML routine, the simulated draws follow a Halton sequence with 50 draws, as suggested by Bhat (2003). We experimented with a range of starting values and Halton draws, and our results did not change substantially from one trial to the next, so we conclude that our estimates are relatively stable. In the Results and Discussion section that follows, we present results from a number of alternative specifications in order to establish the validity of our maintained model.

Identifying pricing conduct in structural models is nearly always problematic. First, our analysis uses pricing data from only two retailers in a single market. Although this would seem to be a small sample from which to draw inferences on pricing, the retailers interact weekly over a three-year period, so time series variation, combined with differences in pricing among categories, should be sufficient to estimate both marginal cost and the conduct parameters. Moreover, others in this literature have used either fewer retailers (Villas-Boas and Zhao 2005) or the same number (Sudhir 2001) and report no issues in identifying retailer pricing behavior. Second, in the pricing model (24), the markup term on the right-side is clearly endogenous. Consequently, we estimate the pricing model using a control-function method similar to that used for the demand-side model. Correcting for endogeneity in this

way requires a set of instruments that are likely to be correlated with retail margins, but mean-independent of the error term. Intuitively, identifying retailer pricing conduct requires instruments that shift the demand curve facing retailers. For this purpose, we use marketing mix variables for the other store in the market, average values across the sample for each demographic variable, and a set of category dummy variables. A first-stage regression of this set of instruments on the retail margin yields an F-value of 168.7, so our set of instruments is, again, not weak in the sense of Staiger and Stock (1997). Our identification strategy is well-accepted in the literature (Richards and Hamilton 2015) so should at least yield results that are comparable to others.

We estimate the structural model, which consists of the demand model in (25) and the supply model in (24), sequentially, first estimating the demand model and then the supply, or pricing, model conditional on the demand estimates. While simultaneous estimation is preferred on efficiency grounds, estimating sequentially does not introduce any source of bias and, in a multi-category model as complex as this one, is the only tractable econometric approach.

4 Data Summary and Stylized Facts

Our sample is comprised of purchase-occasion based household-scanner data from the IRI Academic Data Set (Bronnenberg, Kruger, and Mela 2008) in the Eau Claire, WI market over the period 2009 - 2011. For each household in our sample, we construct a shopping basket consisting of milk, cereal, carbonated soft drinks, and salty snack purchases.¹² Generalizing the analysis of consumer purchasing behavior to consider demand relationships among these four categories allows us to examine the impact of expected demand complements (milk and cereal; soft drinks and snacks) as well as anticipated substitutes (milk and soft drinks) on supermarket pricing of shopping baskets.

We chose these four categories because they are representative of a typical shopping cart,

¹²While typical shopping baskets for households in our sample contain more than four items, the MVL model quickly becomes intractable for unrestricted choice sets (Kamakura and Kwak 2012).

include products from only a few large manufacturers, and are well-described by the IRI data. In fact, all four were in the top eight grocery categories in the US in 2013 (#1 soda, #2 milk, #4 snacks, #8 cereal (Supermarket News 2014)), while the others in the top eight were not chosen because they are either dominated by in-store products (#3 bread), or highly fragmented, speciality categories (#5 beer, #6 cheese, and #7 wine). Each of our selected categories has a penetration rate above 70% in the IRI data, resulting in a large number of transactions for all households across the four categories. The high purchase frequency among categories in our sample allows us to capture other forms of complementarity besides use-complementarity, such as when retailers market umbrella brands (Erdem 1998), or when frequently-purchased items happen to follow similar purchase cycles. Moreover, store brands are well-represented in the milk and soft drinks categories, and are purchased in approximately the same frequency, leading to a large number of observations in which products from these categories are purchased together in a household shopping basket.

To ensure a rich, within-subject data set, we only retain households with at least 50 purchase occasions over the 3-year sample period. Focusing on households with a large number of repeat purchases allows us to control for state-dependence in demand using household-varying inventory variables for each cross-sectional observation.

In panel data, it is necessary to have data on prices for not only the product that was purchased, but those that were not purchased as well. For this purpose, we merged the household- and store-level data sets by store, week, and UPC. By combining the household and store-level data, we observe the complete set of prices, and other marketing mix variables, for all UPCs available on a given purchase trip.

Despite the relatively small nature of the Eau Claire retail market, the retail supermarket industry appears to be highly fragmented, with two dominant stores, and a number of smaller, "fringe" competitors. In order to minimize the confounding effect of store-specific loyalties to lesser players in the market, we include only households who purchase cereal from one of the two most popular stores in the data set (IRI keys 257871 and 1085053), which together

account for well over 60% of weekly sales for each category in our sample. By focusing on the two most important stores in the market, our estimates are most likely to capture strategic pricing behavior, rather than decisions taken in isolation by a niche competitor with a highly loyal, price-insensitive, customer base.

We choose to examine shopping basket composition effects on supermarket pricing in Eau Claire, because Eau Claire is a relatively small (approximately 65,000 population), homogeneous city, with two, competing supermarket chains and a relatively small number of alternatives. Nevertheless, there is considerable variation in household demographics across stores in our sample. Table 2 provides a summary of household demographics in our sample, indexed according to store of purchase. Notice that households shopping at store 1 have annual income nearly \$10,000 (20.7%) higher than households shopping at store 2. This difference in average income suggests that consumers shopping at store 1 may have less price sensitivity for products and greater demand for high-quality items relative to consumers shopping at store 2. Shoppers at store 1 are similar in age, but slightly more educated, and have fewer children than shoppers at store 2. Milk and soft drinks are important drivers of basket volume, suggesting that the difference in family size may have important implications for how each supermarket sets shopping basket prices. Across all households in our sample, 73.4% switch among stores, while 19.9% visit store 1 exclusively and 6.7% visit store 2 exclusively. Among households switching among stores, 58.5% of the time these customers visit store 1 and 41.5% of the time they visit store 2. Because store 1 appears to be inherently more attractive to households than store 2, which is potentially due to other factors besides basket-pricing, we include a store fixed effect in the empirical model described below.

[table 2 in here]

Variation in shopping-basket composition is necessary to identify cross-category demand relationships in our data. Ideally, observed shopping baskets in our sample would span the entire set of combinations among the four categories, with some households purchasing one of each item, some purchasing two items together, some three, and some all four items to-

gether in the shopping basket. Table 3 provides a summary of the shopping behavior of all households across both stores, and the amount spent on each type of shopping basket. As expected, the most common shopping basket item across the four categories we examine is milk, with 28.1% of all shopping occasions representing a single-purchase “milk run.” Two-item shopping baskets are purchased on 29.4% of all shopping occasions, with the most common basket consisting of milk and soft drinks, followed closely by milk and snacks. The popularity of two-item shopping baskets containing both milk and soft drinks baskets is not surprising given the frequent incidence of both milk and soft drink purchases.¹³ Three-item shopping baskets are purchased on 10.2% of all shopping occasions, with the most common combination being milk, soft drinks and snacks. Customer shopping baskets contain items from all four categories on 1.8% of all shopping occasions. Overall, the distribution of shopping basket purchases in our sample involves sufficient representation from each category-combination to identify parameters of the MVL model.

[table 3 in here]

The summary information in Table 3 reveals several interesting patterns for shopping basket expenditure. First, notice that consumer expenditure is higher for shopping baskets that contain soft drinks than for any other basket with a comparable number of items. This anecdotal evidence suggests that retailers have an incentive to set a mix of prices that promotes soft drink purchases in as many baskets as possible, which may explain why soft drinks are promoted and displayed more often than any other category in our sample (see Table 4). Second, as expected, larger shopping baskets generally entail greater expenditure, even on a per item basis. Third, expenditure on larger shopping baskets is more stable and, hence, more predictable for retailers. The coefficient of variation of shopping basket expenditure among single-item purchases is over 75%, whereas it is only 49% for three-item purchases and 43% for four-item purchases. Larger and more stable patterns of basket expenditure are inherently more attractive for retailers, although whether this leads to more

¹³On each visit to the store, households purchase in anticipation of several consumption occasions, and for several different family members, likely with heterogeneous tastes (Dube 2004).

pricing power is an empirical question.

[table 4 in here]

Examining marketing mix data by store provides some summary insights in this regard. The entries in Table 4 provide a summary of retail prices, marketing mix activity and inventory holding behavior for each category and each store. Notice that there are marked differences in category prices charged by each store, particularly in the milk and soft drink categories. Given that the stores are of the same general format (traditional supermarkets containing approximately 35,000 SKUs each), marketing many of the same national brands, and they are only 1.5 miles apart, this pricing evidence suggests that the stores do indeed compete using different multi-category pricing strategies. In light of the importance of soft drinks in generating larger shopping baskets, for example, the greater frequency in which store 1 displays and promotes soft drinks can partially explain the fact that store 1 attracted over 50% more customers in the Eau Claire market than store 2. Given the lower average household income of customers at store 2, it is surprising that milk and soft drink prices are set higher at store 2, particularly given the greater frequency of milk and soft drinks in larger, multi-item shopping baskets.

In terms of the pricing model, we estimate the departure from the maintained Bertrand-Nash behavioral assumption by estimating marginal costs in equation (21) and inferring any difference between the implied margin and the competitive one as due to some form of imperfect competition. By estimating the conduct parameter, we neither restrict ourselves to a particular mode of conduct, nor do we assume the absence of errors in profit maximization.¹⁴ Input prices include a category-specific primary ingredient price (fluid milk for the milk category; wheat, rice, and oats for the cereal category; sugar for soft drinks; and wheat and salt for the snack category, each from the Bureau of Labor Statistics (BLS)), indices of wages earned by workers in the food retailing and food manufacturing industries (BLS),

¹⁴The criticism of the conduct parameter approach by Corts (1999) is well understood, but the alternative "menu approach" (Gasmi, Laffont, and Vuong 1992) also suffers from conceptual issues in that each assumed form of competition is absolute. That is, the researcher implicitly assumes there are no deviations from each type of assumed conduct. We prefer this more general approach.

and producer price indices for utilities, energy, packaging, advertising, and other business services (BLS). The summary data in Table 2 documents the extent of variation in each input-price series, and suggests that there is sufficient variation to identify the parameters of the marginal cost function.

In the next section, we present the results of our structural pricing model on the link between shopping basket composition and market power.

5 Results and Discussion

Overview

We begin our presentation of the results with the MVL demand estimates, comparing fixed-coefficient versions with a random-coefficient version in order to examine the importance of unobserved heterogeneity. We then compare the MVL estimates with a logical alternative – a binary logit category-choice model that assumes independence among categories. After establishing the preferred demand model, we then present estimates of the supply side, or pricing conduct model, comparing equilibrium prices under the maintained MVL model and the logit alternative. Because the MVL model consists of a mix of complementary and substitute categories, however, and the point of our research is to examine the effect of complementary and substitute relationships among categories on equilibrium prices, we present results from a cleaner simulation experiment in which we compare equilibrium prices under varying levels of full-complementarity and full-substitute relationships. This counterfactual experiment clearly demonstrates the impact of consumers’ shopping-basket purchase behavior on equilibrium prices.

Demand Estimates

Estimates from two versions of the general MVL model for category choice in each store are shown in table 5. In this table, Model 1 does not allow for unobserved heterogeneity in the response to any variable as all coefficients are assumed to be fixed. We relax this assumption in Model 2 by allowing for random category-interaction terms, as well as price-

response coefficients. Because Model 1 is nested within the more general version, likelihood ratio (LR) tests are appropriate tests of model mis-specification. The LR test statistic used to compare Model 1 to Model 2 is Chi-square distributed with 14 degrees of freedom (there are 13 restrictions involved in nesting Model 1 in Model 2), which implies a critical Chi-square value of 22.36. The calculated LR statistic is 37.96 for the nested model, so we easily reject Model 1 in favor of Model 2, and conclude that unobserved heterogeneity is an important factor driving category choice. Consequently, we interpret the results from Model 2 for category choice.

[table 5 in here]

Among the parameter estimates in Table 5, the own-price coefficient is less than zero and strongly significant.¹⁵ With respect to the marketing mix variables, we find that display and promotion activity are statistically significant, and apparently highly effective, but feature has a statistically-significant and negative impact on utility. This finding is perhaps due to the fact that most featured items are also promoted, and often on display, so much of the benefit of featuring is captured by these other variables. Each of the category-store preference terms is statistically significant and, interestingly, exhibit different patterns of preference between the stores. Because these preference terms are estimated while controlling for each of the other marketing-mix and market-composition variables, this finding likely represents a manifestation of the different strategies used by the two retailers. Among the demographic variables, we find that income and education have positive effects on the utility derived from category choice, but household size and age are statistically significant and negative. Overall, the model provides a relatively good fit to the data.

Our primary interest in the MVL model is our θ_{ij} estimates, which define the extent and significance of substitution effects among categories in consumer shopping baskets. These estimates are shown in the second part of table 5. In this table, we impose symmetry as a condition for identification, resulting in six estimated parameters for each store. Among

¹⁵Because these estimates are scaled by different units of measure, the magnitudes of each of these coefficients is not interpretable *per se*.

these, we find a significant positive demand relationship in Store 1 between all pairs of items, except for between milk and soft drinks, and cereal and snacks. Although the magnitudes of the estimates differ between the two models, the pattern of substitution is the same, as milk / soda and cereal / snacks are the only substitute pairs of the six relationships of interest. In general, our estimates for both stores suggest that complementary demand relationships among shopping-cart items are both more frequent and stronger than previously believed. Finding that cereal and snacks are net substitutes is perhaps due to shifting use patterns for cereal, as more consumers choose cereal in the evening and less at breakfast time. Overall, categories in consumer shopping baskets are more likely to be complements than they are substitutes, even after controlling for the traditional price-driven notion of complementarity. Therefore, complementarity cannot be ignored in models of equilibrium pricing.

To this point, however, our within-store estimates are only able to capture the effect of complementarity on the intra-store margin, or how a retailer internalizes demand relationships. With estimates of a store-substitution model, taking the demand relationships within each store into account, we are also able to determine the effect of complementarity on the inter-retailer, or store-choice, margin. Estimates of the store choice model are shown in the lower part of table 5 above. Because store choice is nested in the MVL model, the random-coefficient version is again preferred, of course. As expected, price and distance have a negative effect on store choice, while loyalty and variety are significant, positive determinants of which store is chosen. Most important from this table, however, is the estimate of store substitution, λ . Because this estimate is significantly below 1, our estimates imply that consumers do not readily substitute among stores, but are still sufficiently willing to visit another store that substitution remains a possibility. Recall from the theoretical model above that complementarity is likely to have a net positive effect on equilibrium prices only if the stores are highly substitutable, or are strong structural competitors. With an estimate of $\lambda = 0.37$, these two stores appear to be only weak competitors. Whether the estimated substitution parameter is sufficient to affect equilibrium prices, however, remains an empirical

issue that we address below.

We examine the robustness of our MVL model by comparing our estimates to those obtained from a logit model of store-and-category choice. The first step in this comparison is determining whether a nested, multi-product, multi-store model is preferred to one in which all choices are binary. Clearly, the nature of the demand model used to condition the equilibrium pricing estimates depends on the nested MVL model as a viable description of demand at the shopping basket level. Conventional models of purchase incidence, or category choice, assume categories are purchased according to a discrete, typically binary logit process (Bucklin and Lattin 1992). If consumers make unwavering purchases from a “shopping list” irrespective of posted prices, then there is no clear multinomial choice among categories on a given shopping occasion, but rather a binary decision as to whether or not to make a purchase from each category. For this reason, the logical alternative to our MVL model is a binary logit model applied to each category; however, the error assumption in the logit model implies that each category decision is independent of any other category decision, which implicitly assumes a shopping basket demand structure with independent goods.¹⁶ The MVL model nests the binary logit category choice model as a special case with all θ_{ij} parameters jointly restricted to zero, which allows us to compare these specifications using LR tests.

To remain consistent with the MVL estimation strategy, we estimate several versions of the binary-logit category-choice model, accounting for differing characterizations of unobserved heterogeneity. Models 3 and 4 table 6 represent fixed-coefficient, random category-preference, and fully random-parameter versions of the binary logit category choice model, respectively. As in the case of our MVL model, the preferred logit model is the more comprehensive version (model 2). Comparing this specification with the most general MVL model, the LR test statistic with 12 degrees of freedom (recall we are restricting all of the θ_{ij} parameters to equal zero) has a critical Chi-square value of 21.03, whereas the estimated test

¹⁶Alternatively, if we were to apply a multinomial logit model to category choice, the implication would be that categories are strict substitutes.

statistic value is 29.38. The LR test clearly rejects the binary logit model in favor of the maintained MVL alternative.

[table 6 in here]

Given widespread use of the logit model in the purchase incidence literature, it is worthwhile to compare the parameter estimates from the binary logit model to those from the MVL. This comparison is marked by several notable features. First, with respect to the estimated price coefficients, there is a clear pattern on the direction of mis-specification bias in the logit model as the price parameter is over-estimated by the logit model in each case (less negative). Second, a similar pattern emerges for the category-preference pattern estimates, as each is significantly over-estimated by the logit model. This outcome is intuitive when product categories are complementary, for instance purchasing cereal with milk increases the value of milk in a shopping basket, additional value that is unaccounted for in the binary logit model under the restriction of independent categories. Whether these biases have any implications for equilibrium prices, however, is addressed by comparing equilibrium prices between the MVL and binary logit models of category-and-store choice.

Pricing Model Estimates

Our structural pricing model implies retailers set prices in oligopoly equilibrium, conditional on estimates from a model of consumer demand. How category demand changes with variation in category prices, therefore, is key to deriving the resulting equilibrium prices. We first describe the likely state of retail equilibrium between our two sample stores using our preferred MVL model, and then compare these results to counterfactual experiments that show what equilibrium prices would be like in alternative environments with independent product categories, perfect complements, and perfect substitutes. Our MVL estimates necessarily contain a mix of complementary and substitute relationships among product categories, and for this reason, we motivate our examination of the effect of shopping basket composition on retail pricing behavior by focusing on extreme cases.

Table 7 shows our estimates from the MVL pricing model. Model 1 presents OLS esti-

mates that do not correct for the endogeneity of retail margins, whereas Model 2 presents estimates using the control function approach. Correcting for endogeneity is necessary in both cases, as a Hausman (1978) test rejects the notion that margins are exogenous.¹⁷ Based on these estimates, the primary input is the most important input cost, as expected, but retailing and food manufacturing wages also have a substantial impact on equilibrium prices. Most importantly, the estimated conduct parameter, ϕ , in the MVL model reveals that our retailers are more competitive than the maintained Bertrand-Nash assumption. Specifically, the estimated value of $\phi = 0.39$ implies that retailers are more competitive than under Bertrand-Nash behavior, but less than perfectly competitive, as to be expected when retailers are differentiated both by space, the variety of items on offer, and other unobserved store-quality factors.

[table 7 in here]

Table 8 shows how estimated retail conduct changes when the model is restricted to impose demand independence among categories in consumer shopping baskets. In this case, there are no pricing externalities of the type described by Smith and Thomassen (2012) that can be internalized either to increase margins (if substitutes) or to enhance cross-category sales (if complements). Based on our control-function estimates, notice that estimated market conduct is now $\phi = 0.50$, substantially greater than for the MVL pricing model. This outcome suggests that retailers would behave in a considerably less competitive fashion when setting category-level prices under independent category demand relative to setting prices that internalize demand relationships between categories in the shopping basket. Indeed, the average margin across all four categories implied by the binary logit conduct parameter is approximately \$1.14, compared with an implied equilibrium basket price in the MVL case of only \$0.88.¹⁸

¹⁷The Hausman (1978) specification test compares one estimator that is consistent under the null hypothesis (exogeneity) with one that is efficient under the null, but inconsistent under the alternative hypothesis (control function). The resulting test statistic is Chi-square distributed with one degree of freedom.

¹⁸In grocery retailing, a 30% difference in category-level margins is indeed substantial when net profit margins average less than 2%.

Our comparison so far considers retailer pricing behavior in shopping baskets with a mix of substitute and complementary categories to the outcome under independent goods. Our finding of higher market power in the case of independent goods may be an artifact of the particular product categories we consider, rather than a general outcome for equilibrium prices when retailers stock categories comprised of complementary goods. For this purpose, we extend our analysis to numerically consider the extreme cases of perfect substitutes and perfect complements.

[table 8 in here]

Simulation Exercise

Recall that the logit model alternative describes category purchases as completely independent. While the nested structure of the MVL model is instructive in revealing the specification bias in models of discrete category choice, it does not provide insight on how substitute or complementary relationships among categories affect equilibrium retail prices. For this purpose, we conduct a simple simulation exercise in which we fix the θ_{ij} parameters in the MVL model that govern substitute and complement relationships to values that represent varying levels of product relationships from perfect substitutes to perfect complements. We then calculate equilibrium price indices across all four categories – a "shopping basket index" – that reflect how changes in the structure of demand alone affect likely price outcomes. The result provides a clean test of how different relationships among categories affect equilibrium prices within the context of our maintained model, resulting in a clear *reducto ab absurdum* demonstration of the importance of shopping basket composition effects on equilibrium prices when retailers provide complementary categories of goods.

Table 9 shows the results of various simulation exercises. In this table, we allow the θ_{ij} parameters to take one of three levels, ranging from the base, or estimated, case to high substitutability ($\theta_{ij} = -2$) to high complementarity ($\theta_{ij} = 2$). We also allow the "store substitutability" parameter (λ) to vary from the estimated level (0.37) to one in which there is very little store-substitution ($\lambda = 0.10$) to a case where consumers readily substitute among

stores ($\lambda = 0.90$). In the base case, high substitutability implies an equilibrium average price of approximately \$4.95, while prices are more competitive than the MVL model estimates based on a mix of demand relationships, as the equilibrium price is slightly lower (\$4.91). Under strong complementarity, equilibrium prices fall to roughly \$4.89 on average, indicating more competitive behavior and correspondingly lower equilibrium prices than in the case of mixed category relationships. This finding is consistent with the prior literature (Smith and Thomassen 2012; Rhodes 2015) as prices are lower under complementarity. However, when stores are more competitive, for instance if we were able to pick them up and move them closer together, the store-substitution parameter rises ($\lambda = 0.90$). In this case, pure complementarity implies higher prices than in either the base case or the pure-substitutes case (\$4.68 versus \$4.48), which is consistent with our theoretical model above. If stores are even less substitutable than in the base case ($\lambda = 0.10$), we get the expected result that equilibrium prices again fall with the degree of complementarity.

[table 9 in here]

Our findings provide counterpoint to the recent theoretical literature on how shopping-basket demand relationships affect retail prices (Smith and Thomassen 2012; Rhodes 2015). In these models, the authors focus on the intra-retailer margin, which suggests that internalizing demand complementarity between categories in a shopping basket reduces retail prices; however, these analyses ignore the counteracting incentives to raise prices of complementary goods on the inter-retailer margin. When the interaction of oligopoly retailers is included, we find that complementarity between goods in consumers' shopping baskets can increase equilibrium retail prices if stores are regarded as close substitutes.

6 Conclusions and Implications

In this study, we investigate the role of category-level complementarity on equilibrium retail prices. An emerging theoretical literature argues that the inherent complementarity associated with purchasing groceries by the shopping-basket leads to more competitive pricing

than would otherwise be the case. However, current empirical models are insufficient to test this theory as they implicitly assume categories are independent in demand. We derive an empirical model that is able to accommodate a full range of complementarity and substitute relationships at the category level, and use this model to test whether complementary goods in consumer shopping baskets results in increased retail market power.

Our findings have important implications for our understanding of retail pricing, and the competitiveness of supermarket retailers. Take the growth of store brands, for example. Perhaps the dominant retail trend of the last 20 years, the rise of store brands has been attributed to a host of causes, from increasing bargaining power over manufacturers (Mills 1995) to building brand loyalty for the store (Corstjens and Lal 2000). Our findings suggest another, fundamental reason for introducing store brands. If a retailer can market private labels under one umbrella brand (Erdem and Sun 2002; Erdem and Chang 2012), then the resulting complementarity created within the store can reduce the intensity of price competition on the inter-retailer margin. Retail strategies designed to induce cross-category purchases, such as erecting potato displays in the meat aisle, or offering salad dressings in the produce aisle, are easily explained in terms of our findings as retailers have a clear motive to create as many opportunities for complementarity purchases in a shopping basket as possible.

From a broad, policy perspective, the food retailing sector is generally regarded as being highly competitive. However, the growth of super-center retailing through the likes of Walmart and Target, and the emergence of online retail giants (e.g., Amazon) are based on cross-selling over multiple categories. To the extent that these firms have been, and will likely continue to be, successful in expanding the scope of their customers' shopping baskets, our findings suggest that supermarket retailing may become decidedly less competitive.

The MVL model used at the core of this study is currently the state-of-the-art for analyzing shopping basket demand. However, it is not inherently scalable to the level required to understand an entire shopping basket. Because the MVL becomes intractable for any more

than four or five items, a more general model is necessary. We leave this for future research.

References

- [1] Ainslie, A., and P. E. Rossi. 1998. Similarities in choice behavior across product categories. *Marketing Science* 17: 91-106.
- [2] Bell, D. R., and J. M. Lattin. 1998. Shopping behavior and consumer preference for store price format: Why “large basket” shoppers prefer EDLP. *Marketing Science* 17: 66-88.
- [3] Bell, D. R., T. H. Ho, and C. S. Tang. 1998. Determining where to shop: Fixed and variable costs of shopping. *Journal of Marketing Research* 35: 352-369.
- [4] Bhat, C. R. 2003. Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. *Transportation Research Part B: Methodological* 37: 837-855.
- [5] Besag, J. 1974. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36: 192-236.
- [6] Briesch, R. A., P. K. Chintagunta, and E. J. Fox. 2009. How does assortment affect grocery store choice? *Journal of Marketing Research* 46: 176-189.
- [7] Bronnenberg, B. J., M. W. Kruger, and C. F. Mela. 2008. Database paper-The IRI marketing data set. *Marketing Science* 27: 745-748.
- [8] Bucklin, R. E., and J. M. Lattin. 1991. A two-state model of purchase incidence and brand choice. *Marketing Science* 10: 24-39.
- [9] Bucklin, R. E., and J. M. Lattin. 1992. A model of product category competition among grocery retailers. *Journal of Retailing* 68: 271.
- [10] Chevalier, J. A., A. K. Kashyap, and P. E. Rossi. 2003. “Why Don’t Prices Rise During Periods of Peak Demand? Evidence from Scanner Data.” *American Economic Review* 93: 15-37.

- [11] Chib, S., P. B. Seetharaman, A. Strijnev. 2002. Analysis of multicategory purchase incidence decisions using IRI market basket data. *Advances in Econometrics* 16: 55–90.
- [12] Corstjens, M., R. Lal. 2000. Building store loyalty through store brands. *Journal of Marketing Research* 37: 281-291.
- [13] Cox, D. R. 1972. The analysis of multivariate binary data. *Journal of the Royal Statistical Society Series C* 21: 113–120.
- [14] Cressie, N. A.C. 1993. *Statistics for spatial data*. New York: John Wiley and Sons.
- [15] Erdem, T. 1998. An empirical analysis of umbrella branding. *Journal of Marketing Research* 35: 339–351.
- [16] Erdem, T., and B. Sun. 2002. An empirical investigation of the spillover effects of advertising and sales promotions in umbrella branding. *Journal of Marketing Research* 39: 1–16.
- [17] Erdem, T., S. R. Chang. 2012. A cross-category and cross-country analysis of umbrella branding for national and store brands. *Journal of the Academy of Marketing Science* 40: 86-101.
- [18] Gentzkow, M. 2007. Valuing new goods in a model with complementarity: Online newspapers. *The American Economic Review*, 97(3), 713-744.
- [19] Hamilton, S. F. and T. Richards. 2009. Variety competition in retail markets. *Management Science* 55: 1368-76.
- [20] Hausman, J. A. 1978. Specification tests in econometrics. *Econometrica* 46: 1251-1271.
- [21] Hosken, D. S., L. Olson, and L. Smith. 2012. Do retail mergers affect competition? Evidence from grocery retailing. Federal Trade Commission, Bureau of Economics, Working paper no. 313.

- [22] Hotelling, H. 1929. Stability in competition. *The Economic Journal* 37: 41-57.
- [23] Kamakura, W., and K. Kwak. 2012. Menu-choice modeling. Working paper, Rice University, Department of Marketing.
- [24] Kwak, K., S. D. Duvvuri, and G. J. Russell. 2015. An analysis of assortment choice in grocery retailing. *Journal of Retailing* 91: 19-33.
- [25] Lal, R. and Matutes, C., 1994. Retail pricing and advertising strategies. *Journal of Business*: 345-370.
- [26] Manchanda, P., A. Ansari, S. Gupta. 1999. The shopping basket: A model for multi-category purchase incidence decisions. *Marketing Science* 18: 95–114.
- [27] McFadden, D. 1978. Modelling the choice of residential location. Institute of Transportation Studies, University of California. pp. 75-96.
- [28] Mills, D. E. 1995. Why retailers sell private labels. *Journal of Economics & Management Strategy* 4: 509-528.
- [29] Moon, S. and G. J. Russell. 2008. Predicting product purchase from inferred customer similarity: an autologistic model approach. *Management Science* 54: 71–82.
- [30] Niraj, R., V. Padmanabhan and P.B. Seetharaman. 2008. A cross-category model of households' incidence and quantity decisions. *Marketing Science* 27: 225–35.
- [31] Pavlidis, P., and P. B. Ellickson. 2012. Switching costs and market power under umbrella branding. Working paper, University of Rochester, Department of Economics, Rochester, NY.
- [32] Petrin, A. and K. Train. 2010. A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research* 47: 3-13.
- [33] Rhodes, A. 2015. Multiproduct retailing. *The Review of Economic Studies* 82: 360-390.

- [34] Richards, T. J., and S. F. Hamilton. 2006. Rivalry in price and variety among supermarket retailers. *American Journal of Agricultural Economics* 88: 710-726.
- [35] Richards, T. J. and S. F. Hamilton. 2015. Variety pass-through: An examination of the ready-to-eat cereal market. *Review of Economics and Statistics* 97: 166-180.
- [36] Russell, G. J. and A. Petersen. 2000. Analysis of cross category dependence in market basket selection. *Journal of Retailing* 76: 367-92.
- [37] Smith, H. 2004. Supermarket choice and supermarket competition in market equilibrium. *The Review of Economic Studies* 71: 235-263.
- [38] Smith, H., and D. Hay. 2005. Streets, malls, and supermarkets. *Journal of Economics and Management Strategy* 14: 29-59.
- [39] Smith, H., O. Thomassen. 2012. Multi-category demand and supermarket pricing. *International Journal of Industrial Organization* 30: 309-314.
- [40] Song, I. and P. K. Chintagunta. 2006. Measuring cross-category price effects with aggregate store data. *Management Science* 52: 1594-609.
- [41] Staiger, D., and J. H. Stock. 1997. Instrumental variables with weak instruments. *Econometrica* 65: 557-586.
- [42] Sudhir, K. 2001. Structural analysis of manufacturer pricing in the presence of a strategic retailer. *Marketing Science* 20: 244-264.
- [43] Supermarket News. 2014. Data table: Supermarket categories by dollar, unit sales. (<http://supermarketnews.com/%5Bprimary-term%5D/data-table-supermarket-categories-dollar-unit-sales>). Accessed on Nov. 16, 2015.
- [44] Train, K. 2003. *Discrete choice methods with simulation*. Cambridge, U.K.: Cambridge University Press.

- [45] United States Department of Labor. Bureau of Labor Statistics. Inflation and prices. (<http://www.bls.gov/data/>) Accessed on July 31, 2015.
- [46] Villas-Boas, S. B. 2007. Vertical relationships between manufacturers and retailers: Inference with limited data. *Review of Economic Studies* 74: 625-652.
- [47] Villas-Boas, M. and Y. Zhao. 2005. Retailer, manufacturers, and individual consumers: Modeling the supply side in the ketchup marketplace. *Journal of Marketing Research* 42: 83-95.

Table 1. Retail Prices with Complementarity

| Complementarity | Retail Price Level | |
|-----------------|--------------------|------------|
| | $t = 1$ | $t = 0.01$ |
| $\delta = -0.8$ | \$0.528 | \$0.523 |
| $\delta = -0.6$ | \$0.562 | \$0.524 |
| $\delta = -0.4$ | \$0.605 | \$0.517 |
| $\delta = -0.2$ | \$0.661 | \$0.512 |
| $\delta = 0$ | \$0.733 | \$0.509 |

Table 2. Data Summary: Sample Buyers and Input Prices

| | | Store 1 | | | Store 2 | | |
|--------------------|------------|---------|-----------|-------|---------|-----------|-------|
| | | Mean | Std. Dev. | N | Mean | Std. Dev. | N |
| Income | \$,000 | 57.610 | 21.116 | 85062 | 47.728 | 30.793 | 51560 |
| Family Size | # | 2.458 | 0.028 | 85062 | 2.319 | 1.100 | 51560 |
| Age | Years | 58.224 | 0.616 | 85062 | 58.579 | 9.986 | 51560 |
| Education | Years | 12.305 | 0.030 | 85062 | 11.845 | 2.566 | 51560 |
| Number of Children | # | 2.049 | 0.971 | 85062 | 2.207 | 0.854 | 51560 |
| Trips | # | 149.845 | 75.716 | 85062 | 158.138 | 90.429 | 51560 |
| Single Store | % | 19.948 | 39.981 | 85062 | 6.702 | 25.081 | 51560 |
| Multi-Store | % | 58.459 | 34.607 | 85062 | 41.541 | 34.607 | 51560 |
| Primary Ingredient | \$ / lb. | 0.090 | 0.046 | 1664 | 0.090 | 0.046 | 1664 |
| Retail Wage | \$,00/wk. | 4.087 | 0.073 | 1664 | 4.087 | 0.073 | 1664 |
| Food Mfg. Wage | \$,00/wk. | 6.727 | 0.189 | 1664 | 6.727 | 0.189 | 1664 |
| Energy | Index | 0.528 | 0.092 | 1664 | 0.528 | 0.092 | 1664 |
| Business Services | Index | 1.103 | 0.019 | 1664 | 1.103 | 0.019 | 1664 |
| Packaging | Index | 1.412 | 0.063 | 1664 | 1.412 | 0.063 | 1664 |

Note: Single store refers to the % of households choosing only one store, while Multi-Store is defined as the % of households choosing two stores.

Table 3. Data Summary: Shopping Basket Composition

| | Share % | Std. Dev. | Expenditure | Std. Dev. |
|-----------------------------------|---------|-----------|-------------|-----------|
| Milk Only | 0.281 | 0.449 | 3.478 | 1.979 |
| Cereal Only | 0.052 | 0.222 | 5.344 | 3.951 |
| Soft Drinks Only | 0.139 | 0.346 | 8.158 | 7.607 |
| Snacks Only | 0.116 | 0.320 | 4.664 | 3.591 |
| Milk, Cereal | 0.048 | 0.215 | 9.945 | 5.694 |
| Milk, Soft Drinks | 0.080 | 0.272 | 11.023 | 6.387 |
| Milk, Snacks | 0.078 | 0.267 | 8.871 | 4.580 |
| Cereal, Soft Drinks | 0.014 | 0.117 | 13.369 | 7.568 |
| Cereal, Snacks | 0.018 | 0.132 | 10.814 | 6.183 |
| Soft Drinks, Snacks | 0.055 | 0.229 | 14.334 | 8.771 |
| Milk, Cereal, Soft Drinks | 0.019 | 0.135 | 17.139 | 8.323 |
| Milk, Cereal, Snacks | 0.024 | 0.152 | 15.861 | 7.477 |
| Milk, Soft Drinks, Snacks | 0.049 | 0.216 | 18.407 | 9.161 |
| Cereal, Soft Drinks, Snacks | 0.010 | 0.100 | 20.411 | 10.229 |
| Milk, Cereal, Soft Drinks, Snacks | 0.018 | 0.134 | 24.699 | 10.504 |

Note: Data are pooled over both stores in the data set.

Table 4. Data Summary: Category Marketing Mix by Store

| | Units | Store 1 | | | Store 2 | | |
|----------------------|-----------------|---------|-----------|-------|---------|-----------|-------|
| | | Mean | Std. Dev. | N | Mean | Std. Dev. | N |
| Milk Price | \$ / gallon | 3.072 | 1.244 | 85062 | 3.653 | 0.932 | 51560 |
| Cereal Price | \$ / 16 oz box | 3.052 | 0.571 | 85062 | 3.075 | 0.603 | 51560 |
| Soft Drink Price | \$ / case | 4.775 | 1.784 | 85062 | 5.093 | 1.609 | 51560 |
| Snack Price | \$ / 16 oz unit | 3.760 | 0.955 | 85062 | 3.749 | 0.934 | 51560 |
| Milk Feature | % | 0.046 | 0.197 | 85062 | 0.054 | 0.179 | 51560 |
| Cereal Feature | % | 0.188 | 0.255 | 85062 | 0.270 | 0.283 | 51560 |
| Soft Drink Feature | % | 0.239 | 0.307 | 85062 | 0.346 | 0.330 | 51560 |
| Snack Feature | % | 0.193 | 0.274 | 85062 | 0.195 | 0.271 | 51560 |
| Milk Display | % | 0.071 | 0.218 | 85062 | 0.103 | 0.230 | 51560 |
| Cereal Display | % | 0.278 | 0.273 | 85062 | 0.374 | 0.267 | 51560 |
| Soft Drink Display | % | 0.556 | 0.312 | 85062 | 0.446 | 0.331 | 51560 |
| Snack Display | % | 0.596 | 0.303 | 85062 | 0.498 | 0.328 | 51560 |
| Milk Promotion | % | 0.393 | 0.426 | 85062 | 0.227 | 0.323 | 51560 |
| Cereal Promotion | % | 0.424 | 0.282 | 85062 | 0.588 | 0.254 | 51560 |
| Soft Drink Promotion | % | 0.705 | 0.288 | 85062 | 0.567 | 0.324 | 51560 |
| Snack Promotion | % | 0.682 | 0.282 | 85062 | 0.624 | 0.296 | 51560 |
| Milk Inventory | oz | 36.687 | 139.600 | 85062 | 44.740 | 138.157 | 51560 |
| Cereal Inventory | oz | 3.002 | 10.224 | 85062 | 4.134 | 13.742 | 51560 |
| Soft Drink Inventory | oz | 5.047 | 30.638 | 85062 | 15.480 | 111.814 | 51560 |
| Snack Inventory | oz | 3.833 | 12.238 | 85062 | 4.366 | 12.984 | 51560 |

Table 5. Nested MVL Demand Model

| | | Model 1 | | Model 2 | |
|------------------|------------------|----------|----------|----------|-----------|
| | | Estimate | t-ratio | Estimate | t-ratio |
| Category Choice | Milk - Store 1 | 1.5977* | 74.5840 | 4.4293* | 190.4100 |
| | Cereal - Store 1 | 2.5133* | 42.1897 | 3.7462* | 161.5982 |
| | Soda - Store 1 | 3.9899* | 164.6573 | 5.9020* | 226.9375 |
| | Snacks - Store 1 | 4.4874* | 177.3829 | 5.3707* | 230.1559 |
| | Milk - Store 2 | 4.3570* | 135.1304 | 3.0582* | 232.5555 |
| | Cereal - Store 2 | 3.0895* | 34.1030 | 3.7542* | 206.9590 |
| | Soda - Store 2 | 2.3702* | 98.7929 | 2.8450* | 230.6713 |
| | Snacks - Store 2 | 1.2185* | 50.0318 | 2.6628* | 211.3943 |
| | Price | -0.0613* | -26.3238 | -0.0755* | -46.9989 |
| | Feature | -0.9034* | -61.3994 | -0.2980* | -31.3349 |
| | Display | 1.3744* | 97.3812 | 0.6475* | 71.0794 |
| | Promotion | 0.7804* | 61.5470 | 0.2022* | 23.4328 |
| | Income | 0.0057* | 39.0977 | 0.0056* | 61.1400 |
| | HH Size | -0.0700* | -15.2375 | -0.5015* | -155.6133 |
| | Age | -0.0115* | -36.5833 | -0.0356* | -171.7872 |
| | Education | -0.0773* | -40.7680 | 0.0442* | 37.7964 |
| Control Function | 0.2540* | 164.8684 | 0.2363* | 280.5640 | |
| Store Choice | Store 1 | 0.4807* | 14.1047 | 0.9325* | 150.9170 |
| | Price Index | -0.6283* | -28.0840 | -0.1960* | -79.0190 |
| | Variety | 0.8139* | 33.1673 | 0.7317* | 178.5162 |
| | Loyalty | 0.1010* | 8.4441 | 0.7785* | 64.3945 |
| | Distance | -1.1464* | -83.2577 | -0.4281* | -27.8655 |
| | λ | 0.6494* | 76.6257 | 0.3719* | 48.6775 |
| Scale Parameters | Price | | | 0.0071* | 8.4020 |
| LLF | | -35.8568 | | -16.8747 | |
| AIC | | 141.7136 | | 129.7493 | |

Table 5 (cont). Nested MVL Demand Model

| | | Model 1 | | Model 2 | |
|-----------------------|--------------|----------|----------|----------|-----------|
| | | Estimate | t-ratio | Estimate | t-ratio |
| Interactions: Store 1 | Milk,Cereal | 4.3659* | 72.6436 | 4.3145* | 119.9480 |
| | Milk,Soda | -0.2327* | -8.4196 | -0.2043* | -8.9964 |
| | Milk,Snack | 0.2897* | 10.8096 | 0.3095* | 94.1415 |
| | Cereal,Soda | 3.4871* | 56.9116 | 1.7398* | 34.9232 |
| | Cereal,Snack | -0.7602* | -17.5996 | -0.7536* | -143.5799 |
| | Soda,Snack | 4.9574* | 124.0265 | 1.7385* | 7.7338 |
| Interactions: Store 2 | Milk,Cereal | 5.6155* | 59.6213 | 1.5273* | -5.5431 |
| | Milk,Soda | -5.3158* | -65.2138 | -5.7789* | 28.3780 |
| | Milk,Snack | 1.3506* | 37.0470 | 3.0963* | 139.7952 |
| | Cereal,Soda | 3.1402* | 36.0106 | 1.3066* | 73.1178 |
| | Cereal,Snack | -5.3422* | -41.0168 | -2.1234* | -56.5276 |
| | Soda,Snack | 0.3531* | 8.5265 | 0.4580* | 116.4153 |
| Scale Parameters | | | | | |
| Store 1 | Milk,Cereal | | | -2.3226* | -94.4239 |
| | Milk,Soda | | | -1.3701* | -103.6604 |
| | Milk,Snack | | | -3.7140* | -134.4287 |
| | Cereal,Soda | | | -9.4934* | -89.2125 |
| | Cereal,Snack | | | -8.2304* | -104.1456 |
| | Soda,Snack | | | 1.2838* | 81.9128 |
| Store 2 | Milk,Cereal | | | 5.3186* | 106.8234 |
| | Milk,Soda | | | 5.9169* | 123.1076 |
| | Milk,Snack | | | 0.2920* | 16.3306 |
| | Cereal,Soda | | | -4.2122* | -46.9111 |
| | Cereal,Snack | | | -7.2260* | -100.3938 |
| | Soda,Snack | | | -1.6324* | -54.3580 |

Note: A single asterisk indicates significance at a 5% level. Table is a continuation of Table 5 on previous page. See that table for goodness-of-fit measures.

Table 6. Independence Model of Multi-Category Demand

| | | Model 3 | | Model 4 | |
|-----------------|------------------|----------|-----------|----------|-----------|
| | | Estimate | t-ratio | Estimate | t-ratio |
| Category Choice | Milk - Store 1 | 3.8693* | 199.5189 | 6.0899* | 428.2706 |
| | Cereal - Store 1 | 2.3635* | 124.7093 | 2.3117* | 114.5853 |
| | Soda - Store 1 | 3.5389* | 189.4969 | 6.0264* | 474.1122 |
| | Snacks - Store 1 | 4.0843* | 188.7712 | 6.6813* | 472.0414 |
| | Milk - Store 2 | 5.8880* | 104.9600 | 6.7451* | 296.2387 |
| | Cereal - Store 2 | 2.4457* | 103.1939 | 4.9435* | 340.1391 |
| | Soda - Store 2 | 3.6140* | 153.8643 | 5.7562* | 397.1262 |
| | Snacks - Store 2 | 2.2266* | 120.1131 | 2.1650* | 125.9147 |
| | Price | -0.1162* | -41.2550 | -0.3158* | -195.8018 |
| | Feature | 1.0221* | 62.3515 | 2.8918* | 182.1945 |
| | Display | 0.3842* | 24.8565 | 0.7689* | 72.0316 |
| | Promotion | 0.4545* | 30.0865 | 0.2534* | 24.2178 |
| | Income | 0.0035* | 21.7397 | 0.0026* | 24.6470 |
| | HH Size | -0.3735* | -70.6377 | 0.3943* | 126.2639 |
| | Age | -0.0614* | -156.6495 | -0.0403* | -171.9137 |
| | Education | 0.1250* | 59.7847 | 0.1412* | 103.6831 |
| | Control Function | 0.2203* | 139.7129 | 0.1872* | 195.6020 |
| Store Choice | Store 1 | 1.0721* | 31.5167 | 3.3574* | 253.4612 |
| | Price Index | -0.6859* | -57.3127 | -1.9774* | -363.6076 |
| | Variety | 0.3435* | 15.2965 | 0.2581* | 247.3269 |
| | Loyalty | 2.0965* | 166.8166 | 0.1496* | 23.6430 |
| | Distance | -0.1941* | -61.7659 | -0.1619* | -108.3058 |
| | λ | 0.2449* | 46.6492 | 0.1336* | 156.2895 |
| Scale Parameter | Price | | | 0.0142* | 17.0791 |
| LLF | | -33.3378 | | -31.5638 | |
| AIC | | 110.6756 | | 109.1277 | |

Note: A single asterisk indicates significance at a 5% level. Model 3 is fixed coefficient, Model 4 is random coefficient version.

Table 7. Equilibrium Pricing: MVL Demand Model

| | Model 1: OLS | | Model 2: CF | |
|-------------------|--------------|---------|-------------|---------|
| | Estimate | t-ratio | Estimate | t-ratio |
| Milk | -5.0794* | -2.2877 | -5.2707* | -2.3742 |
| Cereal | -5.5096* | -2.4918 | -5.6967* | -2.5768 |
| Soda | -3.9984 | -1.8116 | -4.1848 | -1.8964 |
| Snacks | -5.7326* | -2.5875 | -5.9203* | -2.6727 |
| Primary Input | 6.3393* | 5.7130 | 6.3549* | 5.7327 |
| Retail Wage | 0.3284 | 1.1084 | 0.3224 | 1.0895 |
| Food Mfg Wage | 0.4441 | 1.5282 | 0.4232 | 1.4568 |
| Energy | -0.2596 | -0.7185 | -0.3049 | -0.8432 |
| Business Services | 11.4974* | 3.8175 | 11.5432* | 3.8365 |
| Packaging | 1.1571* | 2.5963 | 1.1211* | 2.5162 |
| ϕ | 0.3873* | 13.7434 | 0.3871* | 13.7464 |
| Control Function | | | -0.0012* | -2.1017 |
| LLF | -1,484.0792 | | -1,481.9011 | |
| AIC | 1.7970 | | 1.7956 | |

Note: A single asterisk indicates significance at at 5% level.

"Margin" coefficient is the conduct parameter, CF is control function.

Table 8. Equilibrium Pricing: Logit Demand Model

| | Model 3: OLS | | Model 4: CF | |
|-------------------|--------------|---------|-------------|---------|
| | Estimate | t-ratio | Estimate | t-ratio |
| Milk | -5.8059* | -2.6195 | -5.9946* | -2.7051 |
| Cereal | -6.2336* | -2.8247 | -6.4183* | -2.9090 |
| Soda | -4.7301* | -2.1464 | -4.9137* | -2.2302 |
| Snacks | -6.4638* | -2.9219 | -6.6486* | -3.0060 |
| Primary Input | 6.3304* | 5.7116 | 6.3460* | 5.7312 |
| Retail Wage | 0.2953 | 0.9982 | 0.2896 | 0.9796 |
| Food Mfg Wage | 0.5335 | 1.8355 | 0.5124 | 1.7636 |
| Energy | -0.2925 | -0.8105 | -0.3375 | -0.9345 |
| Business Services | 12.4966* | 4.1509 | 12.5393* | 4.1691 |
| Packaging | 1.2090* | 2.7152 | 1.1730* | 2.6349 |
| ϕ | 0.5019* | 13.8945 | 0.5012* | 13.8922 |
| Control Function | | | -0.0012* | -2.0847 |
| LLF | -1482.1483 | | -1479.9848 | |
| AIC | 1.7946 | | 1.7933 | |

Note: A single asterisk indicates significance at a 5% level.

Table 9. Equilibrium Prices with Simulated Demand

| | Subs: $\theta_{ij} = -2$ | | Base Case | Comps: $\theta_{ij} = 2$ | |
|------------------|--------------------------|----------|-----------|--------------------------|---------|
| | Estimate | t-ratio | Estimate | Estimate | t-ratio |
| $\lambda = 0.37$ | 4.9461* | 9.0066 | 4.9137 | 4.8924* | -5.9230 |
| | 0.5780 | | 0.5772 | 0.5776 | |
| $\lambda = 0.90$ | 4.4876* | -30.4567 | 4.5979 | 4.6788* | 22.7280 |
| | 0.5872 | | 0.5757 | 0.5673 | |
| $\lambda = 0.10$ | 5.0258* | 9.4623 | 4.9892 | 4.9656* | -6.1033 |
| | 0.6214 | | 0.6207 | 0.6210 | |

Note: A single asterisk indicates significance at 5%.

T-ratios defined relative to Base Case.