



Marketing Science

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

An Empirical Analysis of Assortment Similarities Across U.S. Supermarkets

Minha Hwang, Bart J. Bronnenberg, Raphael Thomadsen,

To cite this article:

Minha Hwang, Bart J. Bronnenberg, Raphael Thomadsen, (2010) An Empirical Analysis of Assortment Similarities Across U.S. Supermarkets. Marketing Science 29(5):858-879. <https://doi.org/10.1287/mksc.1100.0564>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2010, INFORMS

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

An Empirical Analysis of Assortment Similarities Across U.S. Supermarkets

Minha Hwang

UCLA Anderson School of Management, University of California, Los Angeles, Los Angeles, California 90095,
mihwang@anderson.ucla.edu

Bart J. Bronnenberg

CentER, Tilburg School of Economics and Management, Tilburg University, 5000 LE Tilburg,
The Netherlands, bart.bronnenberg@utv.nl

Raphael Thomadsen

UCLA Anderson School of Management, University of California, Los Angeles, Los Angeles, California 90095,
raphael.thomadsen@anderson.ucla.edu

This paper examines pairwise assortment similarities at U.S. supermarkets to understand how assortment composition and size are related to underlying factors that describe local store clientele, local competitive structure, and the retail outlets' characteristics. The top-selling items, which cumulatively make up 50% of sales, are sold at nearly every store, but other items are viewed as optional. We find that, within states, supermarkets owned by the same chain carry similar assortments and that the composition of their clientele and the presence of competing stores have effects on assortment similarity that are an order of magnitude smaller than ownership structure. In contrast, we find that, across states, supermarkets owned by the same chain do version their assortment. We explain this difference using extant work on the minimal efficient scale of supermarkets and on local demand effects. Furthermore, we investigate the distribution and role of regional brands. We find that regional brands are primarily distributed by small regional chains or independent stores. "Value" regional brands are primarily distributed by supermarket firms without store brands, whereas the distribution of "premium" regional brands is unrelated to the presence of store brands. We discuss our findings in the context of modeling assortment decisions and manufacturers designing distribution policies.

Key words: retailing; assortment; supermarket; dyadic data

History: Received: September 19, 2008; accepted: January 8, 2010; processed by Robert Meyer. Published online in *Articles in Advance* April 8, 2010.

1. Introduction

Providing product assortment is one of the most important tasks for supermarkets (Levy and Weitz 2008). The assortment a retailer carries has a large impact on consumers' store patronage decisions, sales, and gross margin, and hence, assortment planning has been considered as one of the high-priority tasks for retailers, consultants, and software providers (Briesch et al. 2009, Kök et al. 2009). However, relatively little is known empirically about how product assortment composition varies across stores within a supermarket chain or across supermarket chains. Most existing empirical studies have focused on assortment size rather than assortment composition (Bayus and Putsis 1999, Putsis and Bayus 2001, Kadiyali et al. 1999, Draganska and Jain 2005, Watson 2009, Richards and Hamilton 2006,

Misra 2008).^{1,2} Indeed, Kök and Fisher (2007, p. 1002) observe that they "are aware of no papers that provide empirical information about how assortment planning works in practice." Because of the central role assortment planning plays in retailing, this paper seeks to fill this gap in the literature by documenting how supermarket assortment similarities between stores are related to similarities in clientele, ownership structure, store competition, and geographic markets. The absence of basic empirical findings is remarkable in the context of ongoing disagreements among

¹ In Misra's (2008) study on product assortment composition, products are sorted based on demand and cost parameters in such a way that the assortment decision effectively becomes one of assortment size.

² Another relevant study is Draganska et al. (2009), who investigate product line offerings by ice cream manufacturers. However, we examine the assortment choice of retailers.

experts about how supermarkets approach assortment planning. On one hand, Kök et al. (2009, p. 139) state that the most common approach for assortment planning among practitioners is “to decide on a single common assortment that is carried by all stores of the chain, except [...] in smaller stores.” On the other hand, there is a literature asserting that retailers do, in fact, localize assortment to better meet the heterogeneous needs of consumers across store locations (Mantrala et al. 2009, Grewal et al. 1999, Misra 2008). This observation is in line with suggestions from literature on micromarketing (Fox and Sethuraman 2006, Iyer and Seetharaman 2008, Levy and Weitz 2008). Kroger’s fact book also suggests that assortment is localized.³ However, there has been no empirical study to our knowledge that has attempted to rigorously document how similar assortments are across stores within a chain and across chains, or that has tried to explain the drivers of local variation in assortment when it arises. Our paper conducts such an analysis.

This study is also related to research on chain-level decisions about other marketing instruments such as pricing. For example, past studies based on Dominick’s *Finer Foods* data on pricing demonstrate that some marketing mix decisions are made at a chain level, whereas others are made at a store level. Besanko et al. (2005) and McAlister (2007) demonstrate that different regular prices (i.e., unprompted prices) are set for different pricing zones, which are determined by local competition and price sensitivities of local clientele. In contrast, Hoch et al. (1995) show that the same promotional retail price is set for all stores within the chain. Similarly, we seek to empirically establish the extent to which assortment compositions and sizes are determined at a corporate level and how much these decisions are determined according to the demographic and competitive environment of the specific stores.

A systematic empirical study on assortment planning is ideally based on observations of assortments for a large cross section of stores across multiple chains, markets, and clienteles. Until very recently, such data did not exist.⁴ Indeed, historically, academic researchers have had access only to account-level data or store-level data from a single chain, which precluded them from empirically investigating assortment-planning practice. In this study, we use

a unique large store-level database of retail assortment for U.S. supermarkets across multiple chains and independent stores that allows us to investigate the factors that explain retail assortments.

Our analysis of U.S. retail assortment across a broad cross section of supermarket stores consists of four parts. We first describe some basic patterns in the data. We show that, across several product categories, the top-selling items, which cumulatively make up 50% of sales, are viewed as core products that are sold at nearly every store; however, once a product is not a top seller, stores start to view stocking the product as optional.

Second, we analyze the dependence of assortment composition on underlying factors describing each of the “3 Cs”: customer (local demographic characteristics), company (ownership or store size characteristics), and competition (proximity to competing stores). Rather than analyzing assortment composition directly, we propose an approach in which we focus on modeling assortment similarity across pairs of stores. Using a summary measure of assortment similarity allows us to avoid challenges associated with direct measurement and characterization of assortment. For example, estimating profits for all possible assortment compositions would not be a feasible task, because the number of possible assortment compositions grows exponentially with the number of products. With typically 100 to 500 unique products available in any given consumer packaged goods (CPG) category, there are 2^{100} to 2^{500} possible assortment compositions to evaluate, which is an impractically large number.⁵ In contrast, the number of assortment similarity comparisons does not grow exponentially, but quadratically. Moreover, by using dyadic analysis, we utilize a measure that allows us to investigate factors that describe a relation between a pair of stores (e.g., common ownership, distance between competing stores). For example, if local demand matching is a key driver for assortment composition, two stores that face similar local demographics would offer similar assortments. Our proposed method of analysis accounts for the influence of these dyads on inferences and for the dependencies implied by the fact that a single store is part of multiple store pairs. We use the output of our dyadic analysis to conduct a variance decomposition of assortment similarity. Based on analysis of four large product categories—cereals, coffee, colas, and toothpaste—we find that corporate structure and

³ Kroger’s 2007 fact book (p. 15) notes that “Kroger’s decentralized structure places substantial authority for merchandising and operating decisions in our supermarket divisions. Divisional managers are able to respond quickly to changes in competition and customer preferences within each local market.” However, the relevant definition of a local market is not clear from this statement.

⁴ For instance, it was the policy of both IRI and Nielsen not to release store-level information (see, e.g., Boatwright et al. 2004).

⁵ As an example of the scale of this number, there are approximately 2^{263} atoms in the universe. Put another way, imagine that one had a supercomputer that could calculate 400 trillion profit numbers per second. It would take this computer over 100 million years to calculate the profits of 2^{100} items.

store size explains between 25%–48% of the within-state differences in assortment composition, whereas clientele-related variables explain only 4%–7% of the variation and competition-related factors explain only 1%–4% of the variation. Thus, ownership structures and store size of retail stores are an order of magnitude more powerful in explaining variation in assortment composition than either clientele effects or store competition. We also find that these results extend to assortment size. Store ownership structure and store size accounts for approximately 16%–28% of explained variance of assortment-size similarity, but clientele and competition effects explain only 0%–4% of the variation in assortment size.

Third, we analyze whether the pooling of assortments across stores of the same chain continues to hold across geographic markets by comparing mean assortment similarities. We find that assortments of stores under common ownership differ substantially across states and that multiregional supermarket firms localize assortment by regional divisions. Assortment similarity of stores from the same chain but in different markets can be as low as the assortment similarity of stores from different chains within the same market.

Finally, we analyze the determinants of regional brand distribution by conducting binary logistic regressions where store-level distribution of any regional brand is taken as a dependent variable. In the cola and toothpaste categories, we find that regional brands are generally distributed by small-scale supermarkets such as small regional chains or independent stores. We also find that the presence of private-label programs mostly negatively impacts the distribution of low-price “value” regional brands but not high-price “premium” regional brands. We interpret this to mean that chains without a private-label program view such value regional brands as an alternative to these private labels.

The remainder of this paper is organized as follows. Section 2 describes the data. Section 3 presents potential determinants of assortment composition and provides descriptive evidence of our main conclusion. Section 4 describes the assortment similarity measure and our analytical framework. Section 5 presents the main results. Section 6 presents the results on regional brand distribution. Section 7 concludes with academic and managerial implications and avenues for further study.

2. Data

Our analysis utilizes three main sources of data: weekly store-level scanner data drawn from four categories of consumer-packaged goods from ACNielsen, store characteristics data from Trade Dimensions, and store trading area data from Spectra.

2.1. Store-Level Scanner Data

ACNielsen provided us with weekly store-level scanner data for the cola, ready-to-eat cereal, ground coffee, and toothpaste categories. Supermarkets are the main channel of sales for these products (Inman et al. 2004). The cola data are provided at the “brand–flavor–package size” level. The data for the other categories were provided at a natural “item aggregate,” although we do not have package size data for those categories. For example, a typical item in each category might be “Coca-Cola/Classic/12-pack cans,” “General Mills/Cinnamon Toast Crunch/Reduced Sugar,” “Folgers/Coffeehouse Series/Decaf,” or “Colgate/Sparkling White/Multi-Benefit,” respectively. In the ready-to-eat cereal category, multiple-product packages such as variety packs are excluded because of difficulties in assigning unique brand–flavors to the items.⁶ In all of the categories, the data aggregate private-label products into a single brand. However, the flavor (and package size for cola) dimension distinctions are still maintained for private-label products.

We used 52 weeks from the year 2005 as a basis for constructing assortment composition and size similarities. In the original data, there are 3,040 sample supermarket stores. We focus on 21 states with more than 30 sample stores in each state to ensure that we can reliably infer a set of products available in the state. In addition, we have excluded stores with missing data, outliers, interruption in operation, and problems in matching store characteristics data with store trading area data. Therefore, the final number of sample stores is 2,017 in the cola category, 2,011 in the ready-to-eat cereal category, 2,001 in the ground coffee category, and 1,759 in the toothpaste category.

There are 244 unique products in the cola category, 722 unique products in the ready-to-eat cereal data, 809 unique products in the ground coffee category, and 212 unique products in the toothpaste category.⁷ To check the robustness of the results against different aggregation levels, we also analyze the cola data using the brand–flavor-level aggregation by aggregating across different package sizes. This aggregation yields 66 unique brand–flavor combinations in the cola data.

2.2. Store Characteristics Data

The store characteristics are obtained from the 2006 Trade Dimensions data. These data include average weekly store sales, physical store size measured

⁶ The total unit market share of multiple-product packages is 0.69% over the sample stores.

⁷ We measure a store’s assortment through the presence of products. Ideally, one may also use other measures of assortment, such as the number of facings or the positioning of products on a shelf (e.g., Drèze et al. 1994). We do not use these measures because they are not present in our data.

by square footage, ultimate corporate parent owner of the store, marketing group (local chain banner) of the store, local operating division, store name, store zip code, and longitude and latitude of the store, which provides the basis to compute distance between stores. In addition, the same data set is utilized to calculate retailer market shares within a state based on average weekly store sales and a number of competitors of each type (e.g., supermarket, supercenter) around the stores.

2.3. Store Trading Area Data

We characterize the clientele of each store from data by Spectra on the characteristics of each store's trading area. As described by Hoch et al. (1995), Spectra defines a trading area by expanding a polygon around each store to enclose an area large enough to support the all-commodity volume (ACV) of the store. These store trading areas are defined through a proprietary model that takes into account population density, competition, road conditions, and various regional differences. We use the marginal distributions of local demographics Spectra provides for each of these trading areas, including income, household size, head of household age, ethnic group, presence of children in each age group, home ownership, and education level. Because the total number of households in the trading area is inferred from (physical) store size and store sales, we do not use this variable in our analysis.

3. Potential Determinants of Assortment Similarity

In this section, we first draw from the previous literature and describe factors organized according to the 3 Cs—company, customer, and competition—that can affect a supermarket's assortment choice. We then provide some descriptive evidence that supermarket ownership is the most important factor in determining an outlet's assortment.

3.1. Company Factors

3.1.1. Ownership. Common ownership can affect assortment choices of multioutlet retailers in a number of ways. If there are economies of scale or scope, which can arise from quantity discounts or operational efficiencies, we can expect stores belonging to the same chain to carry relatively homogeneous assortments. These economies of scale and scope could be reinforced by the presence of chain-level slotting allowances or by contracting with category captains to make assortment decisions (Dhar et al. 2009). Similarly, Ellickson (2007) offers an account of the cost

savings that can be gained through an efficient distribution system. Menu costs associated with store-level customization of assortment would have similar effects.

Empirical studies in other industries examine the link between ownership and product homogeneity. Chisholm et al. (2006) study first-run movie theaters in Boston and find that theater pairs under common ownership tend to make more similar programming choices. Sweeting (2006, 2010) investigates how common ownership affects the programming and listenership of contemporary music radio stations, and he finds that common ownership in the same market leads to more differentiation and an increase in listenership. However, he also finds that common ownership of stations in different markets is associated with playlist homogenization, consistent with economies of scale and scope in offering similar programming in different markets. One commonality between these two studies is that the presence of spatial differentiation allows multioutlet firms to benefit from economy of scope by having similar assortments. Given the strong role of spatial differentiation in supermarkets, we posit that a pair of supermarkets with common ownership would make more similar assortment choices compared to the assortment choices for a pair of stores belonging to different owners.

It is not clear which level of ownership should matter for assortment choices. For example, Vons, Pavilions, and Safeway are separate chains operating under different "banners," but all of them have the "Safeway Group" as their ultimate corporate owner. Homogenization of assortments could be expected either at the banner level or at the corporate ownership level. If economies of scale and scope are important, homogeneity could appear at the corporate ownership level. On the other hand, the franchising literature has pointed out that the value of a chain's brand name comes from its consistency across outlets (Kaufmann and Eroglu 1999). If consistency in shopping experiences across stores within a same chain matters, stores belonging to the same chain banner could show stronger homogenization of assortments. Moreover, many supermarkets have local divisions, which typically designate separate buying offices.

There are also other types of chain stores such as affiliated independent chains, which are owned by multiple independent owners. One benefit of belonging to this type of chain store could be economies of scale and scope in procurement. On the other hand, we would not expect independent stores to have selections as coordinated as those in chain stores.

3.1.2. Store Size. Physical store size directly affects the shelf space available for each product category and its associated opportunity costs. Although

stores could use the additional shelf space to add more product categories or to stock a greater quantity of each item (e.g., they could have several columns of 12-packs of Diet Coke), it seems reasonable to expect that larger stores have larger assortment size. In Figure 1, we create a scatter plot of store size and assortment size. The scatter plot reveals that larger stores do indeed carry a greater number of products on average but that this relationship is far from linear and that the effects of either stocking a greater quantity of each item or using the extra space to expand to other categories is also a significant effect.

3.2. Customer Factors

Supermarket stores in different geographical regions face different clientele in their trading area. If different demographics have different tastes, it is possible that it would be profitable for supermarkets to tailor their assortment offerings to these demographics.

Past empirical studies provide evidence of a relationship between preferences and demographics for the product categories we consider: cola and ready-to-eat cereal. Dubé (2004) finds that demographics such as income, family size, and the presence of children partially explain observed differences in tastes for product attributes in the carbonated soft drink category. He shows that larger households purchase larger quantities of soda and that households with higher income have a higher taste for quality. In the ready-to-eat cereal category, Nevo (2001) finds that a similar set of demographic variables such as income, age, and the presence of children in the household explain a significant proportion of brand-preference heterogeneity. In another study, Hoch et al. (1995) find that age, family size, income, wealth, and ethnic diversity are correlated with an individual's price sensitivity. Finally, Dhar and Hoch (1997) find that demographic variables such as wealth, age, education, and ethnic diversity explain some of the variation in store brand penetration across supermarket chains. Therefore, we test the role of the following set of local demographic variables on the assortment decisions of supermarkets: income, household size, ethnic diversity, age of head of household, presence of children in each age group, education, and home ownership.

3.3. Competition Factors

3.3.1. Distance. Theoretical research has investigated how competition affects product offerings in differentiated industries. One of the longstanding debates in this literature is whether firms offering horizontally differentiated products choose product designs with minimum differentiation (Hotelling 1929) or maximum differentiation (d'Aspremont et al. 1979). Neven and Thisse (1990) and Irmen and Thisse (1998) extend

the theoretical analysis to competition in a multi-characteristics space. Both papers suggest that firms should differentiate themselves along one product dimension—horizontal or vertical—and agglomerate along the other dimensions. A study by Fischer and Harrington (1996) provides a similar prediction—a greater degree of product differentiation and a larger amount of consumer search are related to a greater degree of agglomeration of stores. This research collectively suggests that supermarkets should choose to differentiate their assortments from those of nearby supermarkets.

A few empirical studies investigate product design issues of competing firms in specific industries. Chisholm et al. (2006) find that theater pairs located close together show different movies. This result might be partially driven by contractual obligations specific to the movie industry, however. In the supermarket industry, Stassen et al. (1999) find that assortment overlap and interstore distance are determinants of shared patronage in a market of 27 stores. They also find that supermarkets located closer together tend to duplicate beef assortment but differentiate consumer packaged goods assortment.

3.3.2. Number of Competitors. Competition can affect both the size and the specific products chosen by the supermarket. Note that competition can lead to either an increase or a decrease in the number of products a supermarket carries, depending on how decrease in the number of consumers balances with the need to be more attractive to consumers in order to get them in the door. Johnson and Myatt (2003) demonstrate such a trade-off in a theory paper. In the price discrimination literature, Seim and Viard (2004) demonstrate that competition can increase the number of calling plans offered by phone companies, whereas Borzekowski et al. (2009) find that competition leads mailing-list companies to offer more versions of their products.

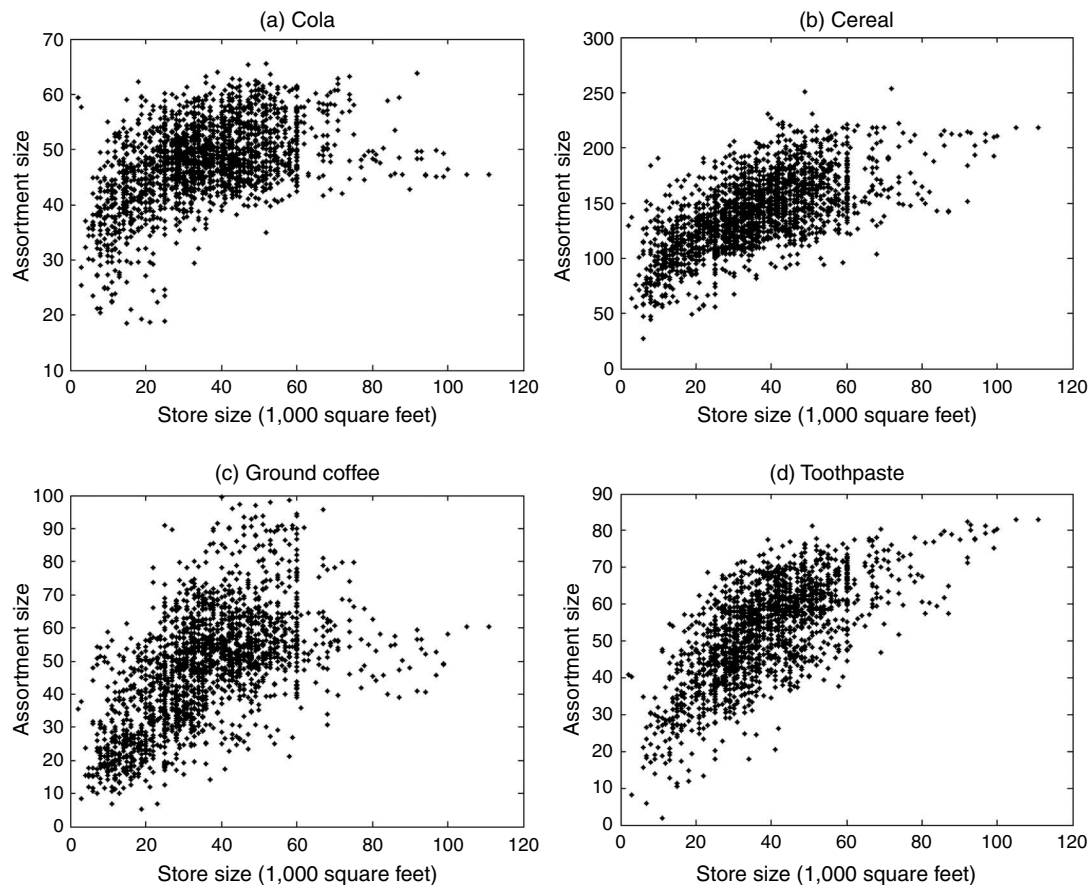
3.4. Descriptive Evidence

Although each of the factors mentioned previously affects the assortments that supermarkets offer, some factors are more influential in assortment choice than others. In §5, we present regression results that demonstrate that company factors affect product assortment choice more than local demographic or competitive variables. In this subsection, we present a model-free analysis that is also suggestive of this outcome.

Figure 2 presents a graphic of the cola products stocked by each of the California stores in our data set.⁸ The solid lines separate different owners,

⁸ Figure 2 excludes small chain stores with only one to three sample stores.

Figure 1 Store Size vs. Product Assortment Size



whereas the dashed lines separate different chain banners. Whereas there are sporadic differences between stores within the same banner, visual inspection reveals that the variance across owners is greater than the variance within an owner (or especially within the same banner). For instance, the graph reveals that Kroger has made Shasta a part of its cola assortment in both of its Californian retail chains (Food 4 Less and Ralphs). In contrast, Kroger's main competitors have not, even though these competitors frequently operate in the same shopping areas and serve the same local clientele. Similar observations can be made for selected varieties of RC Cola. Taken together, ownership thus appears to be a large driver of assortment choice for a retail organization. Figure 2 also reveals that there are certain products—especially selected Coke and Pepsi products—that almost all stores carry, whereas there are others that are carried by a much smaller number of stores.

Figure 2 also shows that regional brands have a hard time getting onto store shelves. In general, regional cola companies are only able to get shelf space in independent stores. The exceptions are Shasta, which is carried by many Kroger and Stater Bros. stores, and St. Nick's Cola, which is carried by

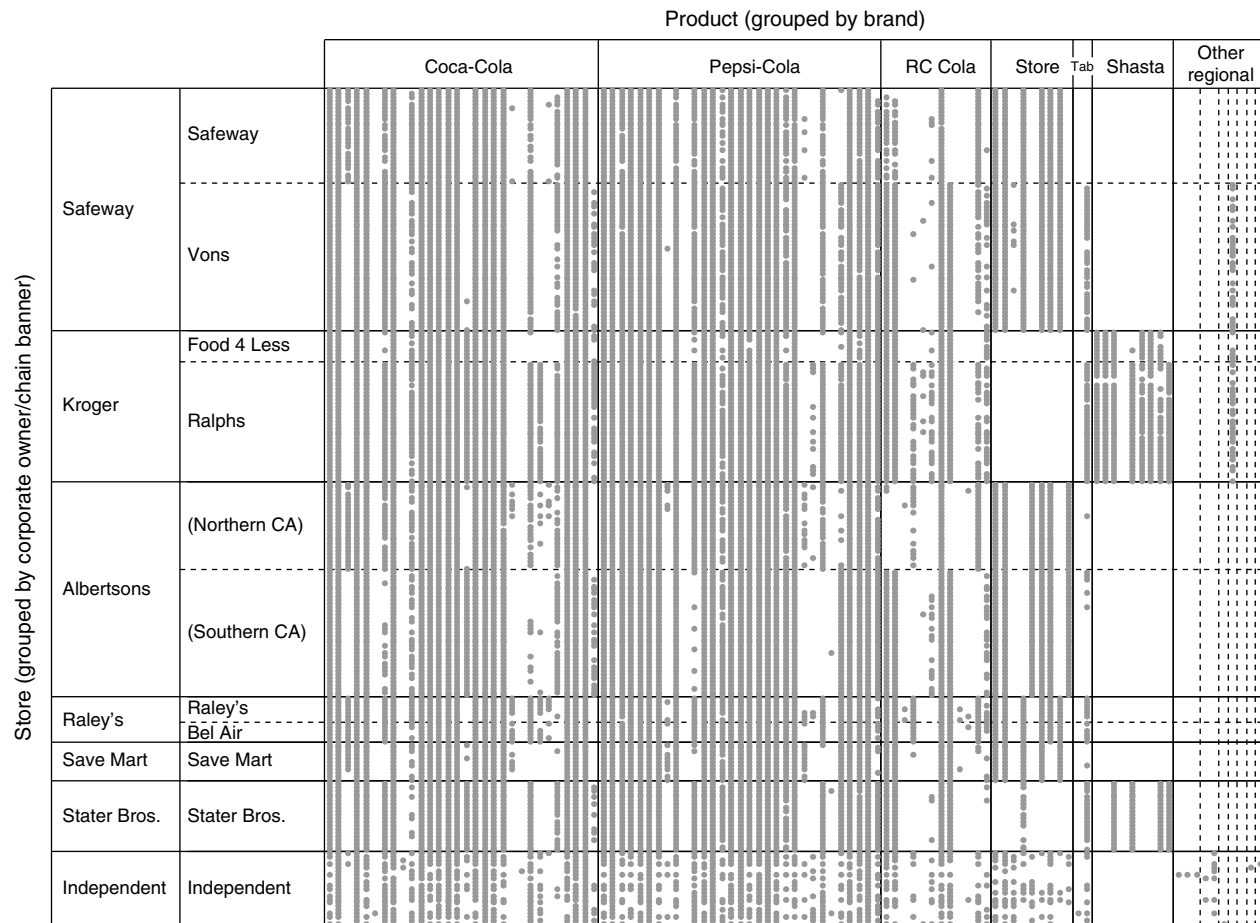
Safeway and Kroger as a fundraiser for St. Jude's Hospital.⁹ In addition, Figure 2 also suggests that large chains distribute regional brands when their store brands are not present or weak. For instance, two large chains in California that distribute Shasta—Kroger and Stater Bros.—either do not have store brands or have very limited store brands in cola. This suggests that regional brands may often play a role of an alternative to store brands. We formally test this conjecture by conducting binary logistic regressions in §6.

Figure 3 presents a plot of cumulative shares of the products—sorted by their national market share¹⁰—against the percentage of stores that stock them. All of the panels show that the higher-selling items are distributed more. However, there is no strict ordering in these graphs. The lack of monotonicity in this graph allows us to reject the hypothesis that retail assortment is obtained by ranking products based on

⁹ In some ways, even Shasta and St. Nick's Cola should be treated as national brands, because they are both owned by the National Beverage Corporation, which is the fifth-largest soft drink company, after Coca-Cola, Pepsi, Cadbury Schweppes, and Cott (which makes RC Cola).

¹⁰ Defined over 21 sample states.

Figure 2 Store-Level Assortments for Cola in California



a combination of national demand and common cost factors.

Figure 3 also reveals that across the four categories we consider, stores appear to consider products that are in the top 50% of cumulative sales as being core products, which almost all stores carry. However, there is a fair amount of heterogeneity in the decision to stock products in the bottom half of the cumulative sales, and the stores appear to treat stocking these products as optional.

4. Measurement Details

4.1. Assessing Assortment Composition and Size

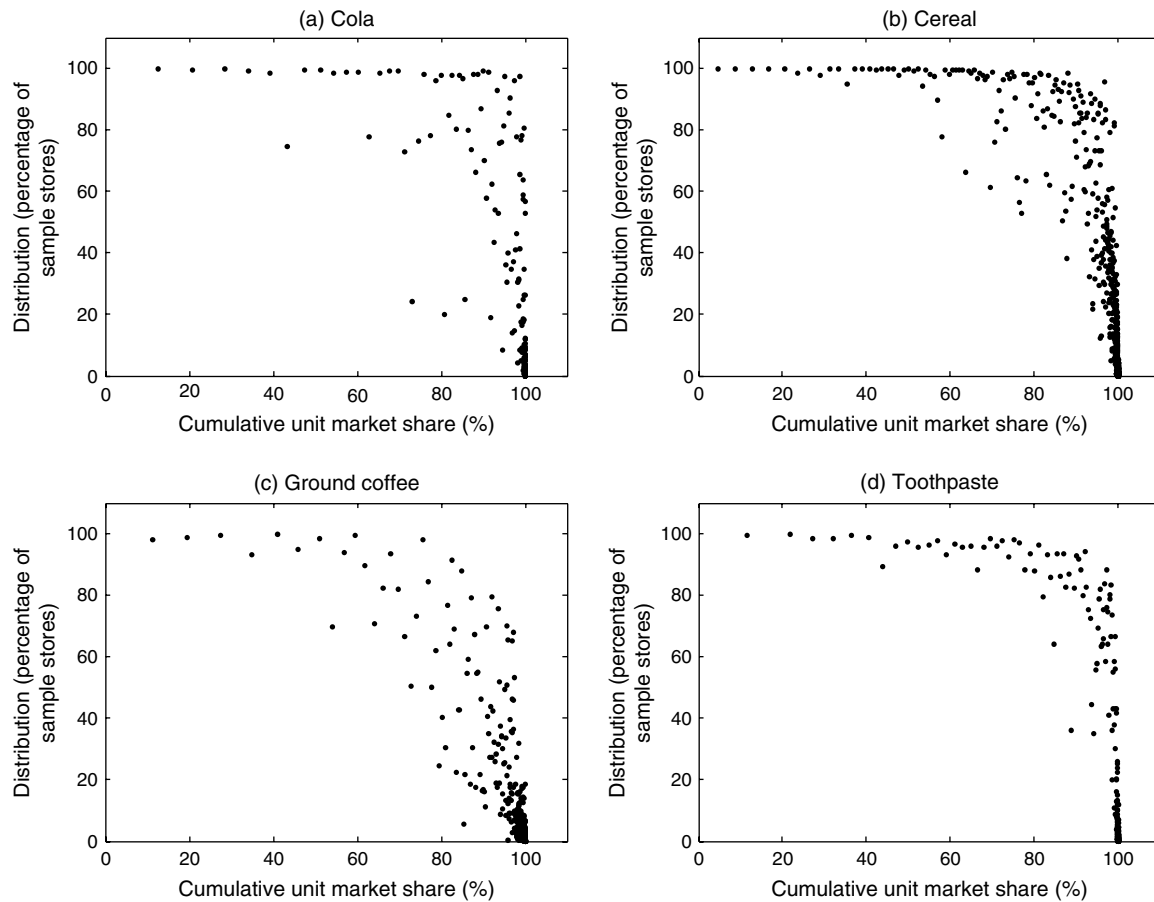
We analyze the similarity of assortment decisions between pairs of stores by comparing the similarity in the stores' assortments with the similarity in the other variables described in §3. We first present how we construct assortment vectors, and we then present a metric of assortment similarity. By relating assortment similarities to distances of underlying factors between a pair of stores, we can study the relative impact of various effects.

4.1.1. Assortment Vector Construction. To construct an assortment vector, we start by counting the total number of available products (N) in a given geographic market (state) for a given observation window.¹¹ For notational convenience, we suppress the subscripts for the market in the explanation of the assortment vector construction and assortment similarity measures below. The assortment vector A_i has length N , where the n th element, a_{in} , is each product's proportion of time in our data for which the product is on the shelf of a supermarket i .

We illustrate the assortment vector through an example. Suppose that there are only four available products in a market. During an observation window of a year, product 1 was stocked by a given supermarket for the entire period of 52 weeks, product 2 was stocked for 26 weeks, product 3 was also stocked for 26 weeks, and product 4 was never

¹¹ We assume that availability of products from manufacturers is specific to a geographic market (state), not to supermarkets. The total number of available products in a given geographic market is then the set of products that are offered by at least one supermarket in the state.

Figure 3 Distribution Density by Cumulative Unit Market Share Over All the Sample Stores



stocked by supermarket 1. The assortment vector A_i can be calculated as

$$\begin{aligned} \text{Product 1: } & 52/52 = 1.0, \\ \text{Product 2: } & 26/52 = 0.5, \\ \text{Product 3: } & 26/52 = 0.5, \\ \text{Product 4: } & 0/52 = 0.0. \end{aligned}$$

This definition makes clear that we rely on sales data to infer product-stocking decisions. However, unfortunately, store-level scanner data do not directly provide information on product availability. One concern is that relying on sales to infer assortment could potentially lead to underdetection problems of slow-moving products. Below, we propose three alternatives to alleviate this problem. The results are robust to all three approaches.

First, in the main analysis of the paper, we smooth the weekly sales status with a four-week window.¹²

¹² As an example, if a product was observed in the data as being purchased in week 1 and in week 5, but not in weeks 2, 3, or 4, the smoothing procedure would reflect that assumption that the product was being offered by the stores in weeks 2–4 but that customers did not purchase the item during those few weeks.

This procedure helps the detection of slow-moving products because products are counted as part of the assortment if there is only one sale over the duration of approximately one month. We eliminate products with less than three total weeks of sales in the year because those products are likely to be trial products, leftover inventories, or data input errors.

Second, to gauge the robustness of our findings, we replicate the analysis using the assumption that the store always stocks any item that sells at least once in the year. In this alternative definition, products are considered as a part of store assortment unless there are no sales of products over a year. Logically, this definition is free from undercounting.

Third, we also replicate the analysis using a vector that has a value of one for a given product if a store has sales in at least 13 out of 52 weeks, and zero otherwise. In this measure, only relatively fast-moving products are considered as a part of store assortment. Thus, this definition tends to warrant against overcounting.

4.1.2. Assortment Similarity Measures. There are several candidate measures to parsimoniously capture similarities and differences in assortment composition

decisions between pairs of supermarkets. Because we are interested in assortment differentiation between a pair of supermarkets, an ideal similarity measure is the one that captures the fraction of assortment overlap and is robust to the size of assortment. We use a cosine similarity measure motivated by Jaffe (1986) in the context of firms' research and development portfolios. Applied to our retailing context, the similarity measure between supermarket i 's and j 's assortments can be calculated from a cosine of an angle between the assortment vectors A_i and A_j ,

$$S_{ij} = \frac{A_i' A_j}{\|A_i\| \cdot \|A_j\|}. \quad (1)$$

The cosine similarity measure in Equation (1) increases when two supermarket stores make more similar or less differentiated assortment choices relative to each other. If two supermarket stores stock identical products, the cosine similarity measure will be one. The cosine similarity measure will be zero if there are no overlapping products between the pair of supermarkets.

The cosine similarity measure has the nice property of being unaffected by the variation in the number of available products for different geographic markets, although it is appropriately sensitive to differences in assortment composition from differences in assortment sizes between a pair of supermarkets.

A potential drawback of the measure in Equation (1) is that if the assortments for each store were chosen randomly, pairs of larger stores generally would have more similar assortments than pairs of smaller stores.¹³ To ensure the robustness of our results against different stocking rates, we will also analyze assortments using a second similarity measure that corrects for these stocking rates. Specifically, in addition to the "noncentered" correlation coefficient proposed by Jaffe (1986), we propose a centered correlation measure defined as

$$C_{ij} = \frac{(A_i - \pi_i)'(A_j - \pi_j)}{\|A_i - \pi_i\| \cdot \|A_j - \pi_j\|}. \quad (2)$$

¹³ This can be seen as follows. Suppose that the assortment vectors A_i and A_j for supermarkets i and j are completely random assignments of 0 and 1 with a mixing probability of π_i and π_j . Under this stocking rule, it can be shown that the expected value of $A_i' A_j$ is equal to $N\pi_i\pi_j$, whereas the expected value of $A_i' A_i$ is equal to $N\pi_i$. The expected value of the assortment similarity measure S_{ij} therefore equals

$$E[S_{ij}] = \frac{N\pi_i\pi_j}{N\sqrt{\pi_i\pi_j}} = \sqrt{\pi_i\pi_j}.$$

As Jaffe (1986) notes, the measure in Equation (1) is not dependent in expectation on N . However, it does depend on the fraction of products in the assortment π_i and π_j . In the special case of equally sized assortments, $\pi_i = \pi_j$, the expectation of the assortment similarity is equal to π_i , or the fraction of products on the shelf. The size of the assortment, and therefore the fraction of items stocked, tends to increase with store size.

This measure is always between -1 and 1 , with 1 meaning that two assortments are identical, 0 meaning they are completely uncorrelated, and -1 meaning they are perfectly differentiated.¹⁴

Finally, we supplement the analysis of assortment composition by an analysis of assortment size. Specifically, we define assortment-size similarity as the negative absolute difference¹⁵ in yearly average assortment sizes between a pair of stores, where assortment size is defined as a simple count of a number of products.¹⁶

4.2. Operationalization of Measures

4.2.1. Measures of Ownership. We incorporate the ownership structures of the outlets through a series of indicator variables. For corporate chains, three dummy variables that indicate different levels of hierarchy of the "same ownership" are created as follows: same corporate owner (*SOWN*), same chain name or chain banner (*SCHAIN*), and same division (*SDIVISION*). For example, *SOWN* is 1 for a pair consisting of one Vons and one Safeway outlet, both of which belong to the Safeway Group. However, Vons and Safeway are different banners. Thus, the banner-indicator *SCHAIN* would equal 1 for a pair of Vons stores, but *SCHAIN* would equal 0 for a pair consisting of one Vons and one Safeway store. Finally, some stores have local divisions within a banner that have separate buying offices (e.g., Randalls Food Market has separate Austin and Houston divisions). We define a third ownership indicator accordingly: *SDIVISION* is 1 if two stores are served by the same buying office and 0 otherwise.

For affiliated independent chains, we create a separate indicator variable, *SAIC*, that equals 1 if both stores in the pair belong to the same affiliated independent chain. Examples of such affiliated independent chains are independent retailer cooperatives such as Piggly Wiggly and ShopRite, or wholesale groups such as Shop 'n Save. The independent grocer alliance (IGA) represents another form of coordination; thus, we create a separate indicator variable for IGA stores, *SIGA*, which is 1 if both stores in a store pair are part of the independent grocer alliance.

¹⁴ Our similarity measures can be used with alternative dimensions of variety of an assortment. For instance, Hoch et al. (1999) relate assortment variety to underlying attributes. In their context, it is possible to define A_i using $K \leq N$ unique combinations of (any subset of) attributes on which variety is measured, instead of the N products.

¹⁵ We use the negative absolute value to create a measure of size *similarity* and stay consistent with the analysis of assortment composition.

¹⁶ The difference in assortment size can be a fractional number if some products are not stocked for the entire duration of an observation window.

Finally, we create two indicator variables to capture systematic differences for nonaffiliated independent stores. *ONE-IND* equals 1 if exactly one member of the store pair is a nonaffiliated independent store. *BOTH-IND* equals 1 if both stores in the pair are nonaffiliated independent stores.

In summary, seven indicator variables are created to capture different hierarchies and types of ownership: an indicator for the same corporate owner, an indicator for the same banner, an indicator for the same division, an indicator for the same affiliated independent chain, an indicator for the IGA stores, an indicator for pairs with only one nonaffiliated independent store, and an indicator for nonaffiliated independent store pairs.

4.2.2. Measure of Store Size. The difference in physical store sizes between a pair of supermarkets (*SSDIFF*) is measured by computing (absolute) difference in physical store sizes for the pair of stores. The difference is expressed in thousands of square feet.

4.2.3. Measures of Customer Factors. We measure demographic differences through cosine demographic distance measures.¹⁷ We do this for seven different demographic measures: income (*D(INCOME)*), household size (*D(HSIZE)*), head of household age (*D(AGE)*), ethnic diversity (*D(ETHNIC)*), presence of children in each age group (*D(CHILDREN)*), education (*D(EDUCATION)*), and home ownership (*D(HOWN)*).

4.2.4. Measures of Competition Factors. We measure the geographic distance between a pair of stores by calculating great circle distances using the longitudes and latitudes of stores in the pair (*DIST*). We also calculate the differences in the number of competitors located within three miles (a typical size of trading areas for supermarkets) of each store. When we count the number of competitors,

¹⁷For supermarkets i and j , we define vectors of marginal demographic distributions as D_i and D_j . A cosine demographic distance measure between supermarket i and j is defined as $1 - ((D_i D_j) / (\|D_i\| \|D_j\|))$. The bins for marginal demographic distributions are defined as follows:

- (a) Income: households with income (1) under \$10,000, (2) \$10,000–\$19,999, (3) \$20,000–\$29,999, (4) \$30,000–\$39,999, (5) \$40,000–\$49,999, (6) \$50,000–\$74,999, (7) \$75,000–\$99,999, (8) \$100,000 and over.
- (b) Household size: (1) 1 person, (2) 2 persons, (3) 3 persons, (4) 4 persons, (5) 5+ persons.
- (c) Head of household age: (1) 18–24, (2) 25–34, (3) 35–44, (4) 45–54, (5) 55–64, (6) 65 and over.
- (d) Ethnic diversity: Householder's ethnic origin is (1) white, (2) black, (3) Hispanic, (4) other.
- (e) Presence of children in each age group: (1) No children, (2) children < 6 only, (3) children 6–17 only, (4) children < 6 and 6–17.
- (f) Education: (1) grade school, (2) some high school, (3) high school graduate, (4) some college, (5) college graduate.
- (g) Home ownership: (1) owned, (2) rented.

we distinguish supermarkets (*D(FRNUMSCOMP)*) and supercenters from discounters such as Walmart (*D(FRNUMSCCOMP)*). We have also used an alternative definition of a local market (the number of outlets located in the same five-digit zip code) as a robustness check.

Table 1 lists the descriptive statistics of the measures used in this study.

4.3. Models and Estimation

We estimate the following linear regression model, where observations are pooled across geographic markets:

$$y_{ijm} = \beta' x_{ijm} + e_{ijm}, \quad (3)$$

where y_{ijm} is the observed assortment composition similarity (or assortment size similarity) between store i and store j in a geographic market m , x_{ijm} are independent variables that characterize the pair of stores (i, j) in a geographic market m , and e_{ijm} is the error term. We estimate this model using ordinary least squares (OLS). However, because our unit of analysis consists of pairs of stores, we expect dependence among store pairs that share a common store. This prevents us from calculating our standard errors in the standard way, because we have correlation in the errors across different pairs. We are not able to use traditional procedures to deal with these correlations and obtain standard errors (e.g., Wooldridge 2003) because the sets of pairs that share common stores are nonnested. Cameron et al. (2008) recently proposed a method to compute robust standard errors under a nonnested clustering structure. However, this approach is not very scalable when the number of nonnested clusters becomes large. Mizruchi (1989) suggests a fixed-effect approach, where the researcher inserts dummy variables for the $N - 1$ members. However, this approach requires estimating a huge number of dummy variables.

We use a method called the quadratic assignment procedure (QAP) that circumvents these drawbacks and was suggested by Krackhardt (1988).¹⁸ The QAP does not require estimation of a large number of dummy variables and is robust against potential misspecification of autocorrelation structure in the dyadic data. In the QAP, an empirical sampling distribution of coefficients under the null hypothesis (i.e., no statistical association between independent variables and dependent variable) is generated by swapping the dependent variables between observations and conducting OLS regressions with the permuted data set. The permutations of the data are created in a way that keeps the correlation structure in the original data. An empirical sampling distribution of

¹⁸For the details of the QAP, refer to the appendix.

Table 1 Descriptive Statistics of the Data

Variable	Description	Mean	Std. dev.
<i>ACOS</i> (Cola)	Cosine assortment similarity measure in cola category	0.83	0.10
<i>ACORR</i> (Cola)	Correlation assortment similarity measure in cola category	0.70	0.15
<i>ASIZE</i> (Cola)	Assortment size similarity measure in cola category	−6.40	6.16
<i>ACOS</i> (Cereal)	Cosine assortment similarity measure in cereal category	0.78	0.09
<i>ACORR</i> (Cereal)	Correlation assortment similarity measure in cereal category	0.67	0.13
<i>ASIZE</i> (Cereal)	Assortment size similarity measure in cereal category	−35.75	28.15
<i>ACOS</i> (Coffee)	Cosine assortment similarity measure in ground coffee category	0.69	0.13
<i>ACORR</i> (Coffee)	Correlation assortment similarity measure in ground coffee category	0.60	0.15
<i>ASIZE</i> (Coffee)	Assortment size similarity measure in ground coffee category	−16.68	13.20
<i>ACOS</i> (Toothpaste)	Cosine assortment similarity measure in toothpaste category	0.82	0.11
<i>ACORR</i> (Toothpaste)	Correlation assortment similarity measure in toothpaste category	0.71	0.14
<i>ASIZE</i> (Toothpaste)	Assortment size similarity measure in toothpaste category	−13.65	11.41
<i>SOWN</i>	Same ultimate corporate owner	0.14	0.35
<i>SCHAIN</i>	Same (local) chain banner (nested in <i>SOWN</i>)	0.11	0.32
<i>SDIVISION</i>	Same local division (nested in <i>SCHAIN</i>)	0.08	0.28
<i>SAIC</i>	Same affiliated independent chain	0.003	0.06
<i>SIGA</i>	IGA member stores	0.002	0.04
<i>ONE-IND</i>	Indicator(one store in the pair = independent store)	0.16	0.37
<i>BOTH-IND</i>	Indicator(both stores in the pair = independent stores)	0.01	0.11
<i>SSDIFF</i>	Physical store size difference ('000 square feet)	18.15	15.23
<i>BAND1</i> × <i>SOWN</i>	Indicator(distance ≤ 3 miles) × same ultimate corporate owner	0.001	0.02
<i>BAND1</i> × (1 − <i>SOWN</i>)	Indicator(distance ≤ 3 miles) × competing owners	0.004	0.06
<i>DIST</i> × <i>BAND1</i> × <i>SOWN</i>	Indicator(distance ≤ 3 miles) × distance (in miles) × same ultimate corporate owner	0.001	0.05
<i>DIST</i> × <i>BAND1</i> × (1 − <i>SOWN</i>)	Indicator(distance ≤ 3 miles) × distance (in miles) × competing owners	0.01	0.12
<i>D(FRNUMSCOMP)</i>	Difference in a number of supermarket competitors in a circular area within a 3-mile radius	8.61	17.79
<i>D(FRNUMSCCOMP)</i>	Difference in a number of supercenter competitors in a circular area within a 3-mile radius	0.42	0.61
<i>D(INCOME)</i>	Income distribution dissimilarity (cosine distance)	0.10	0.12
<i>D(HSIZE)</i>	Household size distribution dissimilarity (cosine distance)	0.04	0.06
<i>D(AGE)</i>	Head of household age distribution dissimilarity (cosine distance)	0.04	0.06
<i>D(ETHNIC)</i>	Ethnic group distribution dissimilarity (cosine distance)	0.12	0.19
<i>D(CHILDREN)</i>	Presence of children in each age group distribution dissimilarity (cosine distance)	0.01	0.02
<i>D(EDUCATION)</i>	Education level distribution dissimilarity (cosine distance)	0.10	0.11
<i>D(HOWN)</i>	Home ownership distribution dissimilarity (cosine distance)	0.06	0.09

Notes. Units are store pairs; $N = 131,376$. Values listed are for the cola category; the values for the other categories are similar.

coefficients under the null hypothesis is then used to assess the statistical significance of coefficients. Krackhardt's suggestion has been widely applied (Gulati and Gargiulo 1999, Khanna et al. 2006). It is important to realize that the QAP provides more conservative estimates of statistical significance compared to inferences under assumed independence, and it controls for the impact of network autocorrelations on inferred standard errors.

Because we restrict pairwise comparisons within the same geographic market in our data and pool the data across different geographic markets, we implement the QAP for multiple groups where permutations of the dependent variable are limited only within a group. A STATA subroutine developed by Simpson (2001) is utilized to implement the QAP.

5. Results

5.1. Results on Assortment Similarity

Table 2 presents regression results that analyze our assortment similarity measures for four product categories: cola, ready-to-eat cereal, coffee, and toothpaste.

Model (1) is a base specification that takes a cosine assortment similarity as the dependent variable and includes all the variables discussed previously. In model (2), the centered correlation measure is used as the dependent variable to ensure robustness of results against potential biases from differences in stocking rates and assortment size. Finally, in model (3), assortment size similarity is used as the dependent variable. To control for state-level differences in mean assortment similarities, state dummies are included in all

the specifications. The estimated state effects are suppressed to avoid cluttered tables.

5.1.1. Ownership. We find that ownership has a strong impact on the assortment similarities of supermarkets in all four product categories. Having the same corporate owner is the largest factor in determining the degree of assortment similarity, increasing assortment overlap by 7.9% in the cola category, 11.1% in the ready-to-eat cereal category, 12.1% in the ground coffee category, and 8.3% in the toothpaste category. These effects are both statistically significant and economically substantial. Two stores being in the same division further increases the observed assortment homogeneity between the stores in the cola, cereal, and ground coffee categories, and chain banners are important in all but the cola category. However, these effects are much smaller than the same-owner effect. The fact that the effect of having the same corporate owner is greater than the effect of having the same chain banner suggests that economies of scope and scale in the operations of the company is a larger factor in determining a store's assortment than the brand value of consistency across stores.

These same-ownership variables also impact the total assortment size, although the relative importance of each of the different levels of ownership differs to some extent. The largest coefficient for assortment composition is with same-parent ownership, *SOWN*, whereas the largest coefficient for assortment size is with same-chain ownership, *SCHAIN*.

We also find a strong homogenization effect of chain membership for affiliated independent stores. When two stores become members of the same affiliated independent chains, assortment overlap increases by 4.7% (cola) to 18.5% (ground coffee). This suggests that affiliated independent chains also try to benefit from economies of scale and scope by distributing similar sets of products across its member stores. Common membership in an independent chain does not have a statistically significant impact on the total number of products that are on the store's shelf except for the ground coffee category. Combined, the results of the regressions for assortment composition similarity and assortment size similarity may reflect that belonging to an independent chain makes a certain set of products more easily and cheaply available to member supermarkets, leading to greater similarity. However, the stores are not coordinated on how many of these items to stock.

Finally, independent stores are shown to make dissimilar assortment decisions. On average, having at least one independent store in the pair decreases assortment overlap by around 5%–7%, with an even larger effect in the toothpaste category.

5.1.2. Store Size. We find that stores with greater differences in physical size have fewer assortment similarities. This effect is both statistically and economically significant. With the maximum observed variation in the data—107,000 square feet—the calculated decrease in assortment similarity from the coefficient estimates can be as high as 9% (cola) to 17% (ground coffee).

5.1.3. Local Demographics. We find some evidence that supermarkets tailor their assortments to local demographics. Supermarkets facing more similar local demographics are generally shown to make more similar assortment choices, both in terms of assortment size and composition. The only statistically significant exceptions are home ownership in the cola category and head of household age in the ground coffee category.

5.1.4. Distance Between Stores. We estimate separate impacts of distance on assortment choices for stores with the same owner and for stores with different owners. The impact of geography is separated into two components: a fixed component (denoted by *BAND1*) if the stores are located within three miles of each other and a component that varies with distance within that band. In general, the effects of distance are small (see also the next section), and the only consistent pattern that we find is that the impact of geographic distance has less influence on assortment offerings than other factors.

5.2. Relative Importance of Factors

We assess the relative importance of each of the factors in our study (corporate ownership, customer characteristics, and competition—the 3 Cs) by comparing the fit of regressions analogous to those in Table 2 when we include only some of the independent variables. As above, again, state dummies are included in all of the regression equations.

Table 3 presents the results of a series of regression equations with their adjusted R^2 . The first column shows the model with only state dummies as regressors. Columns 2 to 4 present each set of factors from the 3 Cs, added separately to state dummies. Finally, column 5 shows the regression results with all factors.

Consistent with the discussion of the results in the previous section, we find that company-specific factors are most influential in explaining assortment similarities.¹⁹ Table 3 demonstrates that the addition of the company factors to the base model, which only includes state dummies, results in increases in adjusted R^2 that are an order of magnitude larger

¹⁹ This is consistent with Dhar and Hoch (1997), which also finds weak effects of demographics on the variability of store-brand market shares at the chain level.

Table 2 Estimation Results for Overall Assortment Similarity Regressions

Independent variables	Assortment composition 1 (Cosine similarity)		Assortment composition 2 (Pearson correlation)		Assortment size similarity	
	Coefficient	(p-Value)	Coefficient	(p-Value)	Coefficient	(p-Value)
(1) Cola						
<i>SOWN</i>	0.079	(0.000)***	0.147	(0.000)***	0.595	(0.102)
<i>SCHAIN</i>	−0.0005	(0.942)	−0.002	(0.852)	1.521	(0.006)***
<i>SDIVISION</i>	0.034	(0.000)***	0.064	(0.000)***	0.737	(0.024)**
<i>SAIC</i>	0.047	(0.000)***	0.094	(0.000)***	−0.141	(0.786)
<i>SIGA</i>	−0.022	(0.056)*	−0.010	(0.592)	−1.425	(0.122)
<i>ONE-IND</i>	−0.064	(0.000)***	−0.079	(0.000)***	−4.475	(0.000)***
<i>BOTH-IND</i>	−0.071	(0.000)***	−0.072	(0.000)***	−4.593	(0.000)***
<i>SSDIFF</i>	−0.001	(0.000)***	−0.001	(0.000)***	−0.087	(0.000)***
<i>BAND1</i> × <i>SOWN</i>	−0.025	(0.428)	−0.030	(0.590)	−2.935	(0.146)
<i>BAND1</i> × (1 − <i>SOWN</i>)	0.035	(0.000)***	0.069	(0.000)***	−0.239	(0.736)
<i>DIST</i> × <i>BAND1</i> × <i>SOWN</i>	0.012	(0.412)	0.018	(0.446)	0.939	(0.304)
<i>DIST</i> × <i>BAND1</i> × (1 − <i>SOWN</i>)	−0.005	(0.316)	−0.008	(0.286)	−0.021	(0.936)
<i>D(FRNUMSCOMP)</i>	−0.001	(0.000)***	−0.001	(0.000)***	−0.008	(0.218)
<i>D(FRNUMSCOMP)</i>	−0.0002	(0.924)	−0.003	(0.422)	−0.289	(0.052)*
<i>D(INCOME)</i>	−0.043	(0.002)***	−0.075	(0.000)***	−0.997	(0.344)
<i>D(HSIZE)</i>	−0.047	(0.408)	−0.065	(0.424)	−1.434	(0.666)
<i>D(AGE)</i>	−0.003	(0.848)	−0.029	(0.490)	2.078	(0.338)
<i>D(ETHNIC)</i>	−0.050	(0.002)***	−0.066	(0.002)***	−3.426	(0.000)***
<i>D(CHILDREN)</i>	0.056	(0.790)	0.155	(0.584)	0.990	(0.948)
<i>D(EDUCATION)</i>	−0.029	(0.084)*	−0.041	(0.104)	−1.606	(0.158)
<i>D(HOWN)</i>	0.040	(0.034)**	0.065	(0.018)**	−0.582	(0.680)
STATE DUMMIES	YES		YES		YES	
Sample size	131,376		131,376		131,376	
<i>R</i> ²	0.477		0.466		0.242	
Adjusted <i>R</i> ²	0.477		0.466		0.241	
(2) Cereal						
<i>SOWN</i>	0.111	(0.000)***	0.169	(0.000)***	4.532	(0.000)***
<i>SCHAIN</i>	0.025	(0.004)***	0.028	(0.012)**	6.395	(0.004)***
<i>SDIVISION</i>	0.020	(0.000)***	0.037	(0.000)***	−1.758	(0.196)
<i>SAIC</i>	0.103	(0.000)***	0.172	(0.000)***	3.816	(0.216)
<i>SIGA</i>	0.034	(0.002)***	0.080	(0.000)***	1.437	(0.670)
<i>ONE-IND</i>	−0.052	(0.000)***	−0.047	(0.000)***	−10.409	(0.000)***
<i>BOTH-IND</i>	−0.033	(0.000)***	0.0001	(0.986)	−6.516	(0.004)***
<i>SSDIFF</i>	−0.001	(0.000)***	−0.002	(0.000)***	−0.663	(0.000)***
<i>BAND1</i> × <i>SOWN</i>	−0.086	(0.038)**	−0.106	(0.018)**	−8.442	(0.430)
<i>BAND1</i> × (1 − <i>SOWN</i>)	−0.029	(0.012)**	−0.041	(0.010)**	−7.809	(0.036)**
<i>DIST</i> × <i>BAND1</i> × <i>SOWN</i>	0.025	(0.124)	0.031	(0.170)	3.036	(0.506)
<i>DIST</i> × <i>BAND1</i> × (1 − <i>SOWN</i>)	0.006	(0.268)	0.009	(0.226)	2.938	(0.090)*
<i>D(FRNUMSCOMP)</i>	−0.0001	(0.186)	−0.0002	(0.230)	0.129	(0.000)***
<i>D(FRNUMSCOMP)</i>	−0.0002	(0.912)	−0.0018	(0.494)	−0.194	(0.764)
<i>D(INCOME)</i>	−0.006	(0.684)	−0.019	(0.284)	1.809	(0.680)
<i>D(HSIZE)</i>	−0.096	(0.122)	−0.142	(0.074)*	−17.241	(0.268)
<i>D(AGE)</i>	0.002	(0.960)	−0.038	(0.342)	7.457	(0.342)
<i>D(ETHNIC)</i>	−0.016	(0.086)*	0.009	(0.440)	−7.104	(0.010)***
<i>D(CHILDREN)</i>	0.107	(0.574)	0.258	(0.326)	20.808	(0.664)
<i>D(EDUCATION)</i>	−0.148	(0.000)***	−0.203	(0.000)***	−29.097	(0.000)***
<i>D(HOWN)</i>	−0.009	(0.616)	−0.009	(0.672)	−12.986	(0.026)**
STATE DUMMIES	YES		YES		YES	
Sample size	130,748		130,748		130,748	
<i>R</i> ²	0.539		0.528		0.226	
Adjusted <i>R</i> ²	0.539		0.528		0.225	

than the increases that are gained from adding competition or customer variables. It also appears that in most categories, customer factors impact assortment more than competition factors do. State dummies also explain a sizable variance, especially in the cola category. Given that there are differences in the

availability of regional brands and the number of independent stores across states, state dummies are expected to pick up variance as a result of different market structure across states.

As an illustrative example of the relative impact of corporate ownership and customer characteristics,

Table 2 (Cont'd.)

Independent variables	Assortment composition 1 (Cosine similarity)		Assortment composition 2 (Pearson correlation)		Assortment size similarity	
	Coefficient	(p-Value)	Coefficient	(p-Value)	Coefficient	(p-Value)
(3) Ground coffee						
<i>SOWN</i>	0.121	(0.000)***	0.162	(0.000)***	2.161	(0.000)***
<i>SCHAIN</i>	0.067	(0.000)***	0.076	(0.000)***	5.501	(0.000)***
<i>SDIVISION</i>	0.060	(0.000)***	0.087	(0.000)***	0.140	(0.850)
<i>SAIC</i>	0.185	(0.000)***	0.261	(0.000)***	7.248	(0.000)***
<i>SIGA</i>	0.059	(0.004)***	0.110	(0.000)***	3.535	(0.024)**
<i>ONE-IND</i>	−0.056	(0.000)***	−0.037	(0.000)***	−7.175	(0.000)***
<i>BOTH-IND</i>	0.009	(0.306)	0.057	(0.000)***	0.528	(0.678)
<i>SSDIFF</i>	−0.002	(0.000)***	−0.002	(0.000)***	−0.203	(0.000)***
<i>BAND1</i> × <i>SOWN</i>	−0.081	(0.080)*	−0.090	(0.082)*	−3.554	(0.534)
<i>BAND1</i> × (1 − <i>SOWN</i>)	0.019	(0.180)	0.026	(0.134)	0.500	(0.732)
<i>DIST</i> × <i>BAND1</i> × <i>SOWN</i>	0.024	(0.262)	0.028	(0.278)	0.458	(0.880)
<i>DIST</i> × <i>BAND1</i> × (1 − <i>SOWN</i>)	−0.006	(0.422)	−0.009	(0.268)	−0.420	(0.562)
<i>D(FRNUMSCOMP)</i>	−0.0002	(0.106)	−0.0002	(0.052)*	0.066	(0.000)***
<i>D(FRNUMSCCOMP)</i>	−0.0014	(0.606)	−0.0027	(0.334)	−0.388	(0.160)
<i>D(INCOME)</i>	−0.049	(0.006)***	−0.064	(0.000)***	−3.488	(0.040)**
<i>D(HSIZE)</i>	−0.023	(0.766)	−0.029	(0.688)	2.026	(0.744)
<i>D(AGE)</i>	0.080	(0.036)**	0.071	(0.084)*	8.524	(0.010)***
<i>D(ETHNIC)</i>	−0.052	(0.000)***	−0.035	(0.008)***	−4.156	(0.002)***
<i>D(CHILDREN)</i>	−0.010	(0.982)	0.025	(0.880)	−15.648	(0.446)
<i>D(EDUCATION)</i>	−0.133	(0.000)***	−0.144	(0.000)***	−17.605	(0.000)***
<i>D(HOWN)</i>	−0.025	(0.294)	−0.032	(0.224)	−6.177	(0.008)***
STATE DUMMIES	YES		YES		YES	
Sample size	128,574		128,574		128,574	
<i>R</i> ²	0.593		0.592		0.235	
Adjusted <i>R</i> ²	0.593		0.591		0.234	
(4) Toothpaste						
<i>SOWN</i>	0.083	(0.000)***	0.140	(0.000)***	0.836	(0.134)
<i>SCHAIN</i>	0.044	(0.000)***	0.074	(0.000)***	2.192	(0.004)***
<i>SDIVISION</i>	−0.003	(0.658)	−0.0001	(0.978)	1.165	(0.012)**
<i>SAIC</i>	0.070	(0.000)***	0.151	(0.000)***	2.392	(0.122)
<i>SIGA</i>	−0.049	(0.084)*	−0.017	(0.630)	1.425	(0.500)
<i>ONE-IND</i>	−0.126	(0.000)***	−0.133	(0.000)***	−8.892	(0.000)***
<i>BOTH-IND</i>	−0.133	(0.000)***	−0.111	(0.000)***	−2.051	(0.176)
<i>SSDIFF</i>	−0.002	(0.000)***	−0.002	(0.000)***	−0.342	(0.000)***
<i>BAND1</i> × <i>SOWN</i>	−0.249	(0.000)***	−0.299	(0.000)***	−11.494	(0.062)*
<i>BAND1</i> × (1 − <i>SOWN</i>)	−0.027	(0.086)*	−0.025	(0.164)	−1.851	(0.224)
<i>DIST</i> × <i>BAND1</i> × <i>SOWN</i>	0.085	(0.000)***	0.102	(0.000)***	4.297	(0.104)
<i>DIST</i> × <i>BAND1</i> × (1 − <i>SOWN</i>)	0.001	(0.940)	−0.001	(0.894)	0.141	(0.792)
<i>D(FRNUMSCOMP)</i>	−0.001	(0.002)***	−0.001	(0.004)***	−0.073	(0.002)***
<i>D(FRNUMSCCOMP)</i>	0.002	(0.650)	0.001	(0.752)	0.121	(0.676)
<i>D(INCOME)</i>	0.022	(0.310)	0.029	(0.238)	−0.488	(0.780)
<i>D(HSIZE)</i>	−0.063	(0.462)	−0.084	(0.390)	−4.551	(0.440)
<i>D(AGE)</i>	0.019	(0.676)	0.010	(0.876)	1.481	(0.656)
<i>D(ETHNIC)</i>	−0.036	(0.046)**	−0.029	(0.124)	−2.639	(0.046)**
<i>D(CHILDREN)</i>	0.124	(0.669)	0.180	(0.562)	12.109	(0.516)
<i>D(EDUCATION)</i>	−0.113	(0.000)***	−0.149	(0.000)***	−5.845	(0.000)***
<i>D(HOWN)</i>	−0.030	(0.284)	−0.039	(0.244)	−3.372	(0.120)
STATE DUMMIES	YES		YES		YES	
Sample size	105,136		105,136		105,136	
<i>R</i> ²	0.491		0.539		0.372	
Adjusted <i>R</i> ²	0.490		0.539		0.371	

*Significant at the 0.10 level; **significant at the 0.05 level; ***significant at the 0.01 level.

Figure 4 presents assortment similarities for two selected Ralphs–Ralphs store pairs on the left and two Ralphs–Vons store pairs on the right. The store pairs are selected to allow for a contrast between the pairs that are catering to trading areas with the

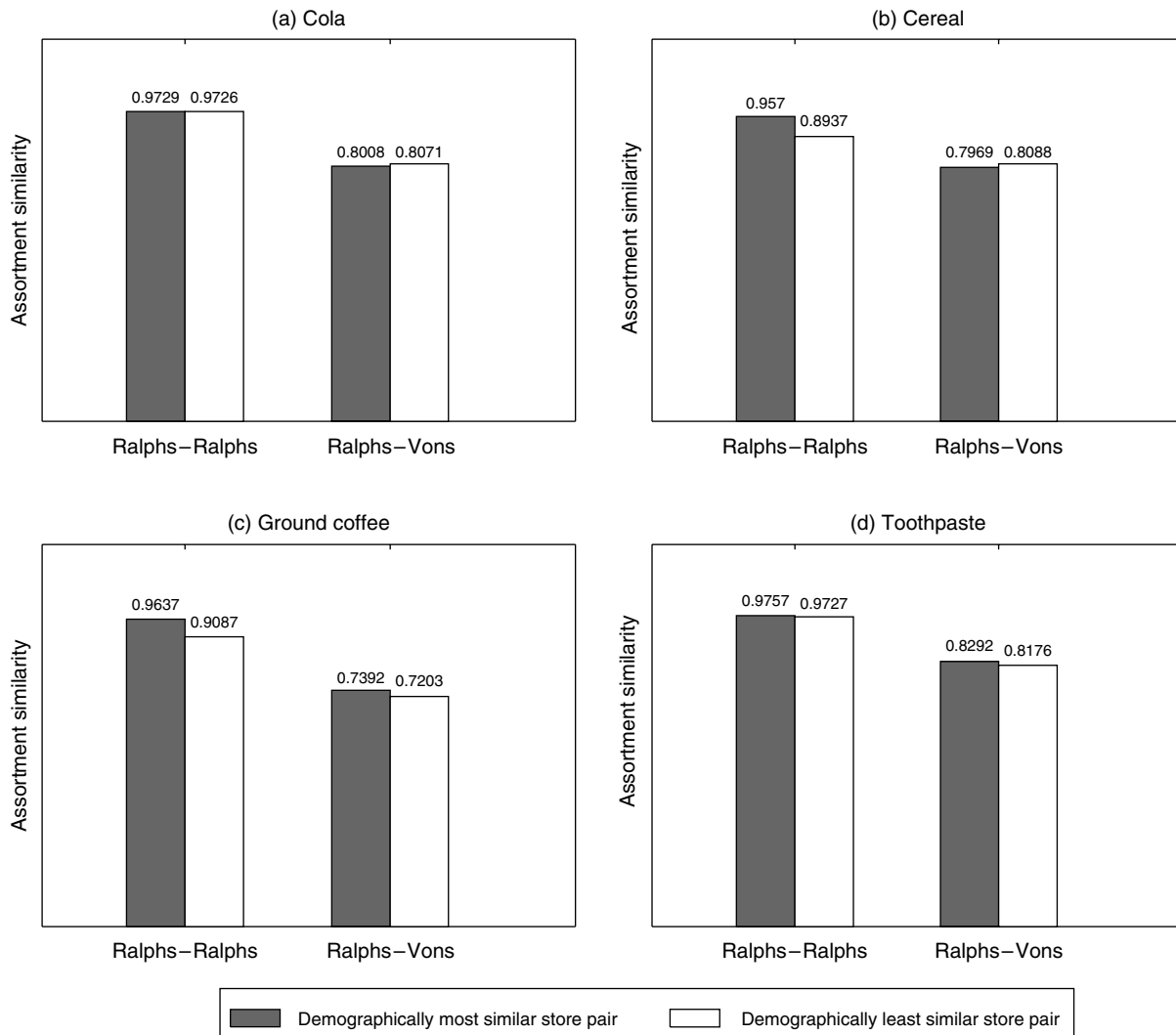
highest and the lowest degrees of demographic similarity. Thus, on the left-hand side of the graph, we graph the pair of Ralphs stores located in neighborhoods with the most similar demographics along with the pair of Ralphs stores located in neighborhoods

Table 3 Decomposition of Explained Variance

Variable	Category	State dummies	State dummies + Competition factors	State dummies + Customer factors	State dummies + Company factors	State dummies + Company + Customer + Competition factors
Assortment composition	Cola	0.185	0.209	0.224	0.437	0.477
	Cereal	0.053	0.065	0.126	0.489	0.539
	Ground coffee	0.077	0.088	0.133	0.555	0.593
	Toothpaste	0.044	0.084	0.107	0.439	0.490
Assortment size	Cola	0.058	0.062	0.081	0.223	0.241
	Cereal	0.028	0.029	0.053	0.204	0.225
	Ground coffee	0.033	0.034	0.076	0.198	0.234
	Toothpaste	0.064	0.090	0.096	0.348	0.371

with the least similar demographics with respect to a focal Ralphs store. On the right-hand side of the graph, we have the Ralphs–Vons store pairs chosen with the same criteria with respect to the focal Ralphs store. Figure 4 graphically illustrates that the impact of demographics on a store’s assortment is

much smaller than the impact of outlet ownership. In addition, consistent with the variance decomposition results, the impact of demographics on a store’s assortment is larger in the cereal and ground coffee categories, especially for store pairs within the same owner.

Figure 4 Assortment Similarity and Demographics

5.3. Robustness Checks

We conduct several robustness checks of our results. First, we replicate our analysis using the three different assortment vector constructions described in §4.1, which demonstrate that our results are robust to common missing data issues.

We also confirm the robustness of our results to the discrepancy in aggregation levels between product categories in our data. Reanalyzing the cola data at the same aggregation level as the other categories yields results that do not change qualitatively from those presented above.

Next, we use alternative demographic distance measures based on differences in percentage of households over thresholds (e.g., difference in percentage of households with more than \$75,000 income) and an alternative local market definition (i.e., a same five-digit zip code instead of a fixed radius of three miles) to compute the difference in a number of competitors. The results are also robust to these changes.

Finally, one might argue that the cola category is only a subcategory and not a complete category, e.g., that the soft drink category would be a better category to use. We replicate our analyses using the complete soft drink category and find similar results.²⁰

5.4. Interstate Analysis

Until now, we have restricted our analyses to assortment similarity comparisons between stores located within the same state. We did this to control for product availability across markets.²¹ Nevertheless, it is worth quantifying how different these interstate assortment similarities are. To accomplish this, we focus on supermarkets with multistate territories, which we operationalize as the top 10 supermarket firms. We compare the mean assortment similarities across stores belonging to the same firm within a state versus across states with the mean assortment similarities across firms within a state versus across states. The results are presented in Table 4. We find that there is a reduction that is economically and statistically large in mean assortment similarities within a firm when we compare assortments across states versus within a state. Relative to within-chain and within-state assortment similarity, the typical drop in assortment similarity for stores of the same chain across

²⁰ We have not included the tables of these results in the paper for space reasons. The results of the robustness checks are available from the authors upon request.

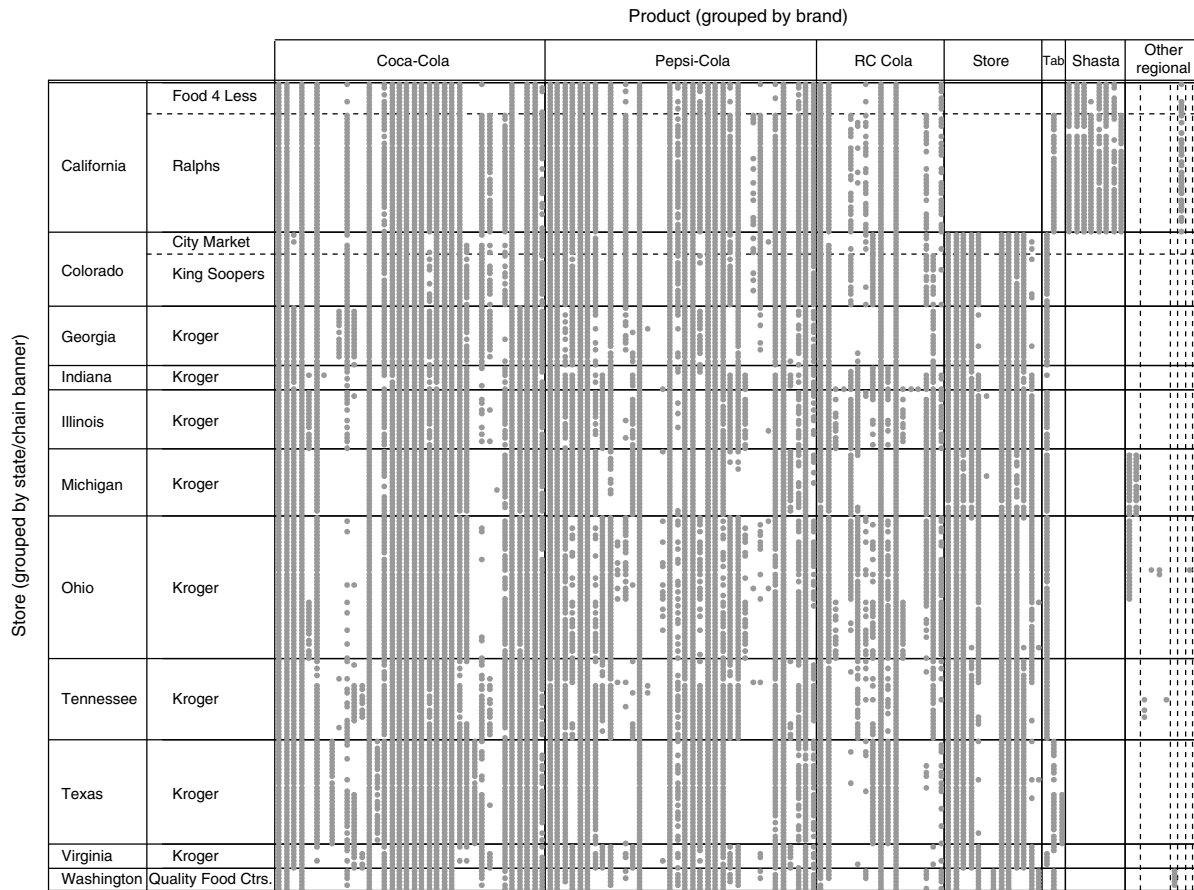
²¹ For example, finding that Jolly Good soda is available in one supermarket in Madison, WI but not another in Milwaukee, WI is meaningful in the sense that Jolly Good is available to both sets of stores. However, comparing the choice of whether to stock Jolly Good between a store in Wisconsin and a store in California, where Jolly Good is not distributed, is *prima facie* less meaningful.

Table 4 Mean Assortment Similarities Within a State and Across States

Firm name	Within a state		Across states	
	Mean	Standard error	Mean	Standard error
(1) Cola: Assortment similarities				
Within owner				
Kroger	0.939	0.003	0.809	0.006
Albertsons	0.945	0.004	0.831	0.004
Safeway	0.945	0.003	0.868	0.004
Ahold USA	0.942	0.006	0.874	0.004
Publix	0.960	0.002	0.928	0.003
SUPERVALU	0.926	0.014	0.810	0.009
Delhaize America	0.947	0.004	0.874	0.008
H-E-B	0.843	0.013	N/A	N/A
Winn-Dixie	0.964	0.003	0.920	0.007
A&P	0.920	0.011	0.827	0.012
Cross-owners	0.813	0.003	0.763	0.002
(2) Cereal: Assortment similarities				
Within owner				
Kroger	0.893	0.008	0.816	0.007
Albertsons	0.895	0.006	0.799	0.005
Safeway	0.913	0.003	0.893	0.003
Ahold USA	0.914	0.007	0.850	0.004
Publix	0.910	0.003	0.917	0.005
SUPERVALU	0.895	0.015	0.805	0.006
Delhaize America	0.960	0.012	0.898	0.013
H-E-B	0.866	0.016	N/A	N/A
Winn-Dixie	0.953	0.003	0.953	0.004
A&P	0.894	0.012	0.879	0.009
Cross-owners	0.753	0.003	0.742	0.002
(3) Ground coffee: Assortment similarities				
Within owner				
Kroger	0.867	0.011	0.668	0.010
Albertsons	0.873	0.010	0.617	0.005
Safeway	0.883	0.007	0.748	0.006
Ahold USA	0.917	0.013	0.772	0.005
Publix	0.945	0.006	0.935	0.006
SUPERVALU	0.872	0.028	0.661	0.012
Delhaize America	0.914	0.016	0.793	0.018
H-E-B	0.807	0.023	N/A	N/A
Winn-Dixie	0.872	0.011	0.799	0.019
A&P	0.913	0.016	0.832	0.019
Cross-owners	0.652	0.003	0.551	0.002
(4) Toothpaste: Assortment similarities				
Within owner				
Kroger	0.921	0.006	0.881	0.007
Albertsons	0.938	0.008	0.915	0.008
Safeway	0.938	0.004	0.938	0.004
Ahold USA	0.942	0.010	0.883	0.009
Publix	0.934	0.007	0.942	0.009
SUPERVALU	0.903	0.015	0.853	0.011
Delhaize America	0.945	0.007	0.894	0.008
H-E-B	0.892	0.014	N/A	N/A
Winn-Dixie	0.977	0.004	0.965	0.004
A&P	0.892	0.015	0.904	0.010
Cross-owners	0.792	0.005	0.794	0.003

different states is less than the drop for stores of different chains in the same state; however, both drops are sizable.

These findings suggest that multiregional supermarket firms version assortment by regional divisions,

Figure 5 Store-Level Assortments for Cola in Kroger Stores Across States

which are largely structured at the state level except for the largest states. However, the fact that mean assortment similarities within a firm across states are still larger than mean assortment similarities across firms across states suggests that some pooling of assortments for stores of the same chain continues to exist even across states.

As an illustrative example of the moderating effects of geography, Figure 5 presents a graph of the cola products stocked by each of the Kroger stores across several states in our data set.²² The solid lines separate different states, whereas the dashed lines separate different local chain banners within a state. Although there are sporadic differences between stores within the same state, visual inspection reveals that the variance across states is greater than the variance within a state, especially within the same local banner. Comparing Figures 3 and 5 reinforces the finding in Table 4 that assortment variation across chains within a state is of a similar order of magnitude as assortment variation across states within a chain.

²² Figure 5 excludes local chain banners with a small number of sample stores.

6. Regional Brand Distribution

6.1. Models and Estimation

The results in the previous section suggest the existence of some degree of regional assortment customization. On the one hand, this may allow regional brands to gain access to large chains. On the other hand, the fact that assortments are not strongly adjusted for demographics raises the question about which retail chains provide the greatest opportunity for regional-brand distribution. In this context, we explore the determinants of regional brand distribution in this section.

We aggregate assortment vectors to the brand level and focus on regional brands, which are defined as brands that do not have distribution in each of the 21 states in our data.²³ We also categorize each regional brand as belonging to one of two price tiers. By taking a median price between an average national brand price and an average store brand price as a reference point, we define “value regional brands”

²³ In the toothpaste category, we also exclude two brands from national manufacturers (Viadent from Colgate and Pearl Drops from Church & Dwight). Both brands are distributed in 19 of 21 sample states.

Table 5 Estimation Results for Regional Brand Distribution

Independent variables	Model (1) (Total regional brand)		Model (2) (Value regional brand)		Model (3) (Premium regional brand)	
	Coefficient	(p-Value)	Coefficient	(p-Value)	Coefficient	(p-Value)
(1) Cola						
Store brand presence	−2.625	(0.000)***	−2.485	(0.000)***	−0.559	(0.126)
Retailer market share	−9.189	(0.002)***	−8.622	(0.004)***	−3.935	(0.518)
Retailer market share ²	14.196	(0.006)***	13.987	(0.010)***	−4.164	(0.798)
Store size	0.004	(0.590)	0.006	(0.377)	−0.014	(0.192)
High-income households (%)	−2.675	(0.001)***	−2.925	(0.000)***	−4.131	(0.001)***
Large-family households (%)	1.088	(0.599)	−0.203	(0.924)	−3.505	(0.465)
Old householders (%)	0.045	(0.974)	−0.140	(0.919)	−2.024	(0.308)
Nonwhite households (%)	0.734	(0.172)	0.582	(0.304)	1.516	(0.095)*
log(1 + Number of superettes)	−0.363	(0.000)***	−0.345	(0.000)***	−0.686	(0.000)***
log(1 + Number of supermarkets)	−0.135	(0.229)	−0.136	(0.211)	0.203	(0.234)
log(1 + Number of supercenters)	−0.289	(0.088)*	−0.214	(0.205)	−0.181	(0.632)
STATE DUMMIES	YES		YES		YES	
Number of adoptions	958		917		235	
Sample size	2,017		2,017		1,807	
McFadden R ²	0.315		0.306		0.351	
(2) Cereal						
Store brand presence	0.695	(0.300)	1.539	(0.019)**	−0.922	(0.197)
Retailer market share	−3.777	(0.129)	−5.251	(0.182)	−3.654	(0.175)
Retailer market share ²	9.046	(0.107)	8.916	(0.214)	10.880	(0.067)*
Store size	0.051	(0.000)***	−0.009	(0.241)	0.073	(0.000)***
High-income households (%)	6.140	(0.000)***	0.932	(0.317)	3.993	(0.000)***
Large-family households (%)	−8.497	(0.000)***	−5.909	(0.008)***	−4.207	(0.023)**
Old householders (%)	2.265	(0.132)	2.915	(0.043)**	−0.080	(0.948)
Nonwhite households (%)	1.371	(0.028)**	2.458	(0.000)***	−0.156	(0.808)
log(1 + Number of superettes)	0.124	(0.292)	−0.070	(0.490)	0.096	(0.350)
log(1 + Number of supermarkets)	0.013	(0.909)	−0.426	(0.004)***	0.296	(0.012)**
log(1 + Number of supercenters)	0.143	(0.532)	0.062	(0.783)	0.137	(0.474)
STATE DUMMIES	YES		YES		YES	
Number of adoptions	1,610		550		1,359	
Sample size	1,959		1,959		1,959	
McFadden R ²	0.218		0.310		0.280	
(3) Toothpaste						
Store brand presence	0.336	(0.135)	−1.174	(0.003)***	0.593	(0.021)**
Retailer market share	−14.164	(0.000)***	−12.315	(0.000)***	−9.091	(0.002)***
Retailer market share ²	23.306	(0.000)***	17.915	(0.007)***	16.747	(0.003)***
Store size	0.063	(0.000)***	0.023	(0.044)**	0.069	(0.000)***
High-income households (%)	2.103	(0.001)***	−2.891	(0.006)***	2.997	(0.000)***
Large-family households (%)	−1.444	(0.408)	9.575	(0.000)***	−6.631	(0.001)***
Old householders (%)	2.939	(0.003)***	2.721	(0.211)	2.301	(0.040)**
Nonwhite households (%)	0.861	(0.096)*	−0.723	(0.405)	1.101	(0.044)**
log(1 + Number of superettes)	−0.051	(0.584)	0.282	(0.102)	−0.144	(0.123)
log(1 + Number of supermarkets)	0.021	(0.854)	−0.613	(0.000)***	0.173	(0.196)
log(1 + Number of supercenters)	−0.068	(0.719)	0.543	(0.019)**	−0.209	(0.265)
STATE DUMMIES	YES		YES		YES	
Number of adoptions	818		202		685	
Sample size	1,759		1,489		1,759	
McFadden R ²	0.188		0.253		0.230	

*Significant at the 0.10 level; **significant at the 0.05 level; ***significant at the 0.01 level.

as regional brands that have below-median prices. In contrast, “premium regional brands” are defined as regional brands that have above-median prices.

To analyze the store’s decision to distribute a regional brand, we estimate the following model:

$$\Pr(y_i = 1 | x_i) = \frac{\exp(\beta'x_i)}{1 + \exp(\beta'x_i)}. \quad (4)$$

In this model, y_i is a binary indicator of whether store i distributes *any* regional brand. The x_i are

independent variables, describing store characteristics, assortment characteristics, demographics, and local competition for each store i (details are listed in Table 5).²⁴ For this analysis, we only use stores located in states where there is at least one regional brand available for distribution (which is why the sample

²⁴ Retailer market share reflects the market share of a particular retailer in a particular state. Thus, this gives a measure of the retailer’s presence in the local market.

sizes reported in Table 5 vary across the different categories).

6.2. Results

Table 5 presents the logistic regression results for three product categories: cola, ready-to-eat cereal, and toothpaste. We omit the ground coffee category because regional brands are distributed by all the sample stores.²⁵ The standard errors reported in Table 5 reflect robust clustered standard errors.²⁶

Model (1) is a base specification, where the dependent variable takes a value of 1 if the store distributes any regional brand. In model (2), the dependent variable takes a value of 1 if the store distributes any value regional brand (i.e., low-price tier regional brand). In contrast, in model (3), the dependent variable takes a value of 1 if the store distributes any premium regional brand (i.e., high-price tier regional brand). State dummies are included in all models but are suppressed to avoid cluttered tables.

6.2.1. Average Adoption Rates. As a first step, we examine the average percentage of stores that distribute a regional brand. This is done by comparing the number of adoptions listed in Table 5 with the sample size. We see that the average regional brand adoption varies widely, from 47% in the cola and toothpaste categories to 82% in the cereal categories.²⁷ In addition, whether this adoption is occurring among low-price value regional brands or high-price premium regional brands also varies significantly across categories. In the cola categories, 45% of stores adopt a value regional brand, whereas only 13% of stores adopt a premium regional brand. In contrast, most of the regional brand adoption in the cereal and toothpaste categories is premium regional brand adoption. Sixty-nine percent of stores adopt a premium regional cereal, and 39% of stores adopt a premium regional toothpaste. These results suggest that it is possible for quality regional products to get access to shelves in a significant percentage of stores in many categories.

6.2.2. Retailer Market Share. We find that small-scale retailers such as regional chains and independent stores have the highest probability of distributing regional brands in the cola and toothpaste categories. We find that a model with a squared term for retailer market share fits the data better than a specification with only a linear term. These results continue to hold when analyzing only value regional brands or only premium regional brands. The results are similar in the cereal category, albeit not statistically significant.

6.2.3. Store Brand Presence. We find that supermarkets with store brands are less likely to distribute low-price value regional brands in the cola and toothpaste categories. In contrast, store brand presence has a much smaller impact on the distribution of high-price premium regional brands, and it even has a positive association with premium regional toothpaste distribution. In the cereal category, the only significant result is that retailers with store brands are more likely to also stock value regional brands. The results in this category may reflect the greater heterogeneity of flavors and cereal types in the category. Thus, it is possible for the store to have a store brand for some types of cereals but to use other value regional brands to act as a proxy for a store brand in other cereal types.

The combination of the effects of store brand presence and retailer market share suggests that value regional brands are used as an alternative to store brands—especially by small regional chains or independent stores that do not have enough scale to have their own store brands.

7. Discussion and Conclusions

Despite its importance as a profit driver both for retailers and manufacturers, much is still unknown and undocumented about how retail assortments vary across stores. In this context, the present study presents a descriptive analysis of assortment similarities across supermarkets. Our main result is that store ownership and other characteristics of retail stores are an order of magnitude more powerful at explaining variation in assortment composition than either local clientele effects or store competition.

Our results are consistent with a market where the costs of varying assortment are larger than the benefits from demand-side customization. Our finding that the effect of having the same corporate owner is greater than the effect of having the same chain banner reinforces this interpretation, because firms could change their assortments across their different banners without disrupting any value that would be associated with having consistent product offerings within a chain. Most of these results generalize to the analysis of assortment size, although different banners under common ownership sometimes have different assortment sizes.

Whereas stores from the same chain and state offer highly similar assortments, we also find that supermarket chains customize their assortments across states. Because the minimal efficient scale of supermarket distribution networks is approximately equal to the surface area of a state (see Ellickson 2007), it is possible that the cost savings from assortment standardization diminishes across states. At the same

²⁵ This likely reflects the high presence of local favorites in the coffee market.

²⁶ The clusters are defined as chain banner–state combinations.

²⁷ In the cereal category, some of regional brands are imported brands.

time, existing research demonstrates that the preference differences for brands across states can be large even for close substitutes (see Bronnenberg et al. 2007). Thus, our findings of scarce within-state assortment versioning and ample across-state assortment differences can potentially be explained from basic cost-benefit principles. However, we also acknowledge that there can be potential alternative explanations such as easier strategic—perhaps tacitly collusive, even—coordination across chains when assortments are selected by regional divisions.²⁸

Our empirical analysis of regional brand distribution suggests that there are two different roles played by regional brands: store differentiation with premium regional brands that may reflect local favorites, and the use of value regional brands as a substitute for store brands. High-price premium regional brands are distributed by small-scale retailers in the cola and toothpaste categories regardless of store brand presence, which suggests that premium regional brands may be used as a differentiation device. In the ground coffee category, some regional brands are local favorites; consequently, they are distributed widely in a regional market, including larger supermarket chains. On the other hand, in the cola and toothpaste categories, low-price value regional brands are used as an alternative to store brands—especially by small regional chains or independent stores, which do not have enough scale to have their own store brands. Consistent with this dichotomy, many regional brand manufacturers actually manufacture both regional brands and store brands. For instance, Personal Care and United Exchange Corp. in the toothpaste category have their own regional brands, but they also provide store brands to large chains, and each of their websites position the regional brands as cost-effective solutions for small businesses that want their own brands without large expenses.

Our results lead to several managerial and academic insights. First, our finding that assortment is largely set at the regional division (defined by state-chain combination) level suggests that there is a fair level of discreteness in the benefit that a manufacturer gains when they convince a large chain to sell their product. For example, the three largest corporate owners—Kroger, SUPERVALU, and Safeway—account for 31% of overall supermarket sales. Our

study suggests that manufacturers should spend a significant amount of their sales effort towards securing the adoption of their products by the large chains. However, we find that low-end brand manufacturers should focus their distribution efforts to chains without store brands.

Second, our results suggest that models of account-level adoption are perhaps more relevant to this industry than models of store-level adoption. The account-level benefits of assortment similarity across stores involve reduction in menu costs, economies of scale, and economies of scope in procurement and operation.

Furthermore, our results refine the meaning of assortment over extant empirical work. For instance, Misra (2008) assumes that products are strictly preference-ordered so that the question of assortment offerings is dual to the question of assortment size. Instead, in all four categories studied, we find that different products are observed in similarly sized assortments, even holding demographics constant. This emphasizes that assortment size and assortment composition are different concepts.

Last, we find that local competition is not a strong driver of assortment composition. This suggests that one can model assortment selection decisions based on demand and supply primitives pertaining to the chain itself, and one can perhaps view competition as a secondary effect. In that same vein, the direction of the small impact of distance we do measure varies category by category. Thus, the conventional wisdom built upon the Stassen et al. (1999) finding that supermarkets located close together tend to differentiate their consumer packaged goods assortment, which was measured across several categories, should be modified to reflect that this effect varies category by category.

Acknowledgments

The authors are grateful to Andrew Ainslie, Robert Zeithammer, Jun Kim, and the editorial team for their insightful comments. They are also grateful for comments received from seminar participants at McGill University, Montreal, Canada; the University of Washington, Seattle; National University of Singapore, Singapore; Singapore Management University, Singapore; and the AMA doctoral consortium. The authors thank the ACNielsen company for the use of their data.

Appendix

QAP Details

An OLS regression with dyadic data set produces consistent coefficient estimates, but OLS standard errors can be too small because the pairwise observations are not independent. This can lead to too-optimistic *p*-values in the assessment of statistical significance of coefficients. Therefore, we use the QAP to assess the statistical significance of regression coefficients. Because we are using a symmetric dyadic

²⁸ The presence of slotting allowances could also be interpreted as contributing to assortment similarity across stores. However, the institutional practice of negotiating slotting allowances at the chain level over the store level is an outcome of the profitability of customizing the assortment at the outlet level. For example, movie contracts are negotiated with exhibition companies in ways that lead to different movies being shown at different theaters or even different markets, reflecting underlying profit maximization (Corts 2001, Gil 2007).

data set, our pairwise dependent and independent variables each form the upper triangle of a matrix where the rows and columns are stores. The QAP method permutes the dependent variable only (i.e., Y permutation) and randomly reassigns the permuted dependent variable to a different observation. The permutation of the dependent variable is such that we permute rows and columns of a dyadic data matrix, e.g., the assortment similarity measures, so as to preserve the network autocorrelation. This is accomplished by swapping for pairs of stores, columns, and corresponding rows of the dyadic matrix. Given proper randomization, the resulting permuted matrix corresponds to the null hypothesis of no statistical association between permuted dependent and nonpermuted independent variables, yet it preserves any row and column dependence of both permuted dependent and nonpermuted independent variables. The main idea of the QAP is to obtain an empirical coefficient distribution under the null hypothesis of no statistical association while allowing for spurious regression effects as a result of network autocorrelation.

The OLS regression is run on the new data set of permuted dependent variables and nonpermuted independent variables, and new coefficient estimates are generated. These new coefficient estimates will be values from the empirical sampling distribution under the null hypothesis, but the sampling distribution correctly takes into account the correlation among observations. Repeating this process many (say, 1,000) times generates an empirical sampling distribution of the coefficient estimates under the null hypothesis. Statistical significance of regression coefficients can be assessed by comparing the actual coefficient estimate with the empirical null distribution. If the actual coefficient falls in the upper or lower 2.5% of the empirical distribution, this coefficient is significant at the 5% level. This statistical test can be thought as a pseudo (two-sided) t -test. Because the computation of the probability value is based on the comparison of simulated values with an actual value, the test is nonparametric. Therefore, the QAP is robust against potential misspecification of error structure. Ideally, one can use all the possible permutations of the original dependent variable matrix, but the number of such permutations is very large ($N!$). Therefore, a random sample from the population of $N!$ permutations is used instead.

Because we restrict pairwise comparisons within a same geographic market in our data and pool the data across different geographic markets, we implement the QAP for multiple groups where permutations of the dependent variable are limited only within a group. This can be done by assigning a group indicator for the data set and limiting permutations only within a group. A STATA subroutine developed by Simpson (2001) is utilized to implement the QAP.

References

- Bayus, B. L., W. P. Putsis Jr. 1999. Product proliferation: An empirical analysis of product line determinants and market outcomes. *Marketing Sci.* **18**(2) 137–153.
- Besanko, D., J.-P. Dubé, S. Gupta. 2005. Own-brand and cross-brand retail pass-through. *Marketing Sci.* **24**(1) 123–137.
- Boatwright, P., S. Dhar, P. E. Rossi. 2004. The role of retail competition, demographics, and account retail strategy as drivers of promotional sensitivity. *Quant. Marketing Econom.* **2**(2) 169–190.
- Borzekowski, R., R. C. Thomadsen, C. Taragin. 2009. Competition and price discrimination in the market for mailing lists. *Quant. Marketing Econom.* **7**(2) 147–179.
- Briesch, R. A., P. K. Chintagunta, E. J. Fox. 2009. How does assortment affect grocery store choice? *J. Marketing Res.* **46**(2) 176–189.
- Bronnenberg, B. J., S. K. Dhar, J.-P. Dubé. 2007. Consumer packaged goods in the United States: National brands, local branding. *J. Marketing Res.* **44**(1) 4–13.
- Cameron, A. C., J. B. Gelbach, D. L. Miller. 2008. Robust inference with multi-way clustering. *J. Bus. Econom. Statist.* Forthcoming.
- Chisholm, D. C., M. S. McMillan, G. Norman. 2006. Product differentiation and film programming choice: Do first-run movie theaters show the same films? NBER Working Paper W12646, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=940595.
- Corts, K. S. 2001. The strategic effects of vertical market structure: Common agency and divisionalization in the U.S. motion picture industry. *J. Econom. Management Strategy* **10**(4) 509–528.
- d'Aspremont, C., J. J. Gabszewicz, J.-F. Thisse. 1979. On Hotelling's "Stability in competition." *Econometrica* **47**(5) 1145–1150.
- Dhar, S. K., S. J. Hoch. 1997. Why store brand penetration varies by retailer. *Marketing Sci.* **16**(3) 208–227.
- Dhar, S. K., J. S. Raju, U. Subramanian, Y. Wang. 2009. The competitive consequences of using a category captain. Mimeo, University of Chicago, University of Pennsylvania; and Fudan University, Shanghai, China.
- Draganska, M., D. C. Jain. 2005. Product-line length as a competitive tool. *J. Econom. Management Strategy* **14**(1) 1–28.
- Draganska, M., M. Mazzeo, K. Seim. 2009. Beyond plain vanilla: Modeling joint product assortment and pricing decisions. *Quant. Marketing Econom.* **7**(2) 105–146.
- Drèze, X., S. J. Hoch, M. E. Purk. 1994. Shelf management and space elasticity. *J. Retailing* **70**(4) 301–326.
- Dubé, J.-P. 2004. Multiple discreteness and product differentiation: Demand for carbonated soft drinks. *Marketing Sci.* **23**(1) 66–81.
- Ellickson, P. B. 2007. Does Sutton apply to supermarkets? *RAND J. Econom.* **38**(1) 43–59.
- Fischer, J. H., J. E. Harrington Jr. 1996. Product variety and firm agglomeration. *RAND J. Econom.* **27**(2) 281–309.
- Fox, E. J., R. Sethuraman. 2006. Retail competition. M. Krafft, M. K. Mantrala, eds. *Retailing in the 21st Century: Current and Future Trends*. Springer, Berlin, 239–256.
- Gil, R. 2007. "Make or buy" in movies: Integration and ex-post renegotiation. *Internat. J. Indust. Organ.* **25**(4) 643–655.
- Grewal, D., M. Levy, A. Mehrotra, A. Sharma. 1999. Planning merchandising decisions to account for regional and product assortment differences. *J. Retailing* **75**(3) 405–424.
- Gulati, R., M. Gargiulo. 1999. Where do interorganizational networks come from? *Amer. J. Sociol.* **104**(5) 1439–1493.
- Hoch, S. J., E. T. Bradlow, B. Wansink. 1999. The variety of an assortment. *Marketing Sci.* **18**(4) 527–546.
- Hoch, S. J., B.-D. Kim, A. L. Montgomery, P. E. Rossi. 1995. Determinants of store-level price elasticity. *J. Marketing Res.* **32**(1) 17–29.
- Hotelling, H. 1929. Stability in competition. *Econom. J.* **39**(153) 41–57.
- Inman, J. J., V. Shankar, R. Ferraro. 2004. The roles of channel-category associations and geodemographics in channel patronage. *J. Marketing* **68**(2) 51–71.
- Irmen, A., J.-F. Thisse. 1998. Competition in multi-characteristics spaces: Hotelling was almost right. *J. Econom. Theory* **78**(1) 76–102.
- Iyer, G., P. B. Seetharaman. 2008. Too close to be similar: Product and price competition in retail gasoline markets. *Quant. Marketing Econom.* **6**(3) 205–234.

- Jaffe, A. B. 1986. Technological opportunity and spillovers of R&D: Evidence from firms' patents, profits, and market value. *Amer. Econom. Rev.* **76**(5) 984–1001.
- Johnson, J. P., D. P. Myatt. 2003. Multiproduct quality competition: Fighting brands and product line pruning. *Amer. Econom. Rev.* **93**(3) 748–774.
- Kadiyali, V., N. Vilcassim, P. Chintagunta. 1999. Product line extensions and competitive market interactions: An empirical analysis. *J. Econometrics* **89**(1–2) 339–363.
- Kaufmann, P. J., S. Eroglu. 1999. Standardization and adaptation in business format franchising. *J. Bus. Venturing* **14**(1) 69–85.
- Khanna, T., J. Kogan, K. Palepu. 2006. Globalization and similarities in corporate governance: A cross-country analysis. *Rev. Econom. Statist.* **88**(1) 69–90.
- Kök, A. G., M. L. Fisher. 2007. Demand estimation and assortment optimization under substitution-methodology and application. *Oper. Res.* **55**(6) 1001–1021.
- Kök, A. G., M. L. Fisher, R. Vaidyanathan. 2009. Assortment planning: Review of literature and industry practice. *Retail Supply Chain Management*. Springer, New York, 1–55.
- Krackhardt, D. 1988. Predicting with networks: Nonparametric multiple regression analysis of dyadic data. *Soc. Networks* **10**(4) 359–381.
- Kroger Co. 2007. The 2007 Kroger fact book. Kroger, Cincinnati. http://www.thekrogerco.com/finance/documents/2007_KrogerFactBook.pdf.
- Levy, M., B. A. Weitz. 2008. *Retailing Management*. McGraw-Hill, New York.
- Mantrala, M. K., M. Levy, B. E. Kahn, E. J. Fox, P. Gaidarev, B. Dankworth, D. Shah. 2009. Why is assortment planning so difficult for retailers? A framework and research agenda. *J. Retailing* **85**(1) 71–83.
- McAlister, L. 2007. Comment—Cross-brand pass-through: Fact or artifact? *Marketing Sci.* **26**(6) 876–898.
- Misra, K. 2008. Understanding retail assortments in competitive markets. Mimeo, Northwestern University, Evanston, IL.
- Mizruchi, M. S. 1989. Similarity of political behavior among large American corporations. *Amer. J. Sociol.* **95**(2) 401–424.
- Neven, D., J.-F. Thisse. 1990. On quality and variety competition. J. J. Gabszewicz, J.-F. Richard, L. Wolsey, eds. *Economic Decision Making: Games, Econometrics, and Optimization*. Elsevier, North-Holland, Amsterdam, 175–199.
- Nevo, A. 2001. Measuring market power in the ready-to-eat cereal industry. *Econometrica* **69**(2) 307–342.
- Putsis, W. P., B. L. Bayus. 2001. An empirical analysis of firm's product line decisions. *J. Marketing Res.* **38**(1) 110–118.
- Richards, T. J., S. F. Hamilton. 2006. Rivalry in price and variety among supermarket retailers. *Amer. J. Agricultural Econom.* **88**(3) 710–726.
- Seim, K., V. B. Viard. 2004. The effect of entry and market structure on cellular pricing tactics. NET Institute Working Paper 03-13, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=618221.
- Simpson, W. B. 2001. QAP—The quadratic assignment procedure. Presentation, First North American STATA Users' Group Meeting, March 12, StataCorp, College Station, TX.
- Stassen, R. E., J. D. Mittelstaedt, R. A. Mittelstaedt. 1999. Assortment overlap: Its effect on shopping patterns in a retail market when the distributions of prices and goods are known. *J. Retailing* **75**(3) 371–386.
- Sweeting, A. 2006. Too much rock and roll? Station ownership, programming and listenership in the music radio industry. Mimeo, Northwestern University, Evanston, IL.
- Sweeting, A. 2010. The effects of mergers on product positioning: Evidence from the music radio industry. *RAND J. Econom.* **41**(2) 372–397.
- Watson, R. 2009. Product variety and competition in the retail market for eyeglasses. *J. Indust. Econom.* **57**(2) 217–251.
- Wooldridge, J. M. 2003. Cluster-sample methods in applied econometrics. *Amer. Econom. Rev.* **93**(2) 133–138.