NORTHWESTERN UNIVERSITY

Differentiated Products Competition in Supermarket Product Categories

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Economics

By

Jean-Pierre Dubé

EVANSTON, ILLINOIS

June 2000

UMI Number: 9974278

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

ABSTRACT

Differentiated Products Competition in Supermarket Product Categories

Jean-Pierre Dubé

The increasing availability of supermarket scanner data permits much richer studies of consumer demand for supermarket products. Most of the product categories consist of a large number of differentiated goods. Using a comprehensive supermarket data set, I explore several alternative methodologies for modeling consumer demand in the presence of differentiated products. I provide three applications in which consumer demand provides a basis for understanding and developing optimal firm strategy and competition policy.

In the first chapter, I develop and apply the popular discrete choice model to characterize consumer segmentation and to develop retailer strategy that exploits the knowledge of these segments. This model applies to any product category in which consumers systematically purchase a single unit of a single product alternative on any given trip. In the second chapter, I present an alternative model of demand that allows for multiple-unit shopping. I develop the importance of accounting for multiple-item purchases in terms of the predictions for substitution patterns and managerial strategy. Finally, in the third chapter, I use the multiple-unit purchase model to investigate the consequences of

mergers in the carbonated soft drink industry. In this final section, I compare my results to those obtained from using the discrete choice model.

DEDICATION

I would like to dedicate the following work to my family, whose continued support has been essential for the successful completion of my doctoral research. Most importantly, I would like to dedicate my thesis to my wife, Teresa, who has shown tremendous patience and kindness, even during my most stressful moments these past two years.

ACKNOWLEDGMENTS

I am very grateful to Tim Conley, David Besanko, Sachin Gupta and Rob Porter for their comments and for numerous useful discussions. I would also like to thank Arie Beresteanu, Michaela Draganska, Chris Jepsen, Paulo Santiago and John Woodruff for comments and suggestions. Finally, financial support from a Northwestern University Dissertation Grant and from the Center for Studies in Industrial Organization are gratefully acknkowledged.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1.
INTRODUCTION

The increasing availability of supermarket-level scanner data permits richer studies of consumer demand. In the current context, I demonstrate how consumer demand provides a basis for understanding and developing optimal firm strategy and competition policy. From the strategy perspective, studying demand provides important insight into effective price discrimination and product positioning. From the policy perspective, modeling the demand and supply environment permits studying mergers and their consequences on prices and welfare.

The basic methodology only requires estimating consumer demand. Combining demand with a theoretical model of firms' strategic pricing behavior yields the insights into strategy and competition. This methodology, the *structural approach*, also provides behavioral interpretations of the estimated parameters.

Recent specifications of demand account for increasingly sophisticated patterns of consumer taste heterogeneity. This heterogeneity complicates the ability to aggregate individual shopping behavior. A simple solution to this aggregation problem is to assume discrete choice behavior: consumers choose a single unit of a single product alternative in a given market. Aggregating demand reduces to integrating over the set of consumers that choose a given product, producing a smooth total demand function. These smooth aggregate discrete choice models (DCM) are particularly convenient in that they can be applied to aggregate data. In several instances, the discrete choice assumption can be validated empirically. For instance, demand for products such as Ketchup generally exhibit single-unit purchases at the consumer shopping-trip level. A growing literature in both industrial organization and marketing has developed increasingly-sophisticated models of aggregate demand and pricing based on the discrete choice assumption.

Unfortunately, many product categories do not exhibit discrete choice behavior. Categories such as carbonated soft drinks, canned soups and packaged cookies exhibit a high incidence of multiple-unit purchasing. The technical ease of the discrete choice models make them a tempting modeling approach, even when simple empirical observation clearly violates the underlying behavioral assumptions. However, the resulting specification error could lead to incorrect strategic and policy implications.

Using a unique, comprehensive supermarket database, generously provided by AC-Nielsen, I develop several applications with supermarket data. The data itself consist of both a panel of supermarkets in a single city-market, including 9 quarters of weekly sales and prices for over 15000 products, and a panel of households that shop in these stores during that same period. My analysis consists of three chapters. In the first chapter, I develop and apply the discrete choice model to study consumer segmentation and potential retailer strategy that exploits the knowledge of these segments. In the second chapter, I present an alternative model of demand that allows for multiple-unit shopping. I develop the importance of accounting for multiple-item purchases in terms of the predictions for substitution patterns and managerial strategy. Finally, in the third chapter, I use the multiple-unit purchase model to investigate the consequences of mergers in the carbonated soft drink industry. In this final section, I compare my results to those obtained from trying to fit an aggregate DCM to the store-level data.

The first section studies product categories for which the discrete choice assumption is empirically-validated. Based on joint research with Sachin Gupta and David Besanko, I develop a model of aggregate store-level demand that allows for heterogeneity in consumer preferences. In this study, we make explicit use of the heterogeneity structure, rather than incorporating it as a control. We combine the demand framework with a model of channel-pricing that allows for wholesalers who sell their product to the supermarkets. Before applying the model to actual data, we run a simulation exercise to demonstrate the ability of store-level data to identify heterogeneity. With the support of the simulations, we then apply the model to actual yogurt sales data. We use the model to predict store-level pricing and to decompose aggregate responses to prices and feature advertising across consumer types. Finally, we use the model to develop profit-enhancing strategic price discrimination strategies for retailers.

In the second section, I develop a model, based on Hendel (1999), that generalizes the DCM to allow for multiple-unit purchases. The complexity of the model does not permit a simple smooth aggregation, as in the DCM. Consequently, I use household-level purchase data to estimate demand. Aggregating the predicted household purchases, I demonstrate how ignoring quantity information, as in the DCM, leads to underpredicted demand. Ignoring quantity information also leads to underpredicted price and marketing-mix elasticities, leading to incorrect managerial implications regarding the soft drink category. I use the proposed model to investigate the implications of product lines in terms of producer market power and consumer valuation of variety.

In the final section, I apply the multiple-unit purchase model to study mergers in the soft drink industry. I compare my findings to the aggregate DCM. I find that the latter produces much lower measures of firms' market power. In the merger context, these low measures of market power translate into unrealistic merger predictions. I use these findings as support for the importance of modeling multiple-unit shopping behavior for industries like soft drinks.

CHAPTER 2.
Recovering Segment Heterogeneity From Aggregate Retail Data: An Equilibrium Approach

Introduction

It[1] is widely recognized in the marketing science literature that consumer heterogeneity plays a critical role in the modeling of brand choice behavior. Empirical studies assert that consumers differ strongly in their brand preferences and in their responsiveness to marketing-mix variables such as prices, in-store displays, and feature advertisements. Such heterogeneity has important theoretical, methodological, and substantive implications. For instance, homogeneous models of brand choice when consumers are heterogeneous are known to provide biased estimates of the market's responsiveness to price and promotional changes. It is also well known that homogeneous models place unrealistic restrictions on substitution patterns between brands. When firms use these estimates to make pricing decisions, they obtain sub-optimal profits. Furthermore, many marketing applications such as price discrimination, differentiated product offerings to market segments, and targeted communications and promotions programs require knowledge of the distribution of consumers' tastes (Allenby and Rossi 1999). Homogeneous models are obviously inadequate for these decisions.

---

[1] I wrote this chapter in collaboration with David Besanko and Sachin Gupta, of the Kellog Graduate School of Management.

In a separate stream of work, several recent studies in marketing have recognized the potential for price endogeneity to bias parameter estimates of logit brand choice models (see Table 2.1 for a summary). These studies adapt techniques developed in the Industrial Organization literature (Berry 1994; Berry, Levinsohn and Pakes 1995; and Nevo 2000) to account for the potential endogeneity of prices, mostly in weekly supermarket data. In several of these studies, one objective is to study the price-setting mechanism on the supply-side, in addition to determining demand. For example, retail prices are modeled as the equilibrium outcomes of a competitive game between multiple manufacturers. In some of these studies, the role of a strategic retailer is modeled as well (see the column titled Channel Structure in Table 2.1). In these marketing studies, interest centers on the manufacturers' and retailer's equilibrium mark-ups (Besanko, Gupta and Jain 1998), on alternative game-structures in the channel (Karunakaran 2000), and on understanding retail category management practices (Chintagunta 1999a). With the exception of Chintagunta (1999b), the dominant assumption in this literature is that demand is characterized by a homogeneous logit model. This assumption places unattractive restrictions on the manufacturers' and retailer's pricing behavior as well as on vertical channel relationships. For example, the vertical Nash model implies that equilibrium retail margins are equal for all brands in a category, an assumption not supported by empirical evidence (Blattberg and Neslin 1990). Similarly, the standard homogeneous logit model implies that, in a vertical Nash game, retailers pass between zero and 100 percent of any changes in wholesale prices through to the consumer. In fact, em-

Study	*Tastes*	*Level of Aggregation of Data*	*Channel Structure*
Besanko, Gupta and Jain (1998)	Homogeneous Logit	Store-level weekly brand shares	Vertical Nash
Villas-Boas and Winer (1999)	Homogeneous Probit	household panel	Not modeled
Chintagunta, Kadiyali, and Vilcassim (1999)	Homogeneous Logit	National weekly brand shares	Not modeled
Chintagunta (1999a)	Homogeneous Logit	Store-level weekly brand shares	Exogenous manufacturers
Karunakaran (2000)	Homogeneous Logit	Chain-level weekly brand shares	Vertical Nash vs. Stackelberg
Chintagunta (1999b)	Heterogeneous Logit (normal)	Chain-level weekly brand shares	Vertical Nash conduct parameters
This paper	Heterogeneous Logit (finite mixture)	Store-level weekly brand shares	Vertical Nash

Table 2.1. Marketing Literature Incorporating Logit Demand with Endogenous Prices

pirical evidence (Chevalier and Curhan 1976, Walters 1989, Armstrong 1991) suggests that retail pass-through rates exceed 100 percent for many manufacturer trade deals. For example, Armstrong (1991) analyzes 605 manufacturer trade promotions for a period of two years for a large supermarket chain and finds that average pass-through exceed 100 percent in all four categories studied.

In this paper, we develop a heterogeneous logit model of consumer demand jointly with a structural model of pricing by competing manufacturers and a retailer. Our model allows for flexible substitution patterns between brands and also results in more realistic outcomes with respect to manufacturer and retailer mark-ups and pass-through rates. Our approach requires the use of aggregate store-level data for estimation. Consumer heterogeneity is modeled as a finite number of latent segments. These two important aspects of our modeling approach merit further discussion.

First, we use aggregate store-level data to estimate our model. The tradition in the marketing literature has been to use household panel data to model unobserved consumer heterogeneity (e.g. Kamakura and Russell 1989, Chintagunta, Jain and Vilcassim 1991). While household-level data offer some advantages, we believe that there are several reasons why, in practice, store-level data might be attractive to use in econometric analyses of consumer demand. Some marketing researchers have questioned the representativeness of purchase behavior of panelist households. In a recent article that reviews commercial advances in the use of scanner data, Bucklin and Gupta (1999) note that "(practitioners) were quick to point out severe limitations of panel data analysis. Sampling problems are one reason for the reluctance (of practitioners) to rely on panel data analysis." Gupta *et. al* (1996) reports that brand choice price elasticities of panelist households are statistically different from those of non-panelist households, although the magnitude of the difference is small. Another reason is that sample sizes in panel data are small at the level of an individual market. Aside from their superior sampling quality, store data are more widely and easily available, especially to retailers. Finally, we believe that it is an interesting and unexplored academic question as to whether consumer heterogeneity can be recovered in the kinds of aggregate store-level data that are typically available to retailers.

Second, we use a discrete representation of consumer heterogeneity (Berry, Carnall and Spiller 1997). Specifically, consumers are assumed to belong to a finite number of latent classes that differ in terms of their brand preference and marketing-mix responsiveness parameters. The membership of consumers in latent classes is probabilistic and to be determined from the data. This approach has been termed a finite mixture or latent class approach in the literature (Wedel and Kamakura 1999). Such models have received considerable attention from practitioners as well, since the latent classes correspond closely with managers' notion of market segments. An alternative approach is to specify continuous parametric distributions of consumer heterogeneity (Hausman and Wise 1978, Gonul and Srinivasan 1993). Berry, Levinsohn and Pakes (1995) developed a methodology for estimating the parameters of a continuous random coefficients logit model with aggregate data while accounting for price endogeneity. Chintagunta (1999b) demonstrates the application of this technique to marketing problems. The advantage offered by continuous random effects models is that they tend to outperform finite mixture models in terms of fit to the data in estimation samples and predictions in holdout samples. From an applications standpoint however, often managers can only address a finite number of market segments when designing pricing or promotion strategies. In these situations, finite mixture models offer a managerially appealing and useful representation of the marketplace. Our approach also does not rely heavily on parametric assumptions to recover the heterogeneity distribution. Furthermore, when applied to aggregate data, latent class models are also more computationally tractable relative to continuous models of heterogeneity since we do not have to simulate high-dimensional integrals.

We demonstrate the usefulness of our model in several ways. We show how knowledge of the consumer segments allows retailers to devise price discrimination strategies that could increase profits above the levels obtained from simple uniform pricing. We consider a scenario in which the retailer is able to assign consumers to their respective segments using some information on their purchase history, and offer segment-specific discounts through Catalina-type coupons (third degree price discrimination). We also consider a scenario in which the retailer is not able to assign consumers to segments, but observes the segment structure. In this case, the retailer strategically combines feature ads (which have a positive marginal utility that varies across segments) and prices to target specific products to specific segments (similar to second degree price discrimination).

We organize the paper as follows. In section two, we develop our model of demand and the supply-side wholesale channel. In section three, we describe the econometric procedure used to estimate the model. In section four, we present demand estimation results using data on the yogurt market. In section five, we demonstrate how segment-specific prices and coordinated featuring and pricing strategies can improve the profitability of the retailer. In section six, we summarize and conclude with a discussion of potential extensions. In an appendix, we also describe a simulation experiment in which we use synthetic data to demonstrate the accuracy with which the model recovers the underlying segment structure.

Model

Utility and Demand

As stated in the introduction, the estimation procedure that we employ below uses aggregate store-level sales data within a given product category. Despite this use of aggregate data, we derive our econometric model from a theory of individuals maximizing their utilities. For simplicity, we assume that the underlying consumer behavior derives from a discrete choice framework (McFadden 1981). We assume the level of utility a consumer derives from a given product is a function of the product's underlying attributes, which may not be perfectly observed by the econometrician. Since we do not observe individual behavior, we aggregate the individual choices for each product to obtain a system of demand equations.

Formally, we assume that, on a given shopping trip, consumers select one of J products (with a typical product indexed by i or j) and T weeks (with a typical week indexed by t). For each week, there are three attributes (x, p, ξ) for each product. The vector x denotes brand-specific indicator variables (thus allowing for the brand-specific constants) as well as marketing mix variables (e.g., feature or display), and p denotes the shelf-price. We also allow for an unobserved (to the econometrician) attribute, ξ to account for other factors generating the choice process such as television advertising and coupon incidence (BLP 1995, Besanko, Gupta, and Jain 1998). In addition to the J

products, we also assume that consumers may select the no-purchase option, which we denote as product 0 and whose utility we normalize to zero.

For a shopping trip during week t, the conditional utility consumer h derives from purchasing product j is given by:

$$u_{hjt} = x_{jt}\beta_h - \alpha_h p_{jt} + \xi_{jt} + \omega_{hjt},$$
$$h = 1, ..., H, j = 1, ..., J, t = 1, ..., T.$$

The coefficients (β_h, α_h) capture consumer h's tastes for attributes, x, and price, p (i.e., the marginal utilities for the underlying attributes). The term ω_{hjt} is an i.i.d. mean-zero stochastic term capturing consumer h's idiosyncratic utility for alternative j during week t. As explained above, the term ξ_{jt} captures a product attribute that is observed by the consumer, but not by the econometrician. If this attribute is observed by producers, then this will influence the producers' pricing decisions, generating a simultaneity bias (BLP 1995, Besanko, Gupta, and Jain 1998). Examples of such variables are coupon availability, exposure to television advertising, and product availability at retail. These variables not only affect brand choices of consumers but may also influence price-setting by firms. For example, empirical evidence suggests a positive correlation between coupon availability and retail prices (Vilcassim and Wittink 1987, Levedahl 1986). Balachander and Farquahar (1994) show that competing firms may find it optimal to limit product availability in order to soften price competition, thereby supporting higher regular prices. The conditions in which this occurs are that the strategic effects of lower price competition outweigh the direct effect of lost sales due to reduced availability. Similarly, national advertising by manufacturers, which can enhance brand salience and image, is positively correlated with wholesale prices (Lal and Narasimhan 1996). We defer the discussion of our treatment of this endogeneity problem to the econometric section below.

Adding the assumption that ω_{hjt} has a type I extreme value distribution, the conditional probability q_{hjt} that consumer h chooses a particular product j in week t has the following form:

$$q_{hjt} = \frac{\exp(x_{jt}\beta_h - \alpha_h p_{jt} + \xi_{jt})}{1 + \sum_{i=1}^{J} \exp(x_{it}\beta_h - \alpha_h p_{it} + \xi_{it})},$$
$$h = 1, ..., H, j = 1, ..., J, t = 1, ..., T.$$

One of the primary difficulties with this specification is the incorporation of the consumer-specific tastes. Assuming homogeneous tastes, removing the subscript from the taste vectors (β_h, α_h), leads to the typical independence of irrelevant alternatives (IIA) problem. In particular, the implicit assumption that the utilities for each product are independent leads to very restrictive substitution patterns. The homogeneity assumption also restricts pricing behavior, which we discuss in the next section. We could use an error components structure (Cardell 1995) such as the nested logit (Besanko, Gupta, and Jain 1998) or the generalized extreme value (Bresnahan, Stern and Trajtenberg 1997). However, these approaches still exhibit the IIA problem within a nest. Alternatively, we could treat the parameters themselves as random variables. For a parametric continuous distribution, such as the normal, we could use the procedure outlined by BLP (1995). However, this approach generally requires a substantial amount of data in order to generate sufficient variation to identify the heterogeneity[2]. A convenient compromise is to use a discrete approximation to the parameter distribution, an aggregate analogue to the latent-class models used for household purchase data (Kamakura and Russell 1989). This is the approach that we take in this paper.

The discrete approximation to modeling consumer heterogeneity has two important advantages for researchers in marketing. First, from a practical point of view, its implementation is not as computationally-intensive as the BLP methodology. As we show below, the model has a simple analytic form, eliminating the need to evaluate a high-dimensional integral as with continuous random coefficient distributions. Second, from a conceptual point of view, the notion that a market consists of a relatively small number of discrete segments fits very well with how marketing practitioners view real-world markets. Indeed, for many smaller retail categories with fairly homogeneous products, we expect that the predominant form of heterogeneity will be in the form of price-sensitive consumers seeking value and relatively price-inelastic consumers seeking quality or responding to retail marketing initiatives such as displays or feature ads. We assume that consumers belong to one of K segments, where each segment k is characterized by its own taste vector, (α^k, β^k).

In particular, we assume that segment membership follows a logit distribution, where we denote the size of segment k as λ^k.[3] Since we do not observe within-segment

---

[2]BLP suggest using data for several markets. For instance, Nevo (1998) uses data for 68 city-markets.

[3]We model the probability of belonging to segment k as $\lambda^k = \frac{exp(\gamma^k)}{1+\sum_{k=1}^{K-1} exp(\gamma^k)}$, where we normalize

activity, we estimate expected behavior. Therefore, the expected probability q_{hjt} that consumer h purchases product j on a trip during week t has the following form:

$$q_{hjt} = \sum_{k=1}^{K} \lambda^k q_{jt}^k$$

where q_{jt}^k is the probability for segment k, $k = 1, ..., K$. Aggregating these expected probabilities, the model predicts the following expected market shares S_{jt} for product j in a given week:

$$S_{jt} = \sum_{k=1}^{K} \lambda^k S_{jt}^k \quad (2.1)$$

$$= \sum_{k=1}^{K} \lambda^k \frac{\exp(x_{jt}\beta^k - \alpha^k p_{jt} + \xi_{jt})}{1 + \sum_{i=1}^{J} \exp(x_{it}\beta^k - \alpha^k p_{it} + \xi_{it})},$$

$$j = 1, ..., J, t = 1, ..., T \quad (2.2)$$

In the econometric section, we show how we use (2.1) to estimate the model parameters.

The Channel Structure

We postulate a vertical channel model similar to the one in Besanko, Gupta, and Jain (1998). Assuming non-cooperative behavior, the product shelf prices exhibit a double-marginalization including both a wholesale margin and a retail margin. Our explicit modeling of the vertical channel contrasts with most of the IO studies cited above which tend to focus solely on manufacturers, treating retailers as exogenous. Following most of the empirical marketing literature accounting for the channel structure, we assume

---

the probability of belonging to segment K as $1 - \sum_{k=1}^{K-1} \lambda^k$. The logit assumption is merely a normalization to ensure that the estimated segment probabilities lie in the interval $(0, 1)$.

that manufacturers only set wholesale prices (Besanko, Gupta, Jain 1998, Karunakaran 2000, Chintagunta 1999b), ignoring the potential for more complicated contracts.[4]

In our model, oligopolistic manufacturers set wholesale prices and sell through a monopoly retailer. The key elements of the model are as follows:

1 The retailer acts as a monopolist in its local area. This assumption is broadly consistent with retailer conventional wisdom that most consumers shop at the same store week after week, often the one closest to their home or workplace (Slade, 1995). Besanko, Gupta, and Jain (1998) provide further support for this assumption.

2 There are K segments, and the retail chain cannot price discriminate across segments. Let M be the total market size .

3 There are J brands each one indexed by j or i. There is also a $J+1$ dummy brand, which represents the no-purchase alternative.

4 Consumers act as utility-maximizing price-takers, as described in Section 2. The game between manufacturers and the retail chain unfolds as follows:

The manufacturers and the retailer move simultaneously.

Manufacturers take retail margins as given, and choose wholesale prices w to maximize their

---

[4] Shaffer and Zettelmeyer(1999) demonstrate that when manufacturers and retailers are allowed to bargain over a two-part tariff, then the equilibrium prices are such that wholesalers price at cost (profiting solely from a fixed fee) and the retailer sets monopoly prices. Although we do not consider two-part tariffs or the potential for bargaining, our pricing model can be thought of as a more general framework where the bargaining outcome is a special case in which the wholesale margin is restricted to be zero.

profits.

The retailer takes wholesale prices as given, and chooses a retail price p to maximize its overall profits.

5In our data, each manufacturer offers a single brand. Therefore, we specify a model in which manufacturers offer just a single brand. This implies that there are J manufacturers.

To derive the equilibrium conditions, note that $S_j^k \lambda^k M$ as the demand for brand j by segment k in a given store-week (we drop the time subscript for convenience), where S_j^k is brand j's (unconditional) expected market share:

$$S_j^k = \frac{\exp(x_j\beta^k - \alpha^k p_j + \xi_j)}{1 + \sum_{i=1}^J \exp(x_i\beta^k - \alpha^k p_i + \xi_i)}.$$

Let $\mu_j = p_j - w_j$ the retail margin on brand j, where w_j is the wholesale price of brand j. The manufacturer takes this as given. Manufacturer j's derived demand is thus:

$$\begin{aligned} X_j &= \sum_{k=1}^K S_j^k \lambda^k M \\ &= \sum_{k=1}^K \frac{\exp(x_j\beta^k - \alpha^k \mu_j - \alpha^k w_j + \xi_j)}{1 + \sum_{i=1}^J \exp(x_i\beta^k - \alpha^k \mu_i - \alpha^k w_i + \xi_i)} \lambda^k M \end{aligned}$$

The manufacturer's total profit is thus:

$$\pi_j = (w_j - mc_j)X_j,$$

where mc_j is the manufacturer's marginal cost. Taking the wholesale prices of other manufacturers and the retail margins of all retailers as given, manufacturer j's profit-maximization condition is:

$$\frac{\partial \pi_j}{\partial w_j} = 0 \Rightarrow (w_j - mc_j)\frac{\partial X_j}{\partial w_j} + X_j = 0.$$

Noting that

$$\frac{\partial S_j^k}{\partial p_j} = -\alpha^k S_j^k(1 - S_j^k),$$

this profit-maximization condition can be shown to imply:

$$w_j = mc_j + \frac{\sum_{k=1}^K S_j^k \lambda^k M}{\sum_{k=1}^K \alpha^k S_j^k(1 - S_j^k)\lambda^k M}. \tag{2.3}$$

There are J such conditions, one for each manufacturer.

The retailer takes the wholesale prices as given, and acts a monopolist in pricing the whole category. The retailer's problem is thus:

$$\max_{p_1,\ldots,p_J} \sum_{j=1}^J \left[(p_j - w_j) \sum_{k=1}^K S_j^k \lambda^k M\right].$$

The first-order condition for brand j is:

$$\begin{aligned} & (p_1 - w_1) \sum_{k=1}^K \frac{\partial S_1^k}{\partial p_j} \lambda^k M \\ & + \ldots + (p_J - w_J) \sum_{k=1}^K \frac{\partial S_J^k}{\partial p_j} \lambda^k M \\ & + \sum_{k=1}^K S_j^k \lambda^k M \\ = {} & 0. \end{aligned}$$

Noting that $\frac{\partial S_i^k}{\partial P_j} = \alpha^k S_i^k S_j^k$ for $j \neq i$, the system of first-order conditions for brands $1, \ldots, J$ can be written in matrix form as:

$$\mathbf{\Omega}(\mathbf{p} - \mathbf{w}) + \mathbf{v} = \mathbf{0}, \tag{2.4}$$

where:

$$\mathbf{p} - \mathbf{w} \equiv \begin{bmatrix} p_1 - w_1 \\ \vdots \\ p_J - w_J \end{bmatrix}_{J \times 1}$$

$$\Omega \equiv \begin{bmatrix} -\sum_{k=1}^{K} \alpha^k S_1^k (1 - S_1^k) \lambda^k M & \cdots & \sum_{k=1}^{K} \alpha^k S_1^k S_J^k \lambda^k M \\ \vdots & \ddots & \vdots \\ \sum_{k=1}^{K} \alpha^k S_J^k S_1^k \lambda^k M & \cdots & -\sum_{k=1}^{K} \alpha^k S_J^k (1 - S_J^k) \lambda^k M \end{bmatrix}_{J \times J}$$

$$\mathbf{v} = \begin{bmatrix} \sum_{k=1}^{K} S_1^k \lambda^k M \\ \vdots \\ \sum_{k=1}^{K} S_J^k \lambda^k M \end{bmatrix}_{J \times 1}.$$

As we mentioned in the demand section, the assumption of homogeneous tastes leads to restrictive pricing behavior by retailers. When consumers are homogeneous, the equilibrium retail prices in (2.4) become:

$$\begin{aligned} p_j &= w_j + \frac{1}{\alpha S_0} \\ &= \left(mc_j + \frac{1}{\alpha (1 - S_j)} \right) + \frac{1}{\alpha S_0}, \end{aligned}$$

where $S_0 = 1 - \sum_{j=1}^{J} S_j$ is the share of the no-purchase alternative. When consumers are homogeneous, the amount by which a retailer sets its mark-ups over the wholesale prices is the same for all the products carried. Moreover, the product with the highest market share will also exhibit the highest total mark-up of shelf-price over marginal cost. Chintagunta(1999a) also discusses the implications of homogeneous preferences for the optimal retail margins. He specifies the indirect consumer utility in terms of the natural logarithm of prices, rather than the shelf-prices themselves, using product-specific price response parameters. While this treatment of prices allows for more flexible margins, it does not alleviate the restrictive substitution patterns on the demand side. For instance, the implied cross-elasticities of all the products with respect to product j depend only on the share of product j, hence they are equal. Our use of heterogeneous preferences, as in (2.4) , eliminates both the restrictive margins and cross-elasticities without requiring an *ad hoc* transformation of the prices.

Let's now summarize the full vertical equilibrium. The retailer's first-order conditions constitute J equations, while the manufacturer's first-order conditions constitute J equations. Thus, the supply side of the model entails $2J$ conditions. The demand side of the model consists of KJ equations: for each of the segments there is one demand equation for each of the J brands, which we can express in log form as:

$$\begin{aligned} \ln S_j^k &= \ln(1 - \sum_{i=1}^{J} S_i^k) + x_j\beta^k - \alpha^k p_j + \xi_j, \\ k &= 1, ..., K, j = 1, \ldots, J. \end{aligned}$$

Thus, in total, we have $(K+2)J$ equations. Similarly, there are $(K+2)J$ unknowns:

J wholesale prices: $w_1, \ldots w_J$

J retail prices: $p_1, \ldots, p_J$.

KJ market shares: $(S_1^1, \ldots, S_J^1), \ldots, (S_1^K, \ldots, S_J^K)$.

The full vertical equilibrium is the solution to this set of $(K+2)J$ equations in $(K+2)J$ unknowns. As is the case with all existing static models of differentiated multiproduct oligopoly, we are unable to formally prove the existence of the Bertand-Nash price equilibrium. In our model, the problem lies with the retailer who maximizes the joint profits from all products. If we only focused on the single-product manufacturers, the proof of the existence of equilibrium would be analogous to Caplin and Nalebuff(1991). Therefore, we must assume the existence of a Bertrand-Nash equilibrium with strictly positive prices.

Estimation

We now outline the estimation procedure for the equilibrium model with heterogeneous preferences developed in the previous section. Recall that heterogeneity enters the model in the form of a random coefficients specification. Unlike the BLP (1995,1998) approach, we use a discrete approximation so that we estimate the actual parameter distribution instead of the mean and variance. In economic terms, we estimate the aggregate analogue of Kamakura and Russell's (1989) latent class model. Our econometric model is similar to that of Berry, Carnall and Spiller (1997). Defining $X_{jt} \equiv [x_{jt}\ p_{jt}]$ and $\widehat{\beta}^k \equiv (\beta^k, \alpha^k)$, we let $\delta_{jt} \equiv X_{jt}\widehat{\beta}^1 + \xi_{jt}$ denote Segment 1's mean utility for product j, and $\delta_t \equiv (\delta_{1t}, \ldots, \delta_{Jt})$ denotes the entire vector of Segment 1's mean utilities. Furthermore, let $\widehat{\beta}^{k*} \equiv \widehat{\beta}^k - \widehat{\beta}^1$, $k = 2, ..., K$ denote the difference in tastes relative to Segment 1. Using this notation and the assumption that segment participation derives from the logit distribution, we can rewrite the share equations in (2.1)as:

$$\overline{S}_j(X_t, \delta_t; \Theta) = \lambda^1 \frac{\exp(\delta_{jt})}{1 + \sum_{i=1}^J \exp(\delta_{it})} + \sum_{k=2}^K \lambda^k \frac{\exp\left(X_{jt}\beta^{k*} + \delta_{jt}\right)}{1 + \sum_{i=1}^J \exp\left(X_{it}\beta^{k*} + \delta_{it}\right)} \quad (2.5)$$

$$j = 1, ..., J, t = 1, ..., T$$

where $\Theta \equiv (\widehat{\beta}^{2\bullet}, \ldots, \widehat{\beta}^{K\bullet}, \lambda^1, \ldots \lambda^K)$ denotes the full parameter vector to be estimated, $X_t \equiv (X_{1t}, \ldots, X_{Jt})$ denotes the vector of product characteristics across brands, and $\overline{S}_j(\cdot, \cdot; \cdot)$ denotes the market share *function* for brand j.[5]

In principle, we could estimate (2.5) using a non-linear procedure based on the market shares. However, this approach would be problematic for two reasons. First, we would expect the residuals of the market shares to be highly correlated both over time and across stores. Second, if we expect that a subset of the attributes in X_{jt} are correlated with the unobserved attribute ξ_{jt}, then we would have difficulty finding a meaningful way to instrument the share equations and offset the endogeneity problem. The fact that ξ_{jt} enters these equations in a non-linear fashion complicates typical instrumentation methods.

The extent to which the weekly scanner data typically used in marketing research involves endogeneity problems is an interesting question. In the economics literature, the unobserved component ξ_{jt} is usually thought to correspond to product attributes that vary from year to year (BLP 1995). However, scanner data used in marketing typically involves weekly observations over a period that is usually less than two years in length. The notion that physical product attributes would vary on a week to week basis is implausible. However, another interpretation of ξ_{jt} is that it involves temporary changes in brand valuations that are induced by consumer advertising or couponing.

---

[5]Recall that we do not estimate λ^k directly but rather estimate γ^k, where $\lambda^k = \left(\frac{\exp(\gamma^l)}{1+\sum_{k=1}^{K-1}\exp(\gamma^l)}\right)$. Thus γ^k is an element in Θ.

This would be relevant for the study in this paper which uses data from the yogurt market. Using household-level data for the yogurt market (over the same time period that we employ in this study) Ackerberg (1999) finds evidence that weekly television advertising has a positive effect on brand choice. He also finds a positive correlation between prices and advertising. Since we do not observe television advertising in our aggregate data, we must treat it as part of our residual and postulate that it might lead to an endogeneity bias. We assume that this endogeneity problem only contaminates the prices, but not the other product attributes that we use in this study. Depending on the number and quality of available instruments, however, the estimation approach outlined below could easily be extended to deal with the endogeneity of additional attributes .

Rather than estimate (2.5), we use the inversion procedure proposed by Berry (1994). We begin by partitioning the observed product characteristics as $X_{jt} = [x_{jt}, p_{jt}]$, where by assumption $E\left(x\xi_{jt}|x_{jt}\right) = 0$ and $E\left(p\xi_{jt}|p_{jt}\right) \neq 0$. Following Berry(1994), we invert (2.5) to recover the vector $\delta_t(\Theta)$ of mean utilities of segment 1 as a function of parameter vectors Θ and set up the estimation procedure in terms of the δ terms. Since the inverse of (2.5) does not have a simple analytical form, we resort to numerical inversion. Rather than use a standard numerical inversion procedure, we use the contraction-mapping of BLP (1995). The approach requires, for each t, picking some initial guess of the mean utility vector δ_t and iterating (2.6) until the following J expressions converges:

$$\delta_{jt}^{n+1} = \delta_{jt}^{n} + \ln(S_{jt}) - \ln\left[\overline{S}_{jt}\left(X_t, \delta_t^n; \Theta\right)\right],\ j = 1, \ldots, J. \tag{2.6}$$

where the superscript n refers to an iteration, and S_{jt} is the observed market share for brand j in period t.

This procedure turns out to be much faster than typical projection methods. The advantage of using δ_{jt} for estimation is that the prediction error, $\delta_{jt} - X_{jt}\widehat{\beta}^1$, is simply the unobserved product characteristic, ξ_{jt}. The fact that ξ_{jt} enters (2.6) linearly facilitates instrumentation. Moreover, with some intuition for the source of the unobserved attribute, we are able to impose reasonable covariance restrictions to set up our GMM procedure.

Estimation of the pricing equations is more straightforward. We assume that mc_{jt} is explained by factors prices, c_t, and a random component (possibly unobserved factors):

$$mc_{jt} = c_t\gamma_{jt} + \eta_{jt}.$$

where we assume $E\left(c\eta_j|c\right) = 0$ (the unobserved factor prices of production are conditionally independent of the observed factor prices). Our linear specification implicitly assumes a fixed proportions (i.e., Leontief) production technology. For retail products, such as yogurt, this assumption may not be unreasonable, at least in the short run.[6] Substituting the marginal cost specification into wholesale pricing equilibrium conditions (2.4)yields the wholesale pricing equations that we estimate.

---

[6] We experimented with a log-linearized (Cobb-Douglas) specification, which allows for more curvature than the linear mc :

$$\log(p - margin) = \sum_{f=1}^{F} \gamma_f \log(c_f).$$

However, taking logarithms reduces the variation in the factor prices substantially. This reduced variation inflates the margins (lowers the estimated price response parameter) to the point that they exceed prices, making it impossible to take the log of $(p - margin)$.

To be consistent with cost-minimizing input choices by manufacturers, the specification of mc_j should be homogeneous of degree one. However, we do not observe all factors of production, and for this reason we also include product-specific intercept in mc_j to improve the fit. This assumption is not entirely structural because it does not follow from cost minimization by manufacturers. But we can provide three potential interpretations for these intercepts.

First, we might assume that the unobserved component of the marginal costs consists of both a fixed-effect η_j and a purely random effect, ψ_{jt}:

$$\eta_{jt} = \eta_j + \psi_{jt}.$$

For instance, if the residual is interpreted as unobserved factor prices, then we are assuming that these factors may be decomposed into a stable and a time-varying component. The intercepts in the marginal cost function now capture the stable component of these factors of production.

Second, the intercepts could capture the effects of fixed physical product attributes. Thus, the marginal cost specification represents both the costs associated with some of the factors of production as well as the costs of the product attributes. This attribute component specification is consistent with the standard IO approach of specifying the marginal costs strictly in terms of attributes.

Finally, we could imagine a scenario in which the manufacturers allocate a fixed slotting allowance to the retailer, part of which the retailer may simply pool into his category revenues. The intercepts would represent the retailer's average allocation of the fixed slotting allowances (Chintagunta 1999a). Of course, unless we can justify that this fixed fee is completely independent of the per-unit wholesale prices, the true underlying channel model should embody a wholesale two-part tariff.

We now set up a GMM procedure to estimate the system of price and demand equations. Let $\varepsilon_t = \begin{bmatrix} \eta_t \\ \xi_t \end{bmatrix}$ be a $(2JN \times 1)$ matrix with the prediction error for marginal costs and the unobserved attributes for each of the products in store-week t. Similarly, we define our instruments, Z_t, an N-dimensional vector including the exogenous product characteristics as well as other potential covariates that may be correlated with p_{jt}. Our data-generating process comes from our conditional mean-independence assumption $E(\varepsilon_t \otimes Z_t) = 0$ and $E(\varepsilon_t \varepsilon_t') = \Omega$ a finite $(2J \times 2J)$ matrix. We are now able to construct our moment conditions:

$$h_t(\Theta) = \varepsilon_t \otimes Z_t,$$

where at the true parameter values, Θ_0, $E(h_t(\Theta_0)) = 0$. For estimation, we compute the corresponding sample analogue of these moment conditions:

$$h_T(\Theta) = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_t \otimes Z_t.$$

Our goal is to find values of Θ close enough to Θ_0 to set the sample moments as close as possible to zero. We estimate Θ by minimizing the following quadratic expression:

$$G(\Theta) = (h_{JT}(\Theta))' \, \mathbf{W} \, (h_{JT}(\Theta)).$$

The matrix $\mathbf{W}$ is a $(2JN \times 2JN)$ weight matrix. Hansen (1982) shows that the most efficient choice of $\mathbf{W}$ is a consistent estimate of the inverse of the variance of the moment conditions:

$$\mathbf{W} = E\left\{(h_T(\Theta))\,(h_T(\Theta))'\right\}$$
$$= E\left\{\varepsilon_t\varepsilon_t' \otimes Z_t Z_t'\right\}.$$

Clearly, we will need to make additional assumptions about the underlying distribution of $\{\varepsilon_t\}_t$ in order to estimate $\mathbf{W}$. Note that misspecifying the correlation structure of $\{\varepsilon_t\}_t$ will only affect the efficiency, not the consistency of our estimates.

Identification and Segment Structure

Identification

One important issue regarding the proposed model is whether, with aggregate data, we can identify segment-specific vectors of taste coefficients. To illustrate why such identification is possible, consider the two-segment world in which consumers are either bargain-hunters or quality-hunters. Moreover, suppose there are two periods and three products. Product 1 is a low-quality and low-price good (perhaps a store brand). Products 2 and 3 are both high-price and high-quality. Suppose that in period one, prices and qualities are such that all three goods have the same market shares. In period 2, suppose that product 3 is featured in the newspaper, constituting an increase in perceived quality, and that everything else remains fixed. This change should have two effects: the aggregate market share of product 3 should rise, and the aggregate share of product 2 should fall more than that of product 1 since product 2 is a closer substitute to product

3 than is product 1. In a model of homogeneous tastes, the IIA property would imply that the aggregate shares of 1 and 2 both fall by the same amount. In our heterogeneous tastes model, the aggregate share of product 2 will be allowed to fall more than that of product 1 if the segment of quality hunters have a different taste for features than bargain hunters. Events in the data that give rise to asymmetric substitution patterns of the kind just described will identify differences in taste parameters across segments and will allow the heterogeneous tastes model to provide a better fit of the data.[7]

Segment Structure

We must also address the issue of the appropriate number of market segments. In some instances, one may have such prior information as to the correct number of segments. For instance, Berry, Carnal and Spiller(1997) estimate a two-segment model based on business and non-business airline travellers. We could adopt the same strategy, assuming consumers are either "bargain-hunters" or "quality-sensitive." Instead, we add segments until we are no-longer able to identify additional parameters. Since we are using a GMM procedure, we do not have a theory-based metric for statistical fit.

---

[7] We illustrate this identification issue in greater detail in the Appendix using a simulation experiment in which we apply the econometric model to synthetic data. We show that the model does a good job of recovering the underlying heterogeneity structure. We also show that failure to account for heterogeneity can lead to incorrect parameter estimates. Intuitively, one might expect the homogeneous model to predict the mean values of the price sensitivities (Allenby and Rossi 1991). However, we do not find this result to be true. Since most managerial applications make use of estimated price-response parameters and elasticities, the homogeneous model will lead to incorrect strategic conclusions.

We stop adding segments once the probability of being in the additional segment is not significantly different from zero.

In principle, we could estimate the model using a Full Information Maximum Likelihood (FIML) procedure. For instance, we could assume the unobserved attributes and marginal cost covariates, $[\xi_t, \eta_t]'$, are drawn from an i.i.d. normal distribution with covariance matrix Ω. Estimating with FIML would allow us to use the Bayesian or the Akaike Information Criterion to determine whether additional segments provide additional statistical information, as is typically done with individual-level data. However, FIML is not a convenient estimator for handling the highly non-linear nature of the model or for allowing more arbitrary dependence structures in the error terms. Nonetheless, a FIML estimator would entail searching for the optimal parameter vector β^{FIML} that minimizes the log-likelihood function:

$$L = -\frac{T}{2}\log(\Omega) + \sum_{t=1}^{T}\log\left\|\frac{\partial f_t}{\partial y_t'}\right\| - \frac{1}{2}\sum_{t=1}^{T} f_t'\Omega^{-1}f_t$$

where $f_t = [p_t, \delta_t]'$. We can simplify this expression by concentrating out the Ω term, which can be shown to be $\Omega^{FIML} = \frac{1}{T}\sum_{t=1}^{T} f_t f_t'$. We minimize the concentrated log-likelihood function:

$$L^c = \sum_{t=1}^{T}\log\left\|\frac{\partial f_t}{\partial y_t'}\right\| - \frac{T}{2}\log\left(\frac{1}{T}\sum_{t=1}^{T} f_t f_t'\right).$$

The most complicated part of this objective function is the computation of the Jacobian term, $\frac{\partial f_t}{\partial y_t'}$. While this term does have an analytic solution, it is computationally demanding to evaluate. In practice, we continue to add segments to the model until we are no longer able to improve the statistical fit, according to a metric such as Akaike's

Information Criterion or the Bayesian Criterion. For now, we focus solely on the GMM results.

Demand Estimation Results Using Yogurt Data

Data

To illustrate the ability to recover market segmentation from aggregate data, we use a sample of weekly store-level data for yogurt. We use data on prices, market shares and feature activity for the four largest brands in the category. The data are collected by the AC Nielsen Company in Springfield, Missouri using store checkout scanners. The data come from 9 stores belonging to a single chain during the 102 week period from 1986-1988.

To compute the market shares, we divide the total unit sales of each brand by the total number of store trips in the given week.[8] Thus, our model assumes that consumers select a single brand of yogurt on a given shopping trip. The no-purchase alternative is simply one minus the sum of the brand market shares. The price variable is measured as the retail shelf-price per ounce, net of in-store promotional price cuts. The feature

[8]Since we do not observe the total weekly store traffic, we infer the total number of trips using a panel of 2500 households. We compute the store traffic by projecting the number of panelist trips to each store onto the total population. Many aggregate store-level data sets do include measures of store traffic.

activity is an indicator for whether the product appeared in a weekly newspaper advertisement.

On the supply side, we use factor prices for labor and materials costs collected by the Bureau of Labor Statistics. Labor costs consist of the average hourly earnings of production workers in the dairy products industry. For materials costs, we use the price index for fluid milk. Since these data are reported on a monthly frequency, we use the linear filtering process suggested by Slade(1995) to convert the monthly data to weekly data[9].

The yogurt data include three national brands, Dannon, Yoplait and Weight Watchers, and a regional brand, Hiland. These products account for over 70 percent of the category volume. We report summary statistics for the product attributes of each of these goods along with the factor prices in Table 2.2. We can see that Dannon is the market share leader, with the second-highest price. Yoplait is the high-price brand, while Weight Watchers charges the lowest price.

Results

We now report our findings for the segment model when we apply it to the yogurt data. In Table 2.3, we report the parameter estimates for various heterogeneity specifica-

---

[9]We assign the factor price W_t the value from the corresponding month and then we smooth the series:

$$W_t^s = 0.25W_{t-1} + 0.5W_t + 0.25W_{t+1}.$$

	Share	*Price*	*Feature*	*Factor*	*Price Index*
Dannon	0.18	0.08	0.04		
	(0.09)	(0.01)	(0.20)		
Yoplait	0.09	0.10	0.06		
	(0.08)	(0.01)	(0.24)	labor costs	9.69
W.W.	0.09	0.05	0.01		(0.11)
	(0.08)	(0.01)	(0.08)	milk costs	103.54
Hiland	0.04	0.08	0.03		(0.81)
	(0.05)	(0.01)	(0.18)		
No Purch	0.60	-	-		
	(0.18)	-	-		

Table 2.2. Descriptive Statistics - sample means (standard deviations in parentheses)

tions. The first two columns contain the results for the SUR and the 3SLS regressions respectively. Both approaches assume homogeneity (a single segment). However, the SUR model does not treat the endogeneity of prices. As expected, we find that the SUR results, which do not instrument for the price endogeneity, yield a price response parameter that is much lower than that of the 3SLS. This finding of a downward effect is consistent with the downward bias found in BLP(1995), Nevo(2000) and Besanko, Gupta, and Jain (1998).

We now focus our attention on the remaining columns, which report our findings for the multi-segment models. By inspection, it would appear that the data are only able to identify three segments. Although not reported, our estimates for the four-segment model yield several insignificant parameters, including the probability of membership in the additional segment. Thus, we conclude that the underlying consumer tastes come from three segments and we will focus on the results of the three-segment model. Comparing the three segments, Segment 1 consumers account for roughly 44 percent of the shopping trips. This segment consists of price-conscious shoppers: they are extremely price sensitive, but do not respond much to feature ads. They also have high preference for the two national brands, Dannon and Yoplait.

Segment 2, accounting for about 46 percent of the shopping trips, is much less price-sensitive and responds much more to feature ads. This segment seems to fit the profile of the "time-starved" consumer: a consumer who does not pay that much attention to relative prices when choosing among brands but who can be persuaded to buy

Attributes	Model						
	SUR	3SLS	2-segment		3-segment		
			Seg 1	Seg 2	Seg 1	Seg 2	Seg 3
price	-54.46	-74.69	-60.78	-30.93	-79.11	-32.67	-64.71
	(1.24)	(4.42)	(4.36)	(2.85)	(5.23)	(8.27)	(11.13)
feature	0.34	0.04	0.62	1.54	0.21	1.62	0.33
	(0.07)	(0.11)	(0.12)	(0.30)	(0.26)	(0.65)	(0.67)
Dannon	3.09	4.74	3.39	1.82	4.17	1.31	5.86
	(0.10)	(0.36)	(0.36)	(0.28)	(0.66)	(1.52)	(0.91)
Yoplait	3.48	5.60	4.30	-9.69	6.50	-0.97	5.84
	(0.14)	(0.47)	(0.46)	(1.14)	(0.61)	(0.88)	(1.31)
W.W.	0.69	1.76	0.39	0.69	0.32	0.19	0.61
	(0.08)	(0.24)	(0.23)	(0.14)	(0.41)	(0.79)	(0.85)
Hiland	1.18	2.76	1.76	-1.51	3.14	-0.56	1.19
	(0.11)	(0.35)	(0.37)	(1.67)	(0.67)	(1.34)	(5.24)
prob.	-0.17	-0.27	0.80	0.20	0.44	0.46	0.10
	-	-	(0.00)	-	(0.23)	(0.19)	-

Table 2.3. GMM Results (standard errors in parentheses)

one brand over the others when that brand is promoted through feature advertising. One might imagine that for customers in this segment, featuring serves to increase awareness of particular brands as opposed to simply announcing periodic price cuts (which might be the primary impact of featuring on consumers in segment 1). This segment has high preferences for the two fruit-bottom brands, Dannon and Weight Watchers.

Finally, Segment 3, accounting for about 10 percent of the trips, has intermediate price sensitivity, but low feature response. This segment values the two national brands equally. In Table 2.4, we report the average of the expected shares for the four products in each of the segments, computed using the demand equations and the observed prices and features.

	Market Shares		
	Seg. 1	**Seg. 2**	**Seg. 3**
Dannon	0.09	0.19	0.52
Yoplait	0.17	0.01	0.15
WW	0.03	0.16	0.02
Hiland	0.05	0.04	0.01
No Purch	0.67	0.60	0.31

Table 2.4. Segment Shares

Factors	*Parameter*
Dannon	0.51
	(0.43)
Yoplait	0.44
	(0.11)
W.W.	0.41
	(0.10)
Hiland	0.44
	(0.10)
labor	-0.14
	(0.01)
materials	0.01
	(0.00)

Table 2.5. Marginal Cost Parameters for 3-segment model (standard errors in parentheses)

In Table 2.5, we present the supply-side estimates for the three-segment model. We find that the coefficient on the price of raw materials has the expected positive but that the coefficient for the price of labor is negative. While it is theoretically possible that the price of labor could have a negative impact on marginal cost (this would occur if labor were an inferior input), the negative coefficient on labor might be due to the overly-stringent assumption of constant marginal cost.

We now take our three-segment model and compute the price elasticities which we report as means over all store-weeks in Table 2.6. All of the own-elasticities are greater than one in magnitude, which is consistent with the underlying static oligopoly behavior. Yoplait has the largest price elasticity of demand, while Weight-Watchers has the smallest elasticity. Within a segment, the own-price elasticity for a particular brand is the traditional logit elasticity, $-\alpha^k(1 - S_j^k)p_j$, so Yoplait's large own elasticity reflects the fact that it is the high-price brand. The estimated cross elasticities indicate that all

	Dannon	*Yoplait*	*W.W.*	*Hiland*	*No Purchase*
Dannon	-3.03	0.89	0.54	0.56	0.61
Yoplait	0.56	-6.50	0.18	0.67	0.66
W.W.	0.17	0.10	-1.75	0.17	0.17
Hiland	0.12	0.23	0.12	-4.24	0.19
No Purchase	0.00	0.00	0.00	0.00	0.00

Table 2.6. Mean Price Elasticities for the 3-segment model

products respond relatively highly to changes in the price of Dannon. Hiland and Dannon both respond to price changes in Yoplait, while Weight-Watchers is less responsive. Finally, the prices of both Weight Watchers and Hiland have relatively little effect on the demand for other products. In terms of stimulating consumers to switch away from the no-purchase alternative, changes in the prices of Dannon and Yoplait have the greatest impact, while changes in Weight-Watcher's and Hiland's prices have a relatively more modest impact on category demand..

By examining the segment-specific elasticities, we gain additional insight into the substitution patterns reported above. Because the substitution patterns within a segment exhibit the IIA properties, we do not report the cross-elasticities. Looking at Table 2.7, we see that Yoplait has the highest own-price elasticity of demand in all three segments. Hiland and Dannon exhibit very similar price-elasticities in segments 1 and 2, but Hiland has a much higher price elasticity in segment three, driving its overall elasticity to be higher than that of Dannon. Finally, Weight-Watchers consistently has the smallest price elasticity of demand in each segment.

	Overall	*Segment 1*	*Segment 2*	*Segment 3*
Dannon	-3.03	-5.85	-2.16	-2.57
Yoplait	-6.50	-6.96	-3.35	-5.78
W.W.	-1.75	-4.07	-1.46	-3.33
Hiland	-4.24	-5.85	-2.43	-4.97

Table 2.7. Mean Segment Own-Price Elasticities for the 3-segment model

Table 2.8 shows brand responsiveness to feature ads. Because feature ads are a discrete variable in our model, we can not compute an elasticity. Instead, we compute the market shares with all of the features set to zero. To compute the response, we set one product's feature indicator to 1 and calculate the percentage change in its market share. Overall, we find that the market shares of the two national brands are much less responsive to features than the smaller brands, Hiland and Weight Watchers. Featuring Dannon or Weight Watchers has a much greater ability to generate store traffic (switching away from the no-purchase alternative).

More generally, Table 2.9 illustrates that most of the feature response is driven by Segment 2, which has a much higher responsiveness to features than the other two segments. The low overall responsiveness of Yoplait's share to featuring (0.43) as compared to its high responsiveness among Segment 2 consumers (3.86) is due to Yoplait's extremely low share within the Segment 2 market. Thus, the small gains in this segment achieved from featuring Yoplait yields a large elasticity value.

	Dannon	*Yoplait*	*W.W.*	*Hiland*	*No Purchase*
Dannon	1.05	-0.07	-0.37	-0.21	-0.22
Yoplait	-0.04	0.43	-0.04	-0.04	-0.04
W.W.	-0.18	-0.03	2.00	-0.17	-0.18
Hiland	-0.06	-0.01	-0.10	1.78	-0.06
No Purchase	0.00	0.00	0.00	0.00	0.00

Table 2.8. Mean Feature Response for the 3-segment model

	Overall	seg. 1	seg. 2	seg. 3
Dannon	1.05	0.21	1.98	0.15
Yoplait	0.43	0.19	3.86	0.31
W.W.	2.00	0.23	2.22	0.37
Hiland	1.78	0.23	3.50	0.38

Table 2.9. Mean Segment Feature Responses for the 3-segment model

Retail Strategies to Capture Consumer Surplus

In this section, we explore how knowledge of the brand preferences and price and feature responsiveness in discrete consumer segments can enable a retailer to make more profitable pricing and featuring decisions. The economic model estimated in the paper assumes that the retailer sets a uniform profit-maximizing price, given the wholesale prices. Thus, the retailer and the manufacturers are assumed to know the underlying segment structure of demand, but they are unable to price discriminate. In the section below we consider two alternative pricing strategies aimed at capturing additional consumer surplus: third-degree price discrimination in which the retailer sets segment-specific prices (possibly implemented through Catalina couponing) and targeted pricing-feature strategies in which the retailer uses featuring to stimulate demand in particular segments and captures part of the additional surplus that is created through adjustments in retail prices. We will use the estimated demand model that we described above to forecast the profit implications of each of these strategies.

While examining such counterfactual or what-if decision scenarios, it is important to take into account the competitive response of all players in the model. Thus, each of the price discrimination scenarios implies not only a new set of retail prices, but also a new set of equilibrium wholesale prices. Previous studies in marketing (e.g., Rossi *et al.* 1996) have examined the profit implications of targeted couponing relative to blanket couponing by manufacturers, but assume that the prices of competing manufacturers remain unchanged. This omission could potentially lead to an overstatement of the benefits of the proposed action. By contrast, the vertical Nash model assumed in our analysis implies that each manufacturer takes retail margins and the wholesale prices of other manufacturers as given and chooses wholesale prices to maximize its own profits. Thus, the competitive responses of manufacturers to the new retail margins implied by a price discrimination strategy and to the consequent changes in the wholesale prices of other manufacturers are taken into account. From a computational standpoint, the new optimal prices are obtained by solving the system of simultaneous equations representing the brand shares and the retail prices.

For the uniform pricing case, optimal prices and profits for the yogurt data are computed using the parameter estimates of the three segment model and are shown in Table 2.10 (prices are in cents per ounce and profits are in cents per store trip). Because the estimated model is the uniform pricing model, it is not surprising that the optimal prices are close to the actual average prices shown in Table 2.2.

Optimal Segment-Specific Prices

In this section we explore the extent to which the retailer could increase profitability through third-degree price discrimination. In so doing, we assume that the retailer can determine the segment in which a consumer belongs, and can charge segment-specific prices. In practice, the implementation of this pricing scheme would require the retailer to use information on current and past purchases to infer (probabilistically) an individual consumer's segment membership and assign a tailored coupon based on that inferred membership. For example, recently developed marketing programs such as Catalina Marketing Incorporated's Checkout Coupon allow retailers to use a shopper's current purchase to deliver a customized coupon at the point-of-purchase, to be redeemed on a future purchase occasion. Moreover, retailers have detailed data on consumer purchasing from frequent shopper or loyalty card programs. We envision a system wherein an individual shopper's past purchasing information is combined with the estimated heterogeneous demand model at the point of purchase to assign the shopper to a consumer segment. Given individual-specific past purchasing information, Bayes Rule can be used to update the prior segment-membership probabilities in the estimated model to yield individual-specific posterior segment-membership probabilities (Kamakura and Russell 1989 discuss this for a model estimated with household panel data). Individuals are then assigned to the highest probability segment. Clearly, the longer the purchase history available for individual shoppers, the more accurately one can assign consumers to segments. However, the gains from basing coupons on even a single purchase occasion can be significant.[10]

When it charges segment-specific prices p_j^k, $j = 1, \ldots, J, k = 1, \ldots, K$, the retailer's profits are now:

$$\pi = \sum_{k=1}^{K} \sum_{j=1}^{J} \left(p_j^k - w_j\right) S_j^k \lambda^k M.$$

[10]Rossi et al. (1996) find that gains in profit from using one purchase occasion to develop customized coupons are large relative to a blanket couponing strategy.

The retailer must be at least as profitable as it was under uniform pricing since the retailer can still choose to set a uniform price for each product and earn the same profits as before. Although a manufacturer might charge a different price and face a different demand in this new scenario, the wholesale pricing rule in (2.3) continues to hold. However, the retail pricing rule changes when the retailer can set segment-specific prices. Because demand within a segment is governed by the homogeneous logit model, the optimal retail price for brand j for segment k takes the simple form that we discussed in Section 2.2:

$$p_j^k = w_j + \frac{1}{\alpha^k S_0^k}, j = 1, ..., J, k = 1, ..., K,$$

where S_0^k is the share of the no-purchase alternative in segment k. To determine the new prices and quantities under this price discrimination scheme, we must solve the system of JK retail price equations, JK product share equations and J wholesale price equations.[11].

In Table 2.10, we compare the optimal uniform retail prices and the segment-specific prices. When the retailer sets segment-specific prices — and manufacturers make equilibrium adjustments in wholesale prices in response — we see that Segment 2's prices are raised well above those of Segments 1 and 3 (recall that these are prices per ounce). This reflects the fact that Segment 2 consumers have much lower price sensitivities. Under segment-specific pricing, the retailer raises the price of Dannon above its opti-

[11] We solve this system numerically, using the "fsolve" function in Matlab. The procedure takes about 30 seconds on a 450 MHz PC.

	Uniform Pricing		Discriminatory Pricing			
	Price	*Mfr. Profit*	*Price 1*	*Price 2*	*Price 3*	*Mfr. Profit*
Dannon	0.08	0.027	0.14	0.16	0.14	0.025
Yoplait	0.10	0.017	0.07	0.09	0.07	0.020
W.W.	0.05	0.030	0.04	0.06	0.04	0.024
Hiland	0.08	0.019	0.06	0.08	0.06	0.015
Retail Profit	0.0379		0.0769			

Table 2.10. Third-Degree Price Discrimination

mal uniform price in all three segments, taking advantage of the high valuations of the Dannon brand in all three segments. The price of Weight Watchers is lowered in Segments 1 and 3, due to its relatively low valuation in these segments, but it is increased in Segment 2, due to that segment's strong preference for the Weight-Watcher's brand. In terms of per-unit profits, we see that third-degree price discrimination has a significant impact on retailer profits: the retailer's profits more than double relative to the uniform pricing scheme. Interestingly, except for Yoplait, the manufacturer profits (reported next to the prices in Table 2.10) go down slightly when the retailer engages in price discrimination, indicating a possible channel conflict between retailers and manufacturers on this issue. Overall channel profits, however, go up, so in principle the retailer could compensate manufacturers for their lost profits and still end up better off than before.

To implement the segment-specific prices, the retailer could set the regular price at the level of the highest segment-specific price, and issue coupons to consumers in the other segments such that their price net of the coupon face value is equal to the optimal price for that segment. To illustrate, the optimal prices in Table 2.10 would call for the shelf price to be set at the level of the optimal prices for segment 2, and coupons of different face values to be issued to consumers who are identified to belong to segments 1 and 3. These coupons would reduce the effective price of the products to the level of the optimal prices for the respective segments.

To summarize, the analysis in this section identifies large potential gains to retailers from engaging in price discrimination in a retail category with segmented consumers. These, of course, represent upper bounds on the retailers ability to gain from the capture of additional consumer surplus because in practice we would not expect the retailer to be able to perfectly segment consumers. Still, this analysis suggests possible equilibrium gains from price discrimination are sufficiently large that it might well be worth the cost and effort by retailers to implement imperfect approximations to the scheme in Table 2.10.

Targeted Feature and Pricing Strategies

We assume now that the retailer cannot observe the segment membership of individual consumers (or even infer it with reasonable accuracy) and is thus unable to charge targeted prices. However, the retailer can use information on market segmentation to strategically coordinate feature advertising and retail prices in order to capture additional segments. This is possible because consumers in the three segments differ not only in their price responsiveness but also in their brand preferences and their sensitivity to feature ads. For example, Tables 2.6 and 2.8 indicate that Segment 2 consumers are the most responsive to feature ads but the least sensitive to price changes. Furthermore, the price and feature elasticities for each segment vary across the four brands. This knowledge allows for the possibility of developing a marketing mix in which specific brands are featured and priced in such a way to extract additional surplus from particular segments.

Typically retailers feature at most one brand in a category at any time. The decision scenario we consider is: if a decision has been made to feature one brand of yogurt for an additional week, which brand should be featured, and what are the optimal retail prices? Thus, the retailer must choose feature, $f_j \in \{0, 1\}$, and $\sum_j f_j \leq 1$, and (uniform) prices, p_j to maximize profits. For simplicity we ignore out-of-pocket costs of featuring.

In Table 2.11, we show equilibrium retail prices and retailer and manufacturer profits when each of the brands is featured. The results indicate that featuring Weight Watchers generates the highest profits for the retailer and the highest profits for the manufacturers as well (although manufacturer profits are generally insensitive to the brand that is targeted). This is consistent with the observation from Table 2.8 that Segment 2 is most responsive to this brand's feature advertising and that Segment 2 has the most price-inelastic demand (Table 2.6). Thus, in this strategy advertising Weight Watchers enables the retailer to attract Segment 2 buyers and charge higher prices to enhance profits.[12]

---

[12]In practice, the decision to feature a given product may not be entirely exogenous. In some categories, manufacturers may subsidize the costs of featuring their products. Thus, we interpret the results of the given analysis as preliminary evidence that retailers are able to predetermine which products make

	Feature Dannon		Feature Yoplait		Feature WW		Feature Hiland	
	Price	*Mfr. Profit*	*Price*	*Mfr. Profit*	*Price*	*Mfr. Profit*	*Price*	*Mfr. Proft*
Dannon	0.14	0.03	0.14	0.03	0.13	0.03	0.14	0.03
Yoplait	0.07	0.04	0.07	0.03	0.07	0.04	0.07	0.04
W.W.	0.04	0.03	0.04	0.03	0.06	0.03	0.04	0.03
Hiland	0.05	0.02	0.05	0.02	0.05	0.02	0.06	0.02
Retail *Profit*	0.008		0.008		0.026		0.019	

Table 2.11. Choice of Product to Feature

In practice, of course, retailers often use feature advertising to announce price cuts. (This was true in our Yogurt data set: the prices of each brands was, on average, lower when it was featured than when it was not.) In this sense, feature can be thought of as a form of targeted marketing aimed at consumers in price-sensitive segments, such as Segment 1. A useful result of estimating heterogenous demand models with discrete segments, though, is that we see that there might be price-insensitive segments that respond heavily to feature advertising. When this is true — as it apparently is in our Yogurt market — the retailer can use feature in a different way: not to announce price cuts but rather to temporarily raise brand awareness among price-insensitive consumers. When feature is used in this way, we would expect that it might be correlated with price increases rather than price decreases.

Conclusions

An extensive body of marketing research has documented the importance of accounting for heterogeneity both as a statistical tool to improve model fit and as a practical tool to develop managerial strategy. However, most of the existing techniques require substantial individual panel data sets which may be unavailable to most retailers. We propose a similar heterogeneous model that makes use of aggregate store-level data, which is more widely available and less cumbersome. The underlying consumer behavior is identical to the popular finite mixture model. We assume that consumer tastes are

---

the most sense to feature.

characterized by a finite number of latent segments, each of which we estimate explicitly. From a modeling perspective, the heterogeneity generates more flexible substitution patterns and margins than the homogeneous model. Qualitatively, the heterogeneity provides a deeper understanding of how a retailer's market responds to the marketing environment. Finally, we demonstrate how the knowledge of the specific segments provides valuable information for developing targeted retail strategies to extract additional consumer surplus from particular segments.

There are a number of possible extensions of the techniques that we present here. First, in our analysis, we have assumed that all retailers face local markets with an identical segment structure. It is straightforward to adapt the model to the possibility that retailers face different segment structures (e.g., the relative proportions of the segments differ across local retail markets, perhaps due to demographic differences). If segment structures differ across local retail markets, retail mark-ups will differ across local markets as well. In future work, we intend to explore the extent to which local retail markets exhibit such heterogeneity and whether such heterogeneity can be systematically related to observable demographic differences (e.g., average household incomes) across local retail markets.

A second limitation of the present analysis is that we do not include product attributes in order to explain differences in brand-specific constants. In product categories with sufficient variation in attribute combinations (e.g., sizes, flavors, and so forth), it is possible to estimate brand constants and then project those constants onto attributes. An advantage of this formulation is that it allows us to estimate segment-level differences in the valuation of various product attributes. This analysis, which we plan to pursue using product categories other than yogurt, allows one to explore targeted product strategies by manufacturers, such the repositioning of existing brands to better appeal to particular segments or the introduction of new varieties of existing brand whose attributes are specifically tailored to the tastes of particular segments.

CHAPTER 3.
Multiple Discreteness and Product Differentiation: Strategy and Demand for Carbonated Soft Drinks

Introduction

Since the seminal paper of Guadagni and Little (1983), a large empirical marketing literature has emerged that estimates microeconomic models of consumer demand using scanner data. This *structural* approach to marketing involves estimating demand parameters, that are consistent with a model of utility maximization, to study the effects of such marketing tools as prices, feature advertising and display on the consumer's decision to buy and his/her subsequent choice of products. The use of a structural model, as opposed to using an arbitrary statistical model with a good fit to the data, permits the interpretation of the estimated parameters as behavioral and the computation of valuable economic metrics such as consumer valuations and willingness-to-pay. Recently, many marketing studies have borrowed from the Empirical Industrial Organization literature approach by combining the estimated model of demand with a partial equilibrium model of strategic price-setting by firms to study retail industry conduct without observing costs (Bresnahan 1989). This equilibrium approach to the retail data-generating process provides a useful tool for understanding competitive strategy.

The most notable of these demand models is the standard discrete choice model (DCM), such as the conditional logit and the probit (McFadden 1981). The DCM's restrictive implicit assumption of single-unit purchase behavior makes it a relatively simple model to estimate. For instance, several recent papers in marketing and industrial organization have taken advantage of the DCM's convenient aggregation properties to apply the logit to aggregate data (Allenby and Rossi 1991, Berry, Levinsohn and Pakes [BLP] 1995, 1998, Besanko, Gupta and Jain [BGJ] 1998 and Chintagunta 1998). However, the misuse of this behavioral assumption has an adverse effect on estimated demand and consumer responses.

In reality, for many categories, a consumer seeking contemporaneous variety may purchase a bundle of goods with several products and an integer quantity of each. This phenomenon constitutes a ***multiple discreteness*** problem. I find evidence of this multiple discreteness property in several industries, including soft drinks, ready-to-eat cereals, canned soup and cookies. For each of these categories, over 20% of observed shopping trips involve the purchase of at least two products. In the current study, I focus on carbonated soft drinks (CSDs). Table (3.40) in the appendix shows that, conditioning on the occurrence of a purchase, only 39% of the trips involve a single unit of a single brand. In fact, almost 31% of the trips result in at least two units of a single product and another 31% of the trips result in two or more products. Similar tables are presented in the appendix for the other categories mentioned. Estimating a household model that ignores this important quantity information will underpredict demand as well as consumer response to promotional variables.

I propose a more general model than the DCM that accounts for the fact that consumers may seek variety on a given purchase occasion. Simonson (1991), Hauser and Wernerfelt (1991) and Walsh (1995) explain the need for purchasing several alternatives on a given trip by the separation between the time of purchase and the time of consumption. If the preferences at the future time of consumption are uncertain, the shopper may purchase a basket of several alternatives to maximize expected utility. Empirically, I expect such uncertainty to reflect a single shopper forecasting the varying tastes for several members of a household or for several different types of future consumption environments. This type of variety contrasts with the empirical stochastic variety-seeking literature, which uses the DCM to address dynamic brand-switching across shopping trips (McAlister 1982, Trivedi, Bass and Rao 1994, and Seetharaman and Chintagunta 1997).

I modify the static random profit formulation proposed by Hendel (1999) to suit the dynamic consumer CSD purchase panel. The model predicts the expected future consumption needs of the household, allowing for the purchase of a bundle of products to satisfy these multiple needs. Unlike the recent imperfect substitutes model of Kim, Allenby and Rossi (1999), the proposed model accounts for no-purchase trips. By accounting for the no-purchase decision, I am able to compute expected demand for the product category, as opposed to conditional demand. Since variation in prices may influence consumers' decisions to purchase any CSD at all, the total expected demand is a vital component for the equilibrium analysis of price-determination. The model's explicit use of household attributes presents a new role for observed demographic variables in identifying the joint distribution of total products and total units purchased on a given trip. If I remove the need for variety, the model collapses into the standard DCM. Thus, the conditional logit represents a special case of the proposed model in which the distribution of expected household needs for a given trip is degenerate at one.

Unlike most existing studies that model household-level decisions using DCMs and maximum likelihood techniques, I use a generalized method of moments $[GMM]$ procedure. Since I do not observe the expected consumption occasions in the data, I simulate them. GMM provides a convenient framework to use simulation techniques while still providing unbiased estimates.[13] Using GMM, I am also able to allow for a more general data-generating process. Since GMM does not require assumptions regarding the distribution of the estimation error, I explicitly account for the often-neglected unobserved dynamic effects inherent in a consumer shopping panel (see McCulloch and Rossi (1994) for a discussion of the panel probit). I account for potential unexplained persistence, even after including the standard dynamic controls, such as lagged choice indicators[14] and inventories, as well as heterogeneity. The estimation procedure fore-

---

[13]In particular, simulated GMM is unbiased even with a finite number of simulation draws (McFadden 1989). In contrast, simulated maximum likelihood estimators are biased, and consistency requires a computationally infeasible number of simulation draws, given the dimensions of the estimated model (Lee 1992).

[14]These lagged indicator variables are not entirely structural since I do not solve a dynamic program for each household. For a discussion of the structural interpretation of lagged choice indicators, see Chintagunta, Kyriazidou and Perktold (1997).

casts the expected random vector of total purchases for each alternative on a given household trip.

Econometrically, the large number of relevant products present a dimensionality problem.[15] Within the context of the DCM, Berry (1994) proposes the *characteristics approach* of Lancaster, which redefines products as bundles of their underlying attributes. Consumer demand consists of individuals selecting the utility-maximizing bundle based on their tastes for attributes, where the tastes consist of parameters for estimation. In addition to solving the dimensionality problem, the estimated taste parameters provide a measure of closeness between alternatives, providing insight into the study of product differentiation (Fader and Hardie 1996, Berry, Levinsohn and Pakes 1995, 1998, Goldberg 1995 and Nevo 2000)[16]. In the same spirit, I project consumer demand for goods onto product attributes to allow for a larger array of alternatives in the choice set and to characterize perceived differences between goods. I also allow for heterogeneity in the tastes for attributes across households and expected needs by using a random coefficients specification.

I find that the proposed model provides a much better explanation of CSD demand than the several variants of the DCM, which I use as benchmarks. By ignoring the multi-

---

[15] Suppose we use a linear demand specification for a category with J products. The estimation procedure would involve at least J^2 parameters to account for the cross-price effects alone. For a 20-product category, we would estimate at least 400 parameters.

[16] Many marketing applications assume that prices, marketing mix variables and brand dummies completely characterize the observable differences between goods. A separate literature assumes the relevant attributes are unobserved. These studies use factor models to recover the latent attributes (Elrod 1988, Elrod and Keane 1995 and Erdem and Winer 1998).

ple discreteness problem, these benchmarks predict lower price and marketing responses than the proposed model. The own-price elasticities are almost all below one in magnitude, contrary to most economic theories of static profit-maximization. In contrast, the proposed model yields elasticities that are all greater than one and, thus, consistent with standard economic theories of static competition. In terms of the characterization of multiple-item shopping, I find substantial evidence of both observed and unobserved heterogeneity. Demographics play a significant role in determining differences in tastes in addition to identifying differences in the assortment of total purchases on each trip.

Finally, I apply the model to quantify product differentiation in the CSD industry. The parameter estimates for the marginal utilities of attributes and the controls for heterogeneity provide an overall view of how consumers perceive products. I use estimated demand to study the impact of differentiated product lines on firms' profitability, consumer demand for variety and the market equilibrium. I analyze the degree to which Pepsi's citrus product, Mountain Dew, increases Pepsi's profitability, net of cannibalizing the sales of colas. I also study how the presence of this citrus product affects overall equilibrium prices. I combine estimated demand with a model of supply to simulate the counterfactual scenario in which I remove Mountain Dew from the choice set. I find that carrying Mountain Dew not only adds to Pepsi's overall profits, it also provides unilateral market power to the entire Pepsi product line, allowing it to raise its cola prices. However, Mountain Dew also draws customers away from Coke and Dr. Pepper, driving their prices down and indirectly increasing competition for colas. Overall, the addition of Mountain Dew increases the profitability of Pepsi, mainly at the expense of its competitors.

The paper is organized as follows. The second section describes the model of individual choice and demonstrates its relationship to the standard DCM. In section 3, I discuss the econometric specification and the estimation procedure. Section 4 describes the data. In section 5, I present results including parameter estimates and substitution patterns. In section 6, I use an equilibrium model to discuss the competitive impact of Mountain Dew. Finally, I conclude in section 7.

The Model of Individual CSD Demand

The model

Several aspects of the multiple discreteness problem have been studied individually with the DCM and with alternative models. One line of research has examined the quantity decision for single-brand purchases using the Hanneman (1984) random utility model (Chiang 1991 and Chintagunta 1993), using models for count data (see Chintagunta 1993 for a survey of methods), and using a sequential model of demand (Krishnamurthi and Raj 1988)[17]. In addition to assuming single-brand purchases, these models have other limitations. The Hanneman approach assumes a perfectly divisible product.

[17] Individual demand is broken into two independent steps. First, a consumer chooses a product. Then, conditional on choosing that product, he/she chooses a quantity.

The count data approach does not apply to a multinomial choice framework[18]. Finally, the sequential model treats the brand and quantity choices as independent. Harlam and Lodish(1995) model the simultaneous choice of multiple brands using a variant of the DCM that does not make use of quantity information. None of these approaches address the full multiple discreteness problem.

My model of demand derives from the random profit framework in Hendel (1999). Hendel develops a static random profit model to account for firms' cross-sectional holdings of computers. For firms, the notion of variety derives from the presence of multiple potential computing tasks. For instance, a firm might be divided into several departments. Each department is assumed independently to select an integer quantity of one of the computer brands to fulfill its computing needs.

I modify this model to a random utility framework, suitable to address the consumer shopping problem. Instead of estimating a static product holdings model, I develop a purchase model in which households maximize a utility function subject to a budget constraint. Conditional on making a shopping trip, a household chooses products to satisfy various tasks. The households' tasks are expected future consumption occasions, which are unobserved by the econometrician. The source of these occasions varies from such factors as a large family with varying tastes, the replenishment of overall household CSD inventory, and uncertain future tastes. For each consumption occasion, the

---

[18]The model assumes the alternative chosen is predetermined, so that estimation focuses on the quantity choice. This approach would apply if, for instance, households always purchased some integer quantity of the same product each period.

household selects an integer quantity of one of the products. Since the consumption occasions are not observed, I simulate them. I assume the number of expected consumption occasions derives from a Poisson distribution whose mean is a function of household attributes and CSD inventories. The estimation procedure yields the expected total purchase vector for a shopping trip, aggregating across all the expected needs.

On a given shopping trip, a household h purchases a basket of various alternatives to satisfy J_h different future consumption occasions until the next trip. In fact, the actual number J_h is not observed by the econometrician. Instead, I assume that J_h derives from a distribution characterized by household demographics and its purchase history (inventory). Suppose the household's preferences are separable in its purchases of the I softdrink products available and a composite commodity of other goods. These preferences are assumed to be quasilinear. Finally, suppose the household spends y_h on the given shopping trip and let z denote a composite commodity. Conditional on J_h, the utility of household h at the time of a shopping trip is given by (I suppress the time index to simplify notation):

$$U^h = \sum_{j=1}^{J_h} u_j^h \left(\sum_{i=1}^{I} \Psi_{ij}^h Q_{ij}^h, D_h\right) + z, \tag{3.1}$$

where D_h is a $(d \times 1)$ vector of household characteristics, Q_{ij}^h is the quantity purchased of alternative i for consumption occasion j and Ψ_{ij}^h captures the household's valuation of alternative i 's attributes on consumption occasion j. This specification assumes additive separability of the J_h subutility functions, eliminating any valuation spillovers between consumption occasions. The individual subutility functions treat each of the goods in the product category as perfectly substitutable for a given consumption need. Perfect substitutability ensures that households select some quantity of a single alternative to satisfy each of its J_h expected consumption needs. Since households select an alternative for each expected consumption occasion, their aggregate purchase vector for a trip could still contain a variety of products. The household's expenditure constraint is given by:

$$\sum_{j=1}^{J_h} \sum_{i=1}^{I} p_i Q_{ij}^h + z \leq y_h.$$

where p_i is the price of product i. So long as the subutility functions satisfy the correct shape and continuity properties, the expenditure equation will be binding and may be substituted into the original utility function to give:

$$U^h = \sum_{j=1}^{J_h} u_j^h\left(\sum_{i=1}^{I} \Psi_{ij}^h Q_{ij}^h, D_h\right) - \sum_{j=1}^{J_h} \sum_{i=1}^{I} p_i Q_{ij}^h + y_h. \tag{3.2}$$

Conditional on the number of anticipated consumption occasions, J_h, the household's problem will be to pick a matrix with columns Q_j ($j = 1, ..., J_h$) to maximize (4.2).

The subutility functions for consumption occasions j are defined as:

$$\begin{aligned} u_j^h(\beta_j^h, D_h, X_j) &= \left(\sum_{i=1}^{I} \Psi_{ij}^h Q_{ij}^h\right)^{\alpha} S(D_h) - \sum_{i=1}^{I} p_i Q_{ij}^h \qquad (3.3) \\ \Psi_{ij}^h &= \max(0\ ,\ X_i \beta_j^h + \xi_i)^{m(D_h)} \end{aligned}$$

where X_i is a $(1 \times k)$ vector of brand i's observable attributes, β_j^h is a $(k \times 1)$ vector of random tastes for attributes for consumption need j, and ξ_i is an unobserved attribute which may be correlated with the price. This correlation could generate an endogeneity problem, an issue I discuss below. Since the actual number of consumption occasions, J_h, is not observed, the estimation procedure will only identify the mean and variance of the valuation of the attributes across the needs and household trips. The term Ψ_{ij}^h can be interpreted as the perceived quality of alternative i for consumption need j. The given specification explicitly allows for zero-demand (no purchase). The term $m(D_h)$ captures the taste for quality as function of the household's characteristics, permitting a vertical dimension in consumer tastes. Households with a larger value of $m(D_h)$ perceive a greater distance between the qualities of goods. $S(D_h)$ captures the effect of household characteristics on the scale of purchases. The α determines the curvature of the utility function, implicitly defining the rate of decreasing marginal returns. So long as the estimated value of α lies between 0 and 1, the model maintains the concavity property needed for an interior solution. Using a different model, Kim, Allenby and Rossi (1999) use a product-specific α to study potential differences in rates of diminishing returns across products.

The model captures household-level heterogeneity in several fashions. First, the tastes for quality, scale of purchases and the expected number of consumption needs (mean of the Poisson) are all functions of observed household characteristics. Heterogeneity enters the model in the form of a random coefficients specification in the quality function:

$$\beta_j^h = \widetilde{\beta} + \gamma D_h + \Omega \sigma_j^h$$

where $\widetilde{\beta}$ captures the component of tastes for attributes that is common to all households and consumption needs. The $(k \times d)$ matrix of coefficients, γ, captures the interaction of demographics and tastes for attributes. Finally, Ω is a diagonal matrix whose elements are standard deviations and σ_j^h is a $(k \times 1)$ vector of independent standard normal deviates. Thus, for each household, the taste vector will be distributed normally with, conditional on demographics, mean $\widetilde{\beta} + \gamma D_h$ and variance $\Omega\Omega'$. While I assume these coefficients are normally-distributed, I could have used several alternative specifications. McCulloch and Rossi(1996) outline a procedure using a mixture of normals to approximate a more flexible utility specification in the DCM framework. In fact, McFadden and Train(1998) show that the mixed multinomial logit approximates a fairly general class of parametric DCM utility functions. Given the complexity of the proposed model, I limit my attention to the normal distribution.

I assume households maximize (4.2). For a given expected consumption occasion, the household can compute the optimal quantity of each of the I products. Each of these optimal quantities has a corresponding utility, which is unobserved to the econometrician. Thus, for each household, there exists a vector of latent utilities, $u_j^* = (u_{j1}^*, \ldots, u_{jI}^*)$, where $u_{ji}^* = \max_Q u_j^h(\Psi_{ij}^h Q_{ij}, D_h)$ represents the utility from consuming the optimal quantity of product i for need j. The household selects the product yielding the highest latent utility for each occasion j. So, brand i is chosen to satisfy a given need if $u^{*}_{ji} = \max(u^{*}_{j1}, ..., u^{*}_{jI})$. Assuming that any continuous quantity is permissible, the optimal quantity of brand i for occasion j solves the first order condition:

$$\alpha \left(\Psi^{h}_{ij}\right)^{\alpha} \left(Q^{h}_{ij}\right)^{\alpha-1} S_h - p_i \doteq 0.$$

Rewriting the first order condition in terms of Q^{h}_{ij}, I obtain:

$$Q^{h*}_{ij} = \left(\frac{\alpha \left(\Psi^{h}_{ij}\right)^{\alpha} S_h}{p_i}\right)^{\frac{1}{1-\alpha}} \tag{3.4}$$

which is the optimal quantity of product i for consumption occasion j. To reformulate this problem to deal with integer quantities, I make use of the fact that the subutility functions are concave and monotonically increasing in Q_{ij}. Therefore, I only need to consider the next highest and next lowest integer quantity to Q^{h*}_{ij}. I then compare the $2 \cdot I$ potential quantities, picking the one yielding the highest utility. Households carry out this decision for each expected consumption occasion, selecting an optimal quantity for each. For each trip, I observe the sum of all of these optimal quantities in the form of an aggregate purchase vector.

My objective is to estimate the mean and variance of the distribution of the random coefficients, $\beta^{h} = (\beta^{h}_{1}, ..., \beta^{h}_{J_h})$, which are assumed to be distributed normally. I assume that the number of consumption needs in a given week, J_h, is distributed Poisson with the mean specified as a function of the household's characteristics and its purchase history, $\Gamma(D_h)$:

$$J_h \sim P(\Gamma(D_h)).$$

Given these assumptions, the overall expected total purchase vector for a given trip can be estimated conditional on the observable information and summed over the J_h consumption occasions:

$$EQ_h(D_h, X) = \sum_{J_h=1}^{\infty} \sum_{j=1}^{J_h} \int_{-\infty}^{\infty} Q_j^{h\bullet}(D_h, \beta_j^h, \Theta)\Phi(d\beta | D_h, \Theta) P(dJ_h(D_h)). \qquad (3.5)$$

Estimation will require specifying functional forms for $\Gamma(D_h)$, $m(D_h)$ and $S(D_h)$.

One limitation of this specification relates to the omission of price expectations. Households may defer some of their purchases if they anticipate better prices on a subsequent trip. Solving the dynamic program associated with the proposed utility model and price expectations is computationally infeasible (Gonul 1999 treats price expectations in a standard DCM). A simple, albeit crude, approximation to these expectations involves the inclusion of measures of consumer reference prices in the mean of the Poisson (see Kalyanaram and Winer 1995 for a survey). For instance, I could create a price index for the previous trip as well as for the current trip. I implicitly assume that consumers use past observations of shelf prices to form an internal reference point with which to assess the current prices (Winer 1985). By including the past and current price levels in the Poisson, I capture the fact that, if consumers find the current price level high (relative to their reference prices), then they may defer some of their consumption needs to a later trip. In practice, I expect past prices to have a positive impact and current prices to have a negative impact on the expected number of consumption occasions for which a household makes purchases.

Comparison with the Standard DCM

One of the main features of this random utility framework is that it is a direct extension of the standard discrete choice models. If I disregard the expected consumption occasions and I assume that consumers are restricted to single-unit purchases, then α no longer plays any role and (4.2) reduces to:

$$u_{hi} = X_i\beta S(D_h) - p_i, i = 1, ..., I.$$

In this formulation, I am no longer able to identify $m(D_h)$, so I set it to one for all households. I can divide through by $S(D_h)$ to get:

$$\widetilde{u_{hi}} = X_i\beta - \frac{1}{S(D_h)}p_i, \quad (3.6)$$

where the inverse of $S(D_h)$ is usually interpreted as the price-response parameter. Adding a random disturbance term directly in (4.6) gives the standard DCM (Manski and McFadden 1981).[19]

Since the relationship between the DCM and the proposed model involves removing the Poisson distribution altogether, I am not able to provide a direct statistical test to resolve the treatment of multiple-discreteness versus single-unit-purchasing. Such a test would involve the verification of the validity of the Poisson distribution versus a discrete distribution whose probability mass is degenerate at a constant. If the latter

[19]The random disturbance is generally assumed to derive from either the extreme value distribution, giving rise to the standard multinomial logit, or from the multivariate normal distribution, giving rise to the probit.

were the truth, then a household would always have the same deterministic fixed number of decisions to make each trip.

Although I am unable to test the validity of the DCM versus the proposed model formally, I formulate several variants of the DCM as benchmarks. In theory, I could recast the product space in terms of all the possible bundles of products. For instance, a product might be a unit of Coke and 2 units of Mountain Dew. However, this approach is computationally infeasible since I would have to consider all permutations of the 26 products and any positive integer quantity of each. Instead, I estimate two simple variants of the DCM. I provide a formal description of these models in the appendix.

In the first scenario, I ignore the quantity information and treat multiple product purchases on a given trip as independent draws from a multinomial logit. This model predicts brand holdings, but it underpredicts total demand since it assumes single-unit purchases - it ignores quantity information. The model also underestimates the no-purchase option since a trip on which no CSDs are purchased is always modeled as a single no-purchase. For instance, if a household makes three choices on one trip and then zero on the following trip, the model does not account for the fact that the latter trip might embody three no-purchase decisions. I also expect such a model to underpredict elasticities since the response to a price change is restricted to constitute either no change or a brand-switch. The model does not capture the cases in which a price-increase decreases a household's demand for a given product. While this first benchmark clearly fails in terms of its ability to predict total purchases, it provides a rough sense of the degree to which consumers substitute between products at the brand choice level.

In the second scenario, I use a sequential demand model in which consumers first choose the total number of CSD units and then decide how to allocate these units across brands. This approach is similar in spirit to the sequential model of Krishnamurthi and Raj(1988). I estimate the total number of CSD units purchased with a random effects Poisson model that accounts for household-specific heterogeneity[20]. I include an overall price index for the particular store-week to capture the effects of the price-level on unit sales. I then take the product of the estimated units purchased with the conditional brand probabilities from a multinomial logit brand choice model to assign each unit sale to a brand[21]. The model predicts the aggregate purchase vector of products for each store-trip. Despite the fact that the model may provide a good fit of the data, it is not structural (it does not derive from a theory of utility maximization) and I expect it to yield restrictive substitution patterns. In particular, an increase in any of the prices is constrained to have a negative impact on the total expected quantity decision for every product. This negative effect comes from the fact that a price increase enters the quantity choice as an increase in the overall price index. To illustrate the implications of this problem, suppose a customer who currently purchases three 12-packs of

---

[20]I assume the household-specific random effects are distributed normally, and I estimate the model using Gauss-Hermite polynomials.

[21]For technical convenience, I do not model unobserved heterogeneity in the brand choice model. Incorporating random coefficients in a model of 26 product alternatives is too computationally intensive.

Pepsi now faces a price increase of that product. Suppose the price increase causes the unit purchase decision to fall to two and the brand purchase decision to switch to a 6-pack of Pepsi (the 6-pack of Coke now has the highest brand choice probability). The model does not take into account the fact that two units of a 6-pack are fundamentally different from two units of a 12-pack. This property places a downward pressure on the cross-elasticities. In contrast, the proposed model from the previous section weighs the optimal quantity of a given product against the optimal quantities of the various alternatives. Thus, a customer could respond to the increase in the price of 12-packs of Pepsi by purchasing four units of 6-packs of Pepsi, for instance. Given the recent popularity of the aggregate logit in the empirical marketing and Industrial Organization literatures, I could also estimate a heterogeneous logit using the aggregate store-level data (see Dubé 2000). Unfortunately, the data are not sufficiently detailed to permit the estimation of models like those of Chiang (1991) and Chintagunta (1993a) which require information on the total shopping basket expenditure per trip.

Endogeneity of Prices

A standard estimation problem associated with the characteristics approach is the potential for unobservable attributes which may be correlated with the price. Nevo (2000) alleviates the potential endogeneity of prices by including alternative-specific dummy variables. These dummies enter the model in the quality function, Ψ_{hj}. Given the short time span of the data, 9 quarters, I estimate a single dummy for each product. I am assuming that any unobserved attribute that could be correlated with price does not vary over time. Unlike most papers using the characteristics approach, the inclusion of transaction-specific feature and display activity allows me to proxy for the time-varying store-specific attributes that could influence consumer perceptions of quality.

Recently, two studies have documented evidence of price endogeneity in several product categories at the weekly frequency, even after including the marketing mix variables. Besanko, Gupta and Jain (1998) find a downward bias on price responses using weekly store-level data and Villas-Boas and Winer (1999) find a similar effect in an individual shopping panel. For instance, unobserved changes in package design, television advertising and shelf space could still introduce variation in households' perceptions of a product's quality during the sample period and at the same time drive a product's price up. Adding more time-varying dummies, such as quarterly brand fixed effects, leads to an unmanageable proliferation of estimated parameters. Although not reported, attempts to use cost-shifters to instrument prices, as in Villas-Boas and Winer (1999), had very little effect on the estimated parameters. This outcome could imply that prices are not endogenous after controlling for fixed product effects. Unlike many of the product categories in my data set, the CSDs do not exhibit coupon usage, which could be the driving force of the weekly-frequency endogeneity. However, I suspect the lack of results lies in the poor quality of the instruments, which are unable to provide additional information simply due to the highly non-linear way in which prices enter the model.

For the rest of the paper, I assume the marketing mix variables account for intertemporal variation in quality.[22]

One of the difficulties of using product-specific dummies is the identification of mean tastes for fixed attributes. Since I wish to control for potential endogeneity of prices, I include a product-specific fixed-effect rather than a brand-specific fixed-effect as in Fader and Hardie (1997). In general, the mean taste parameters for the fixed physical attributes and the product dummy variables are not jointly identified in the GMM procedure. Although the taste parameters are not essential for the computation of elasticities or for computing equilibrium prices and quantities, they provide insight into the existing differentiation between products. I now show that I can still identify the tastes for fixed attributes, as in Nevo (2000), by using Chamberlain (1982)'s minimum distance procedure (also see Hsiao 1986).

I project the estimated product dummies onto a subspace spanned by the fixed product attributes. Suppose I partition the $(1 \times K)$ row vector of product attributes into K_1 fixed characteristics, x_{1i}, and K_2 time-varying characteristics $x_{2i} : X_i = [x_{1i}|x_{2i}]$. Corresponding to these vectors are the respective taste coefficient vectors β_1 and β_2. The GMM procedure does not identify the vector β_1. Define the $(J \times 1)$ vector of product dummy variables as $\delta = (\delta_1, ..., \delta_J)'$, the $(J \times K_1)$ matrix of fixed product characteristics as X_1=$[X'_{11}, ..., X'_{1J}]'$ and the $(J \times 1)$ vector of unobserved product attributes as

---

[22]Another possibility would be to devise a means by which to implement Berry's (1994) inversion procedure. I might improve the ability to instrument if, like Berry, I could write the objective function such that ξ enters linearly.

$\xi = (\xi_1, \ldots, \xi_J)'$. Further, suppose that the estimated values of δ have covariance matrix Σ. By construction, I can think of the product dummies as having the following structure in terms of the fixed attributes:

$$\delta = X_1\beta_1 + \xi.$$

The mean taste coefficients are estimated by projecting the estimated product dummy variables onto the subspace spanned by the product attributes:

$$\widehat{\beta_1} = \left(X_1'\widehat{\Sigma^{-1}}X_1\right)^{-1} X_1'\widehat{\Sigma^{-1}}\delta.$$

Implicit in this specification are the assumptions $E\left(\xi|X_1\right) = E\left(\xi X_1|X_1\right) = 0$. So, if I expect the unobserved characteristic, ξ, to be correlated with the observed fixed product characteristics, then I will not be able to identify the mean tastes. For instance, suppose the unobserved characteristic is a measure of the quality of the flavor. Then I would expect the mean independence to fail for such characteristics as diet if consumers feel that diet products taste worse than regular (although they might derive utility from the fact that the product has no calories). Alternatively, if the unobserved characteristic measures perceived quality of the brand name (due perhaps to long-run advertising efforts), then I would not expect to observe correlation with the observed ingredients. Unlike BLP (1995,1998), the validity of this assumption only affects my ability to recover the mean taste coefficients.

Estimation Procedure

During the derivation of the model of demand, I derived (4.5), the expected purchase vector for each household. I compute the equation for the vector of expected household h demand for each alternative at time t, conditional on the $(K \times 1)$ matrix of household/trip attributes, D_{ht}:

$$Q_{ht}(D_{ht},\Theta) = \sum_{i=1}^{\infty}\sum_{j=1}^{J_{ht}} \int_{-\infty}^{\infty} Q^{*}_{jht}(D_{ht},\beta^{h}_{j},\Theta)\Phi(d\beta|D_{ht},\Theta)P(dJ_{h}(D_{ht})),\ \text{h=1,...,H, t=1,...,T}_{h}$$

where $Q^{*}_{jht}(D_{ht},\beta^{h}_{j},\Theta)$ is the $(I \times 1)$ vector of optimal quantities of each alternative for occasion j on trip t, β^{h}_{j} is a vector of random taste coefficients for task j, and Θ is a vector of parameters to be estimated. The random taste coefficients are drawn from the distribution $\Phi(\bullet|D,\Theta)$, conditional on the information D, and the number of expected consumption occasions are drawn from the distribution $P(\bullet|D)$. In the current context, these distributions are assumed to be normal and Poisson, respectively. Thus, the vector of expected soft drink purchases for each household is the sum of the expected purchases across consumption occasions, conditional on a specific number of tasks, J_{ht}, then weighted by the probability that J_{ht} is the true number of tasks at time t.

Using this formulation, I define the prediction error:

$$\varepsilon_{ht}(D_{ht},\Theta) = Q_{ht}(D_{ht},\Theta) - q_{ht} \qquad (3.7)$$

where q_{ht} is the vector of actual purchases of each of the alternatives by household h at time t. If the model represents the true purchasing process, then at the true parameter values, Θ_0 :

$$E\left\{\varepsilon_{ht}\left(D_{ht},\Theta_0\right)\right\} = \overrightarrow{0_I} \text{ for h=1,...,H and t=1,...,T}_h. \tag{3.8}$$

I also assume that:

$$E\left\{\varepsilon_{ht}\left(D_h,\Theta_0\right)\varepsilon_{hk}\left(D_{ht},\Theta_0\right)'\right\} = \Omega_{tk}, \tag{3.9}$$

where Ω_{tk} is a finite $(I \times I)$ matrix. This assumption implies that the households' prediction errors are distributed identically. Following Hansen (1982) and Chamberlain (1987), any function of the observable data, D_{ht}, that is independent of the unobservables must be conditionally uncorrelated with ε_{ht} at $\Theta = \Theta_0$. I am assuming that the process generating the prediction errors, the demand shocks, is uncorrelated with the point-of-purchase marketing environment. For instance, newspaper advertising for a product does not drive the residual process. Given such a function, $Z_{ht} = f(D_{ht})$, I can construct conditional moments:

$$E\left\{Z_{ht} * \varepsilon_{ht}\left(D_{ht},\Theta_0\right)|Z_{ht}\right\} = \overrightarrow{0_I}. \tag{3.10}$$

From these orthogonality conditions, I can construct the moment conditions:

$$h(D_{ht},\Theta) = Z_{ht} * \varepsilon_{ht}\left(D_{ht},\Theta\right)$$

where $\Theta \in R^k$, and (4.12) implies that $E\left\{h\left(D_{ht},\Theta_0\right)|Z_{ht}\right\} = 0$. Let $\mathbf{D}_{HT} \equiv (D'_{1T1}, ..., D'_{HT_H})$ denote the matrix containing all of the household/trip information for the sample of H households, where household h makes T_h shopping trips. Using the notation $T=\frac{1}{H}\sum_{h=1}^{H} T_h$, the sample analogue of the moment conditions has the following form:

$$g(D_{HT},\Theta) = \frac{1}{HT}\sum_{h=1}^{H}\sum_{t=1}^{T_h} h\left(D_{ht},\Theta\right). \tag{3.11}$$

As H and T_h grow large, g(D_{HT}, Θ_0) should approach zero. Hansen's (1982) formulation involves finding a value of Θ_{GMM} that makes g(D_{HT}, Θ_{GMM}) as close as possible to the population moment of zero. Therefore, I choose a value of Θ_{GMM} that minimizes the function J_{HT} given by:

$$J_{HT}(\Theta) = [g(D_{HT}, \Theta)]' W_{HT} [g(D_{HT}, \Theta)] \tag{3.12}$$

where W_{HT} is generally the efficient weighting matrix given by the asymptotic variance of g. The estimation of W is discussed below. This framework gives estimates with the following asymptotic distribution:

$$\sqrt{N}(\Theta_{GMM} - \Theta_0) \implies N(0, \Xi) \tag{3.13}$$

$$\Xi = \left(plim \left\{ \frac{dg(D_{ht}, \Theta_0)}{d\Theta} \right\} W plim \left\{ \frac{dg(D_{ht}, \Theta_0)}{d\Theta} \right\}' \right)^{-1} \tag{3.14}$$

In order to compute the sample moment conditions, I must evaluate an infeasibly large dimensional integral. Moreover, the regions of integration are not easily solved analytically. Following McFadden (1989) and Pakes and Pollard (1989), I use Monte Carlo methods to simulate these integrals. This approach also solves the problem of determining the region of integration. For each household trip, R independent draws are taken from the Poisson distribution to simulate the number of expected consumption occasions. For each of these R draws, $(N + I - 1) \times K$ draws are taken from the normal distribution to simulate the taste coefficients for these occasions, where K is a sufficiently large number to place an upper bound on the number of occasions simulated for each household. These draws are then used to construct R simulations of the expected purchase vector at each trip, $Q^r_{ht}(D_{ht}, \Theta)$ $r = 1, ..., R$. These estimates are then combined to form an unbiased simulator of the expected purchase vector, $\widehat{Q_{ht}}(D_{ht}, \Theta)$:

$$\widehat{Q_{ht}}(D_{ht}, \Theta) = \frac{1}{R}\sum_{r=1}^{R} Q^r_{ht}(D_{ht}, \Theta).$$

By construction, the simulated values, Q^r_{ht}, derive from the same distribution as q_{ht}. So the variance of $\widehat{Q_{ht}}(D_{ht}, \Theta)$ will be $\frac{1}{R}var(q_{ht})$, which goes to zero as $R \to \infty$. I can write $\widehat{Q_{ht}}(D_{ht}, \Theta) = Q_{ht} + \zeta_{ht}$, where ζ_{ht} is the simulation error and $E(\zeta_{ht}) = 0$ and $var(\widehat{Q_{ht}}) = var(\zeta_{ht})$. I now simulate the moment conditions by substituting $\widehat{Q_{ht}}(D_{ht}, \Theta)$ for $Q_{ht}(D_{ht}, \Theta)$ in (4.14):

$$g^s_{HT}(\Theta) = \frac{1}{HT}\sum_{h=1}^{H}\sum_{t=1}^{T_h}\left[Z_{ht} * \left(\widehat{Q_{ht}}(D_{ht}, \Theta) - q_{ht}\right)\right] = \frac{1}{HT}\sum_{h=1}^{H}\sum_{t=1}^{T_h} h^s\left(D_{ht}, \Theta\right) \tag{3.15}$$

So long as H is sufficiently large, the resulting method of simulated moments estimate, Θ_{MSM}, will be consistent and will have asymptotic variance $\Xi = \left(\frac{dg^s(\Theta_0)'}{d\Theta} W_{HT} \frac{dg^s(\Theta_0)}{d\Theta}\right)^{-1}$. In the next section, I discuss the estimation of the weight, W_{HT}.

Estimation of the Weight Matrix, W:

The estimation of W_{HT} will be complicated due to both the simulation error and the panel aspect of the data. The simulation error will simply add extra variation to the procedure, as demonstrated below. The panel aspect of the data requires some additional assumptions regarding both cross-sectional and intertemporal variation of the residual process. I include several state variables, such as temperature and seasonal dummies, to capture contemporaneous aggregate demand shocks that could affect households in a similar fashion. Having included these controls, I assume that the prediction errors are uncorrelated across households. However, most households have fairly long purchase histories, allowing the possibility of persistent unobserved shocks. The source of these shocks could be measurement error. For instance, household-specific reporting errors in the scanning process could generate unobserved serial dependence. By including observed time-varying factors in the mean of the Poisson function, I assume that this serial dependence is independent of the process generating the number of consumption needs. Therefore, only the covariances of the prediction errors need to be corrected.

Hansen (1982) shows that, under certain regularity conditions, the efficient weighting matrix W_{HT} is the inverse of S, the variance of the sample moments:

$$\begin{aligned}
S &= \lim_{HT\to\infty} HT \cdot E\left\{E\left([g(D_{HT}, \Theta_0)]\,[g(D_{HT}, \Theta_0)]'\,|D_{HT}\right)\right\} \\
&= \lim_{HT\to\infty} HT \cdot E\left\{E\left(\left[\frac{1}{HT}\sum_h\sum_t h^s(D_{ht}, \Theta_0)\right]\left[\frac{1}{HT}\sum_h\sum_t h^s(D_{ht}, \Theta_0)\right]'|D_{HT}\right)\right\} \\
&= \lim_{HT\to\infty} HT \cdot \frac{1}{H^2T^2}\sum_{h=1}^{H}\sum_{t=1}^{T_h}\sum_{k=1}^{T_h} E\left\{E\left(\left[Z_{ht}\left(\widehat{Q_{ht}} - q_{ht}\right)\right]\left[Z_{hk}\left(\widehat{Q_{hk}} - q_{hk}\right)\right]'|Z_{ht}, Z_{hk}\right)\right\} \\
&= \lim_{HT\to\infty} \frac{1}{HT}\sum_{h=1}^{H}\sum_{t=1}^{T_h}\sum_{k=1}^{T_h} E\left\{E\left([Z_{ht}\varepsilon_{ht}\varepsilon'_{hk}Z'_{hk} + Z_{ht}\zeta_{ht}\zeta'_{hk}Z'_{hk}]\,|Z_{ht}, Z_{hk}\right)\right\} \\
&= \lim_{HT\to\infty} \frac{1}{HT}\sum_{h=1}^{H}\sum_{t=1}^{T_h}\sum_{k=1}^{T_h} E\left[Z_{ht}\Omega_{tk}Z'_{hk} + Z_{ht}\frac{1}{R}\Omega_{tk}Z'_{hk}\right] \\
&= \lim_{HT\to\infty} \frac{1}{HT}\sum_{h=1}^{H}\sum_{t=1}^{T_h}\sum_{k=1}^{T_h} E\left[\left(1 + \frac{1}{R}\right) Z_{ht}\Omega_{tk}Z'_{hk}\right].
\end{aligned}$$

Similar to the discussion in McFadden (1989), the added simulation "noise" will not affect the consistency of the estimator, but it will reduce the efficiency by a factor of $\left(1+\frac{1}{R}\right)$. As $R \rightarrow \infty$, the estimator approaches asymptotic efficiency. In this paper, I use thirty simulation draws ($R = 30$) and assume that this will be sufficient to eliminate any added simulation noise.

I now address how the panel aspect of the data enters the estimation of S. To model the residual process more formally, I assume the values of a given household's prediction error on a given trip are determined by the values of an underlying random field, ε_s, at location s_{ht} on a lattice H. I index each observation's location by both time and household. I then allow for serial dependence between observations depending on their relative locations on the lattice H. Technically, I could allow for dependence both across households and over time. As discussed above, I only treat intertemporal dependence to simplify the estimation procedure.[23] Conley (1999) provides limiting distributions and covariance estimation techniques for this more general setting. I use Conley's non-parametric, positive semi-definite covariance estimator which is analogous to the spectral time-series estimator of Newey and West (1987). Given a consistent estimate $\widehat{\Theta}$ and a predetermined time L after which the unobserved household-specific shocks die out, the estimator for S is:

---

[23]Intuitively, we do not expect the unobservables generating a given household's choice process to affect other "close" households' choice processes for a given product category. However, we do expect some such "spatial" dependence for the overall shopping choice. For instance, households' store choices may be affected by local convenience stores. This form of dependence is the subject of work in progress, joint with Tim Conley of Northwestern University.

$$\widehat{S}_{HT} = \frac{1}{HT}\sum_{h=1}^{H}\sum_{t=1}^{L}\sum_{k=t+1}^{T_h}\omega(t)\left[h^s\left(D_{h,k},\widehat{\Theta}\right)h^s\left(D_{h,k-t},\widehat{\Theta}\right)' + h^s\left(D_{h,k-t},\widehat{\Theta}\right)h^s\left(D_{h,k},\widehat{\Theta}\right)'\right]$$
$$-\frac{1}{HT}\sum_{h=1}^{H}\sum_{k=1}^{T_h}h^s\left(D_{h,k},\widehat{\Theta}\right)h^s\left(D_{h,k},\widehat{\Theta}\right)'$$

where $\omega(t)$ is a weight with the following form:

$$\omega(t) = \begin{cases} 1-\frac{|t|}{1+L} \text{ if } l \leq L \\ 0, \text{ else} \end{cases}.$$

This scheme, using the Bartlett weight, assigns decreasing weight to the correlation between a given household's purchases as they grow further apart in time.

The resulting estimate for W is $\widehat{W}_{HT} = \widehat{S}_{HT}^{-1}$. Consequently, the estimate for the covariance matrix of the estimated parameters is $\frac{1}{HT}\cdot\left(\frac{dg^s(\Theta_{MSM})}{d\Theta}'\widehat{W_{HT}}\frac{dg^s(\Theta_{MSM})}{d\Theta}\right)^{-1}$.

Identification

I now discuss several identification issues for the proposed econometric procedure. First, I explain how the data identify the joint distribution of the total number of products and the total number of CSD units purchased on a given trip. Then, I explain how I identify the residual process and the GMM weight in the presence of a large number of moment conditions.

Since I do not observe the individual needs on a given trip, I estimate aggregate demand per trip. Despite the fact that I do not observe the specific needs, I am still able to identify the process that generates them. The main identification problem involves the distinction between a household purchasing 5 units of CSDs to satisfy five expected needs versus 5 CSDs to satisfy a single expected need. Since the random tastes are independent across consumption needs, a household with several needs will tend to purchase several different types of CSDs. Alternatively, a household with a single consumption need will only purchase one type. Thus, the number of consumption needs will determine the joint distribution of the total number of units of CSDs purchased and the number of different brands.

For example, I find that both the total number of CSDs and the number of different types of CSDs purchased on a trip increases with the size of the household. Therefore, household size enters both the scale function, $S(D)$, and the mean of the Poisson, $\lambda(D)$. Since the function $S(D)$ enters the per-task optimal quantity choice in (4.4), it helps identify $\lambda(D)$ and total quantity per consumption need. Similarly, the use of demographic variables in determining $m(D)$ in (4.3) enables the joint identification of $\lambda(D)$ and the taste parameters, β. Although several different sets of parameter values could give the same likelihood for expected total purchases, they will not have the same likelihood for the joint distribution of total products and total units purchased. Since the sample households tend to purchase baskets containing several different CSD brands, the data will identify this joint distribution.

The assumed independence of tastes across consumption needs rules out potential externalities. This assumption seems less of a problem for CSDs than for the purchase of computers, for which there could be obvious shared software-related externalities. Nonetheless, the fact that a consumer has already purchased a cola to satisfy one need might increase the likelihood of purchasing a non-cola to satisfy another need. One way in which I could link the choices made during a given trip would be to introduce interaction dummy variables in the utility function. For instance, I could classify all the CSDs in the sample into five flavor groups. While simulating the contemporaneous choices, I would introduce flavor interaction terms that would reflect which flavor combinations have been selected across needs. In addition to providing a link across the consumption needs, these flavor interaction terms would also provide a statistical test for complementarities between flavors. The test would be a simple significance test for whether a given pairwise flavor combination has a positive, negative or zero effect on overall utility.

With regards to the estimated residual process, I find that correcting for a 15-day lag changes some of the reported standard errors by as much as a factor of 1.8. However, given the large number of products and instruments, I end up with a large number of moment conditions. If I estimate the covariance matrix freely, I could run into some trouble with identification. For now, the only restrictions I impose are the second moment independence of the instruments and the errors. Even so, with 26 products I still estimate the (26×26) residual covariance matrix, Ω, and a $(K \times K)$ instrument covariance matrix, $E(Z_{h,t}Z'_{h,t+l})$, for each lag l. The fact that the lag structure has such a large effect on the current results motivates the need for a time-series structure. For precision, I may need to impose some additional restrictions on subsequent estimations. One way to think about valid restrictions is to imagine the source of these shocks. For instance, households may randomly shop at a non-sample store, such as a convenience store. I expect this sort of measurement error to exhibit some persistence. However, the persistence may only be for products of the same size. So, the fact that a household purchases a 67.6 ounce bottle in a convenience store may only affect the prediction of other 67.6 ounce bottles. In this case, I could set some of the off-diagonal terms between different size products in the autocovariance matrices to zero to improve the identification.

Data

The scanner data, collected by A.C. Nielsen, cover the Denver area between January of 1993 and March of 1995. These data includes consumer information for a random sample of 2108 households as well as weekly store-level information for 58 supermarkets with over $2 million *all commodity volume*. The store level information consists of weekly prices, sales, feature and display activity for 26 diet and regular products with a combined share of 51% of the household-level category sales. The list of 26 products consists of all UPCs with at least a 1% share of total sample CSD volume each. The household level data cover all shopping trips for these items. For each trip, I know the date, the store chosen and the quantities purchased. For each alternative available within the store, I know the prices and whether the product was featured in a newspaper or as an in-store display. Combining the store and purchase data sets, I observe the full set of prices and marketing mix variables for all the alternatives on a given trip.

By focusing on such a large scope of products, I include a variety of package sizes. In general, the price differences between sizes are more sophisticated than simple volume differences. I find that the 67.6 ounce bottles generally have a much lower per-ounce price than the larger-volume six and twelve packs of cans. This phenomenon, *quantity surcharging*, may be attributable to lower production costs associated with plastic bottles versus aluminum cans. However, several supermarket studies link such differences between prices and product sizes to producer price discrimination (Agrawal, Grimm, and Srinivasan 1993 and Cohen 2000). For soft drinks, such price discrimination could reflect consumers' heterogeneous valuations for the differences in storability of these different size products. While the study of such price discrimination schemes is beyond the scope of this paper, I will treat separate sizes of a given brand as different products.

For each shopping trip, I construct a quality measure for each product. The quality consists of three components: fixed physical attributes, time-varying attributes and household-specific loyalty. The fixed physical product attributes consist of the ingredients of the product, which I collect from the nutritional information printed on the product packages. These characteristics include total calories, total carbohydrates, sodium content (in mg), and a set of dummy variables that indicate the presence of caffeine, phosphoric acid, citric acid, caramel color and clear. I report these attributes as per-12-ounce-serving, using 3 additional dummy variables to distinguish between package sizes: 6-pack of 12 oz cans, 12-pack of 12 oz cans, and 6-pack of 16 oz bottles (I omit

67.6 oz bottles since I include a constant term). The time-varying attributes are prices and marketing mix variables, feature and display. Finally, the household-specific loyalty variables are two dummy variables indicating whether the same brand and same UPC respectively were chosen during the most recent shopping trip on which a purchase occurred.

In the appendix, I provide summary statistics of the demographic variables and time-varying product attributes used in the estimation. I also provide a discussion of the fixed attributes to illustrate how attributes help identify product differentiation. Tables (3.49 and 3.50) break down the fixed characteristics by flavor group, providing a rough sense of the relative positions of the different products in attribute space. To illustrate the contribution of these fixed characteristics, consider the difference between diet and regular versions of Coke. Diet Coke contains citric acid, whereas regular Coke does not. In eliminating the calorie content, other ingredients have been added to the cola recipe to recreate the flavor. The use of a citric acid dummy will help control for the possibility that households do not perceive Diet Coke as a simple zero-calorie version of regular Coke. The non-diet colas, lemon/lime and pepper drinks are quite similar with around 150 calories on average. The rootbeers and the citrus beverages are substantially higher, with about 170, and the new age are substantially lower, with 120. Phosphoric acid is used in all regular colas, all the diet colas and in all the peppers. While citric acid is found in all the fruit drinks, it is also used in many of the other products. The caramel and clear attributes span all of the products except for the citrus, which are yellow.

Flavor		continuous variable		
Flavor		calories	sodium (mg)	carbohydrates
cola	regular	150 (7.5)	40.5 (7.4)	41 (1.51)
	diet	0 (0)	34.7 (8.7)	0 (0)
lem/lime	regular	143.3 (5)	61.7 (16.4)	38.333 (0.5)
	diet	0 (0)	35 (0)	0 (0)
rootbeer	regular	168.3 (4.1)	44.2 (14.6)	44.8 (1.5)
citrus	regular	170 (0)	70 (0)	46 (0)
pepper	regular	148.6 (3.8)	45.7 (8.9)	35.1 (15.5)

Table 3.12. Continuous Attributes by flavor and diet vs. regular (averages)

		indicators					
Flavor		caffeine	phos.	citric	caramel	clear	#
cola	regular	7	7	4	7	0	7
	diet	6	9	9	9	0	9
lemon\lime	regular	0	0	2	0	2	2
	diet	0	0	1	0	1	1
rootbeer	regular	0	0	0	1	0	1
citrus	regular	3	0	3	0	3	3
pepper	regular	3	3	0	3	0	3

Table 3.13. Indicator Attributes by flavor and diet vs. regular (counts)

Results

Parameter estimates

I now present empirical findings for the proposed model and compare them to the two benchmarks. I correct the reported standard errors of the proposed model for within-household dependence. Since I estimate the benchmarks using maximum likelihood, I am not able to perform a similar correction for these models.

I present parameter estimates for five specifications of the proposed model. In the second model, I add a random intercept to the Poisson process. In the third model, I also add demographic interaction terms. In the fourth and fifth models, I also add current and lagged price indices to capture reference price effects that might determine how customers form expectations (the fourth model does not have demographic interactions). First, I focus on the taste coefficients that enter the quality function, ψ. Table (3.14) presents the tastes for time-varying attributes. The tastes do not change dramatically across the various specifications. However, the relative magnitudes of the means and standard deviations of the tastes for features and displays are not entirely robust across the specifications. As expected, both ad and display have a strong positive influence on perceived product quality. Moreover, there appears to be a substantial dispersion in the degree to which households are influenced by these variables, especially the ad variable. Additional *a priori* information regarding expected interactions between advertising

Time-Varying Attributes	Model 1	Model 2	Model 3	Model 4	Model 5
ad	1.13	0.66	0.74	2.07	1.92
	(0.02)	(0.02)	(0.03)	(0.07)	(0.04)
s.d. ad	0.53	0.03	0.04	0.71	0.74
	(0.02)	(0.02)	(0.03)	(0.06)	(0.04)
display	0.95	3.29	3.12	2.34	1.94
	(0.02)	(0.07)	(0.05)	(0.10)	(0.06)
s.d. display	0.19	0.57	0.62	0.21	0.23
	(0.01)	(0.07)	(0.03)	(0.06)	(0.02)
brand loyalty	2.28	3.56	5.55	3.00	3.62
	(0.04)	(0.06)	(0.06)	(0.13)	(0.10)
prod. loyalty	0.94	1.25	1.19	0.99	1.00
	(0.07)	(0.31)	(0.14)	(0.22)	(0.11)
Obs	169,788	169,788	169,788	169,788	169,788

Table 3.14. Taste Coefficients for Time-Varying Attributes in the Quality Function

and demographics could provide additional interaction terms to help explain part of this dispersion. The results suggest that loyalty to a specific brand might be stronger than loyalty to a given UPC. For instance, consumers are slightly more loyal to Coca-Cola in general than to a specific package size of Coca-Cola.

Table (3.15) presents the estimates of the tastes for fixed product attributes, including means, standard deviations and the demographic interactions. Recall that I estimate the standard deviation and demographic interaction terms directly in the GMM procedure, whereas I estimate the mean tastes using the minimum distance procedure. The qualitative results of these tastes are robust across specifications, with the exception of no-color, which is negative for the third model. Despite some striking difference which turn up below in the non-linear parameters, I find that the model specification does not lead to a drastic change in the predictions for how consumers value product attributes.

The models predict significant unobserved heterogeneity in consumer perceptions of product-specific quality (the intercept term). Ideally, I would interact the product dummies with demographics to try and characterize these differences in perception. However, these interactions would require too many additional parameters. Instead, I focus on specific product attributes to explain some of these differences.

Previous marketing studies have found seemingly contradictory results with regards to the effects of demographics. With a few exceptions, the research generally finds weak and inconsistent effects (discussed in Kalyanam and Putler 1997). However, most studies use low-purchase incidence categories with a small number of fairly homogeneous products, such as canned tuna fish, coffee, saltine crackers and ketchup. Moreover, most studies condition their models on the occurrence of a purchase, ignoring the purchase versus no-purchase decision.

A fair evaluation of the explanatory power of demographics requires a category in which one would expect some form of segmentation. In the current application, I use demographics to explain well-documented demographic taste differences for CSDs. As expected, households with a female head under 35 years old tend to have higher preferences for diet products[24]. In fact, I might find additional explanatory power from dummies such as female head with a college degree.[25] Similarly, larger households place

---

[24]In Europe, Diet Pepsi was reintroduced as Pepsi Max, with twice the caffeine, to overcome its "feminine" image.

[25]This fact is documented in "Just who's buying all these soft drinks, anyway?" *Beverage Industry*, 84(3), 1993.

slightly more weight on products with more 12-ounce servings, such as the 12-pack. Unexpectedly, households with kids place a higher weight on products with caffeine than those without. Part of this effect may be due to the limited scope of products included. In particular, many of the caffeine-free products, such as 7UP and Sprite, tend to appeal more to adults. Despite controlling for demographic interactions, I still find evidence for unobserved heterogeneity in tastes for package size (number of 12 ounce servings) and diet, suggesting that demographics alone are insufficient to explain taste differences. Moreover, the addition of the reference prices reduces the impact of these interactions tremendously.

On average, households dislike diet products, although they exhibit some variation in this taste. Moreover, households vary tremendously in their tastes for the various product sizes.

Now I present the non-linear terms that help determine the other features of the model. I assume a simple linear form for these terms:

$$\begin{aligned}
\lambda_h &= \lambda_0^h + \lambda_1 kids + \lambda_2(family\ size) + \lambda_3(last\ trip) \\
&\quad + \lambda_4(last\ csd\ trip) + \lambda_5 temperature + \lambda_6 holiday \\
&\quad + \lambda_7\,(lag\ fav.\ prods) + \lambda_8(fav.\ prods) + \lambda_9\,(price\ index) \\
scale &= s_0 + s_1(family\ size) + s_2(last\ trip) + s_3(last\ csd\ trip) \\
m &= 1 + m_1 income.
\end{aligned}$$

Table (3.25) presents the estimated coefficients. Similarly to Kalyanam and Putler (1997),

Physical Attributes	Model 1	Model 2	Model 3	Model 4	Model 5
constant	4.56	5.65	16.60	0.48	14.30
	(0.83)	(1.94)	(1.18)	(2.67)	(1.93)
s.d. product dummy	1.47	3.19	3.37	2.70	2.97
	(0.02)	(0.04)	(0.03)	(0.05)	(0.05)
diet	-2.76	-4.64	-13.26	-0.24	-11.15
	(0.70)	(1.65)	(0.99)	(2.27)	(1.68)
s.d. diet	0.79	0.57	0.63	0.70	0.68
	(0.02)	(0.02)	(0.03)	(0.07)	(0.02)
sodium	-0.02	-0.04	-0.06	-0.02	-0.05
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
carbs	-0.05	-0.09	-0.30	0.01	-0.24
	(0.02)	(0.04)	(0.02)	(0.06)	(0.04)
caffeine	0.53	1.42	1.31	1.24	1.07
	(0.02)	(0.04)	(0.03)	(0.07)	(0.04)
phos.	-0.76	-1.40	-2.82	-0.82	-1.82
	(0.11)	(0.28)	(0.18)	(0.37)	(0.27)
citric	0.03	0.12	0.14	0.04	0.06
	(0.02)	(0.06)	(0.04)	(0.10)	(0.08)
caramel	0.48	1.24	1.05	1.33	0.06
	(0.05)	(0.12)	(0.11)	(0.23)	(0.17)
no color	0.18	0.93	-0.63	1.84	-0.29
	(0.13)	(0.28)	(0.18)	(0.41)	(0.30)
cans×6	0.56	1.75	2.21	1.20	1.62
	(0.02)	(0.03)	(0.03)	(0.06)	(0.05)
s.d. cans×6	0.99	1.09	1.02	1.05	0.94
	(0.02)	(0.06)	(0.03)	(0.08)	(0.04)
cans×12	0.11	0.62	0.49	0.09	0.30
	(0.02)	(0.03)	(0.03)	(0.06)	(0.04)
s.d. cans×12	0.58	0.04	0.04	0.53	0.49
	(0.01)	(0.01)	(0.03)	(0.06)	(0.02)
bott×6	-0.09	1.91	2.26	1.39	1.32
	(0.07)	(0.13)	(0.08)	(0.13)	(0.13)
s.d. bott×6	1.68	0.15	0.14	0.21	0.18
	(0.08)	(0.05)	(0.11)	(0.19)	(0.02)
$kids * caffeine$			0.25		0.20
			(0.01)		(0.02)
$(household\ size) * servings$			0.02		0.02
			(0.00)		(0.00)
$(female\ head < 35) * diet$			0.44		0.47
			(0.03)		(0.04)

Table 3.15. Taste Coefficients for Fixed Attributes in Quality Function

I find that demographics identify product holdings. However, I do not find these effects to be very robust across model specifications. Since these terms enter in a highly non-linear way, it is not entirely surprising that small changes in the specification would change how they interact with one-another.

I find that the expected number of needs depends on the presence of kids and on family size. The rationale for these variables is that the number of needs depends on the number of individuals in a household and the fact that kids may have different needs from adults. Thus, larger households with kids should purchase more types of products. However, both these effects decrease in magnitude in the final specification. Holiday weeks (such as Christmas, Labor Day and Memorial Day) exhibit the expected large positive effect on needs. Surprisingly, the time since last trip and the time since last CSD purchase variables do not appear to explain the number of needs on a given trip[26]. The inclusion of the proxies for reference prices have very strong effects. In particular, high prices in the previous week and low prices in the current week should increase the number of expected needs. Although crude as a measure of expectations, this evidence is consistent with the notion that households use high past prices and low current prices to forecast high prices again in the future, leading them to purchase more today.

The scale of purchases are also increasing in the number of people in the household. Therefore, one would expect larger households to purchase more units of a given product. Surprisingly, time since last trip and time since last CSD purchase only become

[26]In a previous version, I found similar insignificant results when I used an inventory measure.

relevant with the inclusion of prices in the mean of the Poisson. The strong positive effects are consistent with the notion that household inventories have depleted and need replenishing. I am unable to explain why these terms only matter in the with the inclusion of proxies for reference prices. Perhaps the ability of these proxies to capture the effects of price expectations on the scale of purchases allows me to distinguish the effects of inventories. Even with depleted inventories, a household may not always purchase on a larger scale if the current price level is high. The vertical component is increasing in income, suggesting that households with higher income perceive more distance between products. Finally, the estimated values of α are positive and below one, which is consistent with the notion that the utility function is concave.

The reported standard errors have been corrected to account for potential time-series dependence. I attempt to control for as much of the observed potential dynamic factors such as timing of trips, loyalty and inventories. Despite these controls, I still find unexplained persistence in the residuals. Accounting for time-series increases some of the standard errors by as much as a factor of 1.8. Nonetheless, almost all the parameters remain significant after this correction, probably due to the extremely large sample. For now, I have not derived an explicit source for this persistence. As an experiment, I recompute the residuals after setting all of the coefficients for the dynamic factors to zero. I find that the standard errors rise about 50% on average, some almost double. Therefore, the dynamic controls are still picking up a fair bit of the intertemporal effect. Next, I take the actual residuals and average them by product for each household over

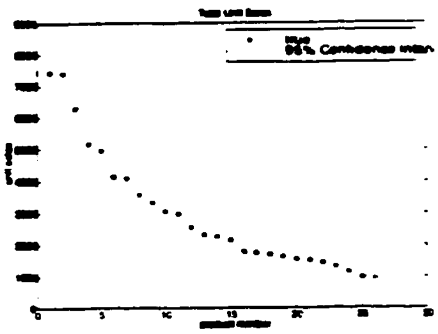
variable	Model 1	Model 2	Model 3	Model 4	Model 5
lambda: constant		0.078	0.083	0.128	0.001
		(0.002)	(0.006)	(0.012)	(0.000)
lambda: kids	0.076	0.139	0.134	0.113	0.001
	(0.005)	(0.002)	(0.005)	(0.016)	(0.001)
lambda: family size	0.060	0.001	0.001	0.032	0.033
	(0.002)	(0.000)	(0.000)	(0.003)	(0.001)
lambda: time since last csd	0.001	-0.001	-0.001	-0.003	-0.002
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
lambda: time since last trip	-0.001	-0.001	-0.001	-0.004	-0.004
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
lambda: temperature	0.001	-0.000	-0.000	0.004	0.004
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
lambda: holiday	0.005	0.170	0.163	0.142	0.136
	(0.002)	(0.003)	(0.006)	(0.010)	(0.005)
lambda: fav. prods				-1.110	-1.010
				(0.261)	(0.118)
lambda: overall prices				-1.016	-1.080
				(0.262)	(0.150)
lambda: lag fav. prods				0.921	0.980
				(0.166)	(0.049)
lambda: random term		0.052	0.050	0.459	0.005
		(0.001)	(0.003)	(0.015)	(0.001)
scale: constant	1.825	-1.008	-0.876	-0.432	-0.004
	(0.069)	(0.066)	(0.022)	(0.067)	(0.019)
scale: family size	1.292	4.690	4.643	2.352	2.234
	(0.077)	(0.154)	(0.105)	(0.199)	(0.074)
scale: time since last trip	0.000	0.000	0.000	0.012	0.012
	(0.002)	(0.003)	(0.001)	(0.003)	(0.003)
scale: time since last csd	0.000	0.000	0.000	0.031	0.030
	(0.002)	(0.004)	(0.005)	(0.010)	(0.005)
vertical: income	2.059	0.751	0.731	1.977	1.722
	(0.129)	(0.019)	(0.059)	(0.116)	(0.046)
alpha	0.031	0.034	0.034	0.014	0.026
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Table 3.16. Non-Linear Coefficients

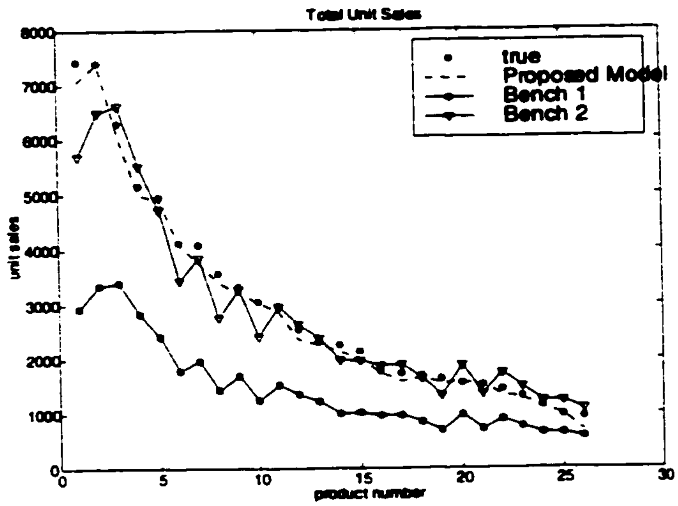
time. If the model is failing to pick up some of the heterogeneity, I should see non-zero values of these averages, much like a household-specific random effect for each product. In fact, for the top 6 products, I observe about 60% of these random effects lying in the interval $(-.01, .01)$. However, I also observe about 20% of the random effects lying in (-.2,.2) in a bell-curve like fashion. I suspect that part of the persistence I pick up derives from mismeasured heterogeneity. Otherwise, this observed persistence must derive from some form of unobserved measurement error. Since most marketing studies do not correct for unobserved time-series, these findings suggest the need for further research into how well existing models capture both heterogeneity as well as choice dynamics.

Aggregate Demand and Substitution Patterns

My goal is to recover aggregate demand for CSDs while accounting for the fact that consumers have heterogeneous shopping patterns and face heterogeneous store conditions. Figure (??) provides a rough idea of how well the benchmarks and the proposed model fit the aggregate data in terms of their predicted aggregate purchases for each product. For the proposed model, I use the second specification from above (model 2). As expected, the logit model clearly underpredicts aggregate sales for each product since it restricts decisions to be single-unit. Consequently, one would expect that the logit will also underpredict aggregate response to such promotional variables as prices, features and displays. The sequential demand model fares better. However, several products are mispredicted by over 20%. In contrast, the proposed model seems to capture the differences in total demand for each product quite accurately. In figure(3.17), I plot the 95% confidence interval of the proposed model to illustrate the degree to which it captures the true differences in total sales of each product.



(3.16)



(3.17)

One potential concern with the current estimates is the possibility of over-fitting. Given the large number of parameters, I run the risk of simply fitting the given sample data, rather than the true purchase behavior. Perhaps a better measure of the accuracy of the model would be to re-estimate demand for a subset of the data, saving the remainder as a hold-out sample. For instance, I could estimate 1993 and 1994 demand. Then, I could use the first quarter of the 1995 to see how well the model predicts out-of-sample. Nonetheless, the figure provides strong support for the improvement in demand predictions from modeling the multiple-discreteness.

Having computed aggregate demand, I am now able to study the aggregate responsiveness of consumers to the prices and marketing mix variables. I use elasticities to measure the responsiveness to prices, advertising and display. One of the main difficulties in computing elasticities is the fact that households do not necessarily face the same mix of prices and marketing variables. One way to recover a summary measure of overall elasticity across consumers and over time is to consider the effect of a uniform percentage change in the price of a good on aggregate demand (Ben-Akiva and Lerman 1985). For a given product, the aggregate observed purchases are (for simplicity I eliminate the household subscript):

$$X_j = \sum_{t=1}^{T} X_{tj}.$$

Assume that everyone experiences the same percent price change:

$$\frac{\partial p_{tk}}{p_{tk}} = \frac{\partial p_{sk}}{p_{sk}} = \frac{\partial \overline{p_k}}{\overline{p_k}}, s, t = 1, ..., T, k = 1, ..., I$$

where $\overline{p}_k = \frac{1}{T} \sum_{t=1}^{T} p_{tk}$. I then compute the price elasticity of total demand in response to a change in the average price level, $\overline{p}_k$:

$$\begin{aligned} \varepsilon^j_{p_k} &= \frac{\partial X_j}{\partial p_k} \frac{\overline{p_k}}{X_j} \\ &= \sum_{t=1}^{T} \frac{\partial X_{tj}}{\partial p_k} \frac{\overline{p_k}}{X_j} \end{aligned}$$

$$= \sum_{t=1}^{T} \frac{\partial X_{tj}}{\partial p_{tk}} \frac{p_{tk}}{X_{tj}} \frac{X_{tj}}{X_j}$$

$$= \sum_{t=1}^{T} \varepsilon_{p_k}^{tj} \frac{X_{tj}}{X_j}$$

which is just the sum of the individual elasticities, weighted by purchase share.

Table (4.7.1) presents estimated own-price elasticities for the proposed model 3 and the two benchmarks. As expected, the two benchmarks predict elasticities which are lower than those of the proposed model. In fact, the second model (the sequential demand model) generates even lower elasticities than those that ignore quantities. Almost all the values of the benchmarks are below one in magnitude. These elasticities are inconsistent with typical models of static oligopoly, in which firms simultaneously set their profit-maximizing prices each period. In such models, the equilibrium prices generate elasticities that are greater than one. In contrast, all of the own-elasticities in the proposed model are greater than one, providing some support for the validity of the proposed model over the benchmarks.

Tables (3.18, 3.19, and 3.20) present the cross-elasticities from the proposed model 3. The predicted substitution patterns show that, most importantly, consumers seem to respond to price changes by switching to another product of the same size. Most of the observed substitution patterns reflect realistic interactions. Almost all products substitute primarily to a cola. Also, 6-packs of caffeine-free diet Pepsi are very substitutable with 6-packs of diet Pepsi. Mountain Dew and Dr. Pepper are generally predicted as likely substitutes. Surprisingly, I find little interaction between Sprite and 7UP, mainly

Product	Model 3	bench 1	bench 2
PEPSI 6P	-2.38	-0.91	-0.68
COKE CLS 6P	-2.11	-1.07	-0.68
PEPSI DT 6P	-2.47	-0.91	-0.71
COKE DT 6P	-3.14	-1.09	-0.69
DR PR 6P	-3.04	-0.97	-0.74
MT DW 6P	-3.56	-0.90	-0.76
PEPSI DT CF 6P	-3.61	-0.93	-0.76
A and W CF 6P	-3.59	-0.90	-0.79
PEPSI 16oz	-2.25	-0.95	-0.91
PEPSI 12P	-2.16	-0.95	-0.89
COKE CLS 12P	-2.13	-0.93	-0.88
COKE DT 12P	-2.50	-0.92	-0.90
PEPSI DT 12P	-2.66	-0.95	-0.90
DR PR 12P	-2.47	-0.93	-0.87
MT DW 12P	-3.02	-0.91	-0.90
COKE DT CF 12P	-2.76	-0.93	-0.91
SP CF 12P	-2.57	-0.91	-0.93
PEPSI DT CF 12P	-2.92	-0.97	-0.92
PEPSI 67.6oz	-2.62	-0.59	-0.49
COKE CLS 67.6oz	-2.80	-0.66	-0.54
PEPSI DT CL 67.6oz	-2.66	-0.59	-0.51
7UP R CF 67.6oz	-2.57	-0.54	-0.48
COKE DT 67.6oz	-2.81	-0.65	-0.55
7UP DT CF 67.6oz	-2.61	-0.54	-0.50
DR PR 67.6oz	-2.94	-0.64	-0.51
MT DW 67.6oz	-3.23	-0.61	-0.52

Table 3.17. Own-Price Elasticities

due to the fact that they do not include comparable sizes in the choice set. The interactions might improve if I add random coefficients for the clear and citric attributes. Although not reported, the cross-elasticities from the first benchmark model (the conditional logit) are much smaller in magnitude. The cross-elasticities of the second benchmark (sequential demand) lie in between those of the proposed model and the conditional logit. As expected, the two benchmarks yield lower cross-price responses overall.

Table (3.21) reports measures of consumer response to features and display. Since these variables are not continuous, I do not in fact compute an elasticity. Instead, I first compute demand with all discrete variables set to zero and with prices set at their means. I then compute the response to ad (display) by computing the average sales-weighted difference in demand with the relevant product's ad (display) set to one. I find that advertising has the largest impact on caffeine-free diet colas, the 12-pack of peppers and on the lemon-lime products. In contrast, regular colas appear to have the lowest advertising and display responses. I also find that advertising has a relatively small effect on the 6-packs of cans and bottles, and a relatively large effect on 67.6 oz bottles. I find similar effects from display. From a retail manager's perspective, these results suggest that the ability to stimulate consumer response from marketing tools such as newspaper advertising and in-store displays will vary for different flavors and for different package types.

Product	PEP	COKE	PEP DT	COKE DT	DR PR	MT DW	PEP DT CF	AW	PEP 16oz
PEPSI 6P	-1.88	0.37	0.21	0.23	0.31	0.32	0.38	0.31	0.15
COKE CLS 6P	0.12	-1.99	0.24	0.28	0.25	0.13	0.45	0.20	0.02
PEPSI DT 6P	0.19	0.14	-2.23	0.17	0.14	0.10	1.07	0.15	0.15
COKE DT 6P	0.13	0.16	0.18	-2.23	0.16	0.02	0.25	0.09	0.02
DR PR 6P	0.20	0.15	0.17	0.07	-2.45	0.11	0.21	0.13	0.05
MT DW 6P	0.07	0.15	0.07	0.03	0.23	-2.30	0.02	0.05	0.01
PEPSI DT CF 6P	0.11	0.05	0.04	0.11	0.07	0.13	-2.57	0.18	0.02
A and W CF 6P	0.02	0.04	0.16	0.02	0.04	0.05	0.00	-2.43	0.05
PEPSI 16oz	0.03	0.04	0.06	0.03	0.02	0.00	0.01	0.04	-1.62
PEPSI 12P	0.08	0.04	0.05	0.04	0.03	0.03	0.10	0.11	0.13
COKE CLS 12P	0.03	0.02	0.01	0.02	0.02	0.04	0.08	0.12	0.14
COKE DT 12P	0.02	0.02	0.05	0.02	0.01	0.06	0.02	0.07	0.03
PEPSI DT 12P	0.02	0.03	0.05	0.07	0.05	0.07	0.04	0.02	0.01
DR PR 12P	0.02	0.02	0.01	0.04	0.00	0.01	0.00	0.04	0.01
MT DW 12P	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.02
COKE DT CF 12P	0.01	0.01	0.01	0.02	0.01	0.00	0.08	0.00	0.01
SP CF 12P	0.02	0.00	0.01	0.01	0.00	0.04	0.00	0.00	0.04
PEPSI DT CF 12P	0.02	0.00	0.01	0.00	0.00	0.02	0.00	0.00	0.02
PEPSI 6P	0.08	0.07	0.08	0.06	0.11	0.15	0.07	0.14	0.09
COKE CLS 6P	0.08	0.06	0.09	0.05	0.05	0.00	0.04	0.02	0.13
PEPSI DT 6P	0.05	0.06	0.06	0.12	0.06	0.03	0.01	0.04	0.13
COKE DT 6P	0.03	0.05	0.02	0.01	0.07	0.04	0.03	0.12	0.05
DR PR 6P	0.02	0.03	0.07	0.05	0.06	0.03	0.06	0.02	0.10
MT DW 6P	0.01	0.02	0.07	0.03	0.02	0.00	0.01	0.08	0.01
PEPSI DT CF 6P	0.03	0.03	0.05	0.03	0.02	0.01	0.03	0.00	0.06
A and W CF 6P	0.02	0.03	0.03	0.00	0.04	0.01	0.00	0.02	0.05

Table 3.18. Cross-Elasticities for 6-packs (with respect to row prices)

Product	PEP	COKE	COKE DT	PEP DT	DR PR	MT DW	COKE DT CF	SP	PEP DT CF
PEPSI 6P	0.07	0.08	0.09	0.09	0.05	0.10	0.19	0.07	0.01
COKE CLS 6P	0.05	0.02	0.02	0.03	0.09	0.05	0.00	0.00	0.01
PEPSI DT 6P	0.03	0.04	0.03	0.01	0.06	0.00	0.02	0.04	0.00
COKE DT 6P	0.02	0.01	0.03	0.10	0.07	0.05	0.00	0.00	0.00
DR PR 6P	0.02	0.00	0.04	0.05	0.06	0.02	0.02	0.00	0.04
MT DW 6P	0.01	0.02	0.01	0.01	0.00	0.00	0.00	0.02	0.02
PEPSI DT CF 6P	0.02	0.01	0.00	0.01	0.02	0.00	0.00	0.00	0.00
A and W CF 6P	0.01	0.02	0.01	0.00	0.04	0.00	0.03	0.00	0.00
PEPSI 16oz	0.04	0.05	0.05	0.05	0.02	0.04	0.06	0.00	0.04
PEPSI 12P	-1.89	0.22	0.16	0.47	0.25	0.24	0.30	0.04	0.16
COKE CLS 12P	0.16	-2.15	0.32	0.22	0.20	0.09	0.25	0.17	0.44
COKE DT 12P	0.06	0.09	-1.75	0.22	0.07	0.06	0.10	0.17	0.12
PEPSI DT 12P	0.09	0.04	0.17	-2.23	0.34	0.29	0.15	0.06	0.22
DR PR 12P	0.06	0.04	0.10	0.06	-1.85	0.09	0.06	0.07	0.00
MT DW 12P	0.08	0.04	0.04	0.05	0.30	-2.39	0.10	0.37	0.13
COKE DT CF 12P	0.01	0.01	0.19	0.02	0.09	0.05	-3.00	0.06	0.05
SP CF 12P	0.04	0.05	0.11	0.05	0.29	0.06	0.09	-3.15	0.05
PEPSI DT CF 12P	0.09	0.08	0.01	0.16	0.14	0.16	0.03	0.04	-3.00
PEPSI 6P	0.16	0.15	0.15	0.13	0.23	0.11	0.13	0.26	0.12
COKE CLS 6P	0.11	0.08	0.07	0.11	0.05	0.23	0.13	0.04	0.04
PEPSI DT 6P	0.07	0.05	0.19	0.12	0.01	0.11	0.18	0.06	0.04
COKE DT 6P	0.07	0.07	0.03	0.03	0.03	0.06	0.03	0.02	0.09
DR PR 6P	0.04	0.04	0.06	0.05	0.00	0.03	0.05	0.04	0.02
MT DW 6P	0.03	0.05	0.07	0.10	0.03	0.06	0.02	0.17	0.11
PEPSI DT CF 6P	0.03	0.00	0.02	0.08	0.07	0.05	0.05	0.10	0.11
A and W CF 6P	0.06	0.02	0.02	0.04	0.03	0.03	0.00	0.04	0.17

Table 3.19. Cross-Elasticities for 12-packs (with respect to row prices)

Product	PEP	COKE	PEP DT	7UP	COKE DT	7UP DT	DR PR	MT DW
PEPSI 6P	0.04	0.05	0.05	0.04	0.08	0.06	0.13	0.05
COKE CLS 6P	0.06	0.03	0.06	0.08	0.05	0.03	0.11	0.04
PEPSI DT 6P	0.04	0.04	0.04	0.06	0.05	0.08	0.04	0.02
COKE DT 6P	0.02	0.04	0.06	0.03	0.10	0.05	0.04	0.06
DR PR 6P	0.03	0.03	0.03	0.06	0.03	0.03	0.06	0.07
MT DW 6P	0.01	0.01	0.02	0.03	0.01	0.00	0.01	0.01
PEPSI DT CF 6P	0.00	0.01	0.00	0.02	0.05	0.00	0.04	0.03
A and W CF 6P	0.03	0.01	0.01	0.07	0.00	0.01	0.09	0.01
PEPSI 16oz	0.02	0.02	0.03	0.07	0.04	0.05	0.06	0.01
PEPSI 12P	0.07	0.09	0.04	0.14	0.06	0.04	0.05	0.08
COKE CLS 12P	0.06	0.08	0.04	0.06	0.03	0.08	0.08	0.06
COKE DT 12P	0.05	0.06	0.07	0.02	0.05	0.09	0.06	0.02
PEPSI DT 12P	0.02	0.06	0.07	0.05	0.03	0.07	0.02	0.03
DR PR 12P	0.01	0.03	0.01	0.03	0.04	0.00	0.03	0.02
MT DW 12P	0.00	0.04	0.02	0.01	0.02	0.00	0.07	0.03
COKE DT CF 12P	0.01	0.02	0.02	0.01	0.00	0.05	0.01	0.03
SP CF 12P	0.02	0.00	0.01	0.01	0.01	0.03	0.03	0.01
PEPSI DT CF 12P	0.01	0.01	0.03	0.01	0.01	0.02	0.03	0.04
PEPSI 6P	-2.05	0.17	0.61	0.15	0.10	0.22	0.10	0.29
COKE CLS 6P	0.11	-2.15	0.09	0.15	0.17	0.04	0.19	0.22
PEPSI DT 6P	0.23	0.11	-2.29	0.07	0.12	0.19	0.05	0.21
COKE DT 6P	0.03	0.02	0.04	-2.09	0.08	0.16	0.08	0.06
DR PR 6P	0.12	0.10	0.12	0.04	-2.50	0.06	0.08	0.05
MT DW 6P	0.04	0.06	0.09	0.12	0.09	-2.30	0.04	0.05
PEPSI DT CF 6P	0.03	0.08	0.04	0.05	0.03	0.05	-2.03	0.05
A and W CF 6P	0.05	0.05	0.02	0.07	0.04	0.03	0.06	-2.08

Table 3.20. Cross-Elasticities for 67.6 oz bottles (with respect to row prices)

Product	Ad elas	Display elas
PEPSI 6P	1.41	1.69
COKE CLS 6P	2.25	2.49
COKE DT 6P	2.32	2.47
DR PR 6P	2.78	2.81
MT DI 6P	2.85	3.06
PEPSI DT CF 6P	2.69	2.87
PEPSI DT 6P	2.01	1.98
A and W CF 6P	3.57	3.84
PEPSI 16oz	1.42	1.50
PEPSI 12P	2.44	2.69
COKE CLS 12P	2.40	2.67
COKE DT 12P	2.33	2.31
PEPSI DT 12P	3.16	3.63
DR PR 12P	3.95	4.20
MT DW 12P	3.09	3.04
COKE DT CF 12P	4.05	3.45
SP CF 12P	4.71	4.97
PEPSI DT CF 12P	6.20	6.14
PEPSI 67.6oz	2.10	2.23
COKE CLS 67.6oz	3.41	3.42
PEPSI DT CL 67.6oz	3.16	3.40
7UP R CF 67.6oz	3.13	3.57
COKE DT 67.6oz	4.65	4.77
7UP DT CF 67.6oz	3.41	3.47
DR PR 67.6oz	3.18	3.32
MT DW 67.6oz	4.11	4.16

Table 3.21. Own Ad and Display Elasticities

Strategic Interactions of Differentiated Products

In addition to looking at the products' perceived qualities, I also study the competitive interaction between types of CSDs. In particular, I study the effects from Pepsi carrying the citrus product, Mountain Dew, in its product line. Traditional theories might rationalize this behavior as a strategic flooding of the market with slightly differentiated products to prevent entry by potential competitors (see Schmalansee 1978). Brander and Eaton (1984) argue that firms can strengthen these barriers even further by carrying similar interlaced product lines. Alternatively, with sufficient within-line differentiation, Pepsi may simply find it more profitable to have several products targeted towards slightly different consumers, despite the potential for internal cannibalization of sales. In fact, Mountain Dew tends to be targeted towards a younger *teen* segment. Quelch and Kenny (1994) provide a non-theoretical discussion of such non-preemptive profit motives for line extension. The joint-ownership of two substitutable products provides Pepsi with unilateral market power. Below, I show how a multiproduct firm can increase its margins by jointly setting prices for competing products. The main question is whether the increased margins make Pepsi more profitable. At the same time, I wish to know the extent to which consumers value this additional product while having to pay higher Pepsi prices.

Empirically, the firm's decision to extend its product line has been studied to evaluate the welfare gains from new goods (Hausman 1995 and Petrin 1999), and to evaluate the ensuing gains in market power (Horsky and Nelson 1992, and Kadiyali, Vilcassim and Chintagunta 1999). I do not study the actual pre and post product introduction analysis explicitly. Instead, I measure the interaction of Mountain Dew with Pepsi's colas as well as all the other products available by simulating the effects of removing Mountain Dew from the choice set. Using the estimated demand and a model of supply, I recompute the equilibrium prices and quantities without Mountain Dew. Unlike most such comparative static analyses in marketing, I explicitly account for the competitive impact of removing a product on the equilibrium prices and quantities. The loss of unilateral market power from Mountain Dew should place a downward pressure on Pepsi prices. Moreover, Mountain Dew consumers switch either to another good or to not purchasing anything at all, potentially shifting demand for other goods. These changes in quantities will alter the equilibrium prices which, in turn, have a subsequent impact on overall quantities. The new equilibrium prices allow me to compute the change in the variable profits of each product. I also use these prices to compute the change in Hicksian welfare, which measures how much consumers would be willing to pay to keep Mountain Dew available. This exercise is similar to the methodology used to evaluate the introduction of new goods (Horsky and Nelson 1992, and Petrin 1999), the effects of mergers (Nevo 1999, Dubé 1999), and the exchange-rate-pass-through (Goldberg 1995). In this study, as in Dubé(2000), I use point-of-purchase prices rather than prices aggregated over time and across retailers in a market.

A Model of the Price-Setting Behavior of Firms

To model the counterfactual scenario, I must first make assumptions regarding the supply-side behavior. I assume the CSD industry is a static oligopoly with multiproduct firms, where a firm is a concentrate manufacturer. I treat the retailers as exogenous, so that I do not model the pricing decision of the retailer explicitly. This treatment of the retailer is a technical convenience. In theory, I could explicitly account for the vertical channel structure, using a variant of the vertical Nash model used in BGJ. However, the panel data provides an insufficient number of observations to model the optimal retail price in a given store-week. Given the relatively short time-span of the data, I use a short-run model in which firms choose prices conditional on their product portfolios. Manufacturers set wholesale prices quarterly and retailers set shelf prices weekly. I assume that the large sunk advertising costs associated with a new brand are prohibitively high to expect entry to occur, even if the elimination of Mountain Dew should raise overall prices.

The assumption of static price-setting is also a technical convenience. The fact that demand contains dynamic variables, such as loyalty, implies that a current pricing decision could influence future demand. The theoretical literature devoted to dynamic differentiated multiproduct oligopolies is still in a preliminary stage. The incorporation of dynamics into the firms' pricing decisions will be a very important issue for future research. For now, I maintain the static assumption, and view the model as an approximation to actual competition. I provide a more detailed discussion of the potential consequences of demand-related dynamics on the firms' intertemporal pricing decisions in Dubé (2000).

I model the manufacturers as static, profit-maximizing multiproduct oligopolists. Thus, in a given quarter (I omit the time subscripts for simplicity), firm f earns expected variable profits:

$$\pi_f = \sum_{i \in S_f} (p_i^w - mc_i) E\{Q_i(\mathbf{p}^w)\}$$

where $E\{Q_i(\mathbf{p}^w)\}$ is the expected demand for product i, which is a function of the wholesale prices of all the products. I use p_i^w to denote the wholesale price of product i, mc_i to denote product $i's$ per-unit costs and S_f to denote the set of products produced by firm f. Assuming the existence of a pure-strategy static Bertrand-Nash price equilibrium with strictly positive prices, each of the prices, p_i i$\in S_f$, satisfies the following first-order conditions:

$$E\{Q_i(\mathbf{p}^w)\} + \sum_{k \in S_f} (p_k^w - mc_k) \frac{\partial E\{Q_k(\mathbf{p}^w)\}}{\partial p_i^w} = 0, i \in S_f, f = 1, ..., F. \quad (3.18)$$

I construct the following $(I \times I)$ matrix $\boldsymbol{\Delta}$ with entries as follows:

$$\widetilde{\Delta_{jk}} = \begin{cases} -\frac{\partial E(Q_{jk})}{\partial p_i^w}, if \ \exists \text{ f s.t. } \{i,k\} \subset S_f \\ 0, else \end{cases}.$$

Stacking the prices, marginal costs and expected quantities into $(J \times 1)$ vectors $\mathbf{Q}, \mathbf{p}$, and $\mathbf{mc}$ respectively, the first-order conditions can be written in matrix form:

$$E(\mathbf{Q}) - \boldsymbol{\Delta}(\mathbf{p}^w - \mathbf{mc}) = 0.$$

From the first-order conditions, I derive the mark-up equations:

$$\mathbf{p}^w - \mathbf{mc} = \mathbf{\Delta}^{-1} E(\mathbf{Q}). \tag{3.19}$$

In the spirit of the Empirical IO literature, I estimate the values of these mark-ups directly from the estimated demand parameters, without using information on costs.

I assume the presence of an exogenous retailer who sets prices as a fixed quarterly markup over the wholesale price plus a weekly mean-zero disturbance. Although this assumption may not be ideal, it reflects the margin-planning strategy in Blattberg and Neslin (1990). They describe supermarket managers setting a long-run total average margin that embodies a fixed mark-up over wholesale costs and an occasional promotional discount. For consistency with the notation in the demand section, I denote retail prices as p, and wholesale prices as p^w. Using data on observed monthly retail margins at the UPC level[27], I assume the retail price in store r (where there are R stores) for product j in week t has the following form:

$$p_{jt}^r = (1 + M_j)p_j^w + \varepsilon_{jt}^r, r = 1, ..., R, j = 1, ..., J, t = 1, ..., T$$

where M_j is the markup and ε_{jt}^r is a store-specific mean-zero deviation accounting for promotions (negative) or coupons (positive). I treat ε as random to account for the fact that the retailer's decision likely embodies a maximization problem at the store level. Following the discussion in Slade (1995), I assume that retailers compete via total offerings, rather than at the individual product level.

---

[27] We discuss the source of these observed margins later

Since I observe retail margins, I am able to recover mc by combining the computed wholesale markup with observed retail prices. First, I define the monthly average retail price (across stores) as $\overline{p_j} \equiv \frac{1}{RT}\sum_{t=1}^{T}\sum_{r=1}^{R} p_{jt}^{r}$ and I note the following implied relationship between weekly retail and monthly wholesale prices:

$$\begin{aligned}\overline{p_j} &= \frac{1}{RT}\sum_{t=1}^{T}\sum_{r=1}^{R}\left[(1+M_j)p_j^w + \varepsilon_{jt}^r\right] \\ &\approx (1+M_j)p_j^w + E(\varepsilon_{jt}^r) \\ &= (1+M_j)p_j^w.\end{aligned}$$

So, I can compute $p_j^w = \frac{\overline{p}}{(1+M_j)}$. Now I can compute the vector of monthly marginal costs:

$$\mathbf{mc} = \mathbf{p}^w - (\boldsymbol{\Delta})^{-1} E(\mathbf{Q}).$$

Now that I have estimates of the demand parameters and the marginal costs, I am able to consider the experiment. I rewrite (3.19) as:

$$\mathbf{p}^* = \mathbf{mc} + \boldsymbol{\Delta}(\mathbf{p}^*)^{-1} E\{Q(\mathbf{p}^*)\}.$$

The complex nature of the multiproduct oligopoly environment prohibits the traditional analytic derivation of comparative statics. Instead, I solve for $\mathbf{p}^*$ numerically. This approach is quite similar to the one used by Horksy and Nelson(1992), although I assume a multiproduct firm. I discuss the details of this procedure in the appendix. A simpler approach to recovering these prices would be to assume that the markups and quantities do not vary with the prices, in which case I would compute:

$$\mathbf{p}^* = \mathbf{mc} + \boldsymbol{\Delta}^{-1} E(\mathbf{Q})$$

using the actual markups and the demand that prevails at the same prices with the restricted product set. This approach has been used to study mergers, where products simply change ownership[28]. In the current application, the approximation probably will not perform well since a product is physically removed from the choice set. Assuming that the quantities of the remaining CSDs do not change implies that all Mountain Dew consumers switch to not purchasing anything. Once I have the new prices, I can compute the percent change in variable profits as:

$$\frac{\pi(\mathbf{p}^{*}) - \pi(\mathbf{p}^{w})}{\pi(\mathbf{p}^{w})} * 100 = \frac{(\mathbf{p}^{*} - \widehat{\mathbf{mc}})E\{Q(\mathbf{p}^{*})\} - (\mathbf{p}^{w} - \widehat{\mathbf{mc}})E\{Q(\mathbf{p}^{w})\}}{(\mathbf{p}^{w} - \widehat{\mathbf{mc}})E\{Q(\mathbf{p}^{w})\}} * 100.$$

Having determined $\mathbf{p}^{*}$ and, thus, the counterfactual retail prices, I compute how much consumers value Mountain Dew. Using Hicksian compensating variation, I literally compute the dollar value that consumers would be willing to pay to keep Mountain Dew in their choice set. Technically, this measure amounts to computing the required change in income for each household trip to ensure that the maximized utilities under the counterfactual prices are still equal to those under the actual prices. Recall that the utility in the proposed model is defined as:

$$U^{h}(p, y_{h}) = \sum_{j=1}^{J_{h}} u_{j}^{h}\left(\sum_{i=1}^{I} \Psi_{ij}^{h} Q_{ij}^{h}, D_{h}\right) - \sum_{j=1}^{J_{h}} \sum_{i=1}^{I} p_{i} Q_{ij}^{h} + y_{h}.$$

I find Δy_{h} such that optimal true and counterfactual utilities are equal:

$$U^{h}(\mathbf{p}, y_{h}) = U^{h}(\mathbf{p}^{*}, y_{h} + \Delta y_{h}).$$

[28]Nevo (1999) finds that, for most instances, the approximation works well. However, he did find some instances in which the approximation was quite different from the numerical solution.

Since utility is linear in y_h, I can compute:

$$\Delta y_h = u^h(\mathbf{p}, y_h) - u^h(\mathbf{p}^*, y_h).$$

Now I can compute the sample compensating variation:

$$\Delta y = \sum_{h=1}^{H} \Delta y_h.$$

Since I do not observe the total size of the market, I am only able to predict the percent change in income rather than the absolute values. The strategic value of this welfare measure is that it reveals how much consumers would be willing to pay to keep Mountain Dew available. From a store manager's point of view, this sort of measure could be useful in determining how much variety to provide (whether to provide the entire Pepsi line or to focus on specific products). Similarly, CSD manufacturers may also consider such an analysis for the long-run determination of which products to distribute to retailers with stringent shelf-space allocations.

The goal of this exercise is to compute the counterfactual equilibrium prices that prevail after removing a product from the choice set. Two factors drive deviations from the actual prices. First, the physical removal of a product from a firm's product line will reduce its unilateral market power for the remainder of its line, leading to lower prices: the market power effect. Second, the omitted product will reduce competition faced by its closest substitutes, leading to higher prices: the competitive effect. To illustrate the market power effect, I re-write the firms' first-order conditions (3.18) as:

$$\frac{p_i - mc_i}{p_i} = -\frac{1}{\varepsilon_{ii}} - \frac{\sum_{k \in S_f} (p_k^w - mc_k)\, Q_k(p^w)\varepsilon_{ki}}{Q_i(p^w)\varepsilon_{ii}}, i \in S_f, f = 1, ..., F$$

where I drop the expectation operator to simplify the notation. Since the own-elasticity, ε_{ii}, is negative and cross-elasticities, ε_{ki}, are positive, I can see that reducing the number of products in a firm's overall line, S_f, will reduce the markups for all products with a non-zero cross-elasticity (assuming the margins are non-negative). At the same time, the elimination of a product from the choice set forces consumers previously consuming the good in question to switch to another product or not to consume. This competitive effect would likely increase the expected quantities of other products, especially close substitutes of the omitted good. Since prices are determined simultaneously, both the market power and the competitive effects interact, making the net effect difficult to predict analytically. Instead, I solve for the prices numerically.

Table (3.22) reports the predicted price changes after removing Mountain Dew from the choice set[29]. The approximations tend to be fairly close for the Pepsi prices. However, by construction, they predict no change for the non-Pepsi products. The numerically-computed prices demonstrate that this latter effect is not always accurate. As expected, all of the Pepsi prices fall due to lost unilateral market power associated with Mountain Dew. The more interesting results are the changes in the non-Pepsi prices. Most products do not change noticeably. However, I observe that the price of 67.6 ounce bottles of Dr. Pepper and 6-packs of regular and diet Coke all rise more than 5%. The intuition

[29]To motivate the importance of modelling the multiple-discreteness, we need to report the equilibrium prices for the logit model. In fairness, we should provide some taste heterogeneity in the benchmark to provide a meaningful comparison. We are in the process of estimating the parameters of a latent-class model.

for this result requires looking back at the cross-elasticities in table (3.18, 3.19 and 3.20). The 6-packs of Coke and diet Coke compete heavily with one-another, insulating them from the downward price-effect of Pepsi. Although not reported in the table, 6-packs of Mountain Dew also substitute substantially to regular Coke. This effect is passed on to 6-packs of Diet Coke. Similarly, the price increase for regular Coke is also passed on to 67.6 ounce bottles of Dr. Pepper. The latter also benefits from the reduced competition with Mountain Dew.

Table (3.23) shows the overall effect on firm profitability. In the first column, I compute the change in profits when I assume prices do not change after removing Mountain Dew. Comparing the first column to the third, in which I compute the new equilibrium, I see that failure to account for the changes in the competitive environment lead to overstating the impact of removing Mountain Dew. Since the first column does not capture Pepsi's aggressive cola pricing, the results overstate the loss in Pepsi profits. They also understate the decrease in competition faced by Coke and Dr. Pepper-7UP, leading to overstating these latter two firms' profits. Using column 3, I still observe Pepsi's profits falling by 3% due to its lost income from Mountain Dew. To compensate, it reduces all of its cola prices to try and draw in more customers. Overall, Coke's profits do not change much. Although it is able to raise its 6-pack prices, it faces much more aggressive pricing from Pepsi overall. Finally, Dr. Pepper-SevenUp raises its profits just over 3%, mainly due to 67.6 ounce bottles of Dr. Pepper. The evidence suggests that Pepsi carries Mountain Dew to increase its overall profits. Mountain Dew steals business

Product	Approx.	Equilib.
PEPSI 12P	-0.89	-1.10
COKE CLS 12P	0	0.21
PEPSI 6P	-1.94	-1.74
COKE DT 12P	0	0.13
PEPSI 67.6oz	-4.67	-3.53
PEPSI DT 12P	-0.53	-0.21
COKE CLS 6P	0	5.50
PEPSI DT 6P	-2.06	-2.66
COKE CLS 67.6oz	0	0.67
PEPSI DT CL 67.6oz	-1.07	-0.65
COKE DT 6P	0	5.08
DR PR 12P	0	0.34
DR PR 6P	0	-2.10
7UP R CF 67.6oz	0	0.27
COKE DT CF 12P	0	0.39
COKE DT 67.6oz	0	0.53
7UP DT CF 67.6oz	0	-0.15
SP CF 12P	0	0.69
PEPSI DT CF 12P	-0.24	0.43
DR PR 67.6oz	0	5.27
PEPSI 16oz	-3.76	-3.50
PEPSI DT CF 6P	-1.26	-1.17
A and W CF 6P	0	0.19

Table 3.22. Percent change in prices after removing Mountain Dew

from Coke's colas as well as Dr. Pepper, while providing Pepsi with additional market power to raise its cola prices.

Turning to consumers, I find that removing Mountain Dew from the choice set only reduces total consumer welfare by 0.1%. If I ignore the change in prices, I estimate a change in consumer welfare of 0.4%. While these changes are quite small, I still find that the failure to account for the change in the competitive environment leads to overpredicting consumer valuation for Mountain Dew by a factor of 4. Using the

Company	No Price Change	Approx.	Equilib.
Pepsi	-5.19	-9.12	-3.23
Coca-Cola	9.00	0	0.44
Dr. Pepper-Seven-UP	6.70	0	3.03

Table 3.23. Percent change in profits after removing Mountain Dew

Hicksian compensated variation argument, I interpret this result as consumers willing to forfeit 0.1% of their income to keep Mountain Dew in the product set. Evidently, consumers do not attribute much value to the extra variety of the citrus group. This low valuation also reflects the fact that adding Mountain Dew allows Pepsi to increase its prices.

Conclusions

While the typical logit and probit DCMs have been found to provide useful predictions for consumer purchases in many product categories, their restrictive single-unit purchase assumption is inappropriate for several categories, such as CSDs. Failure to account for the simultaneous purchase of multiple products and an integer quantity of each results in poor estimates of demand and underpredicted consumer response to prices and marketing variables. In addition, demographic variables, which have typically been found to provide little information in marketing applications of DCMs, appear to be instrumental in determining the joint distribution of total product alternatives and total units purchased on a given trip. Demographics also partially-explain observed differences in tastes for product attributes. I also find that the correction for a 15-day lagged unobserved time series has a substantial impact on estimated standard errors.

The use of the characteristics approach provides an interesting distribution of consumer tastes for attributes. Evidently, the average consumer prefers non-diet caffeine-containing products, especially in the 6-pack of cans size. However, a subgroup of these households with a young female head has preferences for diet products, and larger households tend to prefer the larger product sizes. The random coefficients reveal that, despite these mean tastes, there is a lot of household-level deviation from these means. Since households shop for many independent needs at a given time, this taste heterogeneity is both across households and across expected consumption occasions.

In the analysis, I assume that a 6-pack and a 12-pack of cans are inherently different products. The fact that I do find evidence of high substitutability between goods of the same size supports the assumption that brand/sizes constitute different products, and that differences in per-oz prices reflect quantity surcharging.

The partial equilibrium analysis allows me to use the estimated demand to assess the strategic interaction between certain types of goods. In particular, I simulate the impact of Mountain Dew on the prices and profits of the other goods. I find that the presence of Mountain Dew allows Pepsi to increase its profits by stealing business from some of its competitors and raising its margins. I also find that consumers do not place much value on the extra variety of the citrus product, possibly due to the accompanying increase in Pepsi cola prices.

The findings generate a number of interesting new research questions. Most importantly, I refer to an unobserved need for variety at a given moment. The identifying assumption is the independence between tastes on a given trip. However, if I believe that a contemporaneous form of the attribute satiation model applies, I should try to link choices made for different expected consumption occasions on a given trip. I provide a brief explanation of how one could account for such links to capture the fact that on a given trip, one might expect households to purchase dissimilar products to achieve variety.

I also demonstrate the importance of correcting for unobserved autocorrelation. For now, I characterize unobserved autocorrelation as persistent measurement error. Given that most studies ignore unobserved time-series, future research would benefit from a more in-depth account of the sources of this persistence.

I make use of the method of simulated moments to deal with the random coefficients specification and the unobserved needs. This approach works in this case, in which I am mainly interested in controlling for heterogeneity and then computing aggregate demand. For applications in which one may wish to recover household-level parameters, a Bayesian method might be preferable. The Bayesian approach might facilitate the implementation of maximum likelihood estimation as well. I leave the implementation of a Bayesian approach to this model as an open question.

Finally, the application of the estimated demand in a partial equilibrium model of industry competition might be sensitive to the assumed firm behavior. I make strong assumptions regarding both retail pricing behavior and dynamics. A growing body of research is slowly developing the tools needed to estimate dynamic models of differentiated products industries. The determination of how consumer dynamics (loyalty) could generate a dynamic element in producer behavior would be an interesting contribution to both the economics and the marketing literature.

CHAPTER 4.
Product Differentiation and Mergers in the Carbonated Soft Drink Industry

Introduction

A.C. Nielsen estimates that the carbonated soft drink (CSD) category is the largest in the Dry Grocery Department at US Food Stores, accounting for roughly one tenth of the department's national sales revenue. Today, three companies, Coca-Cola Co, PepsiCo and Dr. Pepper-SevenUp (currently owned by Cadbury-Schweppes) control most of the CSD concentrate market; but their respective shares are spread across a fairly large number of brands, flavors and packaging types.

During the past two decades, the industry has been under heavy scrutiny as the major players attempt to capture higher market shares through aggressive advertising and acquisitions. In 1986, the Federal Trade Commission (FTC) contested the proposed acquisitions of Dr. Pepper by Coca Cola Co., and of SevenUp by PepsiCo, citing the potential for increased concentration and diminished competition. Pepsi immediately called off its merger; however, Coke persisted in bringing its case to trial. Despite several economic arguments put forth by both the FTC and Coca-Cola, the court resorted to a simple legal argument regarding the illegality of the post-merger increase in market share (White 1998). In ruling against Coca-Cola, the Federal District Court found insufficient empirical evidence to make a ruling based on competitive effects.[30] The court's inability to determine the intensity of competition between products in different flavor segments prevented verifications of the economic claims put forth during the case.

The complexity of the CSD product space and the prohibitive cost of suitable data have limited the number of existing academic studies (eg. Gasmi, Laffont and Vuong 1991, Cotterill, Franklin and Ma 1996). I use a unique micro dataset and a new econometric approach to study demand for CSDs in a specific city-market. In addition to the complexity of the CSD product space, I address the complicated manner in which consumers choose their CSD shopping bundles. I observe that households often buy an assortment of products on a given trip. My model of consumer demand reflects realistic choice behavior, while allowing for a meaningful scope of products. Combining estimated demand with a static model of multiproduct oligopoly, I investigate the actual levels of manufacturers' market power and the implications of different CSD mergers for prices. I use data from the nineties. However, for an industry as mature as CSDs, I do not expect this type of extrapolation to be drastic.

Several recent studies have developed economic and econometric models for industries which, like CSDs, consist of a large number of differentiated products. Traditional linear and flexible functional form demand specifications often require estimating more cross-price parameters than the data would likely identify. Berry (1994) proposes the *characteristics approach* of Lancaster, modeling consumer preferences as a function of

---

[30] F.T.C. v. Coca-Cola Co.., 641 F. Supp. 1128 (1986).

the underlying bundle of attributes that characterize the relevant products. Consumers select the utility-maximizing bundle based on their tastes for attributes.

Many recent industry applications of this approach use an econometric model derived from a discrete random utility model at the individual level, such as the conditional logit (McFadden 1975,1981). The typical discrete choice model (DCM) implicitly assumes that each consumer choice consists of a single unit of a single alternative, where one of the alternatives is a no-purchase option. Traditionally, researchers have estimated such models with individual purchase data (e.g. Goldberg 1995). The derivation of aggregate demand amounts to integrating over the set of individuals who make a purchase in a given period. The simplicity of the aggregation has the added benefit of facilitating the direct use of aggregate data, which is easier to obtain and less cumbersome than individual samples. The DCM has been applied to several industries such as automobiles (Berry, Levinsohn and Pakes, [BLP], 1995,1998, Goldberg 1995, Petrin 1999 and Gron, Polson and Viard 1999), micro computers (Bresnahan, Stern and Trajtenberg 1997), ready-to-eat cereals (Nevo 1999, 2000), airlines (Berry, Carnall and Spiller 1997), ketchup and yoghurt (Besanko, Gupta and Jain 1998), movie theaters (Davis 1998) and gasoline stations (Manuszak 1999).

A simple solution for modeling CSD demand would be to apply the aggregate DCM to market-level CSD data. For instance, I could study the share of total weekly supermarket shopping trips resulting in a CSD purchase. While the DCM is convenient, especially for dealing with aggregate market-level data, the underlying consumer behavior is unrealistic for industries such as CSDs. Simple inspection of a panel of individual purchases reveals that households often purchase bundles of CSDs on a given trip (see table(4.40) in the Appendix). Thus, the naive DCM is misspecified and the shares are mismeasured, since the potential market could be larger or smaller than the weekly store-traffic. The biases from these misspecifications could have an adverse effect on substitution patterns and predictions for the competitive behavior in the industry. In particular, I expect the DCM to mismeasure market power, on the supply-side, leading to incorrect predictions for mergers.

I use an alternative model of consumer demand for CSDs that incorporates a large number of products without imposing overly stringent assumptions on how consumers shop. Based on the static random profit model of Hendel (1994), the approach accounts for the *multiple-discreteness* problem explicitly. I recast the Hendel model into a dynamic random utility context and apply it to a unique micro dataset for a specific market. The data set contains information for a panel of households' purchases, as well as point-of-purchase prices (shelf prices) and marketing conditions for all of the available alternatives. The use of household-level data not only allows me to model this richer purchase behavior explicitly, it also allows me to include important point-of-purchase state variables, such as feature advertising and in-aisle displays. These marketing instruments are not typically included in studies using aggregate city-market or national data which would require using an average over store-weeks. The use of panel data also allows me to incorporate lagged choice variables to account for product loyalty, a form of dynamic heterogeneity that is not typically included in studies using aggregate data. Unobserved dynamics also enter the model in the form of persistent demand shocks. I control for heterogeneity by using both individual-specific demographic data and a random coefficients specification.

I use the estimated market-level demand parameters to compute the manufacturer margins and marginal costs that prevail in a static equilibrium model. Combining the estimated market demand and marginal costs with my model of supply, I investigate the effects of several hypothetical mergers. Ultimately, I wish to assess the effects of these mergers on industry prices, thus developing a useful tool for future policy in the CSD industry and any other industry for which the behavioral assumptions of my model are appropriate.

In general, I find that the proposed model yields much lower own-price elasticities than the aggregate DCM. These lower elasticities translate into higher measured market power. Thus, the proposed model predicts that CSD manufacturers have a greater ability to set prices above costs. In contrast, I find the proposed model predicts much higher cross-elasticities than the aggregate DCM, so that the former implies a much more competitive environment and, thus, much higher benefits from joint-pricing of products. Applying the proposed model to study mergers, my evidence supports Coke's claim that the merger with Dr. Pepper would not have been anticompetitive in terms of price increases. However, I do find the merger between Pepsi and 7UP leads to large increases in the prices of the latter, supporting the claim by the FTC. I also find that the merger between Coke and Pepsi results in very large price increases. In contrast, the low market power and level of competition predicted by the aggregate DCM leads to unrealistically small effects from the mergers, even for a merger between Coke and Pepsi.

The paper is organized as follows. Section 2 provides a brief outline of the CSD industry and a summary of the relevant empirical literature. Section 3 develops the proposed model and compares it to the DCM. I also present the model of static multiproduct oligopoly used to describe the CSD manufacturers. Section 4 outlines the estimation procedure. Section 5 describes the scanner panel data and the market. Section 6 presents the empirical results for demand using both the proposed model and the aggregate DCM. Section 7 describes the proposed model's predictions for measured market power and the implications of mergers. Finally, Section 8 concludes and outlines possibilities for further research.

Consolidation and Antitrust in the CSD Industry

The American CSD industry is currently dominated by a small number of firms controlling a substantial number of products consisting of various flavors and package sizes. However, most of the flavor groups are unable to sustain more than a couple of major brands. From the individual brand perspective, the CSD category is highly disaggregated with only a few individual brand/sizes holding more than one percent of the market volume sales, and a large fringe competing for the remainder. By the early eighties, the major brands appeared to have fully exploited the potential differentiation strategies. I discuss these forms of differentiation in the Appendix. Given the huge advertising outlays required for new product entry and the high risks of product failure, the larger companies began entering flavor markets by acquiring existing brands, leading to a dramatic consolidation during the 1980s. By 1989, Cadbury-Schweppes had acquired Canada Dry, Hires Root Beer and Crush; and Hicks and Haas had acquired 7UP, Dr. Pepper, A&W Rootbeer and Squirt. In 1986, at the height of the merger phase, Coke (the number one firm) announced plans to acquire Dr. Pepper (the number three firm) and Pepsi (the number two firm) announced plans to acquire 7-UP (the number four firm).[31] Fearing an anticompetitive effect on industry prices, the Federal Trade Commission contested both mergers. However, Coke persisted, bringing the case to the Federal District Court.

While the Court ultimately rejected the merger on the grounds that it would give Coca-Cola too much market share, the decision was controversial. Both the FTC and Coca-Cola presented several interesting empirical economic questions dealing with the extent of the industry, efficiency in the distribution chain, joint production efficiency and differentiated product competition. Due to a lack of appropriate empirical tools at the time, many of these arguments were not taken into consideration for the Court's final decision and the issues still remain open. More recently, Higgins, Kaplan and

[31] Section 7 of the Clayton Act specifically opposes those mergers and acquisitions "where in any line of commerce in any section of the country, the effect of such acquisition may be substantially to lessen competition or to create a monopoly."

Tollison(1995) investigate the extent of the CSD market and Muris, Scheffman and Spiller(1992,1993) provide a comprehensive treatment of the distribution networks.

I focus on the economic claims regarding market power put forth by both the FTC and Coke. During the case, the court remained undecided over the anticompetitive effects of the mergers. The FTC argued against the relevance of differentiation, claiming that margins were already high due to tacit collusion. Coke argued that differentiation was such that coordination would be virtually impossible even with a merger.[32] Ultimately, Coke argued that only the merger between Coke and Pepsi would lead to an objectionable decrease in competition. As discussed previously, the large number of differentiated CSD products complicates demand estimation, an issue which appears to have prevented valid empirical tests of the arguments put forth by the FTC and Coke.

Since the trial, new methods have been developed. Gasmi, Laffont and Vuong(1992) reduce the dimensionality of the product space by assuming a cola duopoly between Coke and Pepsi, treating the remaining products as a competitive fringe. They find evidence of advertising collusion, not price collusion, using 19 years of annual accounting data between 1968 and 1986. Langan and Cotterill (1994) and Cotterill, Franklin and Ma (1996) extend the scope of products to 9 brands aggregated across package sizes. They assume multi-stage-budgeting at the consumer level, in addition to which the latter

[32]The FTC based their opinion on the relatively hight return on stockholder equity for the major producers. Coca-Cola used reduced form regressions to show an inverse relationship between prices and concentration. (see White1998).

adds price reaction functions[33]. Although neither study finds strong evidence supporting price collusion, the former shows the potential for profit-increasing collusive pricing between several brands. The latter finds that market power may in fact come as much from product differentiation as from collusive pricing once price-reaction functions are considered. The multi-stage-budgeting approach requires prior assumptions about the segmentation of products and the sequential process by which consumers make choices amongst these segments. This approach works well for the limited product set of 9 brands considered above, but the parameter dimensionality problem resurfaces when I disaggregate to the UPC level. Since the current study aims to determine the degree of similarity between different products and how consumers perceive them, I prefer to allow the data to reveal any potential segmentation structure, rather than impose one.

I use an alternative approach to studying the CSD industry that incoporates a richer set of products. Also, by using the characteristics approach, I am able to study the sources of differentiation between alternatives in terms of consumer tastes for CSD product attributes, an exercise which I carry out in Dubé (2000). During the antitrust case against Coca-Cola, the court concluded that local cities constitute separate markets. As such, I use individual household purchases in a single city-market, Denver, to conduct my analysis of the CSD industry.

---

[33]The multi-stage budgeting model has also been applied to beer (Hausman, Leonard and Zona 1994) and ready-to-eat cereals (Hausman 1996).

The model

Individual CSD demand

One of the predominant features of the CSD purchase data is the frequent incidence of multiple-item purchases. In contrast with the behavior of the typical DCM, households do not always select a single unit of a single CSD brand on a given shopping trip. Households appear to be seeking variety in their purchases by selecting a bundle of CSDs. Consequently, a model of CSD demand must allow households to choose an integer number of brands and an integer quantity of each: this phenomenon is the *multiple discreteness* problem.[34]

The model of demand derives from the framework proposed in Hendel (1999). He develops a static random profit model to account for firms' cross-sectional holdings of computers. For firms, the variety of holdings comes from the presence of multiple potential computing tasks, which are not observed by the econometrician. For instance, a firm might be divided into several departments: each department is assumed to select independently an integer quantity of one of the computer brands to fulfill its computing needs. The independence across these decisions allows a firm to hold a distribution of brands rather than concentrating their purchases on a single product.

---

[34] A separate line of research has examined the quantity purchase decision for single-brand purchases using the Hanneman(1984) random utility model (see Chiang 1991 and Chintagunta 1993). However, these models do not account for purchases of multiple brands on the same trip.

I modify this model into a random utility framework to address the consumer shopping problem. Instead of estimating a static cross-sectional CSD holdings model, I develop a purchase model that exploits the panel nature of my data. Conditional on making a shopping trip, a household chooses CSD products to satisfy various needs. The households' needs are expected future consumption occasions, which are not observed by the econometrician. The determinants of these occasions vary and include such factors as family members with diverse tastes, the depletion of overall CSD inventories, and uncertain future tastes, where the latter explanation refers to the separation between the time of purchase and the expected time of consumption (Walsh 1995). For each consumption occasion, the household selects an integer quantity of one of the products. I assume the number of expected consumption occasions is determined by a Poisson process whose mean is a function of household demographics and CSD inventories. Since I do not observe the consumption occasions, I simulate them. The estimation procedure yields the expected total CSD purchase vector for a shopping trip, aggregating across all the expected needs.[35]

On a given shopping trip, a household h purchases a basket of various alternatives to satisfy J_h different future consumption occasions. Suppose the household's preferences are separable in its purchases of the I softdrink products available and a composite commodity of other goods. These preferences are assumed to be quasilinear. Finally,

---

[35]For the special case in which the consumption occasions, or tasks depending on the context, are observed, Hausman, Leonard and McFadden (1995) develop a less complicated model.

suppose the household spends y_h on the given shopping trip and let z denote a composite commodity. Conditional on J_h, the utility of household h at the time of a shopping trip is given by (I suppress the time index to simplify notation):

$$U^h = \sum_{j=1}^{J_h} u_j^h\left(\sum_{i=1}^{I} \Psi_{ij}^h Q_{ij}^h, D_h\right) + z, \tag{4.1}$$

where D_h is a $(d \times 1)$ vector of household characteristics, Q_{ij}^h is the quantity purchased of alternative i and Ψ_{ij}^h captures the household's valuation of alternative i 's attributes on consumption occasion j. The random component of the utility function comes from the treatment of Ψ_{ij}^h as a random variable. This specification assumes additive separability of the J_h subutility functions, eliminating any valuation spillovers between consumption occasions. The individual subutility functions treat each of the goods in the product category as perfectly substitutable for a given consumption need. Thus, households select one alternative to satisfy each of its J_h expected consumption needs. The household's expenditure constraint is given by:

$$\sum_{j=1}^{J_h} \sum_{i=1}^{I} p_i Q_{ij}^h + z \leq y_h.$$

where p_i is the price of product i. So long as the subutility functions satisfy the correct shape and continuity properties, the expenditure equation will be binding and may be substituted into the original utility function to give:

$$U^h = \sum_{j=1}^{J_h} u_j^h\left(\sum_{i=1}^{I} \Psi_{ij}^h Q_{ij}^h, D_h\right) - \sum_{j=1}^{J_h} \sum_{i=1}^{I} p_i Q_{ij}^h + y_h. \tag{4.2}$$

Conditional on the number of anticipated consumption occasions, J_h, the household's problem will be to pick a matrix with columns $Q_j (j = 1, ..., J_h)$ to maximize 4.2.

The subutility functions for consumption occasions j are defined as:

$$u_j^h(\beta_j^h, D_h, X_j) = \left(\sum_{i=1}^{I} \Psi_{ij}^h Q_{ij}^h\right)^{\alpha} S(D_h) - \sum_{i=1}^{I} p_i Q_{ij}^h \qquad (4.3)$$
$$\Psi_{ij}^h = \max(0\ ,\ X_i\beta_j^h + \xi_i)^{m(D_h)}$$

where X_i' is a $(k \times 1)$ vector of brand i 's observable attributes, β_j^h is a $(k \times 1)$ vector of random coefficients for consumption need j, the tastes for attributes, and ξ_i is an unobserved attribute which may be correlated with the price. This correlation could generate an endogeneity problem, an issue I discuss below. Since the actual number of consumption occasions, J_h, is not observed, the estimation procedure will only identify the mean and variance of the valuation of the attributes across the needs and household trips. The term Ψ_{ij}^h can be interpreted as the perceived quality of alternative i for consumption need j. The given specification explicitly allows for zero-demand (no purchase). The term $m(D_h)$ captures the taste for quality as function of the household's characteristics, permitting a vertical dimension in consumer tastes. Households with a larger value of $m(D_h)$, will perceive a greater distance between the qualities of goods. $S(D_h)$ captures the effect of household characteristics on the scale of purchases. The α determines the curvature of the utility function. So long as the estimated value of α lies between 0 and 1, I maintain the concavity property needed for an interior solution.

The model captures household-level heterogeneity in several fashions. I specify the tastes for quality, the scale of purchases and the expected number of consumption needs (mean of the Poisson) as functions of observed household characteristics. I also capture heterogeneity with the random coefficients specification in the quality function:

$$\beta_j^h = \widetilde{\beta} + \gamma D_h + \Omega \sigma_j^h$$

where $\widetilde{\beta}$ captures the component of tastes for attributes that is common to all households and consumption needs. The $(k \times d)$ matrix of coefficients, γ, captures the interaction of demographics and tastes. Finally, Ω is a diagonal matrix whose elements are standard deviations and σ_j^h is a $(k \times 1)$ vector of independent standard normal deviates. Thus, for each household, the taste vector will be distributed normally with, conditional on demographics, mean $\widetilde{\beta} + \gamma D_h$ and variance $\Omega\Omega'$.

The household's problem consists of maximizing (4.2). For each household, there exists a vector of latent utilities, $u_j^* = (u_{j1}^*, ..., u_{jI}^*)$, where $u_{ji}^* = \max_Q u_j^h(\Psi_{ij}^h Q_{ij}, D_h)$ represents the utility from consuming the optimal quantity of product i for need j. The household selects the product yielding the highest latent utility for each occasion j. So, brand i is chosen to satisfy a given need if $u_{ji}^* = \max(u_{j1}^*, ..., u_{jI}^*)$. Assuming that any continuous quantity is permissible, the optimal quantity of brand i for need j solves the first order condition:

$$\alpha \left(\Psi_{ij}^h\right)^\alpha \left(Q_{ij}^h\right)^{\alpha-1} S_h - p_i = 0.$$

Rewriting the first order condition I obtain:

$$Q_{ij}^{h*} = \left(\frac{\alpha \left(\Psi_{ij}^{h} \right)^{\alpha} S_h}{p_i} \right)^{\frac{1}{1-\alpha}}. \tag{4.4}$$

To reformulate this problem to deal with integer quantities, I make use of the fact that the subutility functions are concave and monotonically increasing in Q_{ij}. Therefore, I simply need to consider the next highest and next lowest integer quantity to Q_{ij}^{h*}. I then compare the $2*I$ potential quantities, picking the one yielding the highest utility.

My objective is to estimate the mean and variance of the distribution of the random coefficients, $\beta^h = (\beta_1^h, ..., \beta_{J_h}^h)$, which are assumed to be distributed normally. I assume that the number of consumption needs in a given week, J_h, is distributed Poisson with the mean specified as a function of the household's characteristics and its purchase history, $\Gamma(D_h)$:

$$J_h \sim P(\Gamma(D_h)).$$

Given these assumptions, the overall expected total purchase vector for a given trip can be estimated conditional on the observable information and summed over the J_h consumption occasions:

$$EQ_h(D_h, X) = \sum_{J_h=1}^{\infty} \sum_{j=1}^{J_h} \int_{-\infty}^{\infty} Q_j^{h*}(D_h, \beta_j^h, \Theta) \Phi(d\beta | D_h, \Theta) P(dJ_h(D_h)). \tag{4.5}$$

Estimation requires specifying functional forms for the mean of the Poisson, $\Gamma(D_h)$, the vertical aspect of tastes, $m(D_h)$, and the scale of purchases, $S(D_h)$.

Comparison with the Standard DCM

One of the main features of the proposed random utility model is that it is a generalization of the standard discrete choice models. If I disregard the expected consumption occasions and I assume that consumers are restricted to single-unit purchases, then α no longer plays any role and (4.2)reduces to:

$$u_{hi} = X_i\beta^h S(D_h) - p_i, i = 1, ..., I.$$

I can divide through by $S(D_h)$ to get:

$$\begin{aligned} \widetilde{u_{hi}} &= X_i\beta^h - \frac{1}{S(D_h)}p_i \\ &= X_i\beta^h - \phi^h p_i \\ &= \left(X_i\widetilde{\beta} - \widetilde{\phi}p_i\right) + \left(X_i\Omega\sigma^h - \omega\sigma^h p_i\right) \end{aligned} \tag{4.6}$$

where the inverse of $S(D_h)$, ϕ^h, is the price-response parameter. Adding a random disturbance term directly in (4.6) gives me the standard random utility DCM (Manski and McFadden 1981).[36] The recent popularity of aggregate models for which the underlying consumer behavior reflects (4.6) makes them an interesting comparison model. By comparing the measures of market power and the effects of mergers in the proposed model to those of the aggregate DCM, I illustrate the importance of modeling multiple-discreteness explicitly.

---

[36] The random disturbance is generally assumed to derive from either the extreme value distribution, giving rise to the standard multinomial logit, or from the multivariate normal distribution, giving rise to the probit.

To derive the typical aggregate DCM (eg. Nevo and BLP), I add a Type I extreme value disturbance to (4.6) so that the conditional probability that consumer h chooses product j has the form:

$$P_{j|\beta^h} = \frac{\exp(\widetilde{u_{hj}})}{1 + \sum_{k=1}^{J} \exp(\widetilde{u_{hk}})}.$$

This model is often called the mixed logit model because it mixes the distribution of the random coefficients (which I assume to be normal) with the Type I extreme value error (McFadden and Train 1998). I include a $(J+1)$ product to represent the no-purchase option, the utility of which I normalize to zero. Product j's expected share of aggregate sales is determined by integrating over the set of consumers, A_j, for whom j represents the utility-maximizing choice:

$$S_j = \int_{A_j} P_{j|\beta^h} \Phi(d\beta). \tag{4.7}$$

By using normally-distributed random coefficients, A_j does not have a simple analytic form. Therefore, I use simulation methods to evaluate the integral in (4.7). As in the proposed model, this estimation strategy allows me to identify the mean taste coefficients, $\widetilde{\beta}$, as well as the variance terms, $\Omega\Omega'$. For details on the estimation of the aggregate DCM, I refer the reader to BLP(1995) or Nevo(2000). I provide a brief summary of the estimation techniques in the Appendix.

Unlike the proposed model, this alternative approach relies on aggregate market share data. I use the weekly shares of each product for the 22 stores in the largest chain in my data set. A key aspect of the specification of the underlying model involves the determination of the size of the market. Using similar supermarket data for ready-to-eat cereals, Nevo (2000) assumes consumers choose to eat a serving of cereal each day. He measures the potential market as the total number of per capita daily servings in a metropolitan area during a given quarter. Since CSDs are not a staple food, like breakfast cereals, there is no systematic way in which I expect households to consumer them (such as a daily serving). Instead, I define the potential market as the total number of weekly store-trips. I compute the no purchase share as one minus the sum of the share of store trips for a given week in which one of the CSD products is purchased.

If I only had aggregate data, I could try and justify this assumed shopping behavior as an approximation to the underlying demand process. However, my individual panel data reveals that households frequently purchase multiple units and multiple products on a given store trip. The specification error from assuming a model of single-unit purchases could bias my results. Even if I assume that I can break multiple-item trips into *independent* discrete choices, I am still unable to measure the size of the total potential market, which could be larger or smaller than the total weekly store traffic. For instance, if the definition of market size is too large, then the product shares are too small and the outside share is too large. A downward bias in the market shares generates a downward bias in the magnitudes of the taste coefficients. The combination of these biases generates incorrect elasticities. For instance, if I assume homogeneous tastes (no random coefficients), the cross-price elasticitiy with respect to product k is $\phi S_k p_k$, which will be underestimated. For policy analysis, the downward bias in cross-elasticities understates the unilateral market power associated with jointly pricing a product line, which leads to underestimating market power. The downward bias in substitutibility between goods understates the ability of a merger to internalize competition and raise prices. Therefore, the DCM benchmark could misclassify a potential merger as meeting with the official merger guidelines if the downward bias in post-merger prices is severe.

Endogeneity of Prices

A standard estimation problem encountered with the characteristics approach is the presence of unobserved (by the econometrician) attributes which may be correlated with the price. I alleviate the potential endogeneity of prices by including alternative-specific dummy variables that enter the quality function, Ψ_{hjt}. Given the short time span of my data (9 quarters), I estimate a single dummy for each product. In doing so, I am assuming that any unobserved attribute that could be correlated with price does not vary over time. Unlike most papers using the characteristics approach, my inclusion of transaction-specific features and display activity allows me to proxy for the time-varying, store-specific attributes that could influence consumer perceptions of quality.

Nonetheless, unobserved changes in package design, television advertising and shelf space could still introduce variations in households' perceptions of a product's quality during the sample period. These unobserved attributes bias estimation if they are correlated with any of the observed attributes. For instance, Besanko, Gupta and Jain(1998) find evidence of such high-frequency price endogeneity with weekly store-level data. In a similar study, Karunakaran (1998) finds price endogeneity, but not feature or display endogeneity. Villas-Boas and Winer (1999) document that such price endogeneity can even contaminate estimation with individual data. In the aggregate DCM benchmark model, I am able to resolve such high-frequency endogeneity by using the inversion procedure proposed by Berry (1994) and by using cost-side instruments (factor prices). However, the highly non-linear specification of the proposed model makes it difficult for me to extract much information from cost-side instruments. Although not reported, my results do not change substantially once I include factor prices in the instrument matrix. Two possible alternatives would be either to implement an inversion analogous to Berry (1994), to allow instruments to enter the model in a linear form, or to experiment with series approximations of the optimal instruments (Chamberlain 1987). For simplicity, I assume the marketing variables account for intertemporal variation in perceived quality.

Supply

The softdrink industry is an oligopoly with multiproduct firms. Given the previous empirical findings that prices are not collusive, I assume that each manufacturer maximizes the joint profits associated with its mix of products, which I treat as fixed.[37] I use a static model for technical simplicity (see Pakes and McGuire 1994 for a dynamic

[37]Given that many brands in the larger firms' portfolios were in fact stand-alone companies prior to their acquisition, I could have also modelled a scenario in which brand managers maximize the joint profits of the all the products marketted under the same brand umbrella. For instance, Coke's colas and Sprite would not be priced jointly.

model of differentiated products).[38] The data also include information on store-level advertising and display activity which, for simplicity, I treat as exogenous.[39] I also treat the retailers' pricing decisions as exogenous, which implies either a constant retail mark-up or perfectly competitive retail prices. While I could modify the model below to allow for strategic retailer behavior in the CSD category (as in Besanko, Gupta and Jain 1998), I lack sufficient data to identify weekly store-specific margins. In evaluating mergers, I assume that the large sunk costs associated with a new brand are prohibitively high to expect entry, even if a merger raises overall prices. I also make the standard assumption of the existence of a Bertrand-Nash equilibrium with strictly positive prices.

I model each of the F firms producing some subset, B_f, of the $i = 1, ..., I$ CSD products, making quantity and price decisions at a monthly frequency based on expected demand[40]. Ignoring a monthly time subscript, each firm f has the following cost function:

$$C_f(\{Q_i\}_{i\in B_f}) = C_f + \sum_{i\in B_f} mc_i Q_i$$

---

[38]The incorporation of lagged choice variables imply that current pricing decisions could influence future demand. The size of the effect depends on whether a decrease in current prices increases the number of customers who purchase the good or whether the same number of customers simply purchase more units. For instance, a dynamic model might use consumer penetration as the firms' state variable.

[39]For instance, the retailer makes a weekly exogenous draw from a distribution of store-wide products to advertise and display. This assumption is consistent with the findings of Slade (1995), whereby retailers compete for overall offerings, rather than on a product-by-product basis.

[40]The assumption of monthly price-setting does not seem too unrealistic. However, this assumption is mainly a technical convenience to ensure I have enough observed purchases to compute meaningful markups.

where Q_i is the quantity produced of product i and C_f measures the overall fixed costs incurred by firm f. Thus, firm f earns expected profits:

$$\pi_f = \sum_{i \in B_f} (p_i - mc_i) E\{Q_i(p^w)\} - C_f$$

where $E\{Q_i(p)\}$ is the expected demand for product i in a given month, which is a function of the prices of all the products. Assuming the existence of a pure-strategy static Bertrand-Nash price equilibrium with strictly positive prices, each of the prices, p_i $i \in B_f$, satisfies the following first-order conditions:

$$E\{Q_i(p)\} + \sum_{k \in B_f} (p_k - mc_k) \frac{\partial E\{Q_k(p)\}}{\partial p_i} = 0, i \in B_f, f = 1, ..., F.$$

I construct the following $(J \times J)$ matrix $\mathbf{\Delta}$ with entries as follows:

$$\widetilde{\Delta_{jk}} = \begin{cases} -\frac{\partial E(Q_{jk})}{\partial p_i}, if \ \exists \text{ f s.t. } \{i,k\} \subset B_f \\ 0, else \end{cases}.$$

Stacking the prices, marginal costs and expected quantities into $(J \times 1)$ vectors, $\mathbf{Q}$, $\mathbf{p}$ and $\mathbf{mc}$ respectively, the first-order conditions can be written in matrix form:

$$E(\mathbf{Q}) - \mathbf{\Delta}(\mathbf{p} - \mathbf{mc}) = 0.$$

From the first-order conditions, I derive the mark-up equation:

$$\mathbf{p} - \mathbf{mc} = \mathbf{\Delta}^{-1} E(\mathbf{Q})$$

As in Hausman, Leonard and Zona (1994), Goldberg (1995), and Nevo (2000) I estimate these mark-ups directly from the estimated demand parameters, without using information on costs.

I am able to recover **mc** by combining the computed wholesale markup with observed prices:

$$\mathbf{mc} = \mathbf{p} - (\boldsymbol{\Delta})^{-1} E(\mathbf{Q}). \tag{4.8}$$

Since manufacturers likely set their prices less frequently than weekly, I use the quarterly average price (averaging across store-weeks in a given month) to recover quarterly measures of the marginal costs.

Computing Counter Factual Equilibrium Prices and Welfare

Once I have estimates of demand parameters and marginal costs, I consider the effects of the hypothetical mergers. To back out the equilibrium prices that prevail after a merger, I make use of the first-order condition (4.8). One simple approach to obtaining the new prices is to assume that the markup term is approximately independent of the price (Hausman, Leonard and Zona 1994), and to compute the new price vector using (4.8). The changes in prices are driven entirely by the changes in Δ, which reflect the changes in ownership of the products in question. Alternatively, I could solve the equation:

$$\mathbf{p}^{*} = \mathbf{mc} + \boldsymbol{\Delta}(\mathbf{p}^{*})^{-1} E\{\mathbf{Q}(\mathbf{p}^{*})\}$$

for $\mathbf{p}^{*}$. The large number of prices make it impossible to solve this system analytically. Instead, I must solve the system numerically. Nevo finds that the approximation and the numerical methods generally give similar predictions, unless the mergers lead to drastic changes in prices and quantities. Solving the proposed model numerically requires additional work to deal with the non-smooth demand (see Appendix).

Once I have the new equilibrium prices, I can compute the changes in profits and in aggregate consumer welfare. I measure the change in consumer welfare as the compensating variation of the merger. Effectively, I compute the aggregate change in income that would ensure that consumers maintain their pre-merger utilities. Recall that the utility function has the form:

$$\begin{aligned} U^h(p, Q, y_h) &= \sum_{j=1}^{J_h} u_j^h\Big(\sum_{i=1}^{I} \Psi_{ij}^h Q_{ij}^h, D_h\Big) - \sum_{j=1}^{J_h}\sum_{i=1}^{I} p_i Q_{ij}^h + y_h \\ &= u^h(\mathbf{p}, \mathbf{Q}^h) + y_h. \end{aligned}$$

So, the change in income needed to equate pre and post merger utilities is:

$$\Delta Y = \sum_{h=1}^{H}\sum_{t=1}^{T_h} \Delta y_h = \sum_{h=1}^{H}\sum_{t=1}^{T_h} \left[u^h(\mathbf{p}, \mathbf{Q}(\mathbf{p})) - u^h(\mathbf{p}^*, \mathbf{Q}(\mathbf{p}^*))\right].$$

In fact, this approach is very similar to the one commonly used with DCMs (Trajtenberg 1989). Since I do not observe the entire market, I am unable to compute aggregate absolute changes in profits or welfare. Instead, I report changes as percentages, since the sample should approximate the proportional changes.

Model Estimation

The model of consumer utility from soft drink consumption yields the following equation for the vector of household h's expected demand for each alternative at time t, conditional on the $(K \times 1)$ matrix of household/trip attributes, D_{ht},:

$$Q_{ht}(D_{ht},\Theta)=\sum_{i=1}^{\infty}\sum_{j=1}^{J_{ht}}\int_{-\infty}^{\infty}Q^{*}_{jht}(D_{ht},\beta_j^h,\Theta)\Phi(d\beta|D_{ht},\Theta)P(dJ_h(D_{ht})),\ \text{h=1,...,H, t=1,...,T}_h$$

where $Q^{*}_{jht}(D_{ht},\beta_j^h,C)$ is the $(I\times 1)$ vector of optimal quantities of each alternative for task j on trip t, β_j^h is a vector of random taste coefficients for task j, and Θ is a vector of parameters to be estimated. The random taste coefficients are drawn from the normal distribution, $\Phi(\bullet|D,\Theta)$, conditional on the information D, and the number of tasks are drawn from the Poisson distribution, $P(\bullet|D)$. Thus, the vector of expected soft drink purchases for each household is computed as the sum of the expected purchases for each task, conditional on a specific number of tasks, J_{ht}, and weighted by the probability that J_{ht} is the true number of tasks at time t.

Using this formulation, I can define the prediction error:

$$\varepsilon_{ht}(D_{ht},\Theta)=Q_{ht}(D_{ht},\Theta)-q_{ht} \tag{4.9}$$

where q_{ht} is the vector of actual purchases of each of the alternatives by household h at time t. If the model represents the true purchasing process, then at the true parameter values, Θ_0 :

$$E\left\{\varepsilon_{ht}\left(D_{ht},\Theta_0\right)\right\}=\overrightarrow{0_I}\ \text{for h=1,...,H and t=1,...,T}_h. \tag{4.10}$$

I also assume that:

$$E\left\{\varepsilon_{ht}\left(D_h,\Theta_0\right)\varepsilon_{hk}\left(D_{ht},\Theta_0\right)'\right\}=\Omega_{tk}, \tag{4.11}$$

where Ω_{tk} is a finite $(I\times I)$ matrix. This assumption implies that the households' prediction errors are distributed identically. Following Hansen (1982) and Chamberlain (1987), any function of the observable data, $\mathbf{D}_{ht}$, that is independent of the unobservables must be conditionally uncorrelated with ε_{ht} at $\Theta = \Theta_0$[41].Given such a function, $Z_{ht} = f(D_{ht})$, I can construct a generalized method of moments estimator from the conditional moment restrictions:

$$E\left\{Z_{ht} * \varepsilon_{ht}\left(D_{ht}, \Theta_0\right) | Z_{ht}\right\} = \overrightarrow{0_I}. \tag{4.12}$$

I use (4.12)to construct the moment conditions:

$$h(D_{ht}, \Theta) = Z_{ht} * \varepsilon_{ht}\left(D_{ht}, \Theta\right)$$

where $\Theta \in R^k$, and $E\left\{h\left(D_{ht}, \Theta_0\right) | Z_{ht}\right\} = 0$. Let $\mathbf{D}_{HT} \equiv (D'_{1T1}, ..., D'_{HT_H})$ denote the matrix containing all of the household/trip information for the sample of H households, where household h makes T_h shopping trips. Using the notation $T = \frac{1}{H}\sum_{h=1}^{H} T_h$, the sample analogue of the moment conditions has the following form:

$$g(D_{HT}, \Theta) = \frac{1}{HT}\sum_{h=1}^{H}\sum_{t=1}^{T_h} h\left(D_{ht}, \Theta\right). \tag{4.13}$$

As H and T_h grow large, g(D_{HT}, Θ_0) should approach zero. Hansen's (1982) formulation involves finding a matrix, Θ_{GMM}, that makes g(D_{HT}, Θ_{GMM}) as close as possible to zero. I choose a value of Θ_{GMM} that minimizes the function J_{HT} given by:

$$J_{HT}(\Theta) = \left[g\left(D_{HT}, \Theta\right)\right]' W_{HT}\left[g\left(D_{HT}, \Theta\right)\right] \tag{4.14}$$

where W_{HT} is the efficient weighting matrix, given by the asymptotic variance of g. The estimation of W_{HT} is discussed below. This framework gives estimates with the

[41] I am assuming that the process generating the prediction errors, the demand shocks, is uncorrelated with the point-of-purchase marketing environment. For instance, newspaper advertising for a product does not drive the residual process.

following asymptotic distribution:

$$\sqrt{N}(\Theta_{GMM} - \Theta_0) \implies N(0, \Xi), \tag{4.15}$$

$$\text{with } \Xi = \left(plim \left\{ \frac{dg(D_{ht}, \Theta_0)}{d\Theta} \right\} W plim \left\{ \frac{dg(D_{ht}, \Theta_0)}{d\Theta} \right\}' \right)^{-1} \tag{4.16}$$

In order to compute the sample moment conditions, I must evaluate an integral of infeasibly large dimension. Following McFadden(1989) and Pakes and Pollard (1989), I use Monte Carlo methods to simulate these integrals. For each household trip, R independent draws are taken from the Poisson distribution to simulate the number of tasks. For each of these R draws, (N+I-1)$\times K$ draws are taken from the normal distribution to simulate the taste coefficients for these tasks, where K is a sufficiently large number to be an upper bound on the number of tasks simulated for each household. These draws are then used to construct R simulations of the expected purchase vector at each trip, $Q^r_{ht}(D_{ht}, \Theta)$ $r = 1, ..., R$. These estimates are then combined to form an unbiased simulator of the expected purchase vector, $\widehat{Q_{ht}}(D_{ht}, \Theta)$:

$$\widehat{Q_{ht}}(D_{ht}, \Theta) = \frac{1}{R} \sum_{r=1}^{R} Q^r_{ht}(D_{ht}, \Theta).$$

By construction, the simulations, Q^r_{ht} come from the same distribution as q_{ht}. So the variance of $\widehat{Q_{ht}}(D_{ht}, \Theta)$ will be $\frac{1}{R} var(q_{ht})$, which goes to zero as $R \to \infty$. I can write $\widehat{Q_{ht}}(D_{ht}, \Theta) = Q_{ht} + \zeta_{ht}$, where ζ_{ht} is the simulation error with $E(\zeta_{ht}) = 0$ and var($\widehat{Q_{ht}}$) =var(ζ_{ht}). I now simulate the moment conditions by substituting $\widehat{Q_{ht}}(D_{ht}, \Theta)$ for $Q_{ht}(D_{ht}, \Theta)$ in (4.14):

$$g^s_{HT}(\Theta) = \frac{1}{HT}\sum_{h=1}^{H}\sum_{t=1}^{T_h}\left[Z_{ht} * \left(\widehat{Q_{ht}}(D_{ht},\Theta) - q_{ht}\right)\right] = \frac{1}{HT}\sum_{h=1}^{H}\sum_{t=1}^{T_h} h^s\left(D_{ht},\Theta\right) \tag{4.17}$$

So long as H is sufficiently large, the resulting method of simulated moments estimate, Θ_{MSM}, will be consistent and will have asymptotic variance $\Xi = \left(\frac{dg^s(\Theta_0)}{d\Theta}' W_{HT} \frac{dg^s(\Theta_0)}{d\Theta}\right)^{-1}$.

Estimation of the Weight Matrix, W

The estimation of W_{HT} is complicated due to both the simulation error and the panel aspect of the data. The simulation error simply adds extra variation to the procedure, as demonstrated below. The panel aspect of the data requires some additional assumptions regarding both cross-sectional and intertemporal variation of the residual process. I include several state variables, such as temperature and seasonal dummies, to capture contemporaneous aggregate demand shocks that could affect households in a similar fashion. Having included these controls, I assume that the prediction errors are uncorrelated across households. However, most households have fairly long purchase histories, allowing the possibility of persistent unobserved shocks. The source of these shocks could be measurement error. For instance, household-specific reporting errors in the scanning process could generate unobserved serial dependence. By including observed time-varying factors in the mean of the Poisson function, I assume that this serial dependence is independent of the process generating the number of consumption needs. Therefore, only the covariances of the prediction errors need to be corrected.

In a standard GMM setting, Hansen(1982) shows that the efficient weighting matrix W_{HT} is the inverse of S, the variance of the sample moments:

$$\begin{aligned}
S &= \lim_{HT\to\infty} HT \cdot E\left\{E\left(\left[g\left(D_{HT},\Theta_0\right)\right]\left[g\left(D_{HT},\Theta_0\right)\right]'\left|D_{HT}\right.\right)\right\} \\
&= \lim_{HT\to\infty} HT \cdot E\left\{E\left(\left[\frac{1}{HT}\sum_h\sum_t h^s\left(D_{ht},\Theta_0\right)\right]\left[\frac{1}{HT}\sum_h\sum_t h^s\left(D_{ht},\Theta_0\right)\right]'\left|D_{HT}\right.\right)\right\} \\
&= \lim_{HT\to\infty} HT \cdot \frac{1}{H^2T^2}\sum_{h=1}^{H}\sum_{t=1}^{T_h}\sum_{k=1}^{T_h} E\left\{E\left(\left[Z_{ht}\left(\widehat{Q_{ht}}-q_{ht}\right)\right]\left[Z_{hk}\left(\widehat{Q_{hk}}-q_{hk}\right)\right]'\left|Z_{ht},Z_{hk}\right.\right)\right\} \\
&= \lim_{HT\to\infty} \frac{1}{HT}\sum_{h=1}^{H}\sum_{t=1}^{T_h}\sum_{k=1}^{T_h} E\left\{E\left(\left[Z_{ht}\varepsilon_{ht}\varepsilon_{hk}'Z_{hk}' + Z_{ht}\zeta_{ht}\zeta_{hk}'Z_{hk}'\right]\left|Z_{ht},Z_{hk}\right.\right)\right\} \\
&= \lim_{HT\to\infty} \frac{1}{HT}\sum_{h=1}^{H}\sum_{t=1}^{T_h}\sum_{k=1}^{T_h} E\left[Z_{ht}\Omega_{tk}Z_{hk}' + Z_{ht}\frac{1}{R}\Omega_{tk}Z_{hk}'\right] \\
&= \lim_{HT\to\infty} \frac{1}{HT}\sum_{h=1}^{H}\sum_{t=1}^{T_h}\sum_{k=1}^{T_h} E\left[\left(1+\frac{1}{R}\right)Z_{ht}\Omega_{tk}Z_{hk}'\right].
\end{aligned}$$

As in McFadden (1989), the added simulation "noise" does not affect the consistency of the estimator, but it will reduce the efficiency by a factor of $\left(1+\frac{1}{R}\right)$. As $R\to\infty$, I approach the asymptotically efficient covariance matrix. In this paper, I use 30 simulation draws ($R = 30$) and assume that they suffice to eliminate any added simulation noise.

I now address how the panel aspect of my data affects the estimation of S. To model the residual process more formally, I assume the values of a given household's prediction error on a given trip are determined by the values of an underlying random field,

ε_s, at location s_{ht} on a lattice H. I index each observation's location by both time and household. I then allow for serial dependence between observations depending on their relative locations on the lattice. Technically, I could allow for dependence both across households and over time. As discussed above, I only treat intertemporal dependence to simplify the estimation procedure.[42] Conley(1999) provides limiting distributions and covariance estimation techniques for this more general setting. I use Conley's non-parametric , positive semi-definite covariance estimator which is analogous to the time-series estimator of Newey and West (1987). Given a consistent estimate $\widehat{\Theta}$ and a pre-determined time L after which the unobserved household-specific shocks die out, the estimator for S is:

$$\widehat{S}_{HT} = \frac{1}{HT}\sum_{h=1}^{H}\sum_{t=1}^{L}\sum_{k=t+1}^{T_h}\omega(t)\left[h^s\left(D_{h,k},\widehat{\Theta}\right)h^s\left(D_{h,k-t},\widehat{\Theta}\right)' + h^s\left(D_{h,k-t},\widehat{\Theta}\right)h^s\left(D_{h,k},\widehat{\Theta}\right)'\right]$$
$$-\frac{1}{HT}\sum_{h=1}^{H}\sum_{k=1}^{T_h}h^s\left(D_{h,k},\widehat{\Theta}\right)h^s\left(D_{h,k},\widehat{\Theta}\right)'$$

where $\omega(t)$ is specified as the Bartlett weight:

$$\omega(t) = \begin{cases} 1-\frac{|t|}{1+L} \text{ if } 1 \leq L \\ 0, \text{ else} \end{cases} .$$

This scheme assigns decreasing weight to the correlation between a given household's purchases as they grow further apart in time.

---

[42]Intuitively, I do not expect the unobservables generating a given household's choice process to affect other "close" households' choice processes for a given product category. However, I do expect some such "spatial" dependence for the overall shopping choice. For instance, a local convenience store might affect households in a community in a similar fashion. This form of dependence is the subject of work in progress, joint with Tim Conley of Northwestern University.

The resulting estimate for W is $\widehat{W}_{HT} = \widehat{S}_{HT}^{-1}$. Finally, the estimate for the covariance matrix of the estimated parameters is $\frac{1}{HT} \cdot \left(\frac{dg^*(\Theta_{MSM})'}{d\Theta} \widehat{W_{HT}} \frac{dg^*(\Theta_{MSM})}{d\Theta} \right)^{-1}$.

Identification:

I now discuss several data identification issues for the proposed econometric procedure. First, I explain how the data identify the joint distribution of the total number of products and the total number of CSD units purchased on a given trip. Then, I explain how I identify the residual process and the GMM weights in the presence of a large number of moment conditions.

Since I do not observe individual expected needs on a given trip, I estimate aggregate demand per trip. Despite the fact that I do not observe specific needs, I am still able to identify the process that generates them. The main identification problem involves the distinction between a household purchasing 5 units of CSDs to satisfy five needs versus 5 CSDs to satisfy one single need. Since the random tastes are independent across consumption needs, a household with several needs will tend to purchase several different types of CSD. Alternatively, a household with a single consumption need will only purchase one type of CSD. Thus, the number of consumption needs will determine the joint distribution of the total number of units of CSDs purchased and the number of different brands.

For example, I find in the data that both the total number of CSDs and the number of different types of CSDs purchased on a trip increases with the size of the household. Therefore, household size enters both the scale function, $S(D)$, and the mean of the Poisson, $\lambda(D)$. Since the function $S(D)$ enters the per-task optimal quantity choice in (4.4), it will be instrumental in the identification of $\lambda(D)$ and total quantity per consumption need. Similarly, the use of demographic variables in determining $m(D)$ in (4.3) enables the joint identification of $\lambda(D)$ and the taste parameters, β. Although several different sets of parameter values could give the same likelihood for expected total purchases, they will not have the same likelihood for the joint distribution of total products and total units purchased. Since the sample households tend to purchase bundles containing several different CSD brands, the data will identify this joint distribution.

The assumed independence of tastes across consumption needs rules out potential shopping externalities. Purchasing a 12-pack of colas for one expected need does not influence the choice for another need. This assumption seems less of a problem for CSDs than for the purchase of computers, for which there could be obvious shared software-related externalities. Nonetheless, the fact that a consumer has already purchased a cola to satisfy one need might increase the likelihood of purchasing a non-cola to satisfy another. One way in which I could link the choices made during a given trip would be to introduce interaction dummy variables in the utility function. For instance, I could classify all the CSDs in the sample into five flavor groups. While simulating the contemporaneous choices, I would introduce flavor interaction terms that would reflect which flavor combinations have been selected across needs. In addition to providing a link across the consumption needs, these flavor interaction terms would also provide a statistical test for complementarities between flavors. The test would be a simple significance test for whether a given pairwise flavor combination has a positive, negative or zero effect on utility.

With regards to the estimated residual process, I find that the correction for within-panel serial dependence has a noticeable effect on the standard errors of my parameters. However, given the large number of instruments and products, I could run into some trouble with identification if I estimate the underlying covariance matrix freely. For now, the only restrictions I impose are the second moment independence of the instruments and the errors. Even so, with 26 products I still estimate the (26×26) residual covariance matrix, Ω, and a $(K \times K)$ instrument covariance matrix, $E(Z_{h,t}Z'_{h,t+l})$, for each lag l. For precision, I may need to impose some additional restrictions on subsequent estimations. One way to think about valid restrictions is to consider the source of these shocks. For instance, households may randomly shop at a non-sample store, such as a convenience store. I expect this sort of measurement error to have some persistence. However, the persistence may only be for products of the same size. So, the fact that you purchase a 67.6 ounce bottle in a convenience store may only affect the prediction of other 67.6 ounce bottles. In this case, I could set some of the off-diagonal terms between different size products in the autocovariance matrices to zero to improve the identification.

Data

The Market

During the trial against Coca-Cola, the court ruled that 33 large US cities each constitute an independent CSD market and that the entire country represents yet another market. Although the use of a single city cannot capture the effects of national CSD competition perfectly, the Denver market presents a very interesting basis of study as its demographic base is perfect for CSD consumption. According to a recent article[43] the population of Denver is unusually young, athletic and outdoors-oriented. For the year ending in January of 1995, Denver had a booming economy and the median age was 33.5, one of the lowest in the country. The Consumer Expenditure Survey claims that Americans between the ages of 35 and 44 hold the largest share of soft drink consumption.[44]

On the production side, a local law prohibiting the sales of alcoholic beverages exceeding a proof of 3.2 percent has kept alcohol sales unusually low. Denver is a "shared ad market," meaning that a given retailer may promote the leading products, Coke and Pepsi, at the same time. The consequence of this market structure is high CSD

---

[43]Hilary S. Miller [1995], "Rocky Mountain Fever." *Beverage Industry*, 86, 47-51.

[44]Sarah Theodore (1997), "Soft drink demographics hinge on age and demographics," *Beverage Industry*, 88, 48-50.

demand and intense competition. However, Denver has been one of the few submarkets in which Pepsi outperforms Coca-Cola both in the cola segment and for overall CSDs.[45]

Data

The scanner data, collected by A.C. Nielsen, cover the Denver area between January of 1993 and March of 1995. These data include consumer information for a random sample of 2108 households as well as weekly store level information for 58 supermarkets with over $2 million in "all commodity volume". The store level information consists of weekly prices, sales, feature and display activity for 26 diet and regular products with a combined share of 51% of the household-level category sales. This list of relevant products includes all UPCs with at least one percent of the aggregate sales volume. The household level data covers all shopping trips for these items. For each trip, I know the date, the store chosen and the quantities purchased. For each alternative available within the store, I know the prices and whether the product was featured in a newspaper or as an in-aisle display. Combining the store and purchase data sets, I observe the full set of prices and the in-store marketing environment for all the alternatives on a given trip.

Table (4.40), in the Data Appendix, illustrates the incidence of multiple-item shopping in the sample. Conditioning on the occurrence of a purchase, only 39% of the trips

---

[45] "15 U.S. Markets: Coke Leads in 11. Pepsi Leads in 4. Biggest Gains: Pepsi in NY; Coke in Minneapolis/St. Paul." *Beverage Digest*, July 18, 1997.

involve a single unit of a single brand. In fact, 31% of the trips result in the purchase of at least two units of a single product and another 31% of the trips result in the purchase of two or more products. These findings confirm the fact that the potential market for CSDs exceeds simple store-traffic (which is analogous to assuming single-unit purchasing on shopping trips).

For each shopping trip, I construct a quality measure for each product. The quality consists of three components: fixed physical attributes, time-varying attributes and household-specific loyalty. The fixed physical attributes consist of the ingredients of the product, which I collected from the nutritional information printed on the product packages. These characteristics include total calories, total carbohydrates, sodium content (in mg), and a set of dummy variables that indicate the presence of caffeine, phosphoric acid, citric acid, caramel color and no color. I report these attributes as per-12-ounce-serving, using 4 additional dummy variables to distinguish between package sizes: 6-pack of 12 oz cans, 12-pack of 12 oz cans, 6-pack of 16 oz bottles and 67.6 oz bottles.

The time-varying attributes are the prices and the marketing mix variables, feature and display. Finally, the household-specific loyalty variables are two dummy variables indicating whether the same brand and same UPC respectively were chosen during the most recent shopping trip on which a purchase occurred. While such loyalty variables are typical in the marketing literature, most empirical IO studies have not had sufficient data to include them. Studies that omit these loyalty terms when they matter will suffer from strong unobserved persistence in the residual process. So long as I control for heterogeneity sufficiently, my estimated loyalty coefficients will not be spurious.

In the Appendix, I provide summary statistics of the demographic variables and the product attributes used in my estimation. I also provide descriptive statistics for the data in the 22 stores used for the aggregate DCM.

Results

Parameter estimates

I now present parameter estimates for four specifications of the proposed model. These specifications differ mainly in their inclusion of random coefficients and interaction terms between demographic variables and product attributes. The second model includes a random intercept in the mean of the Poisson process. Adding a random intercept implies that unobserved household-specific random effects also drive expected consumption needs. The third model also includes interaction terms between some of the demographic variables and certain attributes. In the fourth specification, I make the valuation of citric and caramel random, to allow for more heterogeneity in tastes. In general, I find the most striking differences between these models to come from the addition of the random intercept in the Poisson (models 2, 3 and 4), which changes the relative magnitudes of several variables. All four models are estimated with a full set of product-specific fixed-effects, which I do not report to conserve space. I only report the parameter estimates from the GMM procedure. In Dubé(2000), I project the product fixed-effects onto the physical attribute space to study how consumer tastes for these attributes drive substitution patterns. I also experiment with more sophisticated notions of prices, such as reference prices.

Table (4.14) reports the taste coefficients that enter the quality function, ψ. While the inclusion of a random intercept in the Poisson process (models 2 and 3) changes a few of the parameters, the addition of demographic interactions (model 3) does not lead to substantial qualitative differences. Similarly, the addition of the random terms on citric acid and caramel (model 4) does not seem to change the results; although the attributes are significant. The addition of the random intercept in the Poisson process (models 2, 3 and 4) causes both the mean and variance of the taste for feature ads to decrease, while those of in-aisle displays increase. These changes suggest that some of the random response to marketing variables in the first model was proxying for random needs. Despite these changes, marketing variables appear to have a strong positive influence on purchasing, although households differ substantially in their tastes for these terms. I also find that controlling for both the brand and the specific product chosen on the previous trip seems to explain a lot of the perceived quality. The results suggest that loyalty to a specific brand might be stronger than loyalty to a given UPC. For instance, consumers are slightly more loyal to Coca-Cola in general than to a specific package size of Coca-Cola.[46] The demographic interaction terms and the additional random coefficient on citric are significant, suggesting that the first two models do not pick up all of the heterogeneity.

The models predict significant unobserved heterogeneity in consumer perceptions of product-specific quality (the standard deviation of the product fixed effects). Ideally, I would interact the product dummies with demographic variables to try to characterize these differences in perception. However, these interactions would require too many additional parameters. Instead, I focus on specific product attributes to explain some of these differences.

As expected, households with a female head under 35 years old tend to have higher preferences for diet products, a well-documented fact in the CSD industry[47]. In fact, I might find additional explanatory power in dummies such as female head with a college degree.[48] Similarly, larger households place slightly more weight on products with more 12-ounce servings, such as the 12-pack. Unexpectedly, households with kids place a higher weight on products with caffeine than without. Part of this effect may be due to

---

[46]The interpretation of these terms is controversial. If I have controlled for heterogeneity adequately, then these terms suggest there may be habit-persistence. However, they could also be picking up unmeasured heterogeneity. Also, these terms are not entirely structural since the model does not allow consumers to consider the expected future effects of a current purchase. Technically, the model should have consumers solving a dynamic program to resolve the structural nature of the lagged terms.

[47]In Europe, Diet Pepsi was reintroduced as Pepsi Max, with twice the caffeine, to overcome its "feminine" image.

[48]This fact is documented in "Just who's buying all these soft drinks, anyway?" *Beverage Industry*, 84(3), 1993.

the limited scope of products included. In particular, many of the caffeine-free products, such as 7UP and Sprite, tend to appeal more to adults. In contrast with Nevo(2000) who finds little additional random tastes after including demographics variables, I still find evidence for unobserved heterogeneity in tastes for package size (number of 12 ounce servings) and diet, despite controlling for demographic interactions.

Now I present the terms that help determine the other features of the model. For now, I assume a simple linear form for these terms:

$$\begin{aligned} \lambda_h &= \lambda_0^h + \lambda_1 kids + \lambda_2(family\ size) + \lambda_3(last\ trip) \\ &\quad + \lambda_4(last\ csd\ trip) + \lambda_5 temperature + \lambda_6 holiday \\ scale &= s_0 + s_1(family\ size) + s_2(last\ trip) + s_3(last\ csd\ trip) \\ m &= 1 + m_1 income. \end{aligned}$$

Table (4.25) presents the estimated coefficients. Beginning with the mean of the Poisson process, λ, I find heterogeneity in the expected number of household needs. I find that the expected number of needs depends on the presence of kids and, to a lesser extent, on family size. The inclusion of a random intercept increases the importance of kids, while decreasing the role of family size in determining the expected number of consumption needs. Similarly, temperature no longer has much effect on expected needs. In contrast, the second, third and fourth models both exhibit strong positive effects from holidays. Surprisingly, the time since last trip and since last CSD purchase do not appear to explain much of the expected needs, especially in the second and third models. I anticipated that

Quality Function	Model 1	Model 2	Model 3	Model 4
ad	1.13	0.66	0.74	0.67
	(0.02)	(0.02)	(0.03)	(0.04)
s.d. ad	0.53	0.03	0.04	0.03
	(0.02)	(0.02)	(0.03)	(0.09)
display	0.95	3.29	3.12	3.35
	(0.02)	(0.07)	(0.05)	(0.07)
s.d. display	0.19	0.57	0.62	0.57
	(0.01)	(0.07)	(0.03)	(0.04)
brand loyalty	2.28	3.56	5.55	3.58
	(0.04)	(0.06)	(0.06)	(0.08)
prod. loyalty	0.94	1.25	1.19	1.21
	(0.07)	(0.31)	(0.14)	(0.15)
s.d. product	1.47	3.19	3.37	3.15
	(0.02)	(0.04)	(0.03)	(0.04)
s.d. diet	0.79	0.57	0.63	0.10
	(0.02)	(0.02)	(0.03)	(0.13)
s.d. citric				0.15
				(0.14)
s.d. caramel				0.58
				(0.03)
s.d. 6-pack	0.99	1.09	1.02	1.09
	(0.02)	(0.06)	(0.03)	(0.04)
s.d. 12-pack	0.58	0.04	0.04	0.04
	(0.01)	(0.01)	(0.03)	(0.01)
s.d. 16oz	1.68	0.15	0.14	0.10
	(0.08)	(0.05)	(0.11)	(0.04)
$kid * caffeine$			0.25	
			(0.01)	
$(family\ size) * servings$			0.02	
			(0.00)	
$(female < 35) * diet$			0.44	
			(0.03)	
Obs	169,788	169,788	169,788	169,788

Table 4.24. Taste Coefficients for Time-Varying Attributes in the Quality Function (standard errors in parentheses)

these terms would proxy for inventory effects. In a previous version of the model, I found a similar insignificantly small effect from an explicit measure of inventory.

The scale of purchases is also increasing in the number of people in the household, especially in the second, third and fourth models. Once again, the effects of time since last trip and time since last CSD purchase are very small (and insignificant). The vertical component is increasing in income, so that households with higher income perceive more distance between products, although this effect diminishes with the addition of the random intercept. Finally, the estimated values of α are positive and below one, which is consistent with the notion that utility is concave.

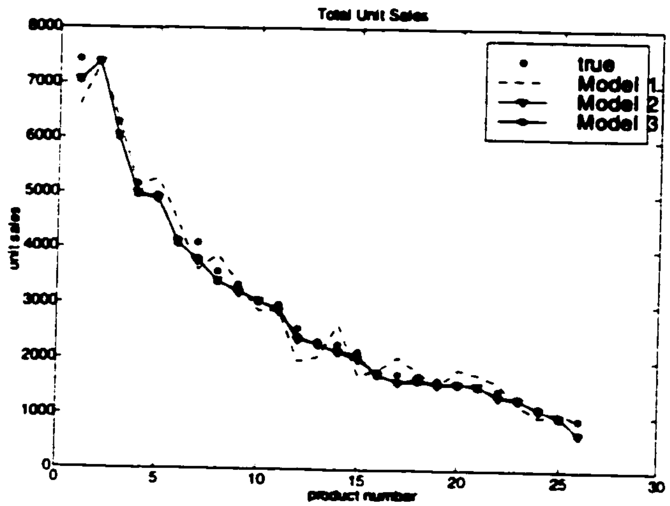
The reported standard errors have been corrected to account for potential serial-dependence. I attempt to control for many of the observed dynamic factors such as timing of trips, loyalty and inventories. Despite these controls, I still find unexplained persistence in the residuals. Accounting for time-series effects increases some of the standard errors by as much as a factor of 1.8. Nonetheless, almost all the parameters remain significant after this correction, probably due to my extremely large sample. As an experiment, I recompute the residuals after setting all of the coefficients for the dynamic factors to zero. I find that the standard errors rise by about 50% on average, some almost double. Therefore, my dynamic controls are picking up a fair bit of the intertemporal effect. Next, I take the actual residuals and average them by product for each household over time. If the model is failing to pick up some of the heterogeneity, I should see non-zero values of these averages, much like a household-specific random

variable	Model 1	Model 2	Model 3	Model 4
lambda: constant		0.078	0.083	0.078
		(0.002)	(0.006)	(0.002)
lambda: kids	0.076	0.139	0.134	0.140
	(0.005)	(0.002)	(0.005)	(0.002)
lambda: family size	0.060	0.001	0.001	0.001
	(0.002)	(0.000)	(0.000)	(0.000)
lambda: time since last csd	0.001	-0.001	-0.001	-0.001
	(0.000)	(0.000)	(0.000)	(0.000)
lambda: time since last trip	-0.001	-0.001	-0.001	-0.001
	(0.000)	(0.000)	(0.000)	(0.000)
lambda: temperature	0.001	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
lambda: holiday	0.005	0.170	0.163	0.170
	(0.002)	(0.003)	(0.006)	(0.003)
lambda: random term		0.052	0.050	0.052
		(0.001)	(0.003)	(0.001)
scale: constant	1.825	-1.008	-0.876	-1.006
	(0.069)	(0.066)	(0.022)	(0.069)
scale: family size	1.292	4.690	4.643	4.699
	(0.077)	(0.154)	(0.105)	(0.152)
scale: time since last trip	0.000	0.000	0.000	0.000
	(0.002)	(0.003)	(0.001)	(0.000)
scale: time since last csd	0.000	0.000	0.000	0.000
	(0.002)	(0.004)	(0.005)	(0.002)
vertical: income	2.059	0.751	0.731	0.746
	(0.129)	(0.019)	(0.059)	(0.022)
alpha	0.031	0.034	0.034	0.034
	(0.001)	(0.001)	(0.001)	(0.001)

Table 4.25. Non-Linear Coefficients (standard errors in parentheses)

effect for each product. In fact, for the top 6 products, I observe about 60% of these random effects lying between $(-.01, .01)$. I also observe about 20% of the random effects lying in (-.2,.2) in a bell-curve like fashion. Thus, I suspect that part of the intertemporal persistence I pick up is from mismeasured heterogeneity. The remainder of this observed persistence must be from some form of unobserved measurement error.

Figure (4.18) provides a rough idea of how well the proposed models fit the aggregate data. I omit the fourth model since it does not appear look much different from the second and third. In general, the first model does not fit the data quite as well as the second and third. Since the GMM procedure does not provide an estimate of the joint distribution of the data, as in maximum likelihood, I am not able to provide a statistical test for the fit of the model. To lend more credence to the proposed model, I could re-estimate the system with the first eight quarters of data, leaving the ninth quarter as a hold-out sample. I would then use this ninth quarter to verify how well the model predicts out-of-sample.



(4.18)

Results from the Aggregate DCM

In Table (4.26) I report the parameter estimates for two specifications of the aggregate DCM using weekly store-level data. In the second specification, I add random coefficients on some of the product attributes to be more consistent with the proposed model. For each model, I report both the mean and the standard deviation of each random taste parameter. For both models, I find substantial heterogeneity in the price response coefficient. However, I do not find the standard deviations of the ad and display coefficients to be significantly different from zero in the first specification. Contrary to the findings of the proposed model, the DCM predicts that consumers do not vary in their responses to marketing instruments. Unexpectedly, the influence of temperature is negative, on average, with a relatively large amount of variation. I suspect this result may be due to the fact that some of the winter holidays generate spikes in sales. Including a holiday dummy in the specification (to distinguish the purchase versus no purchase decision) might absorb this negative temperature effect. I also find variation in the degree to which consumers value diet products as well as the various package sizes. Adding these additional random coefficients reduces the mean response of prices as well as the degree of heterogeneity. I do not report the mean tastes for the fixed attributes, although I could obtain these using a minimum distance procedure as in Nevo(2000).

As with the proposed model, I have no statistical test with which to evaluate the fit of the aggregate DCM. By construction, the shares should fit the data very closely

Product	DCM 1		DCM 2	
	mean	st.dev.	mean	st.dev.
price	-13.60	2.39	-10.08	0.70
	(1.56)	(0.16)	(1.99)	(0.02)
ad	0.28	0.09	0.39	0.54
	(0.05)	(0.25)	(0.06)	(0.08)
display	0.43	0.13	0.56	0.29
	(0.04)	(0.25)	(0.05)	(0.17)
temperature	-0.02	0.02	-0.02	0.02
	(0.00)	(0.00)	(0.00)	(0.00)
diet				1.20
				(0.03)
6-pack				3.09
				(0.06)
12-pack				0.06
				(1.04)
16-oz bottle				1.77
				(1.04)

Table 4.26. Aggregate Random Coefficients Discrete Choice Model (standard errors in parentheses)

since the inversion procedure fits the predicted shares to the actual shares for the given parameter vector. I could re-estimate the model with the first eight quarters of store-weeks, leaving the ninth quarter as a hold-out sample with which to perform out-of-sample predictions.

Measuring Market Power and Mergers

Price Elasticities

If consumers are highly price-sensitive, then manufacturers are less able to raise their prices above costs without sacrificing market shares. The price elasticity is the standard measure of consumer price-sensitivity as it summarizes the price-response without taking into account the units with which prices and shares are measured. Since the accuracy of the predicted price responses of each product will be crucial for my study the price-setting behavior of firms, I compare the results of the proposed model to those of the aggregate DCM.

One of the main difficulties in computing elasticities with point-of-purchase data is the fact that households do not necessarily face the same mix of prices and marketing variables on any given store-trip. One way to recover a summary measure of overall elasticity across consumers and over time is to consider the effect of a uniform percentage change in the price of a good on aggregate demand (Ben-Akiva and Lerman 1985).

For a given product, the aggregate observed purchases are (for simplicity I eliminate the household subscript):

$$X_j = \sum_{t=1}^{T} X_{tj}.$$

Assume that everyone experiences the same percent price change:

$$\frac{\delta p_{tk}}{p_{tk}} = \frac{\delta p_{sk}}{p_{sk}} = \frac{\delta \overline{p_k}}{\overline{p_k}}, s, t = 1, ..., T, k = 1, ..., I$$

where $\overline{p}_k = \frac{1}{T}\sum_{t=1}^{T} p_{tk}$. I then compute the price elasticity of total demand in response to a change in the average price level, $\overline{p}$:

$$\begin{aligned} \varepsilon^j_{p_k} &= \frac{\delta X_j}{\delta p_k}\frac{\overline{p_k}}{X_j} \\ &= \sum_{t=1}^{T} \frac{\delta X_{tj}}{\delta p_k}\frac{\overline{p_k}}{X_j} \\ &= \sum_{t=1}^{T} \frac{\delta X_{tj}}{\delta p_{tk}}\frac{p_{tk}}{X_{tj}}\frac{X_{tj}}{X_j} \\ &= \sum_{t=1}^{T} \varepsilon^{tj}_{p_k}\frac{X_{tj}}{X_j} \end{aligned}$$

which is just the sum of the individual elasticities, weighted by the purchases on the given trip. The elasticities associated with the aggregate DCM are easier to compute directly since the price and marketing-mix are the same for a given store-week:

$$\begin{aligned} \varepsilon^j_{p_k} &= \frac{\delta X_j}{\delta p_k}\frac{\overline{p_k}}{X_j} \\ &= \frac{p_k}{\int_{A_j} S_{j|\beta}\Phi(d\beta)} \int_{A_j} \frac{\partial S_{j|\beta}}{\partial p_k}\Phi(d\beta). \end{aligned}$$

Table(4.7.1) presents estimated own price elasticities for the second specification of the proposed model (model 2) and the second specification of the aggregate DCM (DCM 2). For the DCM, I report the mean across all store-weeks. Roughly half of the aggregate DCM's own-price elasticities are larger than those of the proposed model. This result could be partially model-driven since the small measured shares imply that the relative magnitudes of the elasticities are, to a certain extent, driven by prices. One way I might alleviate this problem would be to include an interaction term between price and product size to account for the fact that consumers might value the various package sizes differently. On average, the DCM yields own-price elasticities that are 10% higher.

Table(4.28) shows that the cross-elasticities are systematically higher for the aggregate DCM, as expected. Rather than report the full (26 $\times$ 26) matrix of cross-elasticities for each model, I compute the median cross-elasticity for each product with respect to the prices of the alternatives. Since the DCM uses small shares, the cross-elasticities are very close for each product. I could alleviate this effect somewhat by including more random coefficients on the product attributes.

Wholesale Mark-Ups and Market Power

Having estimated demand and price-responses for both the proposed model and the aggregate DCM, I am now able to compute the wholesale markups. I measure relative market power using the Lerner index, the margin-to-price ratio: the proportion of the price not attributable to costs. Table(4.29) presents the predicted wholesale margins in dollars and as a percent of the underlying price. For both the proposed model and the aggregate DCM, I compute the margins at a monthly frequency, reporting the medians. Note that the prices in the proposed model are not identical to those of the DCM since

Product	Model 2	DCM
PEPSI 6P	-2.38	-2.57
COKE CLS 6P	-2.11	-2.85
PEPSI DT 6P	-2.47	-2.64
COKE DT 6P	-3.14	-2.87
DR PR 6P	-3.04	-2.61
MT DW 6P	-3.56	-2.67
PEPSI DT CF 6P	-3.61	-2.64
A and W CF 6P	-3.59	-2.68
PEPSI 16oz	-2.25	-4.01
PEPSI 12P	-2.16	-3.51
COKE CLS 12P	-2.13	-3.72
COKE DT 12P	-2.50	-3.76
PEPSI DT 12P	-2.66	-3.55
DR PR 12P	-2.47	-3.66
MT DW 12P	-3.02	-3.52
COKE DT CF 12P	-2.76	-3.75
SP CF 12P	-2.57	-3.76
PEPSI DT CF 12P	-2.92	-3.56
PEPSI 67.6oz	-2.62	-2.21
COKE CLS 67.6oz	-2.80	-2.23
PEPSI DT CL 67.6oz	-2.66	-2.24
7UP R CF 67.6oz	-2.57	-2.34
COKE DT 67.6oz	-2.81	-2.26
7UP DT CF 67.6oz	-2.61	-2.35
DR PR 67.6oz	-2.94	-2.28
MT DW 67.6oz	-3.23	-2.25

Table 4.27. Own Price Elasticities

Product	Model 3	DCM
PEPSI 6P	0.050	0.019
COKE CLS 6P	0.052	0.019
PEPSI DT 6P	0.053	0.019
COKE DT 6P	0.059	0.019
DR PR 6P	0.089	0.019
MT DW 6P	0.067	0.019
PEPSI DT CF 6P	0.076	0.020
A and W CF 6P	0.064	0.020
PEPSI 16oz	0.063	0.020
PEPSI 12P	0.048	0.019
COKE CLS 12P	0.048	0.020
COKE DT 12P	0.061	0.020
PEPSI DT 12P	0.067	0.019
DR PR 12P	0.065	0.020
MT DW 12P	0.073	0.020
COKE DT CF 12P	0.082	0.020
SP CF 12P	0.090	0.020
PEPSI DT CF 12P	0.084	0.020
PEPSI 67.6oz	0.052	0.020
COKE CLS 67.6oz	0.055	0.020
PEPSI DT CL 67.6oz	0.058	0.020
7UP R CF 67.6oz	0.052	0.020
COKE DT 67.6oz	0.057	0.020
7UP DT CF 67.6oz	0.051	0.020
DR PR 67.6oz	0.082	0.020
MT DW 67.6oz	0.059	0.020

Table 4.28. Median Cross-Elasticity of Each Product

the latter only focuses on 22 of the stores. Consequently, I should compare only the percent margins for the two models, not the actual dollar-values.

I find the relatively low margins for 12-packs, especially for the regular colas, particularly interesting. This finding for the highest-selling products is consistent with the common characterization of the industry as "low-margin" and "high-volume." Moreover, this find is further evidence of the long-standing "cola wars." For the proposed model, I find that all sizes of Dr. Pepper and 7UP products exhibit lower margins as a proportion of prices than the same size products carried by Coke and Pepsi, no doubt a reflection of their substitutability with colas. Given these asymmetric substitution patterns, I expect that the proposed mergers between either of the colas and Dr. Pepper or 7UP would result in a larger price increase for the non-colas than for the colas. The mergers would internalize part of the competition that has been keeping the non-colas' margins down. These differences do not turn up in the DCM. Moreover, with a few exceptions, the margins are much lower as a percent of price for the DCM than for the proposed model. The DCM clearly predicts much less market power for the CSD manufacturers. It does not capture the differences in market power for Dr. Pepper and 7UP products relative to Coke and Pepsi. Instead, the predicted margins of the DCM are much more uniform for given package sizes. This effect reflects the low cross-elasticities, which reduce the degree to which a firm benefits from joint-pricing. Evidently, the low cross-elasticities translate into lower predicted market power for the CSD manufacturers. For merger analysis, I expect these pronounced differences in market power for the two models to lead to qualitatively different merger predictions as well.

Mergers

Now that I have estimates of both demand parameters and marginal costs, I have recovered all of the features of the proposed industry model. Using the model, I am now able to study the effects of three hypothetical mergers on equilibrium prices. I report the predicted percent price changes for the proposed mergers between Coke and Dr. Pepper, and Pepsi and 7UP as well as the hypothetical merger between Coke and Pepsi. The values reported are medians of the monthly price changes for the 27 months in the sample using the approximation method. Numerical solutions are not yet available. Unlike the industry structure in 1986, Dr. Pepper and 7UP constitute a joint entity during the sample period (later to be acquired by Cadbury-Schweppes in 1995). Although not reported, I find the approximate effect on prices of breaking apart 7UP and Dr. Pepper to be small (less than 1 percent on average). Since the creation of the joint-entity did not appear to have much effect on industry prices, I assess the proposed 1986 mergers as if Coke or Pepsi were acquiring a division of the Dr.Pepper-7UP company. First I study the effects of mergers using the proposed model. Then, I repeat the analysis for the aggregate DCM.

Table(4.30) reports the predicted percentage price changes for the mergers using the proposed model. The second column predicts the outcome of the Coke-Dr. Pepper

Product	Model 3		DCM	
	MU ($)	MU to price (%)	MU ($)	MU to price (%)
PEPSI 12P	1.11	30.51	1.04	27.38
COKE CLS 12P	1.01	28.88	1.02	27.03
PEPSI 6P	0.58	35.68	0.51	26.81
COKE DT 12P	1.02	29.38	1.00	22.82
PEPSI 67.6oz	0.48	46.89	0.47	22.99
PEPSI DT 12P	1.15	31.53	1.01	28.08
COKE CLS 6P	0.50	30.59	0.50	27.66
PEPSI DT 6P	0.63	40.38	0.50	22.59
COKE CLS 67.6oz	0.44	41.82	0.46	22.54
PEPSI DT CL 67.6oz	0.50	48.74	0.47	30.66
COKE DT 6P	0.53	32.63	0.51	30.93
DR PR 12P	0.96	26.17	0.99	27.51
MT DW 12P	1.11	30.36	1.02	23.12
DR PR 6P	0.48	28.79	0.51	23.22
7UP R CF 67.6oz	0.44	42.81	0.47	30.61
COKE DT CF 12P	1.02	29.53	1.02	30.90
COKE DT 67.6oz	0.45	42.80	0.46	22.56
7UP DT CF 67.6oz	0.45	43.43	0.47	25.12
MT DW 6P	0.58	35.43	0.52	28.71
SP CF 12P	1.01	28.92	1.02	28.48
PEPSI DT CF 12P	1.15	31.65	1.01	25.30
DR PR 67.6oz	0.44	41.42	0.48	25.42
MT DW 67.6oz	0.48	46.87	0.48	25.69
PEPSI 16oz	0.73	27.22	0.76	28.69
PEPSI DT CF 6P	0.63	41.22	0.53	26.63
A and W CF 6P	0.48	28.50	0.51	28.56

Table 4.29. Predicted Mark-UPs

merger. In general, this merger does not seem to have an objectionable effect on prices. Coke prices never rise by more than about 3% and only the prices of 12-packs of Dr. Pepper increase by 8%. Clearly the competition maintained between Coke and Pepsi is still sufficient to keep both Coke and Dr. Pepper prices reasonably low. Next, I look at the merger between Pepsi and 7UP. Now I find an even smaller change in cola prices. But, the price of 7UP rises by almost 16%, for regular, and 19% for diet. Internalizing the competition between Pepsi and 7UP and, moreover, 7UP and Mountain Dew, reduces the competition faced by 7UP sufficiently that prices may be increased profitably. This merger would clearly violate the 10% rule used during the trial against Coke and Dr. Pepper. Moreover, this result may explain why Pepsi did not persist in its proposed merger with 7UP, whereas Coke persisted in its merger with Dr. Pepper. In the final column, I consider the extreme case of a merger between Coke and Pepsi. Now I find substantial price increases, as expected. With the exception of Sprite, all of Pepsi and Coke's products' prices increase by more than 20%. Unfortunately, the approximation method does not capture the competitive response of Dr. Pepper and 7UP. Intuitively, I would expect these firms to begin raising their prices due to the reduced competition in the industry. Moreover, I would expect some of the Pepsi and Coke price increases to be offset by consumers switching to cheaper alternatives. Nonetheless, the approximation suggests that this final merger clearly would violate the criteria for an acceptable merger. Note that this analysis assumes that none of the products are discontinued after merger.

Product	Merger 1 Coke/Dr. Pepper	Merger 2 Pepsi/7UP	Merger 3 Coke/Pepsi
PEPSI 12P	0	2.41	18.51
COKE CLS 12P	2.50	0	31.46
PEPSI 6P	0	1.11	32.73
COKE DT 12P	3.63	0	59.29
PEPSI 67.6oz	0	2.50	45.19
PEPSI DT 12P	0	1.98	44.45
COKE CLS 6P	2.89	0	49.72
PEPSI DT 6P	0	1.72	31.45
COKE CLS 67.6oz	3.15	0.00	46.31
PEPSI DT CL 67.6oz	0	2.93	27.88
COKE DT 6P	0.64	0	68.85
DR PR 12P	8.01	-1.07	0
MT DW 12P	0	3.26	27.79
DR PR 6P	4.68	-0.68	0
7UP R CF 67.6oz	-4.11	14.95	0
COKE DT CF 12P	4.69	0	32.24
COKE DT 67.6oz	2.14	0	36.17
7UP DT CF 67.6oz	-1.74	19.22	0
MT DW 6P	0	0.93	46.80
SP CF 12P	0	0	9.80
PEPSI DT CF 12P	0	1.10	36.53
DR PR 67.6oz	3.87	-1.78	0
MT DW 67.6oz	0	2.19	28.87
PEPSI 16oz	0	3.23	36.26
PEPSI DT CF 6P	0	2.34	35.65
A and W CF 6P	-1.73	-0.06	0

Table 4.30. Approximate Percent Price Change from Mergers (Proposed Model)

I now repeat the merger exercise using the aggregate DCM, reporting my findings in Table(4.31). As expected, the DCM predicts much lower increases for the hypothetical mergers. As with the proposed model, the Coke and Dr. Pepper merger leads to very small changes. However, the DCM also finds very small changes for the Pepsi and 7UP merger. In fact, the DCM predicts that 7-UP price changes do not even come close to the 10% limit. Even the extreme merger between Coke and Pepsi leads to much lower price changes than the proposed model. Only Coke's 12-packs change by more than 8%, still below the permissible limit. This final prediction has little credibility given that Coke specifically claimed that this merger would dampen industry competition substantially. In general, a merger should internalize some of the competition, allowing firms to raise prices. The low cross-elasticities of the DCM imply much lower substitutability between products and, thus, less competition. Thus, mergers do not have a very large impact on prices, according to the DCM specification.

The approximation of equilibrium prices does not take into account the fact that consumers may respond to price changes by shifting their purchases. Thus, I am unable to measure welfare effects of mergers. Changes in consumer welfare must reflect both the lost utility from price increases as well as lost utility from lower consumption. Similarly, the changes in producer profits must reflect both the changes in prices as well as the changes in outputs. By construction, the approximation fixes the non-merging firms' profits at the pre-merger levels. For now, I use the approximations as preliminary evidence that the higher market power predicted by the proposed model, relative

Product	Merger 1	Merger 2	Merger 3
PEPSI 12P	0	0.40	4.50
COKE CLS 12P	0.33	0.00	8.00
PEPSI 6P	0	0.47	5.15
COKE DT 12P	0.87	0	3.05
PEPSI 67.6oz	0	0.34	2.70
PEPSI DT 12P	0	0.50	4.48
COKE CLS 6P	0.18	0	2.63
PEPSI DT 6P	0	0.33	4.48
COKE CLS 67.6oz	0.46	0	4.53
PEPSI DT CL 67.6oz	0	0.33	2.59
COKE DT 6P	0.17	0	2.55
DR PR 12P	0.72	-1.00	0
MT DW 12P	0	0.21	3.94
DR PR 6P	1.14	-0.34	0
7UP R CF 67.6oz	-1.08	2.09	0
COKE DT CF 12P	0.45	0	8.59
COKE DT 67.6oz	0.45	0	4.46
7UP DT CF 67.6oz	-1.08	2.15	0
MT DW 6P	0	0.71	6.99
SP CF 12P	0.56	0.00	8.66
PEPSI DT CF 12P	0.00	0.45	4.44
DR PR 67.6oz	1.64	-0.47	0.00
MT DW 67.6oz	0	0.34	2.70
PEPSI 16oz	0	0.32	2.19
PEPSI DT CF 6P	0	0.32	4.48
A and W CF 6P	-0.39	-0.09	0

Table 4.31. Approximate Percent Change in Prices due to Mergers (Aggregate DCM)

to the aggregate DCM, leads to higher predicted price changes due to mergers. Focusing on the results of the proposed model, I find the blocked mergers from 1986 would not have a tremendous impact on industry prices, while the merger between Coke and Pepsi would have a substantial impact on prices.

Conclusions

More than ten years have passed since the proposed mergers between Coke and Dr. Pepper, and Pepsi and 7UP were successfully opposed by the FTC. During the case against Coke, several sophisticated economic arguments were put forth by both sides. In particular, Coke argued the merger would not increase its ability to raise prices due to the existing differentiation of products, whereas the FTC argued that the merger would lessen competition substantially, leading to more collusive prices. Coke also claimed that only the extreme merger between Coke and Pepsi would have a noticeable effect on industry competition. Finally, Coke argued that the merger would generate joint production efficiencies, especially for Dr. Pepper, which would lead to lower prices. However, a lack of adequate econometric tools prevented an empirical validation of the economic arguments at the time.

I propose and estimate a model of CSD demand that attempts to capture some of the more sophisticated aspects of the industry. In particular, I address multiple-item shopping, heterogeneity and product differentiation. I apply the model to a detailed individual panel of household purchases for a specific market. I compare the results of this model to those of the popular aggregate DCM with random coefficients. The household data reveal that this aggregate model will be misspecified since households often violate the underlying single-unit purchase assumption. While I am unable to predict the specific effect of this specification bias in theory, I do find substantial qualitative and quantitative differences between the predictions of the two models. In particular, I find the proposed model predicts much higher levels of substitutability between products, which imply higher manufacturer market power due to joint-pricing of products in the product line. This difference in market power leads to important differences in the merger analysis.

I use both estimated models in addressing the economic arguments put forth during the merger cases. The results assume that costs do not change in response to the merger. The evidence appears to support Coke's claim that the merger with Dr. Pepper would not have a sizeable effect on industry prices. In fact, both models predicts only modest price increases in this case. However, I do find large increases in the prices of 7UP when it is merged with Pepsi. In contrast, the aggregate DCM predicts very modest price increases for this scenario as well. Using the proposed model, I find that the hypothetical merger between Coke and Pepsi would have a substantial effect on industry competition. I find the merger leads to extremely large price increases for Coke and Pepsi's entire product lines. In contrast, the aggregate DCM predicts much lower price changes for this merger. In fact, the aggregate DCM would likely result in accepting the merger between Coke and Pepsi. Given that Coke claimed that only its merger with Pepsi could dampen industry competition objectionably, the predictions of the DCM must be inaccurate. Therefore, the specification error associated from assuming single-unit purchasing in the DCM potentially leads to incorrect policy conclusions.

For now, the results are still approximate since I do not compute the true post-merger equilibrium prices. Moreover, my analysis assumes a static price-setting environment that assumes away the possibility of entry from new competitors. Despite these simplifications, the evidence favors the position of Coke during the antitrust case in 1986. Moreover, the suspiciously modest results of the aggregate DCM highlight the potential hazards of misspecifying the discrete choice model for products that exhibit multiple-unit purchasing.

References

Ackerberg, D.[1999], "Empirically Distinguishing Informative and Prestige Effects of Advertising." Boston Univeristy, Working Paper.

Agrawal, J., P.E. Grimm and N. Srinivasan [1993], "Quantity Surcharges on Groceries." *Journal-of-Consumer-Affairs*, 27, 335-56.

Ainslie, A., P.B. Seetharaman and P.K. Chintagunta [1998], "Investigating Household State Dependence Effects Across Categories." Cornell University, Working Paper.

Allenby, G.M. and P.E. Rossi [1991], "There is No Aggregation Bias: Why Macro Logit Models Work." *Journal of Business and Economic Statistics*, 9, 1-14.

Allenby, G.M. and P.E. Rossi [1999], "Marketing Models of Consumer Heterogeneity." *Journal of Econometrics*, 89, 57-78.

Andrews, D.W.K. [1991], "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation." *Econometrica*, 59, 817-858.

Armstrong, M.K. [1991], *Retail Response to Trade Promotions: An Incremental Analysis of Forward Buying and Retail Promotion*. Doctoral dissertation, University of Texas at Dallas.

Balachander, S. and P.H. Farquahar [1994], "Gaining More by Stocking Less: A Competitive Analysis of Product Availability." *Marketing Science*, 13, 1, 3-22.

Ben-Akiva, M. and S.R. Lerman [1985], *Discrete Choice Analysis: theory and applications to travel demand*. The MIT Press: Cambridge.

Berry, S. [1994],"Estimating Discrete-Choice Models of Product Differentiation." *Rand Journal of Economics*, 25, 242-62.

Berry, S., M. Carnall and P.T. Spiller [1997], "Airline Hubs: Costs, Markups and the Implications of Customer Heterogeneity." Yale University, Working Paper.

Berry, S., J. Levinsohn, and A. Pakes [1995],"Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841-890.

Berry, S., J. Levinsohn, and A. Pakes [1998], "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market," Yale University, Working Paper.

Besanko, D., S. Gupta and D. Jain [1998], "Logit Demand Estimation Under Competitive Pricing Behavior: An Equilibrium Framework," *Management Science*, November.

Blattberg, R.C. and S.A. Neslin [1990], *Sales Promotion: Concepts, Methods, and Strategies*. Englewood Cliffs, New Jersey: Prentice Hall.

Bresnahan, T. [1990], "Empirical Methods for Industries with Market Power," in *Handbook of Industrial Organization*, ed. R. Schmalensee and R. Willig, North Holland.

Bresnahan, T., S.Stern, and M.Trajtenberg (1997), "Market Segmentation and the Sources of Rents from Innovation: Personal Computers in the Late 1980's." *RAND Journal of Economics*, 28, S28-S44.

Brownstone, D. and K. Train [1998], "Forecasting New Product Penetration with Flexible Substitution Patterns." *Journal of Econometrics*, 89, 109-129.

Bucklin, R.E. and S. Gupta [1999], "Commercial Use of UPC Scanner Data: Industry and Academic Perspectives." *Marketing Science*, 18, 3, 247-273.

Cameron, A.C. and P.K. Trivedi [1986], "Econometric Model Based on Count Data: Comparisons and Applications of Some Estimators and Tests." *Journal of Applied Econometrics*, 1, 29-53.

Caplin, A. and B.Nalebuff [1991], "Aggregation and Imperfect Competition: On the Existence of Equilibrium." *Econometrica*, 59, 1, 25-59.

Cardell, N. Scott [1997],"Variance Components Structures for the Extreme-Value and Logistic Distributions with Application to Models of Heterogeneity" *Econometric Theory*;13(2), 185-213.

Chamberlain, G. [1982], "Multivariate Regression Models for Panel Data." *Journal of Econometrics*, 18, 5-46.

Chevalier, Michel, and Ronald S. Curhan [1976], "Retail Promotions as a Function of Trade Promotions: A Descriptive Analysis." *Sloan Management Review*, 18, 3, 19-32.

Chiang J. [1991], "The Simultaneous Approach to the Whether, What, and How much to Buy Questions" *Marketing Science*, 10, 297-315.

Chintagunta, P. [1993a], "Investigating Purchase Incidence, Brand Choice and Purchase Quantity Decisions of Households," *Marketing Science*, 12, 184-208.

Chintagunta, P. [1993b], "An Empirical Investigation of Purchase Quantity Decisions of Households," University of Chicago, *Working Paper*.

Chintagunta, P. [1999a], "An Empirical Investigation of Retailer Category Profit Maximizing Behavior," University of Chicago, Working Paper.

Chintagunta, P. [1999b], "A Heterogeneous Aggregate Logit Demand Model." University of Chicago, Working Paper.

Chintagunta, P. D. Jain and N.J. Vilcassim [1991], "Investigating Heterogeneity in Brand Preferences in Logit Models for Panel Data." *Journal of Marketing Research*, 28, 417-28.

Chintagunta, P., V. Kadiyali, and N.J. Vilcassim [1999], "Endogeneity and Simultaneity in Competitive Pricing and Advertising: A Logit Demand Analysis." University of Chicago, Working Paper.

Chintagunta, P., E. Kyriazidou and J. Perktold [1997], "Panel Data Anlaysis of Household Brand Choices." University of Chicago, Working Paper.

Cohen, A.M. [2000], "Package Size and Price Discrimination: A Structural Investigation." Northwestern University, *Working Paper.*

Cotterill, R. W., A. W. Franklin, and L. Y Ma. 1996 "Measuring Market Power Effects in Differentiated Product Industries: An Application to the Soft Drink Industry." Food Marketing Policy Center, University of Connecticut, Storrs, CT.

Davis, P. [1997], "Spatial Competition in Retail Markets." Yale University, Working Paper.

Dubé, J.P. [2000], "Product Differentiation and Mergers in the Carbonated Soft Drink Industry." Northwestern University, *Working Paper.*

Elrod, T. [1988], "Choice Map: inferring a product-market map from panel data." *Marketing Science*, 7, 12-40.

Elrod, T. and M. Keane [1995], "A Factor-Analytic Probit Model for Representing the Market Structure in Panel Data." *Journal of Marketing Research*, 32, 1-16.

Erdem, T. and R.S. Winer [1998], "Econometric Modeling of Competition: A multi-category choice-based mapping approach." *Journal of Econometrics*, 89, 159-175.

Fader, P.S. and B.G.S. Hardie [1996], "Modeling Consumer Choice Among SKUs." *Journal of Marketing Research*, 33, 442-52.

Feenstra, R.C. and J.A. Levisohn [1995], "Estimating Markups and Market Conduct with Multidimensional Product Attributes." *Review of Economic Studies*,62, 19-52.

Gasmi,F., J. J. Laffont and Q. Vuong [1992], "Econometric Analysis of Collusive Behavior in a Soft-Drink Market." *Journal of Economics and Management Strategy*, 1, 277-311.

Goldberg, P. (1995), "Product Differentiation and Oligopoly in International Markets: The case of the US automobile industry." *Econometrica*, 63, 891-951.

Gonul, F.F. [1999], "Estimating Price Expectations in the OTC Medicine Market: An Application of Dynamic Stochastic Discrete Choice Models to Scanner Panel Data." *Journal of Econometrics*, 89, 41-56.

Gonul, F. and K. Srinivasan [1993], "Modeling Multiple Sources of Heterogeneity in Multinomial Logit Models: Methodological and Managerial Issues." *Marketing Science*, 12, 3, 213-229.

Guadagni P.M. and J.D.C. Little [1983], "A Logit Model of Brand Choice Calibrated on Scanner Data." *Marketing Science*, 2, 203-38.

Gupta, S., P.K.Chintagunta, A. Kaul, and D.R.Wittink [1996], "Do Household Scanner Data Provide Representative Inferences from Brand Choices: A Comparison with Store Data." *Journal of Marketing Research*, 33, 4, 383-398.

Hanemann W. M. [1984], "Discrete/Continuous Models of Consumer Demand." *Econometrica* 52:541-561.

Hansen, L.P. [1982], "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica*, 50, 1092-1054.

Harlam, B.A. and L.M. Lodish [1995], "Modeling Consumers' Choices of Multiple Items." *Journal of Marketing Research*, November, 404-18.

Hausman, J.A. [1994], "Valuation of New Goods Under Perfect and Imperfect Competition." NBER Working Paper #4970.

Hausman, J. A., G.K. Leonard and D. McFadden [1995], "A Utility-Consistent, Combined Discrete Choice and Count Data Model: Assessing Recreational Use Losses due to Natural Resource Damage." *Journal of Public Economics*,56, 1-30.

Hausman, J.A., G.K. Leonard and J.D. Zona [1994], "Competitive Analysis with Differenciated Products." *Annales d'Economie et de Statistique*, 0, 159-80.

Hausman, J.A. and D.A.Wise [1978], "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences." *Econometrica*, 46, 403-426.

Hauser, J.R. and B. Wernerfelt [1990], "An Evaluation Cost Model of Consideration Sets." *Journal of Consumer Research*, 16, 393-408.

Hellerstein, D. and R. Mendelsohn [1993], "A Theoretical Foundation for Count Data Models." *American Journal of Agricultural Economics*, 75, 604-11.

Hendel, I. [1999], "Estimating Multiple-Discrete Choice Models: An Application to Computerization Returns," *Review of Economic Studies*, 66, 423-446.

Higgins, R. S., D. P. Kaplan, M. J. McDonald and R. D. Tollison [1995], "Residual Demand Analysis of the Carbonated Soft Drink Industry." *Empirica*, 22, 115-26.

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

Horsky, D. and P. Nelson [1992], "New Brand Positioning and Pricing in an Oligopolistic Market." *Marketing Science*, 11, 133-153.

Hsiao, C. [1986], *Analysis of Panel Data*. New York: Cambridge University Press.

Kalyanam, K. and D.S. Putler [1997], "Incorporating Demographic Variables in Brand Choice Models: An Indivisible Alternatives Framework." *Marketing Science*, 16, 166-81.

Kalyanaram, G. and R.S. Winer [1993], "Empirical Generalizations from Reference Price Research." *Marketing Science*, 14, 161-69.

Kamakura W.A. and G.J. Russell [1989], "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure." *Journal of Marketing Research*, 26, 379-90.

Karunakaran, Sudhir [2000], "Structural Analysis of Manufacturer Pricing in the Presence of a Strategic Retailer." New York University, Working Paper.

Kim, J., G.M. Allenby and P.E. Rossi [1999], "Modeling Consumer Demand for Variety." University of Chicago, *Working Paper*.

Krishnamurthi L. and S.P. Raj [1988], "A Model of Brand Choice and Purchase Quantity Price Sensitivities." *Marketing Science*, 7, 1-20.

Kwoka, J.E.Jr. and L.J. White [1999], "The Economic and Legal Context." in *The Antitrust Revolution*, ed. J.E. Kowka Jr. and L.J. White, New York: Oxford University Press.

Lal, R. and C. Narasimhan [1996], "The Inverse Relationship Between Manufacturer and Retail Margins: A Theory." *Marketing Science*, 15, 132-151.

Langan, G. E. and R. W. Cotterill. 1994. "Estimating Brand Level Demand Elasticities and Measuring Market Power for Regular Carbonated Soft Drinks." University of Connecticut, Storrs, CT.

Lee, L. [1992], "On Efficiency of Method of Simulated Moments and Maximum Simulated Likelihood Estimation of Discrete Response Models." *Econometric Theory*, 8, 518-552.

Levedahl, W. J. [1986], "Profit-maximizing Pricing of Cents Off Coupons: Promotion or Price Discrimination." *Quarterly Journal of Business and Economics*, Fall, 56-70.

Manuszak, M. [1999], "Firm Conduct in the Hawaiian Retail Gasoline Indutry." Northwestern University, *Working Paper*.

McAlister, L. [1982], "A Dynamic Attribute Satiation Model of Variety-Seeking Behavior." *Journal of Consumer Research*, 9, 141-150.

McCulloch R. and P.E. Rossi [1994], "An Exact Likelihood Analysis of the Multinomial Probit Model." *Journal of Econometrics*, 64, 207-240.

McCulloch R. and P.E. Rossi [1999], "Bayesian Analysis of the Multinomial Probit Model," in *Simulation-Based Inference in Econometrics*, ed. Mariano, Weeks and Schuermann, Cambridge University Press, Cambridge.

McFadden, D. [1981], "Econometric Models of Probabilistic Choice," in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by C. Manski, and D. McFadden. MIT Press, Cambridge, MA.

McFadden, D. [1989]: "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration." *Econometrica, 57, 995-1026.*

McFadden, D. and K. Train [1998], "Mixed MNL Models for Discrete Response." U.C. Berkeley, Working Paper.

Muris, T.J., D.T. Scheffman and P.T. Spiller [1992], "Strategy and Transaction Costs: The Organization of Distribution in the Carbonated Soft Drink Industry." *Journal of Economics and Management Strategy*, 1, 83-128.

Muris, T.J., D.T. Scheffman and P.T. Spiller [1992], "Strategy and Transaction Costs: The Organization of Distribution in the Carbonated Soft Drink Industry" in *Case Studies in contracting and organization, ed. Scott Masten, Oxford University Press.*

Nevo, A. [1999], "Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry." U.C. Berkeley, *Working Paper*.

Nevo, A. [2000], "Measuring Market Power in the Ready-To-Eat Cereal Industry," U.C. Berkeley, forthcoming, *Econometrica*.

Newey, W.K. and K.D. West [1987], "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica*, 55, 703-708.

Pakes, A. and P. McGuire [1994], "Computing Markov-Perfect Nash Equilibria: numerical implications of a dynamic differentiated product model." *Rand Journal of Economics*, 25, 555-589.

Pakes, A. and D. Pollard [1989], "Simulation and the Asymptotics of Optimization Estimators." *Econometrica* , 57, 1027-57.

Petrin, A. [1999], "Quantifying the Benefits of New Products: The Case of the Minivan." University of Chicago, *Working Paper.*

Quelch, J.A. and D. Kenny [1994], "Extended Profits, Not Product Lines." *Harvard Business Review*, September-October, 153-160.

Rossi, P.E., and G. Allenby [1993], "A Bayesian Approach to Estimating Household Parameters." *Journal of Marketing Research*, 30, 171-182.

Rossi, P. E., R.E. McCulloch and G.M. Allenby [1996], "The Value of Purchase History Data in Target Marketing." *Marketing Science*, 15, 321-40.

Schmalansee, R. [1978], "Entry deterrence in the ready-to-eat breakfast cereal industry." *The Bell Journal of Economics* 9, 305-327.

Seetharaman, P.B. and P.K. Chintagunta [1997], "Do Households Exhibit Similar Brand Choice Dynamics Across Product Categories?" University of Chicago, *Working Paper.*

Shaffer, G. and F. Zettelmeyer [1999], "Bargaining, Third-Party Information, and the Division of Profits in the Distribution Channel." University of Rochester, Working Paper.

Simonson, I. [1990], "The Effect of Purchase Quantity and Timing on Variety-Seeking Behavior." *Journal of Marketing Research*, 27, 150-162.

Slade, M.E. [1995], "Product Market Rivalry with Multiple Strategic Weapons: an analysis of price and advertising competition," *Journal of Economics and Management Strategy*, 4, 445-476.

Sutton, J. [1992], *Sunk Costs and Market Structure*. Cambridge: The MIT Press.

Trivedi, M., F.M. Bass and R.C. Rao [1994], "A Model of Stochastic Variety-Seeking." *Marketing Science*, 13, 274-297.

Villas-Boas, M. and R. Winer [1999], "Endogeneity in Brand Choice Models." *Management Science*, 45, 10, 1324-1338.

Vilcassim, N.J., and D.R. Wittink [1987], "Supporting a Higher Shelf Price Through Coupon Distributions." *Journal of Consumer Marketing*, 4, 29-39.

Walsh, J.W. [1995], "Flexibility in Consumer Purchasing for Uncertain Future Tastes." *Marketing Science*, 14, 148-165.

Walters, R. J. [1989], "An Empirical Investigation into Retailer Response to Manufacturer Trade Promotions." *Journal of Retailing*, 65, 2, 253-72.

Wedel, M. and W.Kamakura (1998), *Market Segmentation: Conceptual and Methodological Foundations*, Kluwer Academic Publishers, Boston, MA.

White, L.J [1998], "The Proposed Merger of Coca-Cola and Dr Pepper." In *The antitrust revolution : economics, competition, and policy*, ed. Kwoka, J.E. and L.J. White (New York : Oxford University Press, 1998), 76-95.

Winer, R.S. [1985], "A Price Vector Model of Demand for Consumer Durables: Preliminary Developments." *Marketing Science*, 4, 74-90.

APPENDIX A.
Identifying the Biases Due to Heterogeneity and Endogeneity Using Simulated Data

In this Appendix, we attempt to isolate and identify the biases due to heterogeneity and endogeneity using simulated data. Our objective is two fold. First, we want to assess the gravity of the bias in parameter estimates that results when we ignore one or both of these issues in a data set that is a plausible representation of those used in demand studies in the marketing literature. In the simulated data set, prices and market shares are generated by a Nash equilibrium among three price-setting firms that are competing for customers who are in one of two market segments. Thus, the "truth" corresponds to a two-segment version of the empirical model described in the paper. Second, we want to show that we can use the econometric model described above to identify the segment heterogeneity from aggregate data, even when prices are endogenous. While the models are identified in theory, the simulation study helps confirm that a simple data set may exhibit sufficient variation to identify the segment structure in practice. In our analysis below, we created three simulated data sets: a "Baseline" case, a "High Endogeneity" case in which the dependence of prices on the unobserved attribute is strengthened, and a "High Heterogeneity" case in which the two segments are more sharply delineated than they are in the baseline case. The baseline case is discussed at length first, followed by the other two cases. Table(4.32) describes the structure of the market in the baseline case. The market consists of two segments, A and B. In each segment, consumers are identical and make logistic choices across three brands,

Parameter	*Segment A*	*Segment B*
Brand constant 1	4	2.5
Brand constant 2	1	2.5
Brand constant 3	-1	1
Price-response coefficient	1	2
Display-response coefficient	4	2
Segment proportions	0.3	0.7

Table 4.32. Parameters for Simulated Market: Demand

each produced by a separate firm, and a no-purchase alternative. The impact of all fixed product attributes and brand reputations are captured by brand-specific constants, which differ across segments. There are two marketing variables, price and display, with price assumed to be endogenous and display assumed to be exogenous.[49] We assume the prices of the three brands derive from the static Bertrand-Nash equilibrium in which each of the firms simultaneously sets its profit-maximizing price by equating prices with marginal costs plus a mark-up term that depends on the inverse own-price elasticity.

In the simulated data set, brands 1 and 2 can be thought of as national brands, while brand 3 is a "value" or store brand. Consumers in both segments generally prefer the national brands to the value brand, as indicated by the generally higher brand-specific constants for the two national brands. We can think of consumers in segment A as consisting of individuals who have a high opportunity cost of time, while consumers in segment B are less time-starved. Consistent with this interpretation, Segment A con-

[49]One interpretation of this set-up is that prices result from national-level wholesale price competition among manufacturers (with wholesale prices fully along by retailers). These decisions are sensitive to week-to- week variations in demand that are driven by national-level advertising or couponing activity and are thus endogenous. By contrast, one can think of display decisions are being retail-level decisions that are sensitive to unmodeled changes in local market conditions. In our empirical work using actual data, we employ a more fully fleshed out model of manufacturer and retailer interactions.

sumers are generally less sensitive to price and more sensitive to display than Segment B consumers.

In addition to price, display, and the fixed product attributes that are embodied by the brand-specific constants, we assume that there is a product attribute ξ_{jt} that varies from week to week. This attribute is exogenous to the model and is unobservable to the econometrician but is observable to the firms. As discussed above, we can think of this attribute as reflecting the impact of national advertising on product demand. In the baseline data set, the unobserved attribute ξ_{kt} is assumed to be drawn randomly each period from a normal distribution with a mean of 0 and a variance of 1 for each brand.[50] We assume that the display variable takes on a value of 1 in each period with an average probability of 0.20. By contrast, we assume that the prices p_{jt} are determined as a Nash equilibrium in a price setting-game between the three producers. The marginal costs of each producer are assumed to be determined by period-to-period exogenous variations in input prices of two factors.

Table(4.33) summarizes the mean values of our simulated data.

[50] We specified covariances of 0.1 between the unobserved attributes of brands, and covariances of 0.01 with the errors in the prices, which are also used to generate the price data.

S_1(Market Share)	0.13
S_2	0.09
S_3	0.18
P_1(Prices in $)	3.95
P_2	2.78
P_3	1.71
d_1 (Display: proportion of weeks)	0.12
d_2	0.13
d_3	0.36

Table 4.33. Summary of Simulated Data (Means)

	Exogenous Prices	Endogenous Prices
Homogeneous Demand	SUR	3SLS
Heterogeneous Demand	GMM (no instruments)	GMM (with instruments)

Table 4.34. Table Caption

Econometric Approach and Results

To identify the biases due to a failure to account for heterogeneity and endogeneity, we conduct four sets of econometric analyses, as described in Table 4.34. These analyses are with and without consumer heterogeneity, and with and without price endogeneity. We begin by estimating the logit demand by assuming no consumer heterogeneity and ignoring the endogeneity problem: we use the seemingly-unrelated regression (SUR). Next, continuing to assume a market of identical consumers, we use 3SLS to account for the endogeneity problem (as in Besanko, Gupta, and Jain 1998). Finally, to account for both endogeneity and heterogeneity, we employ the full aggregate latent class technique described above. In Table 4.35 we show the parameter estimates for each of the four models . As expected, we can see that failure to instrument generates a strong downward bias in the SUR estimate of price response. One might expect that 3SLS should approximately recover the mean level of the two segments' parameters. However, we can see that the 3SLS price parameter is substantially smaller in magnitude than the mean of the segment parameters. GMM without instrumenting for price endogeneity also provides very poor estimates of the segment parameters. Finally, the estimated segment parameters seem to be fairly well identified when we account for both consumer heterogeneity and price endogeneity. The instruments used are display and

	True Values		SUR	3SLS	GMM Estimates no instruments		GMM Estimates (std. errors)	
Parameter	A	B			A	B	A	B
Brand 1	4	2.5	-0.93(0.08)	-0.08(0.11)	-0.00	-0.45	4.62(0.42)	3.66(0.92)
Brand 2	1	2.5	-2.00(0.06)	-1.42(0.08)	1.45	-2.95	1.24(0.51)	2.98(0.69)
Brand 3	-1	1	-1.80(0.03)	-1.54(0.04)	0.30	-2.80	-0.93(0.48)	1.22(0.73)
Price	1	2	0.21(0.02)	0.44(0.029)	1.33	0.19	1.11(0.16)	2.16(0.14)
Display	4	2	1.63(0.04)	1.98(0.056)	2.26	2.26	4.50(0.52)	2.02(1.03)
segment size	0.3	0.7	-	-	0.491	0.509	0.28	0.72

Table 4.35. Parameter Estimates for Simulated Market Demand in Baseline Case

factor prices. The true parameters for both segments are recovered quite well with the exception of the estimated constant for brand 1 for segment B. A possible explanation for this poor estimate is that the conditional market share of brand 1 in segment B is less than 1 percent in the generated data.

In Table 4.36, we report the estimated price elasticities from the 4 estimation procedures. We find that almost all of the estimated own and cross price elasticities from the 3SLS model are less than half the values of the actual elasticities. This result suggests that the homogeneous model will highly understate market power in terms of a firm's ability to profitably raise prices above its marginal costs. Moreover, the model underpredicts the substitutability of goods. In fact, only the correct heterogeneous model yields all own-elasticities greater than one, which is consistent with the oligopoly price-setting behavior used to generate the prices.

The estimation experiments reported suggest that the consequences of ignoring either the endogeneity of prices or consumer heterogeneity are likely to be severe. By contrast, the GMM methodology does an excellent job recovering segment sizes and segment-

Truth			
	Brands		
	1	2	3
1	-2.43	0.64	0.34
2	0.37	-3.98	0.30
3	0.37	0.48	-2.44

	Brands				*Brands*		
SUR	**1**	**2**	**3**	**3SLS**	**1**	**2**	**3**
1	-0.75 (65.3%)	0.04 (85.9%)	0.05 (81.8%)	**1**	-1.51 (29.6%)	0.09 (68.8%)	0.12 (58.3%)
2	0.11 (80.4%)	-0.55 (82.9%)	0.05 (85.0%)	**2**	0.22 (61.0%)	-1.10 (65.8%)	0.09 (70.6%)
3	0.10 (74.6%)	0.03 (85.8%)	-0.30 (85.4%)	**3**	0.19 (50.3%)	0.06 (72.7%)	-0.58 (71.1%)
GMM (No Instr.)	**1**	**2**	**3**	**GMM**	1	2	3
1	-0.64 (73.7%)	0.06 (90.6%)	0.06 (82.4%)	**1**	-2.75 (13.2%)	0.66 (3.1%)	0.41 (20.6%)
2	0.03 (91.9%)	-2.57 (35.4%)	0.31 (3.3%)	**2**	0.39 (5.4%)	-4.47 (12.3%)	0.34 (13.3%)
3	0.05 (86.5%)	0.44 (8.3%)	-1.34 (45.1%)	**3**	0.44 (18.9%)	0.52 (9.2%)	-2.64 (8.2%)

Table 4.36. Own-Price and Cross-Price Elasticities of Market Demand (Percentage deviation of estimates from true elasticities are in parentheses.)

	Baseline	*High Endog.*	*High Heterog.*
S_1(Brand share)	0.13	0.13	0.14
S_2	0.09	0.10	0.08
S_3	0.18	0.19	0.17
P_1(Prices in $)	3.95	3.97	6.57
P_2	2.78	2.80	3.56
P_3	1.71	1.73	1.94
d_1(Display: proportion of weeks)	0.12	0.12	0.12
d_2	0.13	0.13	0.13
d_3	0.36	0.36	0.36

Table 4.37. Summary of Simulated Data

level parameters. This illustrates that in a plausible data set, segment-level micro structure of a market can be recovered solely from aggregate data.

High Endogeneity and High Heterogeneity Cases

To examine the impact of stronger endogeneity and stronger heterogeneity on the econometric estimates, we created two data sets in addition to the baseline case described above. We created a "high endogeneity" data set by increasing the variance of the unobserved attributes from 1 to 2, for each brand. All other demand and cost parameters are maintained at the baseline level. We created a "high heterogeneity" data set by reducing the price response parameter of segment A from 1.0 to 0.5. All other demand and cost parameters are maintained at the baseline level.

Table 4.37 summarizes the mean values of our simulated data for the three cases. Note that the effect of reducing the price response coefficient in segment 1 is to increase the mean equilibrium prices of all brands, especially the national brands.

	True Values		*GMM Estimates (std. errors)*	
Parameter	*A*	*B*	*A*	*B*
Brand constant 1	4	2.5	3.70 (0.81)	4.11 (0.79)
Brand constant 2	1	2.5	0.78 (0.49)	3.78 (1.52)
Brand constant 3	-1	1	-0.81 (0.32)	1.65 (2.91)
Price	1	2	0.89 (0.23)	2.49 (1.82)
Display	4	2	3.73 (0.66)	2.29 (0.91)
Proportion of consumers	0.3	0.7	0.30	0.70

Table 4.38. Parameter Estimates for Simulated Market Demand in High Endogeneity Case

Tables 4.38 and 4.39 show results for the "High Endogeneity" and "High Heterogeneity" cases, respectively:

In both cases the parameters are recovered well with the exception of the brand constants for brands 1 and 2 in Segment B. Once again, these might be explained by the low market shares of these brands in segment B.

	True Values		*GMM Estimates (std. errors)*	
Parameter	*A*	*B*	*A*	*B*
Brand constant 1	4	2.5	4.00 (0.05)	6.12 (0.22)
Brand constant 2	1	2.5	1.04 (0.23)	3.39 (0.48)
Brand constant 3	-1	1	-0.76 (0.31)	1.41 (1.15)
Price	0.50	2	0.47 (0.20)	2.32 (0.77)
Display	4	2	3.89 (0.18)	2.20 (0.72)
Proportion of consumers	0.3	0.7	0.29	0.71

Table 4.39. Parameter Estimates for Simulated Market Demand in High Heterogeneity Case

APPENDIX B.
Multiple Discreteness in other categories

The tables (4.40 to 4.45) present the joint frequency distribution of total products and total units purchased on a given trip. These data motivate the incidence of the multiple discreteness problem in several categories, not just CSDs. For CSDs, I use the data from the econometric analysis. In the non-CSD categories, I combine purchases for a specific brand/size rather than disaggregating to the UPC level. Similar to CSDs, these are all categories with substantial product differentiation and a large number of alternatives.

prods/units	1	2	3	4	5	6	7	8	9	10+	Total
1	20652	11238	1447	2454	245	454	33	282	19	215	37039
2	0	6928	2215	1817	436	464	146	259	45	166	12476
3	0	0	1322	768	302	247	114	109	45	130	3037
4	0	0	0	335	165	109	63	77	28	69	846
5	0	0	0	0	51	69	27	18	16	41	222
6	0	0	0	0	0	7	16	9	8	19	59
7	0	0	0	0	0	0	4	2	2	3	11
8	0	0	0	0	0	0	0	1	2	7	10
9	0	0	0	0	0	0	0	0	0	1	1
10	0	0	0	0	0	0	0	0	0	3	3
Total	20652	18166	4984	5374	1199	1350	403	757	165	654	53704

Table 4.40. Carbonated Soft Drinks

107 Brand/sizes accounting for 76% of category revenues											
brands/units	1	2	3	4	5	6	7	8	9	10 +	Total
1	23222	3698	245	163	16	39	3	7	1	15	27409
2	0	7509	1050	464	59	52	12	13	11	20	9190
3	0	0	2087	401	128	97	14	15	3	16	2761
4	0	0	0	692	161	46	39	30	4	13	985
5	0	0	0	0	165	43	32	9	6	9	264
6	0	0	0	0	0	77	24	17	5	11	134
7	0	0	0	0	0	0	21	10	4	9	44
8	0	0	0	0	0	0	0	11	5	5	21
9	0	0	0	0	0	0	0	0	2	5	7
10 +	0	0	0	0	0	0	0	0	0	11	11
Total	23222	11207	3382	1720	529	354	145	112	41	114	40826

Table 4.41. Ready-to-eat cereal

16 brand/sizes accounting for 96% of category revenues											
brands/units	1	2	3	4	5	6	7	8	9	10	Total
1	15822	4729	657	236	35	55	9	14	7	7	21571
2	0	576	133	34	7	6	2	3	0	1	762
3	0	0	12	6	3	0	1	0	0	0	22
4	0	0	0	0	0	1	0	0	0	0	1
Total	15822	5305	802	276	45	62	12	17	7	8	22356

Table 4.42. Ice Cream

18 brand/sizes accounting for 93% of category revenues											
brands/units	1	2	3	4	5	6	7	8	9	10 +	Total
1	7946	9005	2782	2525	444	723	123	196	57	237	24038
2	0	1779	1635	1392	686	537	217	220	87	263	6816
3	0	0	266	378	278	214	155	114	65	199	1669
4	0	0	0	38	50	62	48	40	21	78	337
5	0	0	0	0	5	1	6	3	3	18	36
6	0	0	0	0	0	0	0	0	0	5	5
Total	7946	10784	4683	4333	1463	1537	549	573	233	800	32901

Table 4.43. Canned Soup

38 brand/sizes accounting for 95% of category revenues											
brands/units	1	2	3	4	5	6	7	8	9	10 +	Total
1	9720	3217	309	216	19	35	4	10	1	9	13540
2	0	690	159	73	10	6	5	1	3	2	949
3	0	0	38	20	6	5	1	0	1	0	71
4	0	0	0	2	1	0	0	1	0	0	4
5	0	0	0	0	0	1	1	0	0	0	2
6	0	0	0	0	0	0	1	0	0	0	1
Total	9720	3907	506	311	36	47	12	12	5	11	14567

Table 4.44. Spaghetti Sauce

84 brand/sizes accounting for 75% of category revenues											
brands/units	1	2	3	4	5	6	7	8	9	10 +	Total
1	16781	3289	243	169	18	32	3	2	2	3	20542
2	0	3354	616	189	34	28	8	1	3	2	4235
3	0	0	467	146	31	19	3	2	1	5	674
4	0	0	0	90	17	18	7	2	3	3	140
5	0	0	0	0	10	10	5	2	1	2	30
6	0	0	0	0	0	1	3	3			7
7	0	0	0	0	0	0	1	0	0	1	2
Total	16781	6643	1326	594	110	108	30	12	10	16	25630

Table 4.45. Cookies

APPENDIX C.
Data

product	abbreviation
PEPSI COLA REGULAR 12 cans	PEPSI 12P
COKE CLASSIC 12 cans	COKE CLS 12P
PEPSI REGULAR 6 cans	PEPSI 6P
COKE DIET 12 cans	COKE DT 12P
PEPSI REGULAR 67.6oz	PEPSI 67.6oz
PEPSI DIET 12 cans	PEPSI DT 12P
COKE CLASSIC 6 cans	COKE CLS 6P
PEPSI DIET 6 cans	PEPSI DT 6P
COKE CLASSIC 67.6oz	COKE CLS 67.6oz
PEPSI DIET CL 67.6oz	PEPSI DT CL 67.6oz
COKE DIET 6 cans	COKE DT 6P
DR PEPPER 12 cans	DR PR 12P
MOUNTAIN DEW 12 cans	MT DW 12P
DR PEPPER 6 cans	DR PR 6P
7UP CAFFEINE-FREE 67.6oz	7UP R CF 67.6oz
COKE DIET CAFFEINE-FREE 12 cans	COKE DT CF 12P
COKE DIET 67.6oz	COKE DT 67.6oz
7UP DIET CAFFEINE-FREE 67.6oz	7UP DT CF 67.6oz
MOUNTAIN DEW 6 cans	MT DW 6P
SPRITE CAFFEINE-FREE 12 cans	SP CF 12P
PEPSI DIET CAFFEINE-FREE 12 cans	PEPSI DT CF 12P
DR PEPPER 67.6oz	DR PR 67.6oz
MOUNTAIN DEW 67.6oz	MT DW 67.6oz
PEPSI REGULAR 6 16oz bottles	PEPSI 16oz
PEPSI DIET CAFFEINE-FREE 6 cans	PEPSI DT CF 6P
A AND W CAFFEINE-FREE 6 cans	A and W CF 6P

Table 4.46. CSD Products Used for Estimation

I list the names of the products as well as the abbreviations I use in the text. The A and W product consists of both non-diet rootbeer and cream soda. These two flavors have almost identical prices and they are indistinguishable in terms of observable attributes. So, I combine them into one generic A and W brand name. Table(4.47) provides sum-

Variable	67.6 oz bottle		6-pack cans		16 oz bottle		12-pack cans	
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
share of unit sales	0.045	0.04	0.05	0.06	0.02	0.05	0.02	0.03
price (cents per 12 oz)	17.83	5.16	21.81	10.96	33.44	10.89	29.83	6.47
feature	0.43	0.50	0.53	0.50	0.14	0.35	0.46	0.50
display	0.46	0.50	0.55	0.50	0.12	0.31	0.50	0.50
temperature (degrees F)	50.67	16.44	50.67	16.44	50.67	16.44	50.67	16.44

Table 4.47. Data Used in Aggregate DCM

mary statistics for the data used in the aggregate DCM.

I provide summary statistics of the demographic variables and the time-varying product attributes for the panel data in table(4.48). Most of these variables are self-explanatory. I compute the time between trips in days. Family size consists of the number of individuals reported for a given household. Temperature is the daily maximum, reported in degrees. Finally, income bracket is divided into 9 groups: 1 indicates less than $10,000, 2 indicates between $10,000 and $20,000, 3 indicates between $20,000 and $30,000, 4 indicates between $30,000 and $40,000, 5 indicates between $40,000 and $50,000, 6 indicates between $50,000 and $60,000, 7 indicates between $60,000 and $70,000, 8 indicates between $70,000 - $100,000, and 9 indicates over $100,000.

Variable	mean	standard deviation
kids	0.3865	0.4870
family size	2.6976	1.4034
income bracket	4.2470	1.9616
female under 35	0.1964	0.3973
time between trips	6.8498	13.7602
inventory	57.6103	86.7435
max. temperature	64.6149	19.8264
spring	0.2411	0.4278
summer	0.2785	0.4483
winter	0.2487	0.4322
price ($)	2.1515	0.3782
ad	0.3203	0.0579
display	0.4174	0.0503

Table 4.48. Descriptive Statistics (averaged across trips)

		continuous variable		
Flavor		calories	sodium (mg)	carbohydrates
cola	regular	150 (7.5)	40.5 (7.4)	41 (1.51)
	diet	0 (0)	34.7 (8.7)	0 (0)
lem/lime	regular	143.3 (5)	61.7 (16.4)	38.333 (0.5)
	diet	0 (0)	35 (0)	0 (0)
rootbeer	regular	168.3 (4.1)	44.2 (14.6)	44.8 (1.5)
citrus	regular	170 (0)	70 (0)	46 (0)
pepper	regular	148.6 (3.8)	45.7 (8.9)	35.1 (15.5)

Table 4.49. Continuous Attributes by flavor and diet vs. regular (averages)

		indicators					
Flavor		caffeine	phos.	citric	caramel	clear	#
cola	regular	7	7	4	7	0	7
	diet	6	9	9	9	0	9
lemon\lime	regular	0	0	2	0	2	2
	diet	0	0	1	0	1	1
rootbeer	regular	0	0	0	1	0	1
citrus	regular	3	0	3	0	3	3
pepper	regular	3	3	0	3	0	3

Table 4.50. Indicator Attributes by flavor and diet vs. regular (counts)

APPENDIX D.
Conditional Logit

To provide a benchmark for the findings of the proposed model, I present the parameter estimates for several specifications of the conditional logit. The advantage of the logit lies in its popularity as well as the fact that it is a restricted case of the proposed model in which I eliminate the tasks and the quantity choice. I treat the contemporaneous purchase of several alternatives as separate transactions and I ignore the quantity choice. This model captures brand choice, not demand. Adding a type I extreme value error to the random utility in (4.6) yields the following probability for a household h to choose product j :

$$P_{hj} = \frac{exp\left(X_j\beta + X_jD_h\gamma\right)}{\sum_{k=1}^{J} exp\left(X_k\beta + X_kD_h\gamma\right)}, j = 1, ..., J$$

where X_i includes all product attributes (including price), and γ captures the interactions between attributes and household demographics, D_h. I report estimates for three specifications of this model, using maximum likelihood, in (4.51). In particular, while the addition of the demographic interactions in model 1 seem to be significant, they do not seem to have much effect on the logit's other parameters. The only noticeable change is a slight decrease in the marginal utility of caffeine, which is clearly related to the strong preferences of caffeinated CSDs by households with kids. I found a similar result with the proposed model. I also found that newspaper advertising does not have a statistically significant effect on the choice problem. Part of this insignificant effect may be due to ignoring quantities. However, I expect that treating simultaneous choices as independent logit decisions inflates the error in the regression, leading to more noise in the estimates.

Variable	Model 1	Model 2	Model 3
price	-3.3829	-3.1058	-3.3061
	(0.0814)	(0.0701)	(0.0791)
ad	-0.0103	-0.0072	-0.0233
	(0.0102)	(0.0100)	(0.0109)
display	0.6064	0.6056	0.6018
	(0.0107)	(0.0105)	(0.0113)
temperature	0.0038	0.0037	-
	(0.0001)	(0.0001)	-
inventory	-0.0004	-0.0004	-
	(0.0000)	(0.0000)	-
spring	0.6127	0.6245	-
	(0.0033)	(0.0032)	-
summer	-0.1293	-0.1294	-
	(0.0068)	(0.0067)	-
winter	-0.0573	-0.0560	-
	(0.0055)	(0.0053)	-
income*price	0.0595	-	-
	(0.0084)	-	-
kids*diet	-0.6526	-	-
	(0.0057)	-	-
kids*caff.	0.3928	-	-
	(0.0131)	-	-
(fem.<35)*diet	-0.1074	-	-
	(0.0057)	-	-
(fem. college)*diet	0.3266	-	-
	(0.0041)	-	-
(family size)*(servings)	0.0004	-	-
	(0.0002)	-	-
Obs	172,351	172,351	172,351

Table 4.51. Conditional Logit Parameter Estimates (standard errors in parentheses)

APPENDIX E.
Random-Effects Poisson

I use the random effects Poisson to provide a benchmark in which I attempt to account for the quantity choice on a given trip. I assume that, independently of the actual products chosen, consumers choose some integer number of CSD units to purchase based on household attributes, D_{ht}, and overall store-conditions, S_t (such as price-level). Formally, I assume that the probability that household h purchases k CSD units on trip t has the following form:

$$P_{ht}(Y = k|\alpha_h) = \frac{\lambda_{ht}^k}{k!} exp\left\{-exp\left(\alpha_h\right)\lambda_{ht}\right\} exp\left(\alpha_h k\right)$$

where $\lambda_{ht} = exp\left(D_{ht}\beta + S_t\omega\right)$ measures the incidence rate and α_h is a household-specific random effect. I estimate the parameters β, ω and the variance of α_h. Thus, the probability of observing the entire T_h-trip purchase string for a household h has the form:

$$P(Y_{h1}, ..., Y_{hT_h}) = \int_{-\infty}^{\infty} \Pi_{t=1}^{T_h} P_{ht}(Y = k|\alpha)\partial F\left(\alpha\right)$$

where I assume α derives from the distribution $F\left(\cdot\right)$. I assume α is distributed as a mean-zero normal distribution and I approximate the integral above using Gauss-Hermite polynomials. I estimate the model using maximum likelihood. Results from the Poisson estimation are reported in table (4.52). As expected, households with kids and higher income purchase more CSD units on average. Also, consumers purchase more as the time since the last trip increases (inventory effects). Surprisingly, the time since last CSD purchase has a negative effect, probably proxying for heterogeneity associated with households that seldom purchase CSDs. Finally, high prices today reduce current

Variable	parameter	standard error
kids	0.136	0.028
household size	0.040	0.010
income	-0.008	0.006
time since last CSD	-0.002	0.000
time since last trip	0.005	0.001
fav. product prices	-2.096	0.116
overall prices	0.046	2.061
lagged fav. product prices	1.043	0.083
constant	-0.635	0.053
variance random effect	0.450	0.038

Table 4.52. Parameter Estimates from Panel Poisson

expected unit purchases, whereas high prices yesterday increase current unit purchases. I also find significant unobserved heterogeneity. To compute the expected purchase vector for each trip, I multiply the expected unit purchases predicted by the Poisson with each of the product probabilities, estimated from a conditional-on-purchase logit.

APPENDIX F
Estimating the Aggregate DCM

The following methodology is analogous to that of BLP(1995) and Nevo(2000). First, I partition the parameters to be estimated into the mean tastes, θ_1, and the standard deviations of the tastes, θ_2. In order to estimate the aggregate DCM, I need to start with the basic share equation for each brand j:

$$S_j = \int_{Aj} \frac{exp\left(u_j + v_{ij}\right)}{1 + \sum exp\left(u_j + v_{ij}\right)} \partial\Phi(\sigma)$$

where u_j is the mean utility of brand j and v_{ij} is the random component of the utility for brand j due to the random coefficients. Estimating the share equation directly introduces several computational problems. Instrumenting becomes quite complicated due to the non-linear fashion in which the explanatory variables enter the model. I also expect the prediction error of this system of equations to be highly correlated within a given store-week and across weeks for a given store. To alleviate these issues, Berry(1994) suggests working with the mean utility, u_j. Using the inversion procedure proposed by BLP(1995), I can solve the share equation for the mean utility:

$$u_j\left(\theta_2\right) = X_j\widetilde{\beta} - \widetilde{\phi}p_j.$$

This equation is much simpler to implement since the mean taste parameters now enter linearly. If the attributes, X_j, are measured with error, ζ_j, then I can construct a Generalized Method of Moments estimator based on the assumption $E\left(X_j\zeta_j|X_j\right) = 0$. Using a matrix of exogenous instruments, Z, which contains X_j as well as supply-side prices of the factors of production, I form the conditional moments $E\left(Z_j\zeta_j|Z_j\right) = 0$.

Estimation amounts to finding the vector $(\theta_1^*, \theta_2^*)'$ that minimizes:

$$G(\theta_1, \theta_2) = \zeta' Z W Z' \zeta$$

where the weight matrix, W, is the inverse of the estimated variance of the conditional moments (Hansen 1982). As discussed in the text, if the outside share is measured with a downward bias and all the product shares, S_j, are measured with an upward bias, the procedure above will fit values of u_j that are biased upwards. Thus, the GMM procedure will yield estimates of $\widetilde{\beta}$ that are too high and values of $\widetilde{\phi}$ that are too low (since they are negative). I demonstrate this result by noting that these parameters enter the estimation problem linearly, so I can compute their values analytically:

$$\theta_1^* = (X' Z W Z' X)^{-1} X' Z W Z' u$$

where u is the vector of mean utilities and X is the matrix of prices and attributes. Clearly, if the vector u is measured as larger than the true underlying mean utilities, then the vector θ_1^* will also be biased upwards in magnitude relative to the true underlying mean tastes.

APPENDIX G.
Computing Equilibrium Prices

Once I have estimates for demand and marginal costs, I am able to back out the equilibrium prices that would prevail after removing a subset of products from the choice set. Assuming the existence of the static Nash equilibrium in prices discussed in the text, the new equilibrium prices must satisfy:

$$p^{*} = mc + \Delta\left(p^{*}\right)^{-1} E\left(Q\left(p^{*}\right)\right).$$

For standard models, such as linear or logit demand, this system of equations is easily solved numerically. For instance, one could use a least squares procedure such as "fsolve" in Matlab. However, the proposed demand model is non-smooth due to the finite number of simulation draws. I define:

$$j(p) = mc + \Delta\left(p\right)^{-1} E\left(Q\left(p\right)\right)$$

where $j(p^{*}) = 0$. To find the equilibrium prices that set $j(p)$ as close to zero as possible, I simply minimize:

$$Q(p) = j(p)' * j(p)$$

which, theoretically, has a minimum when $j(p) = 0$. I minimize $Q(p)$ using the Nelder Meade simplex search (a non-derivative method).

APPENDIX H.
Brief history of the major brands

The earliest roots of the CSD lie in a historic fascination with naturally carbonated mineral water springs. The belief in the healing ability of the water inspired a series of efforts to reproduce the waters artificially as early as the 16th century. While studying the containment of carbonation in waters at the Leeds brewery, Priestley developed the first successful method of artificial carbonation. Recognizing the potential for curing scurvy on naval vessels, he presented his results to the Royal naval society of London in his famous paper, "Observations on Different Kinds of Air." Bewley refined the method, adding a small amount of carbonate of soda and creating the first soda water.[51] Commercially, the carbonated waters became popular with the addition of flavoring, such as lemonade, to mask the unpleasant taste of the sodium bicarbonate in this "soda water". While cultural factors prevented the medicinal beverages from becoming popular in Europe,[52] North American druggists frantically combined various flavor and chemical combinations to promote their immensely successful soda fountains. In 1886, John Styth Pemberton developed a "pick-me-up" syrup, combining powder from the coca leaf with oil from the Kola nut. He masked the bitter taste with sweetened caramel and called the product Coca-Cola, based on its primary ingredients. In 1893, Caleb B. Bradham attempted to replicate the Coca-Cola formula to serve at the fountain in his drug-

---

[51] For a history of these methods, see Joseph Priestley, Directions for Impregnating Water with Fixed Air in order to communicate to it the peculiar Spirits and Virtues of Pyrmont Waters. (Washington DC: reprint by American Bottlers of Carbonated Beverages, 1945).

[52] The British preferred tea and the French and Italians preferred wine and natural mineral water.

store. Using a similar concoction of Kola nut and caramel, he developed Brad's Drink, which he marketed as a cure for stomach dyspepsia and peptic ulcers. After a large initial success, he renamed the drink Pepsi-Cola to reflect these medicinal functions. In Waco Texas, 1885, Charles C. Adderton concocted a mix of phosphorescent waters, fruit juices and sugar. In hopes of amusing his boss into permitting the sale of the drink at the fountain, Adderton named it "Dr. Pepper," in reference to the disapproving Virginia father of the woman his boss was courting at the time. In 1929, Charles Grigg developed a lemon-flavored product, Bib-label, to which he added lithium and marketed the beverage as a cure for hangover and upset stomach. After a substantial commercial success, he renamed the drink 7UP. The origin of the name itself is uncertain.[53]

Many famous brands today owe their widespread appeal to the temperance movement. The name, "soft" drink distinguishes the category from "hard" drinks, beverages containing alcohol. Early marketing of Coca-Cola included such slogans as "The Great Temperance Drink" and "[the] Intellectual Beverage and Temperance Drink" in reference to its non-alcoholic "pick-me-up" quality.[54] Charles Hires ultimately chose the name "root beer" for his product, which he originally called root tea. The term "beer," referring to the brewing process, appealed to the tastes of local Pennsylvanian coal miners seeking an alternative to beer as temperance sentiments grew.[55] Canadian product,

---

[53]These brand descriptions are from Encyclopedia of Consumer Brands. V.1 Ed. Janice Jorgensen. (Detroit: St. James Press, 1994).

[54]For more details see J.C. Louis and Harvey Z. Yazijian, The Cola Wars. (New York: Everest House, Publishers, 1980), p. 14.

[55]Encyclopedia, p.261.

Canada Dry ginger ale, did not become a popular North American CSD until the 1920s, when it was consumed as an alternative to alcohol and as a mixer to mask the flavor of underground liquor, bathtub gin.[56]

[56]Encyclopedia, p. 94.

APPENDIX I.
Origins of CSD product differentiation

Almost all of the major CSD brands today originated as independently-developed flavorings for soda fountains in late 19th and early 20th century apothecary shops. As bottling and distribution technology evolved during the first quarter of the 20th century, the major brands began to compete for market share both in local city markets and nationally. Early on, manufacturers recognized the need to differentiate themselves to increase and to sustain market share. Early differentiation focused on brand awareness. Later in the 20th century, the emphasis moved towards targeting specific consumer groups with different flavors and package sizes and types. From the start, the success of brands like Coca-Cola came from aggressive and original advertising to establish brand identity and awareness. In 1915, Coca-Cola introduced the standardized coke bottle: a $6\frac{1}{2}$ ounce hobble-skirted bottle designed to look like a cola nut. Ironically, the C.J Root Company, which patented the bottle, accidentally copied the cacao bean. In 1960, the $6\frac{1}{2}$-ounce bottle became a Coca-Cola trademark.[57] In 1930, the Coca-Cola name was advertised on the radio, over 500,000,000 letters and sky-written messages, 20,000 walls and 160,000 posters.[58] Pepsi responded by using recycled 12 ounce beer bottles and charging the same price of 5 cents as a $6\frac{1}{2}$ ounce Coca-Cola: "two large glasses in each bottle."[59] In 1939, Pepsi introduced the first fifteen-second radio jingle, rather

---

[57] Louis and Yazijian, p.31.

[58] Taken from "To Pause and Be Refreshed," Fortune, July 1931, p.108-110.

[59] Louis and Yazijian, p.50.

than the standard bland reading of text: "twice as much for a dollar too."[60] The song became an instant hit, played over 6 million times on 469 radio stations in 1941 alone, and orchestrated and marketed as a jukebox hit.[61] With the exception of Pepsi's cheaper bottle, most early industry advertising was intended to increase brand awareness, rather than product differentiation.

As Coca-Cola and Pepsi established their leading positions in the industry by the 1950s, competing brands and new entrants realized the importance of differentiation in acquiring and protecting market shares. Most of the early differentiation occurred amongst colas. During the 1960s, RC focused on ordinal differentiation by introducing the first independent certified taste tests, a method that would later bolster Pepsi sales during the mid and late 1970s. Most of the differentiation consisted of innovative packaging. Pepsi was the first to target specific consumers with its larger, "hostess-sized bottles, intended as a grocery product. Royal Crown Cola (RC) introduced the first nationally distributed CSD in cans (1954), the first 16 ounce bottle (1958) and the first all-aluminum can (1964).[62]

During the mid-to-late 1960s, differentiation took on a new form: market segmentation. Segmentation involved establishing distinct flavors and targeting a specific type of consumer for each. In 1962, RC introduced one of the most influential segments, diet CSDs, with Diet Rite. Coke responded with its own products, Tab (1963), and later

---

[60]Muris et al. (1993), p. 17.

[61]Louis and Yazijian, p.68-69.

[62]Encyclopedia, p.481.

Diet Coke (1982). Later, RC expanded this health-conscious niche to include caffeine-free diet cola, Dccaf.100 (1982). That same year, Coke and Pepsi released their own caffeine-free and caffeine-free diet colas. 7-UP, marketed its lemon-lime flavor as the "uncola."[63] Coke responded with its own lemon-lime product, Sprite, and directed its advertising towards the 18-25 age group.[64] In contrast, Pepsi revitalized its "hillbilly-themed" lemon-lime drink, Mountain Dew, by adding orange to the syrup and establishing a citrus segment targeted towards the teen niche. When Dr. Pepper began its tremendously successful "misunderstood campaign" (1970-1983) with the famous slogan, "Be a Pepper,"[65] Coke entered the new pepper segment with Mr. Pibb. Dr. Pepper also acquired and rejuvenated Canada Dry by establishing an 18-49 year old adult niche with slogans like "Are you ready for Canada Dry?" and "For your tastes grow up". Soon, the growing taste for ginger ales inspired several new clear "New Age" beverages, including Pepsi's clear cola "Crystal Pepsi." Finally, a recent emergence of a quality and health-conscious CSD consumer inspired a line of products containing real concentrated grapefruit juice, Squirt and Coca-Cola's Fresca. The long-run consequence of this targeted advertising has been the creation of specific flavor niches, each with its own primary consumer segment.

Most of these forms of differentiation involve long-term changes in production methods, to accommodate new packaging techniques, and large sunk advertising costs, to

[63]Encyclopedia, p.552.

[64]Encyclopedia, p.552.

[65]Encyclopedia, p.168.

develop brand awareness and to target specific consumer segments. In the current application, these issues are important for the determination of perceived brand quality; but they are treated as fixed. Endogenizing both advertising and fixed brand attributes requires a much longer time period than the available sample data. By the 1980s, CSD producers had, more-or-less, exploited most of the potential differentiation strategies. As the industry matured, consolidation appeared to be the only source of increasing profitability.